

القياس والتقويم التربوي والتقييم
أساليبه وتطبيقاته وتجهيزاته المعاصرة

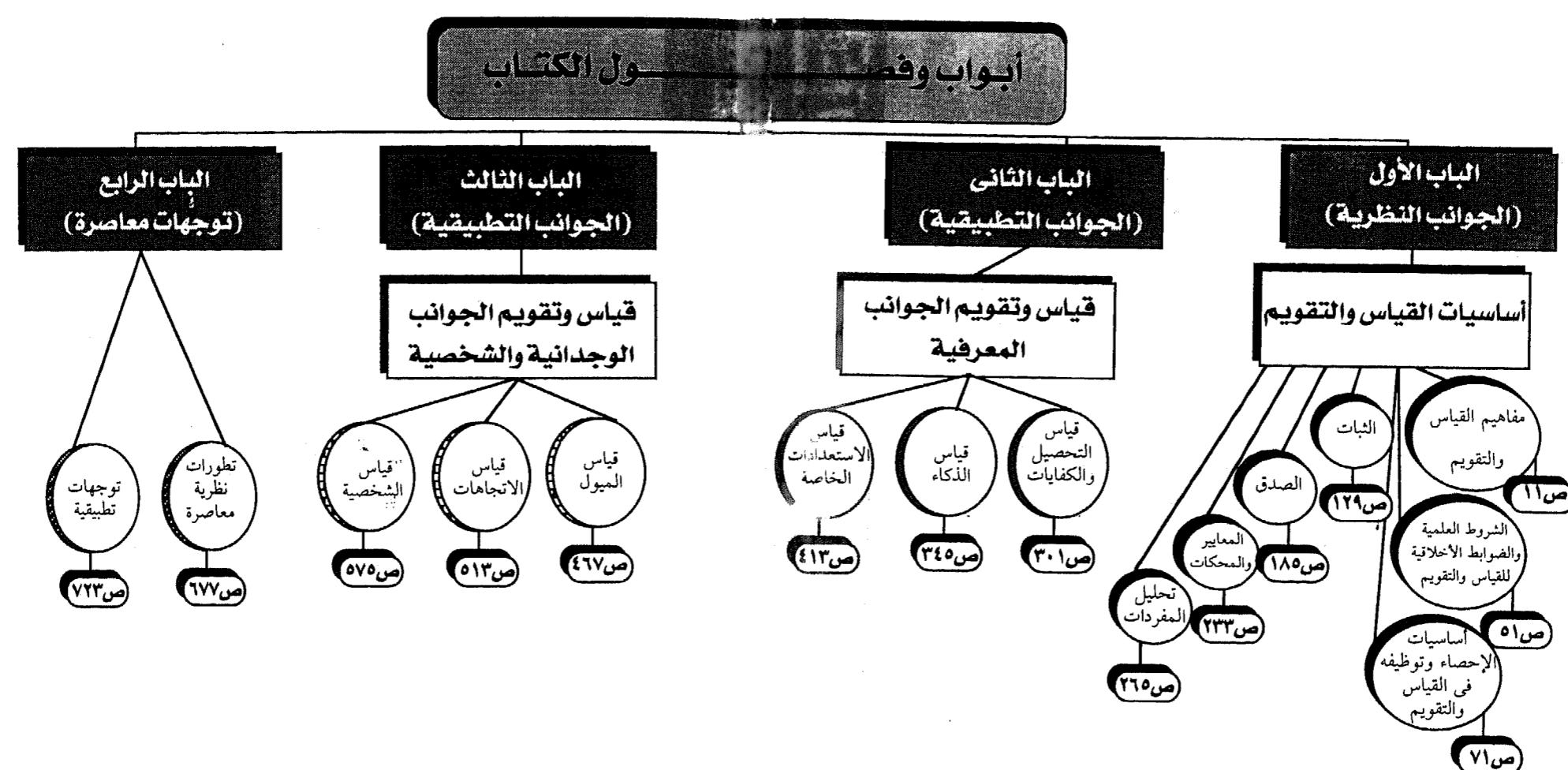
الدكتور صالح الدين محمود علام

دار الفكر العربي

0185-20



Biblioteca Alexandrina



القياس والتقويم التربوي والنفسى

أساسياته وتطبيقاته وتوجهاته المعاصرة

الدكتور صالح الدين محمود لام
أستاذ القياس والتقويم والإحصاء التربوي النفسي
كلية التربية - جامعة الأردن

الطبعة الأولى

م ٢٠٠٠ - هـ ١٤٢٠

ملزم الطبع والنشر
دار الفكر العربي

٩٤ شارع عباس العقاد - مدينة نصر - القاهرة
ت: ٢٧٥٢٩٨٤ - فاكس: ٢٧٥٢٧٣٥

٣٧١,٢٦ صلاح الدين محمود علام.

صلقى القياس والتقويم التربوى والنفسى: أساسياته وتطبيقاته
وتحوياته المعاصرة/ تأليف صلاح الدين محمود علام. -
القاهرة: دار الفكر العربي، ١٩٩٩.

٨١٦ ص: إيض؛ ٢٤ سم.

بليوجرافية: ص ٧٥٣-٧٧٦.

يشتمل على معجم بالمصطلحات الإنجليزى - عربى.

تدمك: ٩٧٧ - ١٠ - ١٢١٠ - ٠.

١- الاختبارات والقياسات التعليمية. ١- العنوان.

تصميم وإخراج فنى

السيد أحمد سيف



أميرة للطباعة عابدين - ت: ٣٩١٥٨١٧

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

مُقْتَسِمٌ

يُعد مجال القياس والتقويم التربوي والنفسى من المجالات الحيوية الأساسية التي لا غنى عنها للدارسين والباحثين في العلوم السلوكية، والمسئولين عن اتخاذ القرارات المتعلقة بالأفراد في مختلف الميادين التطبيقية التربوية والنفسية والاجتماعية والإدارية والصناعية، والعسكرية، وغيرها من الميادين التي تمركز حول الإنسان مصدر الثروة الرئيسة للأمم الوعية.

لذلك حظى مجال القياس والتقويم التربوي والنفسى باهتمام بحثى تطويرى متزايد من جانب علماء النفس والتربية في الأونة الأخيرة. ولعل هذا يبدو واضحاً وملموساً فيما تتناوله أدبيات القياس المعاصر المتخصصة من موضوعات بحثية مستحدثة، وما ترده من إسهامات تطبيقية في مجالات السلوك الإنساني المختلفة. ولم يُعد يُنكر أحد اليوم الدور البالغ الأهمية للقياس التربوي والنفسى في تقدم العلوم السلوكية وإثراء تطبيقاتها الوظيفية المتعددة بما يفهى بالمتطلبات المتجددة للحياة الإنسانية في هذا القرن.

غير أن تنوع مجالات القياس والتقويم التربوي والنفسى وما يندرج تحتها من موضوعات متعددة، وما تتطوى عليه من مفاهيم وأساليب وأدوات متباعدة، يؤدى في كثير من الأحيان إلى إساءة استخدام هذه الأساليب والأدوات وعدم توظيفها توظيفاً مناسباً. وربما يرجع ذلك إلى عدم وضوح أوجه الترابط والتكامل فيما بينها نظراً لحاجة الدارسين والباحثين والممارسين إلى تطوير معارفهم فيما يتعلق بالأسس المنهجية والإحصائية التي تستند إليها، وتنمية مهاراتهم في تطبيق مفاهيم وأساليب وأدوات القياس والتقويم المعاصر بفهم مستثير وفكير مفتوح.

لذلك بدت الحاجة ماسة إلى مرجع أساسى متكامل يكون بمثابة مرشد ينير الطريق للدارسين والباحثين والممارسين في مختلف مجالات القياس والتقويم التربوي والنفسى.

وقد حاولنا في هذا الكتاب أن يكون هناك توازن وتكامل بين الجوانب النظرية والجوانب التطبيقية بحيث يخدم كلاهما الآخر مُراعين الأسس المنهجية والإحصائية التي ترتكن إليها أساليب وأدوات القياس والتقويم، والاستخدامات المناسبة لها، وحدود هذه الاستخدامات، لكي تفي بالغرض الذي صُممت من أجل تحقيقه. واسترشدنا في عرضنا لهذه الجوانب بأهم التطورات التي حدثت في علم القياس والتقويم التربوي النفسي لمواكبة الحركة البحثية المعاصرة، كما أفادنا من خبرتنا الأكademية وممارساتنا الميدانية في مجال القياس والتقويم والإحصاء في كثير من الدول العربية خلال القديمين الماضيين. ويسيراً على القارئ قسمنا الكتاب إلى أربعة أبواب رئيسة متكاملة، تناولنا في الباب الأول الجوانب النظرية المتعلقة بالمفاهيم والمبادئ الأساسية، والأسس المنطقية والإحصائية للقياس والتقويم التربوي النفسي.

وتناولنا في البابين الثاني والثالث أهم التطبيقات المتعلقة بقياس وتقويم الجوانب المعرفية والوجدانية الشخصية. وتناولنا في الباب الرابع بعض التطورات والتوجهات المعاصرة في هذا المجال.

ويشتمل الباب الأول على سبعة فصول، يتناول الفصل الأول المفاهيم والمبادئ الأساسية للقياس والتقويم، وأنواع السمات النفسية والتربوية وخصائصها، واستخدامات أدوات القياس والتقويم، وتصنيفها.

ويتناول الفصل الثاني مشكلات القياس التربوي النفسي وشروطه العلمية، وضوابط وأخلاقيات استخدام الاختبارات والمقاييس.

ويتناول الفصل الثالث أساسيات الإحصاء وتوظيفه في القياس والتقويم التربوي والنفسي متضمناً أساليب تنظيم وعرض البيانات وتمثيلها بيانيًا أو بشكل الأغصان والأوراق، ومقاييس التوزع المركزية والتشتت، ووصف توزيعات درجات الاختبارات، وأنواع المختلفة من الدرجات المحولة، ومقاييس العلاقة بين مجموعتين من الدرجات، واستخدام معادلة الانحدار الخطى البسيط في التنبؤ.

ويتناول الفصل الرابع الخصائص السيكومترية للاختبارات المتعلقة بثبات الدرجات، ومصادر أخطاء القياس، وأنواع المختلفة لمعاملات الثبات، وكيفية تقدير قيمها، وثبات تقديرات المحكمين، وثبات الفروق بين درجتي اختبارين، والخطأ المعياري للقياس، والعوامل المؤثرة في قيم معامل الثبات.

ويتناول الفصل الخامس الخصائص السيكومترية للاختبارات المتعلقة بالصدق وجوداته، وطرق تقدير صدق المحتوى، والصدق المرتبط بمحلك بنوعيه التنبؤي

واللازمى ، والعوامل التى تؤثر فيه ، وكذلك صدق التكوين الفرضي ومنهجيته ، وأساليب جمع أدلة متعلقة به ، ومنظور تكاملى لطرق تقدير الصدق والثبات .

ويتناول الفصل السادس معايير الاختبارات مرجعية الجماعة أو المعيار ، ومحكمات الأداء فى الاختبارات مرجعية المحك ، وأنواع الجماعات المرجعية وخصائصها ، وأنواع معايير الاختبارات ، والصفحات النفسية والتربوية «البروفيل» ، واستخدام محكمات الأداء فى تفسير الدرجات .

ويتناول الفصل السابع أساليب تحليل مفردات الاختبارات وأهمية هذا التحليل ، وكيفية تقدير صعوبة المفردات والمشكلات المتعلقة بذلك ، وأساليب تقييم تميز المفردات استناداً إلى محكمات داخلية أو خارجية ، وتحليل فاعلية بدائل الإجابات ، واستخدام تحليل المفردات في انتقاء مفردات الاختبار ، واستخدم الحاسوب في إجراء هذا التحليل .

ويشتمل الباب الثانى - المتعلق بقياس وتقدير الجوانب المعرفية - على ثلاثة فصول ، يتناول الفصل الثامن قياس التحصيل والكتفيات ، وعلاقة مفهوم التحصيل بمفهوم الاستعداد ، والتمييز بين اختبارات كل منهما ، وتصنيف اختبارات التحصيل واستخداماتها ، وخطوات بناء اختبارات التحصيل مرجعية المحك ، واستخدامات هذه الاختبارات ، واختبارات التحصيل المقنة (جماعية الجماعة أو المعيار) ، وبعض آوجه النقد التي توجّه لاختبارات التحصيل المقنة .

ويتناول الفصل التاسع قياس الذكاء العام ، ومفاهيم الذكاء المتعددة والمشكلات المتعلقة بذلك ، واختبارات الذكاء العام الفردية والجماعية ، وأمثلة متنوعة لكل من نوعى الاختبارات ، وأساليب بناء اختبارات الذكاء ، وبعض القضايا المتعلقة بقياس الذكاء وبعض محددات الذكاء والعوامل المؤثرة فيه ، وثبتت نسبة الذكاء .

ويتناول الفصل العاشر قياس الاستعدادات الخاصة ، وعلاقة مفهوم الاستعداد بمفهوم القدرة ، وتأثير العوامل الوراثية والبيئية في الاستعدادات الخاصة ، واستخدامات اختبارات الاستعدادات ، وبعض الافتراضات الأساسية في قياسها ، وأساليب بناء اختباراتها ، وإجراءات تحليل العمل ، وبطاريات الاستعدادات المتعددة ، واختبارات الاستعدادات المتعلقة بمجالات نوعية محددة ، مثل الاستعداد الميكانيكي والاستعدادات النفسية السحرية ، والاستعداد الكتابي ، والاستعداد الموسيقى ، والاستعدادات الفنية ، والابتكارية ، ومفهوم التفكير الباعدي ، وقياس قدرات هذا النمط من التفكير ، والمشكلات المتعلقة به .

ويشتمل الباب الثالث - المتعلق بقياس وتقدير الجوانب الوجودانية والشخصية - على ثلاثة فصول، يتناول الفصل الحادى عشر قياس الميل ، وعلاقة مفهوم الميل بمفهوم الاستعداد ، والأنماط المختلفة لالميل ، وأساليب بناء استبيانات الميل ، وبعض المشكلات المتعلقة بذلك ، وأمثلة لهذه الاستبيانات واستخداماتها ، وتقدير ميل الطلبة في الصف الدراسي ، وأدوات قياس هذه الميل.

وينتقل الفصل الثانى عشر قياس الاتجاهات ، وعلاقة مفهوم الاتجاه بغیره من المفاهيم الوجودانية ، ومكونات الاتجاهات وخصائصها ، والأسس المنطقية لقياسها ، وتصنيف مقاييس الاتجاهات ، وأساليب بناء موازين الاتجاهات ، مثل أسلوب ثيرستون ، وأسلوب ليكرت ، وأسلوب جتمان ، وأسلوب تمایز معانى المفاهيم ، وأسلوب التصنيف الترتيبى (Q Sort) ، وأسلوب البسط ، وأساليب قياس العلاقات الاجتماعية ، وأساليب بناء موازين الاتجاهات متعددة الأبعاد ، وإرشادات عامة لبناء موازين الاتجاهات ، وأنواع أخرى من مقاييس الاتجاهات ، وبعض المشكلات قياس الاتجاهات واستخدامات مقاييسها .

وينتقل الفصل الثالث عشر قياس الشخصية ، ومفهوم الشخصية وبعض نظرياتها ، وتصنيف مقاييسها ، وطرق وأساليب قياسها ، واستراتيجيات بناء استبيانات الشخصية ، وأمثلة لبعض أهم هذه الاستبيانات التي تستند إلى أساس نظري منطقى ، أو إلى التحليل العاملى ، أو إلى أساس إمبريقى ، وكذلك استبيانات جبرية الاختيار ، وتقسيم استبيانات الشخصية بعامة ، وتصنيف الأساليب الإسقاطية ، وأمثلة لبعض أهم اختباراتها وتقييمها ، وأساليب ملاحظة السلوك ، ومصادر الخطأ فيها وكيفية التغلب عليها ، وموازين تقدير السلوك ، ومصادر الخطأ فيها ومعالجتها ، وأساليب المقابلة وميزاتها وعيوبها وتقييمها ، ومقاييس الأنشطة الفسيولوجية ، ومقاييس الإدراك وتقييمها ، وتعقب عام على طرق وأساليب قياس الشخصية ، واستخدامات هذه المقاييس .

ويشتمل الباب الرابع على فصلين، يتناول الفصل الرابع عشر بعض التطورات النظرية المعاصرة ، ومن أهمها نظرية الاستجابة للمفردة الاختبارية ، والأنواع الرئيسة لنماذجها ، وأساسها المنطقى ، وتصنيفها ، ومفهوم المنهنى المميز للمفردة ، والنماذج الاستاتيكية اللوغاريتمية أحادية البعد ، مثل نموذج راش ، ونموذج البارامترین ، ونموذج البارامترات الثلاثة ، ومزايا هذه النماذج وعيوبها ، ونظرية الثبات متعدد الأوجه ، وبعض مفاهيمها الأساسية ، والدراسات المتعلقة بها ، وتحليل مكونات التباين في هذه الدراسات ، والعلاقة بين هذه النظرية ومفهوم الصدق ، وكذلك يتناول إسهام علم النفس

المعرفي في تطوير أساليب القياس والتقويم، وعلاقة النظرية المعرفية بالعملية الاختبارية، وتصميم الاختبارات في ضوء هذه النظرية.

ويتناول الفصل الخامس عشر بعض التوجهات التطبيقية المتعلقة بأساليب قياس وتقويم التحصيل الدراسي، واستخدام الاختبارات مرجعية المحك في تقويم التحصيل، وتقنيات تفرييد قياس التحصيل باستخدام الاختبارات المُفصّلة، وبناء نظم بنوك الأسئلة، ومراحل إنشاء نظام بنك أسئلة مقترن، والبنك الدولي للأسئلة، وتقويم التحصيل التربوي من منظور دولي، والتقييم التوثيقى الشامل للتحصيل والكافيات Portfolio ، ومكونات الصحف الوثائقية Assessment Portfolios وصيغها ، وبعض المشكلات المتعلقة بها واقتراحات للتغلب عليها.

وقد قدمنا في بداية كل فصل من فصول البابين الثاني والثالث عرضاً موجزاً لتطور قياس السمة التي يتناولها الفصل، لكن يتضح للقارئ هذا التطور من منظور تاريخي .

كما قدمنا العديد من الأشكال التخطيطية التي تيسر استيعاب المفاهيم والأساليب المتعلقة بقياس مختلف السمات الإنسانية، وكذلك قدمنا في بداية الكتاب شكلان تخطيطياً يرشد القارئ إلى أبواب الكتاب وفصوله والصفحات المناظرة لها .

وينتهي الكتاب بمجموعة متقدمة من المراجع التي يمكن الرجوع إليها لمزيد من التفاصيل في موضوعات الكتاب المتعددة .

ويسعدني أن أقدم جزيل الشكر للأستاذ المهندس / عاطف محمود الخضرى مدير دار الفكر العربى على اهتمامه وحفظه للانتهاء من هذا الكتاب وفق جدول زمنى محدد يدل على سداد تفكيره وفاعلية إدارته .

ونرجو من الله سبحانه وتعالى أن يُفيد بهذا الكتاب الدارسين والباحثين والممارسين في مختلف مجالات القياس والتقويم التربوي النفسي، وأن يُمكّنهم من اكتساب المهارات الأساسية المتعلقة بذلك، وتطوير هذه المهارات بحيث تستوعب التطورات المعاصرة المستمرة التي حدثت في هذا العلم الإنساني الح邈 .

والله نسأل ألم يحقق الرجاء

صلاح الدين محمود علام

دكتوراه الفلسفة Ph.D.

في التقويم والقياس والإحصاء النفسي والتربوي
من جامعة ميشيغان الأمريكية



أسسیات القياس والتقویم
التربوي والنفسی



الفصل الأول

المفاهيم والمبادئ الأساسية

- * مقدمة
- * أهمية القياس العلمي
- * المفهوم العلمي للقياس
- * القياس التربوي والنفسى
- * طبيعة القياس التربوي والنفسى
- * أنواع القياس ومستوياته
- * مدخل السمات فى القياس التربوى والنفسى
- * بعض الخصائص الأساسية للسمات
- * أنواع السمات النفسية والتربوية
- * عمليات القياس وأدوات القياس
- * الاختبار التربوى والنفسى
- * التقويم والتقييم
- * استخدامات أدوات القياس والتقويم
- * تصنیف الاختبارات والمقاييس

مقدمة :

يهم علماء النفس والتربية بدراسة السلوك الإنساني دراسة علمية من أجل فهم أفضل لخصائص هذا السلوك وتوجيهه لخير الإنسان، ورفاهية مجتمعه. ونحن عندما نتعامل مع الآخرين، نود أيضًا معرفة تفضيلاتهم ومشاعرهم وإمكاناتهم. ورغبتنا في معرفة الآخرين تكون أحياناً لمجرد حب الاستطلاع، وأحياناً أخرى لتعزف كيفية التعامل معهم في سياق الحياة اليومية. وتعتمد هذه المعرفة في غالبية الأحيان على الملاحظة الذاتية والأراء الشخصية، مما يجعلها تتسم بعدم الاتساق أو الموضوعية وربما التحيز. ولكن المعلم في أدائه لمهامه المهنية، وفي تعامله مع تلاميذه، يهتم بسلوكهم وميولهم وتوجهاتهم وإنجازاتهم وما يمكنهم إنجازه مستقبلاً لكي تتحقق العملية التعليمية أهدافها المرجوة. وكذلك القائمين بدراسة إمكانات القوى البشرية في مختلف مواقع العمل، يهتمون بتعرف خصائص الأفراد من أجل الإفاده القصوى من إمكاناتهم ، ورفع مستوى أدائهم وإنتاجيتهم تلافياً للهدر في الجهد والوقت والمال. وفي مثل هذه الحالات ينبغي اتخاذ قرارات لا يشوّها الحدس، أو الذاتية، أو التحيز، وإنما يجب أن تستند إلى الأساليب العلمية في الانتقاء والتصنيف ، والفحص والتشخيص ، والتسكين والترفيع . ومن هنا تبرز أهمية اكتساب هؤلاء المعنيين المعارف الأساسية ، والمهارات الإجرائية المتعلقة بالقياس والتقويم التربوي والنفسى .

وتحقيقاً لذلك سوف نفرد هذا الباب بفصوله السبعة لمناقشة الجوانب النظرية متضمنة الأسس المنطقية والإحصائية والمنهجية للقياس والتقويم بما يزود القارئ بذخيرة من المعلومات تستند إليها التطبيقات التي سوف تقدمها في البابين الثاني والثالث ، وبذلك يستطيع الإفاده منها في تربية مهاراته المرجوة .

ولعل من المناسب أن نستهل هذا الفصل بعرض أهم المفاهيم والمصطلحات والحقائق الأساسية المتعلقة بمجال القياس والتقويم التربوي والنفسى .

الأهمية القياس العلمي:

بعد القياس أمراً على جانب كبير من الأهمية في أي علم من العلوم، فجميع العلوم تسعى لتطوير أساليب موضوعية دقيقة لقياس الظواهر المتعلقة بها من أجل فهم هذه الظواهر وتفسيرها، والتبنّى بالعلاقات القائمة بين متغيراتها، ومحاولات ضبطها والتحكم فيها. فالتقدم العلمي يعتمد إلى حد كبير على تمثيل الظواهر والأحداث وصياغتها بأساليب موضوعية دقيقة، بحيث تمكن الدارسين والباحثين والممارسين من التواصل فيما بينهم بلغة مشتركة متفق عليها، كما تتمكنهم من التقسيم الموضوعي للنتائج التي يتوصل إليها العلماء، ويدون هذه الأساليب يصبح تعريف المفاهيم والمصطلحات والقواعد المتعلقة بالظواهر المختلفة خاصّاً للأراء الذاتية ، والأحكام الفردية، وبذلك لا تكون هناك أساس متفق عليها في دراسة ويبحث هذه الظواهر.

فالعلم نظام مفتوح نتائجه متداولة مما يستدعي أن تكون الإجراءات والاستنتاجات العلمية خالية من الغموض، بحيث يستطيع علماء مختلفون تكرارها، والتحقق من صدقها، وهذا لا يتأسّى إلا إذا استندت هذه الإجراءات والاستنتاجات إلى أساليب قياس موضوعية.

وربما كان افتقار بعض العلوم في الماضي إلى مثل هذه الأساليب الموضوعية في القياس هو الذي أدى إلى التباين الواضح والجدل المستمر حول تعريف المفاهيم والمصطلحات والقواعد المتعلقة بالظواهر التي تهتم هذه العلوم بدراستها. ولعل علم النفس يعد مثلاً واضحاً لذلك، فدقة موضوعية القياس تسهم إذن في تحديد المفاهيم، وبلورة التفكير من أجل الفهم المستثير لطبيعة الظواهر المختلفة، ويدون ذلك تصبح الدراسة العلمية لهذه الظواهر خاصّة للتأملات العقلية والخبرات الشخصية.

ـ المفهوم العلمي للقياس:

إن القياس Measurement بمفهومه الواسع يشير إلى الجوانب الكمية التي تصف خاصة أو سمة معينة، مثل ارتفاع سائل ، أو حجم كرة، أو ضغط غاز، أو الاستعداد اللفظي لطفل، أو التحصيل الدراسي لطالب، كما يشير إلى عملية جمع المعلومات، وترتيبها بطريقة منتظمة. وبذلك يتضمن مفهوم القياس من هذا المنظور كلًا من عملية جمع المعلومات وتنظيمها، وكذلك نتيجة هذه العملية.

غير أنه يمكن تعريف القياس بصورة أكثر دقة بأنه: تعيين فئة من الأرقام أو الرموز، مناظرة لفئة من الخصائص أو الأحداث، طبقاً لقواعد محددة تحديداً جيداً.

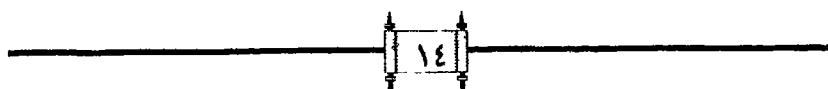
ويشتمل هذا التعريف على ثلاثة عناصر هي : فئات الأرقام أو الرموز ، وفئات الخصائص أو الأحداث ، وقواعد تعين الأرقام أو الرموز للخصائص أو الأحداث . وقد استخدمنا كلمة فئة « الأرقام أو الرموز Numerals » بدلاً من الكلمة « أعداد Numbers » ، لأن الأعداد تشير إلى معانى الأرقام أو الرموز . فالعدد ٣٥ مثلاً يمكن أن يشير إلى رقم لاعب كرة ، أو رقم تليفون ، أو ترتيب طالب في صف دراسى معين ، أو درجة حرارة سائل ، أو طول فرد . وهذه المعانى تتضمن علاقات مختلفة بين الأرقام التي تعد بمثابة رموز ، والقياس يختص بهذه العلاقات .

أما فئة الخصائص أو الأحداث فتعلق بمحنوى المجال العلمي ، فعلم الكيمياء يتناول الغازات والسوائل ، والأجسام الجاسية من حيث الجزيئات والذرارات والإلكترونات والشحنات ، وعلم النفس يتناول السلوك الإنساني من حيث التعلم والتحصيل ، والميول والاستعدادات ، وهكذا في بقية العلوم .

واختلاف فئات الخصائص أو الأحداث يقابلها بالضرورة اختلاف في نوع القياس . وهذا يتطلب تحديد قواعد تعين الأرقام أو الرموز التي تناظر الخصائص أو الأحداث ، بحيث يمكننا التعامل مع الأعداد بطرق معينة للتوصيل من ذلك إلى استنتاجات نعيد تطبيقها على الظاهرة المقاسة .

فعندما نقول : إن طول محمد ١٧٠ سم ، فإننا نعلم تماماً القواعد التي على أساسها أجرى قياس هذا الطول ، فنحن نعلم أن هناك أداة لقياس الطول ذات تداريج موحدة ومتساوية تبدأ من الصفر؛ لذلك فإن العدد ١٧٥ سم الذي يناظر خاصة أو صفة الطول أصبح له مدلول معين متعارف عليه ، ولا يختلف في تفسيره أحد . فالصفر يعني انعدام الطول ، وهناك خصائص معينة للأعداد المدونة على المسطرة أو شريط القياس لها ما يناظرها في ظاهرة الطول المراد قياسه ، وكل من هذه الأعداد يتميز بالفرد ؛ لأن الرقم (٥) يختلف عن الرقم (٢) مثلاً ، والطول المناظر لكل من هذه الأرقام ، يتميز أيضاً بالفرد ، فالشيء الذي طوله ٥ سم يتميز بأنه أطول فعلاً من الشيء الذي طوله ٢ سم .

والأعداد المدونة على الشريط تميز بالترتيب ، وبذلك تؤدى إلى ترتيب الأطوال المناظرة لها ، كما أن هذه الأعداد قابلة للإضافة ، أي إذا جمعنا عددين أو أكثر يتبع عدداً آخر متميزاً ، وكذلك الأطوال المناظرة لأى أعداد يمكن جمعها لتصبح طولاً مساوياً لمجموع الأطوال المقاسة .



وبهذا الشكل تصبح الأعداد المدونة على الشريط المدرج الذى نقيس به ظاهرة الطول تمثل الأطوال تمثيلاً رياضياً له ما يناظره فى الواقع الفعلى . وبذلك يمكننا القول : إن هناك تماثلاً أو تنازلاً أحادى بين خصائص ظاهرة الطول والنظام العددى المستخدم على الشريط المدرج الذى يفترض أنه يقىس الطول .

القياس التربوى والنفسي :

إن تعريف المفهوم العلمى للقياس الذى أشرنا إليه ينطبق أيضاً على القياس التربوى والنفسي ، (إذ يمكن إعادة صياغة هذا التعريف لكي يناسب محتوى المجال التربوى والنفسي بحيث يصبح كالتالى :

ـ القياس : هو تعين فئة من الأرقام أو الرموز تناظر خصائص أو سمات الأفراد طبقاً لقواعد محددة تحديداً جيداً، وهذا يعني أن القياس التربوى والنفسي يعني بتكميم خصائص أو سمات الأفراد، حيث إننا لا نستطيع قياس الأفراد في ذاتهم، وإنما نقيس خصائصهم أو سماتهم).

فلكى نقيس خاصية أو سمة معينة لدى فرد وفقاً لهذا التعريف، فإننا نحتاج إلى تحديد الفرد المراد قياس سمة معينة لديه، وتحديد السمة المراد قياسها، والإجراءات التي سوف تتبع لتعيين أرقام أو رموز Numerals تناظر السمة المقاسة . والأرقام الناتجة التي تكون عادة قيمًا عددية أو درجات Scores ينبغي - كما ذكرنا - أن تحافظ على العلاقات الفعلية القائمة بين مستويات السمة المقاسة لدى الأفراد .

وهذا يعني أنه إذا كان فرد ما يتميز بقدرة لفظية أعلى من فرد آخر، فإن درجة الفرد الأول فى اختبار يقىس هذه القدرة ينبغي أن تكون أعلى من درجة الفرد الثانى فى الاختبار نفسه، أي أن الدرجات أو القيم العددية تقع على متصل تقاس عليه السمة . والهدف من عملية القياس تحديد موقع الفرد والسمة معًا على هذا المتصل ، وهذا لابد أن يعتمد على إجراءات كافية ودقيقة بحيث يمكن اعتبار الدرجة التى يحصل عليها فرد فى اختبار، وهى مجموع النقاط التى ينالها فى أسئلة الاختبار، تكميم أو قياس لقدراته الفعلية مما يتطلب تقييمًا تجريبياً . وسوف نتناول ذلك باستفاضة بعد قليل ، وكذلك فى الفصول اللاحقة .

طبيعة القياس التربوى والنفسي :

لكل توضح طبيعة القياس التربوى والنفسي ربما يكون من المفيد إلقاء مزيد من الضوء على القياس فى العلوم الطبيعية، فالقياس الفيزيائى كما أوضحتنا يمكن إجراؤه بطرق مباشرة، حيث إن هناك عمليات وقواعد مقتنة متقد علىها لذلك ، كما أن هناك موازين Scales معرفة تعرضاً جيداً يتم على أساسها تفسير نتائج القياس .

ف عند قياس طول منضدة مثلاً ما علينا إلا أن نأتي بالمسطرة المدرجة ، ونجعل إحدى حافتي المنضدة مقابلة للنقطة الم対اظرة لصفر التدرج ، ونقرأ الرقم المقابل للحافة الأخرى ، وبهذه العمليات البسيطة نستطيع معرفة طول المنضدة بإحدى الوحدات الخطية المعيارية لقياس الطول مثل المستيمتر أو البوصة .

غير أن هناك ظواهر فيزيائية أكثر تعقيداً ، من مثل السمع والبصر وما شابههما، تتطلب إجراءات تفصيلية للقواعد أو الطرق التي تتبع في قياسها . كما أن هناك ظواهر أخرى يصعب قياسها قياساً مباشرأً مثل الصوت والضوء والحرارة والكهرباء وغيرها نظراً لأنها غير مرئية . ولكن هل لابد من رؤية الظاهرة لكي نتمكن من قياسها؟ بالطبع لا ، فعلماء الفيزياء استطاعوا قياس هذه الظواهر الطبيعية وغيرها ، على الرغم من أنها غير مرئية أو محسوسة ، بل إنهم تمكنوا من قياسها بدقة متناهية بواسطة أدوات قياس متعددة ، فنحن لا نقيس الظاهرة في حد ذاتها ، وإنما نقيس أثرها في الأجسام أو الأشياء بواسطة أدوات قياس أعدت لهذا الغرض . فالصوت مثلاً يحدث اهتزازات وذبذبات في الأثير على شكل موجات ، وتختلف شدة وتكرار الموجات باختلاف نوع الصوت الصادر من مصدر معين ، وأداة القياس أو الجهاز المناسب يقيس هذه التأثيرات ، ولا يقيس الصوت في حد ذاته ، ولكننا نستخدم المقدار الكمي الناتج عن قياس الأثر في الاستدلال على ظاهرة الصوت .

وقد حذا علماء النفس والتربية حذو العلوم الطبيعية ، بل إنهم أكثر اهتماماً بمنطق القياس والأساليب الكمية من علماء العلوم الطبيعية ، ويرجع سبب ذلك إلى أن الظواهر السلوكية ظواهر تتميز بالتعقد والتشابك وتعدد المتغيرات ، وليس ببساطة الظواهر الفيزيائية ، فعالم التاريخ الطبيعي يستطيع أن يثبت زهرة ويقوم بدراستها ، وعالم الكيمياء يمكنه أن يتعامل مع المواد الكيميائية كما يشاء ، وكذلك عالم الفيزياء .

أما عالم النفس فمن الصعب عليه إخضاع الإنسان لتجاربه المعملية ، وثبتت جميع المتغيرات التي يود عزل أثراها .

كما أن كثيراً من الظواهر السلوكية يصعب قياسها على موازین مدرجة تدريجاً معيارياً متفق عليه كما هو الحال في موازین العلوم الطبيعية؛ إذ من الصعب قياس سمات شخصية الفرد ، مثل العدوانية أو الانطوائية ، أو الإقدام أو التسلطية ، أو العاطفة ، بنفس القدر من الدقة كما في حالة قياس درجة حرارة سائل أو شدة تيار كهربائي .

إلا أن هذا لم يقف حائلا دون استمرار علماء القياس التربوي والنفسى فى مواجهة هذه العقبات، وذلك باتباع خطوات مماثلة للخطوات المتتبعة فى القياس الفизيائى، وابتکار طرق وأساليب وأدوات قياس متعددة تميز بخصائص محددة سوف نناقشها فيما بعد.

ولتوضیح ذلك نفترض أننا أردنا قياس «معدل القراءة» لدى تلاميذ أحد صفوف مرحلة التعليم الأساسي، فهنا يجب البدء بتعريف المقصود من الخاصة المراد قياسها ، وهى في هذه الحالة «معدل القراءة».

إذ ربما نُعرّفها بأنها السرعة في القراءة ، أي عدد الكلمات التي يقرأها التلميذ في الدقيقة. وتصبح المشكلة الآن هي تحديد مجموعة من الإجراءات أو القواعد لقياس سرعة القراءة.

فربما نلجم إلى أبسط الطرق وهى أن نطلب من كل تلميذ أن يقرأ قطعة نثرية معينة في مدة زمنية محددة، ثم نوجد عدد الكلمات التي أمكنه قراءتها في هذه المدة، ونقسم هذا العدد على الزمن المستغرق لحصول بذلك على قياس لعدد الكلمات المقروءة في الدقيقة، وبهذا تكون قد استخدمنا أسلوباً مماثلاً لأسلوب قياس الطول الذي أشرنا إليه.

ولكنا هنا يجب أن نأخذ بعين الاعتبار مجموعة أخرى من العوامل من بينها صعوبة المقرء، إذ إن هذه الصعوبة تؤثر بلا شك في معدل القراءة؛ ولذلك فإن عدم معرفتنا درجة صعوبة المقرء يجعل تفسير نتيجة قياس معدل القراءة غير ذي معنى.

والعامل الثاني هو المدة الزمنية المطلوبة للقراءة، فإذا كانت هذه المدة طويلة سوف يشعر التلميذ بالتعب أو الملل ويقل بذلك معدل قراءته. والعامل الثالث هو شروط الموقف الاختباري، فمثلاً إذا كانت الحجرة غير جيدة الإضاءة أو التهوية أو تحيط بها الضوضاء، فإننا نتوقع أن ينخفض معدل القراءة.

كذلك يجب أن نحدد تحديداً دقيقاً المقصود بكلمة «القراءة». وخلاصة ذلك أنه على الرغم من أوجه التشابه بين عمليات وإجراءات القياس الفيزيائى والقياس التربوى والنفسى، إلا أنه يوجد عدد أكبر من المتغيرات التي تؤثر في القياس التربوى والنفسى يجب ضبطها أو التحكم فيها إذا أردنا أن نجري عملية القياس بصورة دقيقة.

أنواع القياس ومستوياته :

تبين لنا أن قياس السمات الإنسانية يتطلب تعريفاً دقيقاً لهذه السمات في ضوء المتغيرات المتعلقة بها والعمليات التي نجريها للحصول على قيم لهذه المتغيرات تكون قابلة للتفسير.

ومما لا شك فيه أن الصياغة الإجرائية لتعريف المتغير ليس أمر هيناً، حيث تتطلب مراجعة فاحصة للأطر النظرية والدراسات والبحوث المتعلقة بهذا المتغير، فتعريف المتغير تعريفاً دقيقاً مرتبط ارتباطاً وثيقاً بالعمليات والإجراءات التي تبعها في قياسه ، وفي كثير من الأحيان نضطر إلى إعادة النظر في تعريف المتغير في ضوء نتائج هذه العمليات .

فالمتغيرات السيكولوجية والتربيوية تعد بمثابة تكوينات فرضية Constructs تحتاج بلورتها إلى أدلة متعددة ودراسات متتجدة وفقاً لما يطرأ على هذه التكوينات من تغيرات نتيجة تطور المعرفة. فمتغير «التكيف Adjustment» مثلاً من بهذه المراحل التطورية في تعريفه، فقد كان يستخدم استخدامات متعددة وفقاً لتعريفات متباعدة ، مما أدى إلى إرباك ملحوظ لدى الباحثين والممارسين ، وقد نتج عن ذلك ضم عدد من السمات أو المتغيرات المنفصلة لتشكل في مجموعة مفهوم التكيف مما تطلب تبريرات فكرية ومنطقية بل وإحصائية لهذا الإجراء .

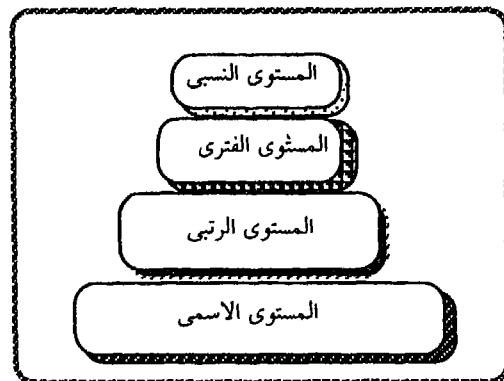
كما أن متغير «التصلب Rigidity» مثلاً لدى الأفراد تبانت بل وتناقضت نتائج البحوث التي اهتمت بدراسته مما أدى إلى مزيد من الغموض حول طبيعة هذا المتغير ، وما إذا كان له وجود فعلى .

لذلك فإن من الضروري العناية بتعريف المتغير المراد قياسه ليكون أساساً لتصميم سلسلة من القواعد والإجراءات التي تكشف عن الخاصية أو السمة المتعلقة به وقياسها.

ومن الجدير بالذكر أن المتغيرات النفسية والتربوية إما أن تكون كيفية أو كمية ، فالمتغير الكيفي أو النوعي تناسبه عمليات وإجراءات التصنيف، ويطلق عليه حينئذ المتغير الاسمي، أما المتغير الكمي فتناسبه عمليات وإجراءات تسمح بالحصول على ترتيب تصاعدي أو تنازلي لقيمه، وعندئذ يطلق عليه المتغير الرتبي، أو تسمح بالحصول على قياسات كمية له، ويطلق عليه المتغير الفترى أو النسبي.

لذلك تباين أنواع القياس ومستوياته وفقاً لنوع المتغير وطبيعته والهدف من عملية القياس ، فلكى نجرى عملية القياس بالدقة المطلوبة يجب أن نراعى مستوى قياس المتغير. و تستند كل من هذه المستويات إلى فرض رياضية ومنطقية ، وتتخذ تنظيماً

هرميًّا تحقق فيه المستويات الأعلى فروض جميع المستويات الأدنى بالإضافة إلى خصائص تميز المستوى الأعلى ، وهذه المستويات موضحة بالشكل التخطيطي التالي :



شكل (١-١) يوضح المستويات الهرمية للقياس

ويتضح من شكل (١-١) أن هناك أربعة مستويات للقياس وفقاً لطبيعة المتغيرات المراد قياسها، وهذه المستويات هي :

المستوى الاسمي : Nominal Scale

وهو أدنى مستويات القياس ويناسب المتغيرات الكيفية أو النوعية التي تتطلب تصنیف الأفراد إلى مجموعات متضمنة للتمیز بينهم في سمة معينة، ويكون الهدف من عملية القياس في هذه الحالة هو التصنيف Classification الذي يراعي الفروق النوعية بين الأفراد. والأعداد المستخدمة في هذا المستوى من القياس تعد بمثابة رموز بسيطة تستخدّم كأسماء لفئات أو مجموعات متضمنة ومتمازية. ومن أمثلة متغيرات هذا المستوى : النوع، والجنسية، والديانة، والحالة الاجتماعية، والانتماء إلى مؤسسات معينة، وهكذا.

وهذه الأعداد لا نستطيع إجراء عمليات حسابية عليها بحيث تكون ذات معنى، فلا معنى لأن نجمع رقم مناظر لنوع معين على رقم مناظر لنوع آخر مثل الذكور والإناث، أو نطرح رقم مناظر لجنسية معينة من الرقم المناظر لجنسية أخرى، ومع هذا يمكن التعامل مع البيانات الناتجة عن القياس الاسمي بأساليب إحصائية مناسبة، سوف نشير إلى البعض منها في الفصل الثالث.

المستوى الارقمي : Ordinal Scale

وفي هذا المستوى يمكن ترتيب الأفراد أو الأشياء تبعاً لخاصية أو سمة معينة. فمثلاً يمكن أن نطلب من الأفراد ترتيب ميلولهم المهنية، إذ ربما يعطي فرد مهنة مبرمج

الحاسب الآلي الرقم ١ ، ويعطى مهنة مهندس الرقم ٢ ، وفني سيارات الرقم ٣ وهكذا ، وهذا يعني أن هذا الفرد قد أجرى نوع من القياس الرتبى لميوله المهنية وفقاً لمحكمات معينة استند إليها في هذا الشأن .

ونظراً لسهولة ويسر هذا النوع من القياس ، فإنه يستخدم بكثرة في القياس التربوى والنفسى وبخاصة في قياس السمات الوجدانية التي تتعلق بالاتجاهات والتفضيلات والأراء والتقديرات . ولكن ينبغي ملاحظة أنه ليس من الضرورى معرفة مقادير أو درجات السمات المقاسة لكي يمكننا إجراء هذا الترتيب ، بل يكفى معرفة ترتيب كل منها بالنسبة للأخر . كما أن القياس الرتبى لا يفترض أن الفروق بين الرتب تساوى الفروق بين درجات السمة المقاسة .

وهذه الأعداد التي تمثل الرتب لا تسمح بإجراء عمليات حسابية عليها شأنها في ذلك شأن القياس الاسمي ، وإنما نكتفى بالقول بأن محمدًا أذكى من أشرف ، أو تحصيل محمد أعلى من تحصيل أشرف ، أو أن محمدًا لديه اتجاه نحو المدرسة أكثر إيجابية من أشرف ، وهكذا .

المستوى الفتري : Interval Scale

وهذا المستوى من القياس يتعلق بتحديد مقدار الفرق بين شيئين ، والمثال الشائع لذلك هو الترمومتر ، فبعض الترمومترات تكون تداريجهها درجات فهرنهايتية ، حيث يمثل العدد ٣٢ درجة فهرنهايتية درجة التجمد ، والعدد ٢١٢ درجة فهرنهايتية درجة الغليان ، والأعداد المحصورة بينهما تشير اعتبارياً إلى تمدد أو تقلص معين في عمود الزئبق أو السائل في الترمومتر ، فتغير قدر معين من هذا السائل ولتكن ١ ، ٢ ملليمتر تعنى تغير في الحرارة مقداره درجة واحدة ، بغض النظر عما إذا كان هذا الفرق بين الدرجتين ٥ ، ٦ أو بين ٦٢ ، ٦٣ درجة فهرنهايتية .

فالفارق المتساوية على طول التدرج تمثل فروقاً متساوية في حركة السائل داخل الترمومتر .

وهنا ينبغي ملاحظة أن الصفر الحقيقي الذى عنده تنعدم الحرارة لا وجود له على مثل هذه الترمومترات ، فالترمومتر يقيس فقط الفرق الحقيقي بين أى درجتين من درجات الحرارة .

وهذا المستوى من القياس يستخدم كثيراً في القياس التربوى والنفسى ، فنحن لا نقيس ذكاء الفرد أو سمات شخصيته أو ميوله قياساً مطلقاً ، وإنما نقيس الفرق الحقيقي بين ذكاء شخصين مثلاً ، أو نوازن بين ذكائه وذكاء زملائه الذين يطبق عليهم نفس

الاختبار ، وتمثل الفروق المتساوية على متصل تدريج السمة (الذكاء) فروقاً متساوية في الذكاء بالفعل .

ولكن لا يوجد صفر مطلق للذكاء أو لأى سمة أخرى من سمات شخصية الإنسان ، فالسمات لا تندم لدى الأفراد .

المستوى النسبي : Ratio Scale

يتميز هذا المستوى من القياس بالخصائص التي تتوافر في المستوى الفترى بالإضافة إلى وجود صفر مطلق على ميزان القياس يناظر بالفعل انعدام المعاشرة أو السمة المقاسة ، ومثال ذلك الأطوال والكتل وغير ذلك . ونظرًا لأنه يمكن إجراء جميع العمليات الحسابية الأساسية في هذا المستوى فإنه يعى أعلى المستويات السابقة . ويستخدم هذا النوع من القياس عادة في العلوم الطبيعية ، ويندر استخدامه في العلوم السلوكية للأسباب السابقة ، فمعظم الاختبارات والمقاييس التربوية والنفسية تؤدى عادة إلى قياسات فترية ، ولكن ينبغي مراعاة العمليات والإجراءات التي سبق أن أشرنا إليها .

مدخل السمات في القياس التربوي والنفسى :

ذكرنا فيما سبق أن تحديد السمات المراد قياسها وتعريفها تعد من الخطوات الأساسية في القياس . فالقياس التربوي والنفسى لا يهدف لقياس الفرد ، وإنما قياس خصائصه أو سماته أو مظاهر معين من مظاهره كالذكاء أو دافعية الإنجاز ، أو الفلق مثلًا . لذلك فإن مفهوم السمة Trait يعد من المفاهيم التي تحتاج إلى مزيد من التوضيح . فقد وجد بعض الباحثين أن هناك ما يزيد عن ٤٠٠٠ كلمة تصف الأفراد . وبالطبع يصعب استخدام هذا القدر من الكلمات أو التعامل معها في القياس التربوي أو النفسي ؛ لذلك حاول علماء النفس اختزال هذا الكم الهائل من الأوصاف إلى مجموعة أقل من السمات في إطار ما يعرف بنظرية السمات Trait Theory .

والسمة مفهوم يستخدم لوصف سلوك الأفراد ، وهي عبارة عن تجمع من السلوك المترابط ، أي الذي يحتمل حدوثه معاً . فالسمة بهذا المعنى ليست صفة منفردة ، وإنما تعد مفهوماً مجرداً ، وليس شيئاً ملموساً ، ويمكن أن تكون السمة عامة مثل سمة «الميل إلى المشاركة في مواقف اجتماعية متعددة» ، أو تكون محدودة مثل سمة «بذوق نوع معين من الطعام» .

السمات لا يمكن قياسها قياساً مباشراً ، وإنما يستدل عليها من أنماط السلوك الملاحظ ، وهذا هو الحال دائمًا في القياس التربوي والنفسى . فعندما يود عالم النفس

قياس إحدى السمات العقلية، ولتكن الذكاء، فإنه عندما يلاحظ طفل معين لا يرى ذكاؤه لأن الذكاء مفهوم مجرد، ولكنه يلاحظ سلوك الطفل في مواقف مختلفة تتطلب الذكاء، أي أن السمات العقلية تعد بمتابة خصائص يفترض أنها تعكس مجموعة مترابطة من السلوك الذي يمكن ملاحظته وتسجيله في مواقف اختبارية مقتنة . ويجب أن تفي السمات ببعض الشروط لكي يتسمى الإفادة منها في وصف السلوك، ومن أهمها:

(١) أن تمثل السمة خاصة محددة يتباين فيها الأفراد مثل السمات العقلية والشخصية والوجودانية .

(٢) أن تكون السمة قابلة للتعريف الإجرائي بحيث يكون هناك قدر معقول من الاتفاق بين القائمين بـ ملاحظة السلوك حول مقدار السمة لدى الأفراد المختلفين ، فمثلاً ما الذي نعنيه بـ «الذكاء» ؟ وما نوع السلوك الذي يميز الشخص الذكي أو يسميه؟ وهل نقصد بالذكاء تناول الأفكار والمفاهيم المجردة ، أم التعامل مع المواقف التي تتميز بالجدة ، أم استجابات الفرد في مواقف مألوفة ، أم سرعة الاستجابة وطلاقتها ، أم إلى مستوى تعقد الاستجابة بغض النظر عن زمنها؟

لذلك ينبغي التوصل إلى تعريف واضح ودقيق ومحبوب للسمة المراد قياسها ، كما ينبغي أن تتضمن الجوانب المهمة التي تؤدي التركيز عليها من حيث فائدتها العملية أو التطبيقية .

(٣) أن تميز السمة بالثبات النسبي بمرور الزمن وباختلاف الموقف ، فسلوك الفرد يصعب التنبؤ به من موقف إلى آخر ، ومن وقت إلى آخر مالم يكن متميزة بالاتساق . ومن الجدير بالذكر أن السمات الإنسانية تتباين في مدى ثباتها واتساقها ، فالسمات العقلية تميز بالثبات النسبي بدرجة أكبر من السمات الشخصية أو الوجودانية .

بعض الخصائص الأساسية للسمات :

أشار كيلي (Kelly , 1974) ، وثورنديك (Thorndike , 1974) ، ونانلى (Nunnally , 1998) إلى مجموعة من الخصائص الأساسية للسمات التي تيسر عمليات قياس السلوك الذي يستدل منه على سمة معينة ، وفيما يلى هذه الخصائص :

(١) يمكن أن تمثل السمة بخط متصل : إن الأخذ بمفهوم السمة الذي أشرنا إليه يجعلنا نفترض أن هناك متصلةً من السلوك نحاول عن طريق عملية القياس أن نحدد موقع الفرد عليه في سمة معينة لديه . ونقصد بالتسلسل Continuum خط مستقيم

يتكون من عدد لا نهائي من النقط الممكنة التي تحدد مواقع مختلفة للسمة المقاسة . فالذكاء مثلاً ليس سمة ثنائية القطب ، أى تمثل بقسمين منفصلين هما الذكى والغنى ، وإنما تمثل بمقادير متصلة بحيث يمكن تبع سمة الذكاء في مستويات مختلفة متدرجة يتمايز فيها الأفراد بحسب درجاتهم في هذه السمة .

(٢) يمكن تدريج متصل السمة بوحدات مناسبة: إن متصل السمة يمكن

تدريجه بحيث يصبح بمثابة ميزان Scale ، ويعتمد هذا التدريج على طبيعة السمة المراد قياسها ، والفرض الذى نستند إليها فى قياسها ، وتعريفنا لوحدات الميزان . إذ ربما نقتصر على الميزان الاسمي الذى يصنف الأفراد إلى مجموعات منفصلة فى سمة معينة مثل انطوائى وابساطى ، وربما نستخدم الميزان الرتبى الذى يصنف الأفراد مع مراعاة الترتيب مثل عبقرى ، وذكى جداً ، فوق المتوسط ، ومتوسط الذكاء ، ودون المتوسط ، أو ربما يفضل الميزان الفترى الذى يراعى تساوى الفروق بين الأقسام المتتالية على متصل السمة المقاسة ، أى تساوى الوحدات .

(٣) يمكن تصوّر متصل السمات على أنه ميزان قياس فترى : فعلى الرغم

من تعدد موازين القياس ، إلا أنه يمكن جعل ميزان كثير من السمات النفسية من المستوى الفترى . فالقياس فى واقع الأمر - كما سبق أن ذكرنا - هو محاولة لإيجاد تناظر بين الأعداد وبين خصائص الظاهرة موضع القياس ، فكلما كان التناظر كبيراً دل ذلك على أن القياس مناسب . وبؤدى ذلك إلى اعتبار مهم هو ما إذا كان مستوى معين من مستويات القياس يناظر بدرجة أكبر واقع الظاهرة عن غيره من المستويات . وينصب عادة اهتمامنا في القياس على ما إذا كان المستوى الفترى يمثل واقع الظواهر النفسية والتربية تمثيلاً مناسباً وكافياً .

ولتوسيح ذلك نفترض أننا نود قياس القدرة الحسابية لدى الطلاب ، فإذا طرق قياس هذه القدرة هو أن نقوم ببناء اختبار جيد يشتمل على العمليات والمهارات الحسابية ، وتكون الدرجة التى يحصل عليها كل طالب بمثابة مؤشر لقدرته الحسابية . فإذا افترضنا أن الاختبار يشتمل على ٥ مفردة تختلف فى صعوبتها ، فهل يمكننا اعتبار أن فرق خمس درجات على جزء من متصل ميزان الدرجات يكفى فرق خمس درجات على أي جزء آخر من الميزان ، كما هو الحال فى درجات الحرارة مثلاً؟ أى هل الفرق بين الدرجتين ٥ ، ١٠ مثلاً على ميزان الدرجات يكفى الفرق بين ٤٥ ، ٥٠ على الميزان نفسه؟

وهنا ربما نفترض أن كل نقطة أو درجة (أى كل مفردة أجاب عنها الطالب إجابة صحيحة) تساوى أى نقطة أو درجة أخرى يحصل عليها فى الاختبار، وبذلك نعتبر الفرق بين الدرجتين $5 - 10$ مكافئاً للفرق بين الدرجتين $45 - 50$ فى مقدار السمة المقاسة وهى القدرة الحسابية. ولكن إذا تمعنا قليلاً ربما نستنتج أن الفرق بين الدرجتين $45 - 50$ أكبر من الفرق بين الدرجتين $10 - 15$ ، إذ إنه لکى يرفع الطالب درجته من 45 إلى 50 ربما يتطلب الإجابة عن خمس مفردات أخرى أكثر صعوبة، بينما يتطلب الأمر الإجابة عن خمس مفردات أقل نسبياً في صعوبتها إذا أراد أن يرفع درجته من 10 إلى 15 ، فإذا قبلنا هذا التبرير كان معنى ذلك أن ميزان الدرجات ليس من المستوى الفترى.

ويرى كرونباك Cronbach أن الاختلاف بين الافتراضين أو التبريرين يعتمد على الأسلوب الذى تتبعه فى تفسير درجات الاختبار وتبيننا لهذا التفسير. فالتبرير الأول المتعلق بتساوي الفروق بين الدرجات يمكن اعتباره تبريراً منطقياً، أما التبرير الثانى فهو تبرير سيكولوجى حيث يأخذ بعين الاعتبار واقع السمة موضع القياس، أى يهتم بمدى تأثير التحسن فى الأداء باختلاف مدى الدرجات وفقاً لصعوبة أو سهولة مفردات الاختبار، حيث إن الصعوبة أو السهولة لها بالطبع جانب نفسي.

وهنا ربما يتساءل القارئ: كيف إذن يمكن أن نحصل على ميزان فترى أو نسبي فى القياس التربوى والنفسى؟ وللإجابة عن هذا التساؤل ينبغى مراعاة متطلبات الميزان النسبي التى أشرنا إليها من قبل، وهى الفترات المتساوية وتوافر صفر مطلق يناظر انعدام السمة. ونظراً لأنه يصعب تصور إنسان ذكاؤه أو قلقه أو تحصيله صفرًا، فإنه لا يوجد صفر مطلق على ميزان قياس السمات الإنسانية، ولكن يمكن الاستعاضة عنه بصفر اعتبارى مثل درجة الحرارة صفر. وهذا الصفر لا يناظر انعدام السمة، وإنما يمكن اعتباره نقطة تقع فى متصف الخط المتصل الذى يمثل السمة، أى أن الصفر يناظر الموقع المحايد على ميزان قياس هذه السمة، كما أن عدد الوحدات أو التداريج التى يقسم إليها الميزان تكون اعتبارية أيضاً، ويمكن جعلها متساوية باستخدام بعض الأساليب الإحصائية التى سنوضحها فى الفصل الثالث.

وهذا الإجراء يمكننا من قياس السمات الإنسانية على ميزان فترى بدلاً من الاقتصار على الميزان الرتبى المميز للدرجات الاختبارات فى كثير من الأحيان. لذلك يهتم علماء القياس بالحصول على ميزان فترى فى المقاييس التربوية والنفسية، وذلك لإمكانية تحويل الدرجات الناتجة عن هذه المقاييس تحويلاً خطياً يسمح بإضافة مقدار

ثابت إلى الدرجات أو طرحة منها دون أن تتأثر العلاقة بين الدرجات، وبهذا نستطيع تحويل الدرجات من ميزان ذي وحدات معينة إلى ميزان له وحدات مختلفة، أي يمكن تحويل الدرجات إلى ميزان مشترك للموازنة بين أداء الأفراد على مقاييس مختلفة. ولعل درجات الحرارة تعد مثلاً جيداً لذلك، فعلى الرغم من أنها نستطيع قياس الحرارة بكل من وحدات الدرجات المئوية، والدرجات الفهرنهايتية، إلا أنه يمكن تحويل كل من الوحدتين إلى الأخرى ووضعهما على ميزان مشترك.

أنواع السمات النفسية والتربوية :

نظرًا لتنوع وتعدد السمات النفسية والتربوية، فقد بذلت محاولات كثيرة لتصنيف هذه السمات لتسهيل دراستها وقياسها والإفادة منها في فهم وتفسير السلوك الإنساني، إلا أن هذه المحاولات لم تتوصّل إلى اتفاق عام في هذا الشأن. فالتساؤل الذي يتردد بين الباحثين والممارسين في مختلف مجالات القياس والتقويم هو : ماذا نقيس؟

وعلى الرغم من أهمية وضرورة هذا التساؤل، إلا أن الإجابة عنه تتطلب معرفة الهدف من القياس، وطبيعة البحث أو الدراسة، ونوع القرارات العملية التي يجب اتخاذها فيما يتعلق بالأفراد الذين نود قياس بعض سماتهم سواء في المدارس أو الجامعات أو المستشفيات أو في مجال الصناعة أو المؤسسات العسكرية.

فنحن بالطبع لا نستطيع قياس الفرد ككل، وإنما نهتم بقياس بعض سماته التي تسهم في اتخاذ قرار معين بشأنه يتعلق بالانتقاء أو التصنيف أو التسكين أو التشخيص مثلاً. وقد ميز كيلي (Kelly, 1974) بين عدة مجموعات من السمات التي تهم الباحثين والمشتغلين بالقياس والتقويم التربوي والنفسي، بعضها يتعلق بالجوانب الجسمية العضوية والفيزيولوجية، وبعض الآخر يتعلق بالجوانب النفسية العقلية والوجدانية والمزاجية والمهارية كما يلى :

(1) **سمات تتعلق بالشكل الخارجي لجسم الفرد :** وهذه السمات يغلب عليها الطابع الجسمى العضوى وليس الطابع النفسي، وإن كانا أحياناً مترابطين. فالسمات التي تميز جسم الفرد مثل الطول والوزن وحجم الكفين والقدمين وشكل الصدر مثلاً ربما ترتبط بجوانب أخرى في الشخصية وبخاصة مفهوم الفرد عن ذاته. ومع هذا يرى بعض علماء النفس أن هذه السمات لا تتنسى إلى مجال الشخصية الإنسانية، ولكنها بلاشك خصائص يتباين فيها الأفراد تبايناً ملحوظاً.

(٢) **سمات فسيولوجية** : وتعلق بالسلوك الداخلي لأعضاء الجسم الحيوية، مثل ضربات القلب، وضغط الدم، ودرجة حرارة الجسم، ونشاط الغدد الدرقية، وغير ذلك. ويتبين الأفراد أيضاً تبايناً ملحوظاً في هذه السمات. وترجع أهمية هذه السمات إلى أنها تؤثر في الجوانب النفسية للفرد وتتأثر بها. فالغيرات في القياسات الفسيولوجية تبايناً كبيراً تبايناً كبيراً تبايناً الموافق الحياتية والضغوط الفيزيائية والنفسية؛ لذلك تستخدم مقاييس مختلفة لأنشطة الفسيولوجية في تقييم الشخصية.

(٣) **الاستعدادات** : وهي القدرة الكامنة لدى الأفراد أو قابلاتهم لأداء عمل معين. ويتبين الأفراد تبايناً كبيراً في هذه الاستعدادات حتى قبل أن تتاح لهم الفرصة للتحصيل أو اكتساب مهارات معينة، ونظراً لاتساع مجال هذه المهارات وما تتطلبها من استعدادات مختلفة، فإنه يصعب حصر هذه الاستعدادات حصرًا دقيقاً، ومن بين هذه الاستعدادات :

الاستعداد الدراسي : ويتصل بالقدرات الضرورية للنجاح في المدرسة وبخاصة القدرات اللغوية والقدرات الرياضية.

الاستعداد المكاني : ويتصل بالقدرة على تصور الأشكال ودورانها في المستوى أو في الفراغ.

الاستعداد الميكانيكي : وهو قدرة الفرد على اكتساب المعرفات والمهارات المتعلقة بفك وتركيب أجزاء الأشياء، والمهارات اليدوية في التعامل مع الأجزاء الدقيقة المكونة للأدوات المختلفة.

الاستعداد لأداء الرياضيات : وهو القدرة على أداء لعبة رياضية معينة أو أكثر بدقة ومهارة، ويطلب ذلك بنية جسمية قوية، وقدرة إدراكية، وقوة عضلية وتناسق.

الاستعداد الموسيقي : وهو القدرة على تعلم العزف على آلة موسيقية أو أكثر بيسر وسهولة، وكذلك مهارة تأليف الألحان.

الاستعدادات الفنية الأخرى : وترتبط بالقدرة على الإنتاج الفني المبدع، مثل الغناء والتمثيل، وتلوّق إنتاج الآخرين.

الاستعداد الكتابي : وهو القدرة على تعلم المهارات الكتابية، مثل استخدام الحاسوب ، والكتابة على الآلة الكاتبة ، والاحتزال ، وأعمال الأرشيف أو حفظ المستندات والوثائق.

(٤) **المهارات والتحصيل** : وهذه تتعلق بالأداء الفعلى للأفراد سواء في المدرسة أو في موقع العمل المختلفة التي تتطلب مهارات معينة. وكما أن هناك تبايناً بين الأفراد في استعداداتهم وإمكاناتهم، كذلك يوجد تباين بينهم في المهارات التي تتطلب

توفر قدرًا معيناً من الاستعدادات المتعلقة بهذه المهارات. وعلى الرغم من أن هذه الاستعدادات تستخدم كمبنيات للسلوك المستقبلي، ويسهل اكتساب المعرف والمهارات اللازمية لأداء عمل معين، إلا أن الأداء الفعلى مشروط أيضًا بالدافعية والمحافز والممارسة والدراسة، أى أن الفروق بين الأفراد في الأداء والتحصيل يمكن أن تعزى إلى الأثر المشترك لكل من الاستعدادات، وهذه المتغيرات الأخرى.

(٥) **الميول**: تتعلق بالجوانب الوجدانية لشخصية الفرد، وينعكس ميله في تعبيره عما يحب ويكره، والأنشطة التي يفضل أن يبذل فيها جهداً وقتاً، مثل قراءة الكتب العلمية، والعنابة بالنباتات، وسماع الموسيقى، وغير ذلك.

(٦) **الاتجاهات**: تعد أيضًا من الجوانب الوجدانية، وتتمثل في آراء الفرد وتفضيلاته فيما يتعلق بجماعات معينة أو مؤسسات أو ممارسات أو سياسات أو مفاهيم، وما شابه ذلك. وينعكس اتجاه الفرد نحو القضايا المختلفة في تعبيره وسلوكه، ومن أمثلة ذلك: الاتجاه نحو حقوق الإنسان، أو تنظيم الأسرة، أو نظام الضرائب، أو أماكن العبادة، وهكذا.

(٧) **القيم**: وهي أيضًا من الجوانب الوجدانية، ولكنها أكثر تأصيلاً وعمقاً من الميول والاتجاهات، وتتعلق بفئات واسعة من الخبرات والمعتقدات والأهداف والسلوكيات التي تميز بالاستغراق والولاء، فالفنان لديه قيم جمالية، ورجل الدين لديه قيم دينية، ورجل الأعمال لديه قيم اقتصادية، وهكذا.

(٨) **السمات المزاجية**: وهي تعد من سمات الشخصية التي نالت قدرًا كبيراً من اهتمام علماء النفس؛ نظراً لأنها من السمات التي تسهم في تكيف الفرد وتفاعلاته مع الآخرين، وهي تعبر عن نزعات الفرد أو طباعه مثل التفاؤل، وحدة المزاج، والثقة بالذات، والاندفاع، والانبساط، والانضباط، والإحساس بالأمان، وغير ذلك من النزعات التي تميز الفرد في مواقف المشاركة الاجتماعية والموافق الانفعالية وغير الانفعالية.

والخلاصة أن هذه المجموعات من السمات يتباين الأفراد فيها تبايناً ملحوظاً. وعلى الرغم من أنها لا تعد المجموعات الوحيدة، إلا أنها تفيد في قياس السلوك الإنساني وتقييمه. وبالطبع يصعب قياسها جميعاً لدى فرد واحد، ولكن ينبغي انتقاء السمات المتعلقة بالهدف من القياس، وبالاستخدامات المناسبة لنتائج القياس. فلكل فرد استعداداته وميوله، وخصائص الشخصية التي تؤهل له للقيام بدور فاعل في مجتمعه، وتحديداً لهذه السمات وقياسها لديه يمكننا من الإفاده الفصوى من إمكانات كل فرد وتنظيم دور كل منهم في الحياة بما يتفق وهذه الإمكانيات.

العمليات القياس وأدوات القياس :

ينبغي أن تفرق بين عمليات القياس التي أوضحتها فيما سبق وأدوات القياس. أي الاختبارات والمقاييس. فأدوات القياس تستخدم في جمع بيانات بطريقة منظمة تقييد في صنع القرارات المختلفة، وهذه الأدوات تشتمل على مطلب أو مجموعة من المطالب أو المهام أو المثيرات التي ترتبط بسمة نفسية أو تربوية.

وعادة تتطلب أدوات القياس من الفرد أن يستجيب لهذا المطلب أو المطلب (مفردات أو أسئلة أو فقرات) بحيث يمكن الاستدلال من استجابته على مقدار السمة المراد قياسها. وبالطبع لا تقتصر أدوات القياس على الاختبارات التي تقيس الجوانب العقلية والتحصيلية، وإنما تشمل الاستبيانات وقوائم الملاحظة والتقارير الذاتية والطرق الإسقاطية وغيرها من أدوات قياس مجموعات السمات السابقة. وتقييد هذه الأدوات في جمع بيانات من مصادر متعددة للاستفادة بها في عمليات التقويم. وسوف نلقى الضوء فيما يلى على مفهوم الاختبارات، ومفهومي التقويم والتقييم وعلاقتها بمفهوم القياس لكي يتسعى توظيف هذه المفاهيم توظيفاً مناسباً.

الاختبار التربوي والنفسي :

إن تعدد السمات واختلاف طبيعتها وتباعين أنواعها يتطلب أن تباين أدوات قياس أو تقدير هذه السمات كما أشرنا، لذلك ربما لا نجد تعريفاً مناسباً ينطبق على جميع هذه الأدوات. فمثلاً يعرف ساكس (Sax, 1974) الاختبار النفسي والتربوي بأنه مطلب أو مجموعة من المطالب تستخدم للحصول على ملاحظات منتظمة يفترض أنها تمثل سمة نفسية أو تربوية، ويعرفه نيتكو (Nitko 1983) بأنه إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد بالاستعانة بميزان قياس أو نظام تصنيف معين. أما براون (Brown 1976) فيعرفه بأنه إجراء منظم لقياس عينة من السلوك. غير أن التعريف الذي اقترحه شيس (Chase 1985) على الرغم من أنه يتفق مع تعريف نيتكو (Nitko) إلا أنه أكثر عمومية لاشتماله على المكونات الأساسية المشتركة بين أدوات القياس. فقد اقترح شيس (Chase) أن الاختبار النفسي والتربوي هو أداة قياس مقننة أو أسلوب منظم يصمم للحصول على قياس موضوعي لعينة من السلوك بهدف موازنة أداء الفرد بمعيار أو بمستوى أداء محدد.

وهذا التعريف يعد تعريفاً شاملأً للأسلوب الذي يتبع في بناء الاختبار، وطريقة تسجيل الملاحظات عن السلوك، وتصحيح أو تقدير درجات الاختبار. ولعل من

ال المناسب إلقاء الضوء على المكونات الأساسية التي يشتمل عليها هذا التعريف، وهي : التقني ، والموضوعية ، وعينة السلوك ، ومعيار أو محك الأداء .

التقنيين Standaradization: ويقصد به أن يكون بناء وتصحيح وتفسير نتائج الاختبار أو أداة القياس مستنداً إلى قواعد محددة بحيث تتوحد فيه وتحدد بدقة مواد الاختبار وطريقة تطبيقه ، وتعليمات إجابته وطريقة تصحيحه أو تسجيل درجاته ، وبذلك يصبح الموقف الاختباري موحداً بقدر الإمكان لجميع الأفراد في مختلف الظروف . والأسلوب الموضوعي المنظم في بناء الاختبار يعني أن يتم اختيار مفرداته أو كتابتها بطريقة منتظمة تتفق مع مواصفات محددة للاختبار ، وأن تمثل هذه المفردات السمة المراد قياسها .

ومفهوم التقنيين ينطبق على الظروف المثالية التي يستطيع فيها مصمم الاختبار أو أداة القياس ضبط جميع المتغيرات ما عدا المتغير الذي يهدف الاختبار لقياسه ، وبالطبع يصعب عملياً تحقيق ذلك بهذا القدر من الدقة .

ولعل مفهوم التقنيين يتعلق بما سبق أن أشرنا إليه في مستهل هذا الفصل وهو أن البحث العلمي يتميز بقابلية التكرار وتدوال النتائج وعموميتها ، فإذا أردنا مقارنة النتائج التي يحصل عليها مختلف الباحثين يجب أن تكون أدوات القياس وتعليمات استخدامها موحدة للجميع لكي يتسعى جمع الملاحظات العلمية في ظروف ماضبوطة ، وهو ما يُعرف في البحث العلمي بالضبط التجربى .

فالاختبار أو أداة القياس التربوى والنفسي يمكن تمثيلها بتجربة علمية مصغرة تُضبط فيها جميع جوانب الموقف الاختباري واستبعاد العوامل المغتربة لجميع الأفراد ، فإذا ما انتظمت هذه الشروط ، فإنه يمكن أن نستنتج بقدر كبير من الثقة أن الدرجة التي يحصل عليها الفرد في الاختبار تمثل مقدار السمة المقاسة لديه .

الموضوعية Objectivity : ويقصد بها ألا يتأثر تطبيق الاختبار أو أداة القياس وتصحيحه وتفسير نتائجه بالحكم الذاتى للقائمين بهذا العمل . وتحتختلف درجة الموضوعية التي يمكن تحقيقها ، فالموضوعية التامة تعد هدفاً يسعى إلى تحقيقه القائمون ببناء الاختبارات وأدوات القياس . ويمكن توضيح ذلك بالاختبار الذى يشتمل على مفردات اختيار من متعدد حيث يطلب من الفرد اختيار إجابة واحدة صحيحة من بين أربع أو خمس إجابات معطاة ، ففي هذه الحالة يمكن تطبيق مفتاح تصحيح معين على أوراق الإجابة؛ وبذلك لا تختلف الدرجة باختلاف الشخص القائم بعملية التصحيح .

أما الاختبارات أو أدوات القياس التي يعتمد تصحيحها على الحكم الذاتي للشخص القائم بعملية التصحيح، فإننا لا نتوقع أن يكون هناك اتفاق تام بين القائمين بذلك.

ومن الجدير بالذكر أن مفهوم الموضوعية في القياس اتخذ اتجاهًا معاصرًا مختلفاً سوف نوضحه في الفصل الرابع عشر.

عينة السلوك Sample of Behavior : فالاختبار أو أداة القياس تشتمل عادة على مجموعة من المفردات التي تقيس عينة من السلوك الذي تستدل منه على السمة المراد قياسها؛ إذ لا يشتمل الاختبار على جميع المفردات التي يحتمل أن تقيس سمة معينة لدى الأفراد؛ ولذلك يجب مراعاة أن تكون عينة مفردات الاختبار مماثلة للسلوك المراد قياسه ولجميع المفردات التي يمكن أن تقيس هذا السلوك. ويعتمد حجم هذه العينة على طبيعة وتكوينات السمة المراد قياسها، والزمن المتاح للإجابة عن مفردات الاختبار، وعلى الكلفة، وغير ذلك من العوامل. ومشكلة تحديد عينة السلوك تعد من مشكلات صدق الاختبار . **Test Validity**

وهناك مشكلة أخرى ترتبط بمشكلة الصدق، وهي مدى اتساق الدرجة التي يحصل عليها الفرد في اختبارين يشتمل كل منهما على عينة مختلفة من المفردات المماثلة للسمة المقاسة، وتعد هذه من مشكلات ثبات درجات الاختبار **Reliability**. ونظرًا لأن هاتين المشكلتين تتعلقان بخصائص الاختبارات التربوية والنفسية، فسوف نتناولهما ياسهاب في الفصلين الرابع والخامس.

معيار أو محك الأداء : فدرجة الفرد في اختبار ما أو أداة قياس معينة يصعب تفسيرها دون تحديد معيار أو محك أداء معين، فإذا حصل طالب في اختبار ما على الدرجة ١٣٠ ، فإن هذه الدرجة ليس لها معنى في ذاتها، وإنما تستمد معناها من معيار **Norm** جماعة معينة، ول يكن متوسط درجات الفصل الذي يتبعه الطالب، أو من ممح **Criterion** أداء مقبول. فمثلاً إذا طبق الاختبار على طلاب الفصل وكان متوسط درجاتهم ١٠٠ ، وأعلى الدرجات ١٢٠ وأدنىها ٥٠ ، فإن الدرجة ١٣٠ التي حصل عليها الطالب يصبح معناها أكثر وضوحاً، أي أن مقارنة أداء الطالب بأداء أقرانه يجعل للدرجة التي يحصل عليها معنى ودلالة؛ لأنها عندئذ تحدد مركزه النسبي بين أقرانه، وبهذا يكون المعيار الذي تفسر الدرجة في ضوئه هو معيار الجماعة، وبعبارة أخرى نقول: أن التفسير مرجعي الجماعة أو المعيار **Norm - Referenced** .

غير أن هذا التفسير لا يفيينا في معرفة مدى إتقان الطالب للمهارات أو المهام التي يقيسها الاختبار، إذ ربما تكون الدرجة ١٣٠ التي حصل عليها الطالب أقل من الحد الأدنى المطلوب للنجاح في أداء المهارات المرجوة على الرغم مما يبدو من مركزه النسبي المرتفع، فالعبرة هنا ليست بمقارنة الدرجة بمعيار جماعته أو أقرانه، وإنما بمقارنتها بمحك الأداء المتعلق بالمهارات أو المهام المراد قياس مدى تمكن الطالب منها بغض النظر عن ترتيبه في الفصل. وبهذا تفسر الدرجة في ضوء محك أداء محدد، أي أن التفسير يكون مرجعى المحك Criterion - Referenced.

ونظراً لأهمية هذان النوعان من التفسيرات في مختلف مجالات القياس والتقويم التربوى والتفسى وفى تصميم وبناء أدوات القياس وتحليل نتائجها، فسوف ينالا قدرأً وافياً من المناقشة فى فصول لاحقة، وبخاصة فى الفصلين الثامن والخامس عشر.

التقويم Evaluation، والتقييم : Assessment

كثيراً ما يحدث خلط بين مفاهيم القياس والتقويم والتقييم، فالقياس كما رأينا يصف السلوك وصفاً كميّاً، ويحدد مقدار سمة معينة أو أكثر لدى الفرد، ولا يتخطى ذلك إلى إصدار حكم معين على السلوك أو السمة المقاسة، فالدرجة ١٣٠ التي حصل عليها الطالب في المثال السابق تعد «قياساً» لسمة معينة يقيسها الاختبار، أي تحدد كم السمة، ولكن إذا أصدرنا حكمًا قيمياً على هذه الدرجة، أو اتخاذنا قراراً بشأن الطالب الذي حصل على هذه الدرجة استناداً إلى معيار أو محك أداء معين، فإننا بذلك نكون قد تخطينا وصف الدرجة إلى تقويمها والحكم على مستوى الأداء. فإذا كان محك الأداء المقبول ١٢٠ مثلاً، فإن الدرجة ١٣٠ تفوق المحك وتحكم عليها بأنها درجة مقبولة، أما إذا قررنا أن الدرجة ١٣٠ تناظر تقدير أداء جيد مثلاً، فإننا بذلك نكون قد أصدرنا حكمًا كيفياً على الدرجة أو الأداء الذي تمثله.

فالقياس Measurement إذن هو الوصف الكمى للموضوعى للأداء Evaluation ، فهو الحكم الكيفي الوصفى على الدرجة مثلاً فى التقدير النوعى للأداء، وهذا الحكم يفيد فى اتخاذ قرار معين بشأن الفرد الذى حصل على الدرجة، أو اقتراح إجراء مناسب له.

وربما يظن البعض أن التقويم بهذا الشكل عملية بسيطة يمكن لأى فرد غير متخصص أو متعرس إجراؤها، ولكن إذا أمعنا النظر نجد أنه عملية منهجية تتطلب جمع بيانات موضوعية ومعلومات صادقة من مصادر متعددة باستخدام أدوات قياس متنوعة في ضوء أهداف محددة بغرض التوصل إلى تقديرات كمية، وأدلة كيفية يستند إليها في إصدار أحكام أو اتخاذ قرارات مناسبة تتعلق بالأفراد.

ومما لا شك فيه أن هذه القرارات لا ينعكس أثراها على الفرد فقط، وإنما على مستوى أدائه ودرجة كفاءته في القيام بأعمال أو مهام معينة. كما أن التقويم لا يقتصر على الأفراد، وإنما يمتد بحيث يشمل تقويم المهام والبرامج والمشروعات والمؤسسات للتحقق من أثرها وفاعليتها في تحقيق أهدافها المحددة.

ومن المفاهيم الأخرى شائعة الاستخدام في مجال قياس الشخصية الإنسانية مفهوم «التقييم أو التقدير Assessment». وعلى الرغم من أن هذا المفهوم يستخدم في كثير من الأحيان مرادفاً لمفهوم التقويم، إلا أنه أكثر خصوصية من مفهوم التقويم، وأكثر اتساعاً من مفهوم القياس.

فمفهوم التقييم أو التقدير Assessment استخدمه في مستهل نشأته مجموعة من علماء النفس الذين كانوا يعملون في مكتب الخدمات الاستراتيجية بالولايات المتحدة الأمريكية، حيث أوكل إليهم انتقاء الأفراد الذين يصلحون للقيام بمهام خاصة فيما وراء البحار في الحرب العالمية الثانية. وقد ظهر هذا المفهوم لأول مرة في كتاب «تقييم الإنسان» الذي أصدره هذا المكتب عام ١٩٤٨. وعقب انتهاء هذه الحرب تبنى هؤلاء العلماء المفهوم نفسه في دراساتهم وبحوثهم مما أدى إلى انتشار المفهوم منذ عام ١٩٥٣، وظهوره كمصطلح سيكولوجي في مراجع علم النفس والقياس النفسي.

ويعرف ساندبرج (Sundberg, 1977) مفهوم التقييم أو التقدير بأنه مجموعة من العمليات التي تستخدم بواسطة أخصائيين متخصصين؛ للتوصل إلى تصورات وانطباعات واتخاذ قرارات واختبار فروض تتعلق بنمط خصائص فرد معين يحدده سلوكه أو تفاعله مع بيئته. وهذا التقييم يمكن أن يتضمن أساليب متعددة وأدوات متنوعة يعتمد بعضها أحياناً على القياس الكمي (مثل درجات اختبارات التحصيل، والذكاء، والاستعدادات، والتفكير الابتكاري، ومقاييس الحاجات والسمبول والاتجاهات والقيم والشخصية)، ويعتمد البعض الآخر عادة على التقديرات الكيفية، والأحكام الوصفية (مثل تقييم استثمارات القبول، وتقييم خطابات التوصية، والمقابلة الشخصية)، وذلك بغرض انتقاء الأفراد، وكذلك الملاحظة الكلينيكية، وتقديرات المعلمين للطلاب وبالعكس.

لذلك يقترن مفهوم التقييم أو التقدير Assessment عادة بالمجالات والنظريات الكlinيكية، حيث يميل الكlinيكيون إلى استخدام هذا المفهوم في فحوصهم وتشخيصاتهم لاضطرابات الشخصية التي يعاني منها بعض الأفراد.

وعموماً، فإن عمليات التقويم والتقييم تعتمد كما ذكرنا على أدوات قياس متنوعة بعضها يؤدي إلى مقادير كمية، والبعض الآخر يعتمد على أحكام الاختصاصيين فيما

يتعلق بسمات الفرد أو شخصيته كفرد متكامل ومتميز عن غيره من الأفراد، وإجراء عمليات التقويم أو عمليات التقييم يعتمد بلا شك على الهدف منها، ونوع القرار المناسب لهذا الهدف وانتقاء أدوات القياس التي تسترشد هذه العمليات بتائجها. وهذا يتطلب معرفة الاستخدامات المختلفة لأدوات القياس، وأنواعها أو تصنيفاتها.

استخدامات أدوات القياس والتقويم :

تستخدم أدوات القياس التربوي والنفسى فى أغراض متعددة يصعب حصرها نظراً لتنوع الظروف والمواقف التى تتطلب تطبيق هذه الأدوات للاستناد إلى نتائجها فى اتخاذ قرارات تتفق وهذه الظروف والمواقف .

فالباحث التربوى والنفسى يحتاج فى تصميم دراسته وتنفيذها إلى العديد من أدوات القياس المناسبة التى تمكّنه من التتحقق من التتحقق من فروض دراسته سواء كانت الدراسة تجريبية أو ارتباطية أو وصفية أو كلينيكية، كما يسّتحاج إليها الممارسون ومستخدمو القرارات فى مختلف المجالات التربوية والنفسية مثل المعلم، والاختصاصى النفسى والاجتماعى، والقائم بالعلاج النفسي، والمسئولون عن انتقاء الأفراد لشغف مختلف أنواع الوظائف والمهن أو تصنيفهم وفقاً لممكبات معينة، والمهتمون بتقدير المشروعات والبرامج والمؤسسات وغيرهم .

ولذلك فإن أدوات القياس تخدم الأغراض البحثية والأكاديمية النظرية، كما تخدم الأغراض العملية والتطبيقية .

وسوف نوضح فيما يلى بعض الاستخدامات الرئيسة لأدوات القياس المختلفة .

بناء وتطوير النظريات التربوية والنفسية : فكثير من النظريات فى العلوم السلوكية اعتمدت فى بنائها أو تطويرها على تطبيق أدوات قياس متنوعة، فنظريات التكوين العقلى ونظريات الشخصية فى علم النفس استندت إلى فروض تم التتحقق منها بتطبيق مجتمعات كبيرة من الاختبارات والمقاييس على عينات مختلفة من الأفراد، وتحليل البيانات المستمدة من هذه الأدوات باستخدام الأساليب الإحصائية، كما أن المفاهيم والتكتوبات الافتراضية التى تتضمنها كثير من النظريات التربوية والنفسية يجرى فحصها والتوصيل إلى تعریفات إجرائية مناسبة لها عن طريق الاختبارات والمقاييس . وكذلك أسهمت أدوات القياس فى تطور علم القياس التربوى والنفسى Psychometry ذاته عن طريق تصميم وبناء اختبارات ومقاييس جديدة متنوعة، وتطوير المقاييس المتاحة ، والتوصل إلى صيغ رياضية وإحصائية تووضح العلاقات القائمة بين المتغيرات والعوامل المتعلقة بأدوات القياس وخصائصها السيكومترية .

النقاء الأفراد : تستخدم أدوات القياس في غالبية الأحيان في تزويد المعينين بشئون الأفراد بيانات مفيدة يمكن الاستناد إليها بقدر محدد من الثقة في اتخاذ قرارات تتعلق بانتقاء الأفراد Selection لشغل وظائف أو مهن معينة، أو انتقاء الطلاب للالتحاق بالكليات والمعاهد ومراكز التدريب المهني وغيرها، إلى جانب ما تحدده هذه المؤسسات التربوية من حد أدنى من المتطلبات، فإن الأماكن الشاغرة فيها تقل عن عدد المتقدمين. فهنا يمكن أن تطبق اختبارات تقيس مدى توافر هذه المتطلبات لدى الطلاب، وتسمى اختبارات مرجعية المحك Criterion - Referenced Tests، كما يمكن تطبيق اختبارات للتبؤ بنجاح الطلاب أو رسوبهم في الدراسة أو العمل بأقل قدر ممكن من الخطأ. والمقصود بالخطأ هنا قبول طلاب يتبيّن رسوبهم خلال دراستهم، أو رفض طلاب كان من الممكّن نجاحهم لو أتيحت لهم فرصة القبول، وهذه الاختبارات التنبؤية ينبغي أن تقلّل احتمال الواقع في أي من هذين النوعين من الخطأ قدر الإمكان، وتسمى هذه الاختبارات مرجعية الجماعة أو المعيار Norm - Reference Tests. وهذا يتطلب مراعاة أن يكون هناك تناقض كبير بين المفردات التي تشتمل عليها الاختبارات التي تستخدم في الانتقاء وبين المتطلبات والمهارات والقدرات المطلوبة للنجاح في الدراسة أو العمل، وإجراء العديد من الدراسات التنبؤية قبل استخدام الاختبارات في هذا الشأن.

تسكين الأفراد : إن انتقاء الأفراد للالتحاق بالدراسة يأخذى الكليات أو لشغلهن وظيفة أو مهنة معينة يعني المزاوجة بين الفرد والعمل، أى وضع الشخص المناسب فى المكان المناسب. ولكن التسکین Placement يعني وضع الطالب الذى تم قبوله بإحدى الكليات مثلاً فى قسم معين يناسبه بحيث يتحقق فيه تفوقاً وتميزاً، أو وضع التلميذ الذى تم قبوله فى المدرسة فى برنامج تعليمي معين يناسبه مثل البرامج العلاجية، أو العاديه، أو الإثرائيه. وكذلك وضع الفرد الذى تم انتقاوه فى أحد الأقسام المتعلقة بالمهنة أو الوظيفة التى شغلها. ويختلف التسکین عن الانتقاء فى أن قرارات الانتقاء تتضمن قبول الفرد أو رفضه، فإذا ما رُفض عليه أن يبحث عن مكان آخر، أما إذا تم قبوله فإن المؤسسة تعمل على تسكيته وفقاً لإمكاناته واحتياجاته. فاللهم الذى يقبل فى مدرسة مهنية معينة ينبغي أن تعمل المدرسة على تسكيته فى قسم مناسب لمهنته، أو تقدم له برنامجاً تدريبياً علاجياً، أو تقوم بترفيهه فى مستوى مستقدم بحسب المعلومات التي تقدمها اختبارات التسکین التي تسمى عادة اختبارات تحديد المستوى .

لذلك فإن الاختبار الذى يستخدم فى اتخاذ قرارات الانتقاء ربما لا يصلح فى اتخاذ قرارات التسكين، إذ ربما يصلح اختبار يقىس الاستعداد الدراسي فى التنبؤ بالنجاح فى الكليات، ولكنه ربما لا يصلح فى تسكين الطالب المقبول فى البرنامج التعليمى الأكثر ملاءمة له.

تصنيف الأفراد: على الرغم من أوجه الشبه بين التصنيف والتسكين والانتقاء إلا أن نيكو (Nitko, 1983) يرى أن التصنيف Classification أكثر عمومية منها. فقرارات التصنيف تتعلق بتعيين الأفراد فى أقسام أو مجموعات مختلفة، فعندما يتم انتقاء الأفراد للالتحاق بالخدمة العسكرية مثلاً يقوم المسؤولون بتصنيفهم فى وحدات عسكرية مختلفة وتسكينهم فى أعمال مناسبة لهم داخل هذه الوحدات، كما أن مرضى الاضطرابات العقلية يجرى عادة تصنيفهم، وكذلك الأطفال من ذوى الإعاقة مثل الأصم، والأعمى، وصعوبة السمع، والتخلُّف العقلى، ومتعدد الإعاقة، وشديد الاضطراب الانفعالي، والمعاق بصرياً وهكذا. لذلك يقترح نيكو أن التصنيف يتعلق بتعيين الأفراد فى أقسام أو أعمال أو برامج. فإذا كانت هذه الأقسام تمثل مستويات مرتبة لنوع معين من البرامج أو الأعمال (مثل المستويات العلاجية والعافية والإثرائية) فإن هذا التعيين يسمى «تسكين» ، وإذا كان أحد هذه الأقسام يمثل الرفض والآخر يمثل القبول، فإن التعيين يسمى «(انتقاء»، أما «التصنيف» فيمثل أقساماً غير مرتبة بطبيعتها ، والقرارات المتعلقة بهما جمِيعاً تعد من نوع القرارات الإدارية.

التشخيص والعلاج : تستخدم أدوات القياس النفسي والتربوى فى تشخيص الاضطرابات النفسية Diagnosis ، وكذلك فى تشخيص جوانب القسوة والضعف لدى الطلاب وأصحاب المهن من أجل اقتراح أساليب علاج مناسبة، فمعلم التعليم الأساسى مثلاً يعمل على تشخيص جوانب الضعف لدى تلاميذه فى المهارات الأساسية المتعلقة بالقراءة والكتابة والخطاب والحساب ، والمرشد النفسي يقوم بتشخيص المجالات السلوكية التى يعاني منها الفرد . فالملعلم يحاول معاونة تلاميذه فى تحسين أدائهم، والمرشد النفسي يعمل على تقديم برنامج علاجى مناسب للفرد الذى يعاني من مشكلات نفسية معينة وهكذا . ومن الجدير بالذكر أن الاختبارات التربوية والنفسية التى تستخدم فى اتخاذ قرارات التشخيص تختلف عن الاختبارات التى تستخدم فى الانتقاء والتسكين والتصنيف فى أنها تكون أكثر تفصيلاً وعمقاً نظراً لتركيزها على جوانب معينة فى السمات المراد قياسها وتعرف جوانب الضعف فيها، ولذلك تسمى الاختبارات التشخيصية Diagnostic Tests ، ويطلب بناء هذه الاختبارات جهداً كبيراً وخبرة فنية عالية فى تقنيات بنائتها. ومع أن مؤلف هذا الكتاب قد أفرد مجلداً خاصاً لهذا النوع من

الاختبارات (علام ، ١٩٩٥) إلا أنها سوف تلقى عليه مزيداً من الضوء في الفصل الثامن، وذلك لأهميته وضرورته في مختلف المجالات التربوية والنفسية وفي كثير من التطبيقات العملية.

الإرشاد والتوجيه التربوي والمهني : فأدوات القياس المتنوعة يمكن استخدامها

استخداماً وظيفياً فاعلاً في عمليات الإرشاد Counseling والتوجيه Guidance التربوي وال النفسي . وعلى الرغم من أن قرارات الانتقاء والتسكنين والتصنيف تعد من القرارات الإدارية المؤسسية، إلا أن قرارات الإرشاد والتوجيه تتعلق بفرد معين شأنها شأن قرارات التشخيص والعلاج . فقد سبق أن ذكرنا أن المؤسسة تتخذ قراراتها في الانتقاء مثلًا بما ينطبق على كثير من الأفراد المتقدمين في ظروف مشابهة ، بحيث تكون هناك درجة معينة من الثقة فيما يتربّط على هذه القرارات ، فاحتمال نجاح الأفراد في دراستهم أو في أعمالهم بدرجة كبيرة هو المحك الذي تستند إليه هذه القرارات ، أما القرار المتعلق بفرد معين فيؤثر على هذا الفرد بذاته ولا يؤثر فيما عداه ، وعليه أن يتحمل المخاطرة في حالة عدم التزامه بهذا القرار .

فالفرد الذي يسعى إلى المرشد التربوي أو النفسي يود أن يتعرف على إمكاناته واستعداداته ويحدد مجال العمل أو الدراسة الأكثر ملائمة له . لذلك فهو يلجأ إليه بمحض إرادته ويتعاون معه تعاوناً متمنراً وبخاصية في الموقف الاختبارية .

وقدرات التوجيه والإرشاد التربوي والنفسى مستعددة ، فالتجيئ والإرشاد التربوى يهتم بتعريف الطالب بما يمكنه دراسته من مقررات والتخصص الدراسي الذى يناسبه . والتوجيه والإرشاد المهني يهدف لمساعدة الفرد على اختيار مجال عمل يستطيع أن يؤديه أداءً متميزاً بما يتفق واستعداداته وسماته الشخصية ، مثل مجال الطب أو الهندسة أو الأعمال الحرة أو الفنية أو الحرفة . والتوجيه والإرشاد الشخصى يركز على المشكلات الشخصية لدى الفرد التي ربما تعكس اضطرابات في الشخصية مثلاً أو تكون ناتجة عن سوء تكيف وقى ، مثل سوء تكيف المراهق مع أقرانه الأكثر تفوقاً ، أو سوء تكيفه مع الجنس الآخر .

وفي جميع الحالات ربما لا يقتصر المرشد التربوي أو النفسي على استخدام الاختبارات ، وإنما يمكنه أن يجمع معلومات ثرية من مصادر متنوعة . إذ ربما يقوم بمقابلات شخصية تفصيلية مع العميل لتعرف وجهات نظره وخلفيته واتجاهاته ورغباته وأهدافه وخصائصه ، وربما يلجأ إلى جمع معلومات عن الفرد من مصادر أخرى ذات علاقة به مثل السجل المدرسي ، أو آراء أقرانه ومعلميه ووالديه وأصدقائه ، وغير ذلك .

وانتقاء الأدوات المناسبة يعتمد على خبرة المرشد النفسي بهذه الأدوات وتمرسه على تطبيقها والإفادة من نتائجها لكي يتخذ قراراً صادقاً فيما يتعلق بالعميل.

وربما يتساءل القارئ عن أوجه الاختلاف بين قرارات الانتقاء والتشخيص والإرشاد والتوجيه . والحقيقة أن الإرشاد والتوجيه يعدان نوعاً من التسخين باستخدام سلسلة متابعة من أدوات القياس المتنوعة التي أشرنا إليها أثناء عملية الإرشاد التي قد تستغرق عدة جلسات.

وعلى الرغم من أن الانتقاء والإرشاد يستندان إلى عمليات أساسية مشابهة حيث يهتم كلاهما بالتبؤ بالنجاح ، إلا أن الانتقاء يركز بدرجة أكبر على عملية التنبؤ ، كما يستندان إلى نتائج أدوات قياس متنوعة على الرغم من اختلاف الإطار المرجعي لكل منها ، فقرارات الانتقاء تنتهي بقبول فرد له خصائص معينة في دراسة أو مهنة معينة مناسبة لخصائصه من بين أعداد كبيرة من المستقدمين للدراسة أو المهنة نفسها بحيث يتوقع أن يكون هذا الفرد أفضل المستقدمين . وعلى العكس فإن قرارات التوجيه والإرشاد تهدف لتحديد مجال العمل المناسب للفرد بحيث يستثمر فيه إمكاناته وقدراته استثماراً فاعلاً . وتتأثر قرارات التوجيه والإرشاد بمدى دقة أدوات القياس في التنبؤ بالآداء في عمل معين ، ولذلك يفضل التنبؤ بالأداء في مجال عمل محدد بدلاً من التخمين ، حيث إن الفرد ربما يمكنه أداء أكثر من عمل واحد بفاعلية ، فالنجاح في أداء الأعمال يعتمد إلى جانب الاستعدادات على المناخ السائد في العمل ، واتجاهات المشرف على العمل ، وسياسة المؤسسة ، والعلاقات بين الزملاء ، وتوافر الأجهزة المناسبة.

تقدير تحصيل الطلاب :

كثيراً ما يقتصر تقويم الطلاب بتقويم التحصيل الدراسي ، ولكن التحصيل الدراسي يعد واحداً من المتغيرات الكثيرة التي يتضمنها هذا التقويم ، فهناك العديد من المتغيرات التي ينبغي أن يأخذها المعلم بعين الاعتبار عند تقويمه لطلابه . فالتحصيل الدراسي يرتبط بمتغيرات عقلية أو معرفية أخرى ، مثل الذكاء والاستعدادات بأنواعها والأنماط المعرفية ، كما يتضمن المتغيرات الوجدانية مثل الميول والاتجاهات والقيم والسمات المزاجية ، والمتغيرات الحركية المهارية .

لذلك فإن أدوات تقويم الطلاب لا تقتصر على الاختبارات التحصيلية بنوعيها الرئيسين ، أي الاختبارات مرجعية الجماعة أو المعيار والاختبارات مرجعية المحك ، وإنما ينبغي أن تسع هذه الأدوات بحيث تشمل اختبارات الذكاء والاستعدادات ومقاييس الميول والاتجاهات وغيرها من الأدوات متضمنة بطاقات ملاحظة السلوك ، وموارين التقدير ، والمقابلات الشخصية ، ومؤشرات الأداء ، وقوائم المراجعة ، واختبارات المواقف .

والبيانات المستمدة من هذه الأدوات المتنوعة لا تفيد فقط في تقويم ومتابعة التقدم الدراسي للطلاب وإنما تفید أيضاً في تقويم عملية التعليم، وبعبارة أخرى فإن عملية تقويم الطلاب تقدم تغذية مرتبطة للطالب والمعلم تساعد في زيادة دافعية الطالب وتشجيعه على تكوين عادات استذكار جيدة، وتعريفه بجوانب القوة والضعف في تحصيله أو أدائه. وبذلك تؤدي عملية تقويم الطلاب إلى تحسين وإثراء عملية التعليم، حيث إن هاتين العمليتين متكمالتان، وهذا يؤكّد أهمية التقويم البنائي الذي سوف نوضحه بعد قليل.

أدوار التقويم وأهدافه :

إن التقويم بمفهومه العلمي - كما أوضحتنا - يختلف بالطبع عن التقويم الذي يمارسه الإنسان العادى في حياته اليومية. فالإنسان العادى يقوم العديد من الأشياء والأمور التي يواجهها في حياته اليومية، فهو يقوم بالأطعمة التي يأكلها، والأشياء التي يشتريها، والكتب التي يقرأها، والأماكن التي يزورها وغير ذلك. ويتضمن هذا التقويم نوعاً من الحكم على الشيء المطلوب على الرغم من أنه ليس تقوينا علمياً بالمعنى المتعارف عليه.

ونحن في المجال التربوي مثلاً نقابل كثيراً من المواقف التي نمارس فيها هذا النوع من التقويم الذي يمكن أن نطلق عليه التقويم غير الرسمي، وهو يتراوح بين حكم المعلم على حسن تهوية الفصل وإضاعته إلى تقدير مدير المدرسة لفاعلية وكفاءة أعضاء الهيئة التدريسية.

ولكن يصعب الركون إلى نتائج هذا النوع من التقويم في اتخاذ القرارات التربوية المختلفة، ومن هنا كان لابد من الاعتماد على التقويم العلمي المنظم بمفهومه الذي قدمناه.

ومفهوم التقويم من هذا المنظور يمكن أن يطبق في مختلف المواقف التربوية سواء كان تقويم أفراد أو تقويم مشروعات وبرامج، فمثلاً يمكن استرشاداً بهذا المفهوم تعرف مدى فاعلية أنشطة تعليمية معينة وفقاً لمحركات محددة، ومعرفة ما إذا كان تأثير نشاط أو برنامج معين أفضل من تأثير نشاط أو برنامج آخر، وكذلك تعرف المتغيرات التي تأثرت من جراء تنفيذ نشاط أو برنامج معين على جماعة ذات خصائص معينة، وما إذا كان استخدام أو تنفيذ برنامج معين يوازي كلفة إعداده وهكذا.

وبذلك تعدد أدوار التقويم في المجالات التربوية، فلتقويم دور في إعداد المعلمين وأنشطتهم، وفي عملية بناء المناهج الدراسية، وفي التجارب الميدانية التجديدية المتعلقة بتحسين عمليات التعليم والتعلم، وفي انتقاء أفضل الاستراتيجيات والتقييمات التربوية، وفي الحكم على استمرار برنامج تربوي معين أو تعديله أو إلغائه. كما أن للتقويم دور بالغ الأهمية في متابعة التقدم الدارسي للطلاب وتحديد مستوى التحصيلي وترجيحهم التوجيه التعليمي والمهني. ولذلك لا ينبغي أن الخلط بين أهداف التقويم وأدواره، فعادة يكون للتقويم أدوار متعددة بل ويجب أن يكون التقويم هكذا بالنسبة لبرنامج تعليمي معين، ولكن ر بما يرتبط بكل دور من هذه الأدوار العديد من الأهداف. فمثلاً في عملية بناء منهج معين، ر بما يكون أحد أدوار التقويم تحسين وتطوير المنهج أثناء عملية بنائه، ويرتبط بهذا الدور العديد من الأهداف أو الأسئلة التي تتطلب إجابة مثل: هل أهداف المنهج واضحة وضوحاً تماماً في أذهان القائمين ببنائه؟ وهل هم متخصصون لبنائه؟ وهل تسير عمليات البناء وفق هذه الأهداف بالفعل؟ وهكذا. ور بما يكون لعملية التقويم في هذا الصدد دور آخر يتعلق بمساعدة المسؤولين عن تنفيذ المنهج في اتخاذ قرار بشأن تبرير استخدامه في مرحلة تعليمية معينة وفقاً للنتائج التي أبرزتها عملية التقويم في الدور السابق.

أى أن دور التقويم يتعلق بالنشاط المراد تقويه، أما أهداف التقويم فتعلق بالأسئلة المراد الإجابة عنها والتي تبلور حول إمكانية وفاعلية إعداد وتصميم وتنفيذ هذا الشامل.

ولعل هذا التمييز بين أدوار التقويم وأهدافه يزيل التخوف والقلق الذي يعتري كل من يسمع كلمة «تقييم Evaluation»، ويؤكد الجانب الإيجابي البناء لهذا المفهوم، فكثيراً ما ارتبط التقويم بالكشف عن مواطن الضعف والقصور سواء في الأدوات أو الأفراد أو في الإمكانيات المادية والبشرية.

لذلك فإن للتقويم دورين رئيسيين هما الدور البنائي التشخيصي Formative Evaluation والدور الختامي التجمعي Summative Evaluation. فالتفريق البنائي التشخيصي يهدف للكشف عن جوانب القوة والضعف في برنامج تعليمي معين أثناء إعداده وتنفيذه في مرحلة التجريب بغرض مراجعة مكونات البرنامج وتعديلها وتطويرها، وكذلك تشخيص مواطن القدرة والضعف في أداء المتعلمين وتحديد الصعوبات التي يواجهها كل منهم أثناء التعليم واتخاذ ما يلزم من أساليب العلاج. ومساعدة المعلم في التعرف على قدراته وإمكاناته، واقتراح سبل ووسائل تحسينها وتنميتها إلى أقصى حد ممكن.

أما دور التقويم الختامي التجمعي فهو الكشف عن مدى تحقيق البرنامج لأهدافه بعد إتمام إجراءات التقويم البنائي المستمر، وكذلك تحديد المستوى التحصيلي العام لل المتعلمين في نهاية مدة دراسية معينة بالنسبة للمعارف والمهارات والاتجاهات المتعلقة بال مجالات الدراسية؛ وذلك لاتخاذ قرارات تتعلق بتقديراتهم ومنحهم الشهادات الدراسية، وهذا يؤكد أيضاً الدور الإيجابي لعملية التقويم الختامي بنفس القدر الذي يؤكد دور عملية التقويم البنائي.

وللتوضيح ذلك نفترض أنه تبين من نتائج تطبيق هذه الأدوات المتعددة أن عدداً قليلاً من الطلاب لم يحقق هدفاً تعليمياً معيناً أو مجموعة من الأهداف المتعلقة بمحظى دراسي معين، فإنه ربما يستنتج المعلم أن عملية التعليم كانت فاعلة، أما إذا تبين أن كثيراً منهم أخفق في تحقيق مجموعة كبيرة من الأهداف التعليمية، فإنه ربما يقرر أن عملية التعليم التي اتبعت كانت فاقصة. وفي كلتا الحالتين ينبغي أن يتخذ المعلم قراراً معيناً فيما يتعلق بأسلوب التعليم الذي اتبعه. كما أن التقويم البنائي Formative Evaluation للطلاب يقدم تغذية مرتدة فيما يتعلق بتحفيظ المعلم المستقبلي للأنشطة التعليمية. فربما يجد المعلم أن معدل سرعته في عرض المادة يجب أن يقل أو يزداد، أو ربما يكشف التقويم عن حاجة المتعلمين إلى تعلم بعض المهارات الالازمة المتابعة موضوع مستقبلي معين، ويطلق على هذه المهارات «المهارات المدخلية Entry Skills» أو المتطلبات الالازمة للبدء في دراسة معينة Prerequisites، وقياس الأهداف المدخلية الالازمة لتعلم الأهداف المرتبطة بها، وتكون عادة من نوع الاختبارات مرجعية المحك Criterion - Referenced Tests ، كما يحتاج المعلم أيضاً إلى اتخاذ قرارات تتعلق بتقديرات الطلاب في نهاية مدة دراسية معينة لتحديد المستوى التحصيلي العام لكل منهم، وهنا تبرز أهمية التقويم الختامي Summative Evaluation ، وبالطبع تعتمد هذه التقديرات على نتائج التقويم المدخلى والتقويم البنائي المستمر أثناء تنفيذ عملية التعليم. ويجب أيضاً أن تُستخدم أدوات التقويم الختامي التي تكون عادة من نوع الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests إذا كان الهدف من هذه التقديرات توجيهه الطلاب إلى مرحلة تعليمية تالية أو دراسة مستقبلية معينة. ويعرف هذا النوع من الاختبارات بالاختبارات التحصيلية المقrite Standardized Achievement Tests التي تحدد المركز النسبي للطالب في مجموعة العمرية أو فرقته الدراسية فيما يتعلق بمستواه التحصيلي العام.

وفي أي الأحوال يجب أن تزود أدوات القياس والتقويم كل من المعلم والطالب ببيانات صادقة ومتسقة عن تحصيل الطلاب بمفهومه الشامل لما لهذه البيانات من أهمية في اتخاذ القرارات التعليمية المختلفة التي تمس حاضرهم ومستقبلهم.

وهذا يوجه النظر إلى أهمية تصميم وبناء الاختبارات التحصيلية التشخيصية والاختبارات الكفايات، والاختبارات المقدرة في مختلف مجالات الدراسة.

تقويم المناهج والبرامج والمشروعات التربوية المستحدثة:

فالاختبارات والمقاييس يمكن أن تستخدم استخداماً فاعلاً في اتخاذ قرارات تتعلق بالمناهج والبرامج والمشروعات التجديدية التي تعمل المؤسسات التربوية على تصميمها وتنفيذها. وقد اختلف العلماء في كيفية تصميم المنهج الدراسي Curriculum وطرق بنائه، فمنهم من اعتبر أن عملية تصميم المنهج تشتمل على: تشخيص الحاجات التربوية وتقييمها، وصياغة الأهداف التعليمية العامة والأهداف السلوكية، وانتقاء المحتوى، وإعداد الخبرات التعليمية وتنظيمها، وتحديد طرق وأساليب تقويم فاعلية ما يتم تعلمه، وهذه العمليات تقدم تحديداً وتعريفاً عريضاً وشاملاً لعمليات تصميم المنهج الدراسي وبنائه وتقويمه.

وفي ضوء هذا التحديد تصبح عملية تقويم المنهج هي عملية تقويم للبرامج التعليمي برمتها، بما يشتمل عليه من أساليب واستراتيجيات تعليم، وكتب مدرسية، وتقنيات تربوية. وهذا التقويم لا يجب أن يتم بطريقة مجردة، وإنما في مواقف مدرسية فعلية بحيث يصبح تقويم المنهج عملية ديناميكية مبدعة. وهذا يتطلب وجود محكّمات داخلية وخارجية للحكم على المنهج. فالمحكّمات الداخلية تفيد في معرفة ما إذا كان المنهج المعين قد حقق أهدافه المرجوة، والمحكّمات الخارجية تفيد في معرفة ما إذا كان المنهج المعين أفضل من المناهج الأخرى المناظرة له من حيث تأثيره واتجاه المعلمين والطلاب نحوه ومردود كلفته وفعاليته.

وهذا النوع من التقويم الذي يعتمد على محكّمات خارجية يصعب إجراؤه نظراً لصعوبة إمكانية تطابق منهجين في الأهداف حتى لو ارتبط كل منهما بالمحظى الدراسي نفسه، كما يصعب وجود أدوات قياس تكون صادقة بالدرجة نفسها لقياس فاعلية كل من المنهجين؛ لذلك يجب البدء بفحص أهداف كل من المنهجين فحصاً دقيقاً إذا أردنا الموازنة بينهما، وبناء أدوات قياس صادقة بالنسبة لكل منها إذا لم تكن متوفّرة.

وفيما يتعلق بتطوير البرامج والمشروعات التربوية المستحدثة، فإن التقويم يلعب دوراً مهماً في كل مرحلة من مراحل تصميم وإعداد وتنفيذ هذه البرامج. ونقصد بالبرنامج أو المشروع التربوي مجموعة من الأنشطة أو العمليات المستحدثة الهدافة والمقصودة التي يمكن تحديدها بدقة ويفترض أنها تؤثر تأثيراً معيناً في مجموعة أومجموعات مستهدفة من الأفراد. ومن أمثلة ذلك برنامج مستحدث لتعليم الأطفال القراءة باستخدام التقنيات السمعية والبصرية، وبرنامج تجديدي لتعليم المواطنة المنتجة الوعية عن طريق التدريبات الحياتية، أو برنامج تربوي مبتكر لتعزيز القيم العالمية والتسامح والديمقراطية داخل المدرسة، أو برنامج مستحدث لإعداد مواد تعليمية ملائمة للمرحلة الثانوية في مجال التسامح والتفاهم الدولي، وغير ذلك من البرامج التربوية التجديدية التي تعد الأفراد للحياة المشرمة في القرن المقبل.

فالتقويم بما يتضمنه من عمليات كشف وجامع للبيانات والمعلومات باستخدام أدوات قياس متنوعة يساعد القائمين على تخطيط هذه البرامج وتنفيذها وإدارتها في توجيه مسار هذه البرامج وتطوير مكوناتها عن طريق التقويم البنائي Formative Evaluation لتحقيق أهدافها المرجوة، إذ ربما تكشف عملية التقويم البنائي عن قصور في نظام تمويل البرنامج أو نظام إعداد المشرفين عليه، أو نظام المواد والأدوات، أو المدة الزمنية اللازمة لتنفيذ كل مرحلة من مراحل البرنامج، أو غير ذلك مما يؤثر في فاعلية البرنامج وفي تحقيقه لأهدافه. فالتقويم البنائي يفيد في اتخاذ قرارات أفضل في كل من هذه المراحل. أما التقويم الختامي Summative Evaluation فيفيد في تحديد فاعلية البرنامج ككل وتأثيره بعد أن يكون قد خضع لعمليات التقويم البنائي المستمر أثناء إعداده وتنفيذها استناداً إلى أدوات قياس متنوعة ومتاسبة.

تقييم المعلم: إن فاعلية العملية التربوية تعتمد اعتماداً أساسياً على ما حققه المعلم في الفصل المدرسي، فجميع المصادر التي يستخدمها المعلم سواء كانت مواد أو أدوات أو تقنيات أو غير ذلك يجب أن تفي في النهاية في إحداث تغيرات في سلوك الطلاب. وقد أكدت حركة المسئولية التربوية Accountability التي انتشرت في بعض دول العالم المتتطور أهمية جمع بيانات باستخدام أدوات قياس متنوعة عن طبيعة هذه التغيرات لتقدير فاعلية المعلم في ضوئها، كما أكدته التوجهات المعاصرة المتعلقة بتطوير التقدم التربوي على المستويات الدولية والوطنية Educational Progress، إذ يرى رواد هذه التوجهات أن المعلم مسئول عن إحداث تغيرات سلوكية محددة مسبقاً لدى الطلاب، وعليه أن ينظم الخبرات التعليمية لهم بأساليب تيسر اكتسابهم للمعارف والمهارات والاتجاهات المتعلقة بالمجالات الدراسية وبمختلف الأنشطة المدرسية.

لذلك من الأهمية بمكان تقويم عمل المعلم تقويمًا بنائياً للتأكد من فاعلية تنظيمه للبيئة التعليمية بما يحقق التغيرات المرجوة في سلوك الطلاب، وبالطبع يكون المعلم قادر من غيره على القيام بمثل هذا التقويم. إذ يجب أن يعتمد على أساليب التقويم الذاتي Self Evaluation للتحقق من ذلك، كما يمكن أن يعتمد على ملاحظات الطلاب فيما يتعلق بالتفاعل القائم داخل الفصل المدرسي. ولجعل البيانات الناتجة عن هذه الملاحظات أكثر موضوعية يمكن تصميم بطاقات خاصة للتقويم تدون فيها عبارات مرتبطة بمختلف الأنشطة التي ينبغي أن يوفرها المعلم، وكذلك بطبيعة التفاعل الذي يتم بين المعلم وطلابه وغير ذلك من الأنشطة.

ويجب أن تكون هناك محكّمات واضحة لفاعلية العملية التعليمية يتم في ضوئها تقويم مدى قيام المعلم بأدواره المتغيرة وفاعلية أدائه . فمن المعلوم أن هناك العديد من المتغيرات خارج الفصل المدرسي تؤثر في النتائج التي يهدف المعلم لتحقيقها ولا يكون له سيطرة عليها. كما أن هناك متغيرات ترتبط بفاعلية المعلم مثل الكفايات التي يمتلكها Competencies وتمكنه من مجاله الدراسي وتحمسه وإدارته للالفصل وقدرته على التنظيم والإبداع وغير ذلك. لذلك ينبغي مراعاة هذه العوامل إذا أردنا أن يكون نظام تقويم أداء المعلم أكثر واقعية . وهنا يقع العبء غالباً على المعلم في إعداد أساليب متطورة لتقدير تحصيل الطلاب وإنجازاتهم باستخدام أدوات قياس مناسبة.

تقدير المدرسة: إن المدرسة تعد مؤسسة معقدة التركيب، فهي تشتمل على

العديد من المكونات المتشابكة والمترادفة التي يعكس أثرها على المخرجات Outputs المتمثلة في تحصيل الطلاب ومكتسباتهم وإنجازاتهم في مجالات متنوعة ومحفلة تتعلق بالمنهج الدراسي بمفهومه الشامل ، وإسهامات المدرسة في البيئة المحلية . وتتأثر هذه المخرجات تأثيراً واضحًا بفريق العمل من معلمين وإداريين وعاملين ، وبالمؤسسات الأخرى التي تشارك المدرسة مهامها ، والمنهج الدراسي وأهدافه ، والمباني التعليمية وغيرها من المصادر . ويطلب تقييم المدرسة تقييم جميع هذه المكونات التي تعد أيضاً مكونات مركبة . فتقدير المدرسة إذن يتضمن تقويمًا للبرنامج التربوي برمته عن طريق جمع بيانات ومعلومات تتعلق بمختلف جوانب البرنامج ووظائفه .

والهدف من ذلك هو تحديد مدى تحقيق المدرسة لأهدافها المرجوة ، وتعرف مجالات القوة والضعف في البرنامج التربوي بعامة . وهذه المعلومات تقدم تغذية مرئية تساعد المسؤولين عن إدارة المدرسة في تحديد الخطط والأنشطة المستقبلية واتخاذ القرارات المتعلقة بتمويل هذه الأنشطة بما يتناسب وأهميتها .

ومن المكونات الأساسية التي تهتم بها كثير من الدول في تقويم المدرسة ما يسمى بالبرامج الاختبارية School Testing Programs التي تكون عادة ببرامج متكاملة وشاملة لمختلف المجالات. فعلى الرغم من أن تقويم التحصيل الدراسي للطلاب يعد من مجالات التقويم المهمة للمدرسة ، حيث إن الهدف الرئيسي للمدرسة هو إثراء تعلم الطلاب ، إلا أن هناك مجالات أخرى لا تقل عن أهمية ينبغي تقويمها كما ذكرنا لارتباطها ارتباطاً وثيقاً بتعلم الطلاب . وتقويم المنهج في ذاته ربما لا يقدم دليلاً كافياً على أنماط تحصيل الطلاب . ولذلك ينبغي تطبيق مجموعات من الاختبارات والمقاييس التي تقدم صورة شاملة عن الطالب من حيث سمات شخصيته واستعداداته وميوله ومفهومه عن ذاته ، حيث إن هذه جميعاً تؤثر تأثيراً بالغاً في مستوى تحصيله وفاعلية تعلمه . وهذه المجموعات من الاختبارات ينبغي العناية بانتقادها أو إعدادها وتطبيقها وتفسير نتائجها في ضوء معايير Norms مناسبة ؛ وذلك للإفادة من نتائجها في مراجعة عمل المدرسة . كما ينبغي الإفادة أيضاً من تقويم كفايات المعلمين والإداريين ، والموارد المادية والمالية للمدرسة ، والمباني المدرسية في اتخاذ قرارات تتعلق بالمنظومة التعليمية وأساليب تفعيلها .

تعريف الاختبارات والمقاييس التربوية والنفسية :

نظراً لتعقد وتشابك الظواهر التربوية والنفسية وتعدد التطبيقات العميدانية لقياس فى مجالات انتقاء الأفراد لمختلف الوظائف والأعمال والمهام وتسكينهم وتصنيفهم ، وفي التوجيه والإرشاد النفسي ، وفي تقويم الأفراد والبرامج ، والمؤسسات وغير ذلك ، كما سبق أن رأينا ، فقد تعددت وتنوعت أدوات القياس والتقويم بحيث تناسب طبيعة الظاهرة المراد قياسها وما تشتمل عليه من متغيرات . لذلك أصبح في متناول الباحثين والممارسين العديد من أدوات القياس التي تمكنتهم أن ينتقلاً منها أكثرها ملاءمة للأفراد الذين ستطبق عليهم ، ولطبيعة السمة المعينة . فأدوات قياس الذكاء والاستعدادات والتحصيل مثلاً تختلف عن أدوات قياس القلق والاكتساب ، أو قياس الميول والاتجاهات، أو الجوانب النفسية الحركية . ومما هو جدير بالذكر أنه لا يوجد تصنيف وحيد لهذه الأدوات ، وإنما يمكن تصنيفها وفقاً للعديد من الأبعاد . وفيما يلى أحد التصنيفات الممكنة .

أدوات قياس الجوانب المعرفية في مقابل أدوات قياس الجوانب الوجودانية ،
فالجوانب المعرفية للشخصية المتمثلة في القدرات العقلية والتحصيل تتطلب أدوات قياس يحاول فيها الفرد المختبر بذل أقصى جهده ليحصل على درجة تعبر تعبيراً صادقاً

عن قدرته الفعلية. فالاختبار الذى يقيس الاستعداد اللغوى مثلاً يتطلب من الفرد أن يجيب عن أكبر عدد من المفردات التى تكون عادة من نوع الاختيار من متعدد فى زمن محدد، أما الجوانب الوجданية والسمات المزاجية للشخصية فإنها تتطلب أدوات لقياس سلوك الفرد الذى يميزه عن غيره من الأفراد ، أى سلوكه المعتمد فى المواقف الانفعالية مثلاً. وهذه المقاييس تشتمل على عبارات ليس لها استجابة نمطية محددة وإنما يستجيب الفرد بما يتفق مع خصائصه وسلوكه اليومى المعتمد فى مواقف مشابهة للموقف الذى تشير إليه العبارة. ومن أمثلة ذلك: أحب تسيق الزهور، أو أحبذ الاختلاط في الجامعة ، أو أعتقد أن الناس لا تجربنني، أو تتبانى وساوس كثيرة، وهكذا. ونظرًا لأن هذه العبارات أو التساؤلات ليس لها إجابة صحيحة أو إجابة خطأ، وإنما تختلف الإجابة باختلاف وجهة نظر أو مشاعر أو معتقدات الفرد، فإن قياس هذه الجوانب يكون أكثر صعوبة من قياس الجوانب المعرفية، حيث يتطلب أدوات قياس تعتمد على أساليب الملاحظة الموضوعية، والتقارير الذاتية، والاستبيانات وغيرها من أساليب التقييم الشخصى . Personal Assessment

أدوات قياس جماعية في مقابل أدوات قياس فردية؛ فأدوات القياس الجماعية تطبق على مجموعة كبيرة من الأفراد في وقت واحد، بينما أدوات القياس الفردية تصمم بحيث تطبق على فرد واحد في وقت واحد، فاختبارات الذكاء الجماعية يمكن أن تطبق على تلاميذ فصل مدرسى في آن واحد، أو على المتقدمين للالتحاق بالخدمة العسكرية، وكذلك اختبارات التحصيل واختبارات الالتحاق بالكليات وغيرها. غير أن هناك موقف تتطلب استخدام أدوات قياس فردية مثل : موقف الإرشاد والتوجيه التربوى والمهنى أو قياس ذكاء الأطفال، وتطبق هذه الأدوات بواسطة أخصائى متخصص ومدرب تدريجياً على استخدامها، حيث يقوم بـ ملاحظة توجه الفرد المستجيب وأداؤه، وكذلك يستطيع أن يتحاور معه حول استجاباته لكي يفهمها الأخصائى جيداً، أى أن أدوات القياس الفردية تسمح بقدر كبير من التفاعل بين الفاحص والمفحوص، وتكوين ألفة بينهما، وبذلك تقدم فرصاً ثرية لللاحظات الكlinيكية. وهناك اختبارات فردية متعددة لقياس الذكاء والشخصية سوفتناولها فى البالىن الثانى والثالث.

ومما لا شك فيه أن أدوات القياس الفردية يستغرق تطبيقها وقتاً طويلاً، وتتطلب غرفة خاصة مجهزة تجهيزاً مناسباً لتطبيق الاختبار في ظروف ملائمة؛ لذلك فإن هذه الاختبارات الفردية لا تستخدم إلا إذا كان هناك ضرورة لاتخاذ قرار مهم يتعلق بفرد ما، كأن يترتب على هذا القرار نقل طفل من الفصول المعتادة إلى فصول المتخلفين عقلياً مثلاً، أو تشخيص اضطرابات النفسية.

أدوات قياس بمثابة عينة في مقابل أدوات قياس بمتابهة مؤشر . تعتمد

بعض أدوات القياس وبخاصة اختبارات التحصيل في بنائها على الانتقاء العشوائي لعينة من المفردات التي تقيس هدفًا أو أهدافًا تعليمية معينة بحيث تمثل هذه المفردات جميع المفردات التي يمكن أن تقيس هذا الهدف أو هذه الأهداف ، ومن ثم يمكن الاستدلال على تحقق الهدف أو الأهداف من هذه العينة من المفردات الاختبارية ، فمثلاً إذا أراد معلم أن يقيس مهارة الطالب في تعرف معنى الرموز الكيميائية ، فإنه ينتقى عشوائياً عينة ممثلة لهذه الرموز ويختبر الطالب بها ، ومن هذه العينة الصغيرة يستطيع المعلم الاستدلال على تتحقق هذا الهدف لدى الطالب ، أي أن الاختبار يقيس عينة سلوكية Sample Test ، أما إذا استخدم اختباراً معيناً للتمييز بين مجموعتين من الأفراد في سمة معينة تختلف بما يقيسه الاختبار بالفعل فإن الاختبار يكون بمثابة مؤشر Sign للتمييز بين المجموعتين في هذه السمة .

فيإذا اشتمل اختبار عقلى معين على مفردات تقيس قدرة التلميذ على تذكر مجموعة من الأرقام في تسلسل معين ، وتوصلت الدراسات إلى أن عدم قدرة الفرد على ذلك تميز بين المتخلفين عقلياً والعاديين ، ففي هذه الحالة لا يصبح قياس تذكر الأرقام هو الهدف من تطبيق الاختبار ، وإنما يستفاد من الاختبار كمؤشر لأغراض معينة .

أدوات قياس لفظية في مقابل أدوات قياس الأداء؛ فمعظم أدوات القياس

تطلب استخدام اللغة سواء في تعليمات القياس أو مفراداته ، لذلك تسمى أدوات قياس لفظية Verbal Tests مثل معظم اختبارات الجوانب المعرفية والوجودانية . ولكن هناك أنواع من أدوات القياس تستخدم مع الأطفال الصغار والأميين والمعاقين لغوياً، تشتمل على أشكال أو رسوم أو صور أو متاهات أو مواد عينية ، فهذه تسمى أدوات قياس غير لفظية Non - Verbal ، إلا أن الفاحص يقرأها جهرياً . ولا تتطلب عادة أدوات قياس الأداء إجابات تحريرية لأسئلة معينة ، وإنما تتطلب أداء مطلوب يتعلق بالمؤشرات العينانية المعطاة ، ومثال ذلك بعض الاختبارات الفردية لقياس الذكاء والاستعدادات المكانية والميكانيكية ، ومقاييس الجوانب النفسية الحركية .

ومن الجدير بالذكر أن كثيراً من الاختبارات الفردية للذكاء تتضمن أجزاء لفظية وأجزاء أدائية ، وبذلك يمكن قياس الجانب غير اللفظي في الذكاء ، وكذلك المقارنة بين درجات كل من الجزئين لتعرف جوانب القصور اللغوي لدى الفرد .

أدوات قياس تتطلب السرعة في مقابل أدوات قياس تتطلب القوة، تباين

أدوات القياس في درجة تحديدها لزمن الإجابة . فهناك اختبارات موقوتة ، أي يتحدد فيها زمن الإجابة تحديداً دقيقاً ولا يسمح للفرد تخطي هذا الزمن ، وعادة يشتمل الاختبار على مفردات سهلة نسبياً ، ولكن نظراً لمحدودية الوقت فإن كثيراً من الأفراد لا يستطيعون إجابة جميع مفردات الاختبار في الزمن المحدد باستثناء النابهين .

ونظراً لأن الهدف من هذه الاختبارات قياس سرعة الأداء ، فإن تباين درجات الأفراد لا ترجع إلى صعوبة مفرداته ، وإنما إلى اختلاف معدل سرعتهم في الإجابة ، وكثير من اختبارات الذكاء والاستعدادات تعد من نوع اختبارات السرعة Speed Tests . أما أدوات القياس التي تهم بما يستطيع الفرد أداءه بغض النظر عن سرعته ، فإنها تسمى اختبارات القوة Power Tests ، وهذه الاختبارات تصمم لقياس قدرة الفرد على حل مشكلات تميز بالصعوبة بغض النظر عن الزمن المطلوب للإجابة . ولعل قياس تحصيل الطالب يعتمد على هذا النوع من الاختبارات ، حيث يكون التركيز على تقسيم ما حققه الطالب من معارف ومهارات ، لذلك يعطى فرصة كافية للتفكير والإجابة عن جميع المفردات قدر إمكانه .

والحقيقة أنه لا توجد حدود قاطعة بين نوعي الاختبارات ، إذ يمكن أن يكون الاختبار سرعة لمجموعة معينة ، وقوة لمجموعة أخرى . فمثلاً اختبار فهم معاني الكلمات يعد اختبار سرعة لطلاب المرحلة الثانوية إذا كانت مفرداته سهلة ومناسبة لطالب المرحلة الابتدائية ، ولكنه يعد بمثابة اختبار قوة لهؤلاء التلاميذ .

لذلك ينبغي مراعاة التحديد الدقيق لمثل هذه الاختبارات التي تعتمد جزئياً على السرعة عند إعدادها .

أدوات قياس موضوعية في مقابل أدوات قياس ذاتية: فالموضوعية

Objectivity والذاتية Subjectivity تتعلقان بطريقة تصحيح الاختبارات والمقاييس التربوية والنفسية وليس بنوع الاختبار أو المقاييس . فقد سبق أن أوضحتنا مفهوم الموضوعية كأحد مكونات تعريف الاختبار ، وذكرنا أنه يعني عدم تأثر تصحيح الاختبار بالأحكام الذاتية للقائمين بها . فالاختبارات التي تشتمل على مفردات اختيار من متعدد مثلاً يتسم تصحيحها بالموضوعية نظراً لوجود مفتاح تصحيح يطبق على أوراق الإجابات ، بل ويمكن تصحيحها بواسطة الحاسوب؛ لذلك لا تختلف درجة الفرد باختلاف القائمين بعملية التصحيح ، ولكن يمكن أن تختلف الدرجة التي يحصل عليها الفرد في اختبار يشتمل على أسئلة مقالية أو أي نوع آخر من الأسئلة التي تتطلب إجابة مفتوحة باختلاف المحكمين أو القائمين بعملية التقدير ، حيث إن عنصر الذاتية يؤدى إلى هذا الاختلاف . ويمكن تلافي ذلك بالعناية بتحديد المطلب الذي يصوغه السؤال

تحديداً دقيقاً لا يشوبه الغموض، وكذلك الاتفاق على محكّات أو عناصر أساسية ينبغي توافرها في الإجابة للاسترشاد بها في تقدير الدرجات مما يقلل من الذاتية، ويزيد من اتساق الدرجات، وبذلك يكون التصحيح أكثر موضوعية. عموماً، فإن تقنيّن عمليات تطبيق الاختبار أو المقياس وتعليماته وطريقة تصحيحه يمكن أن تسهم في تحقيق الموضوعية المرجوة بغض النظر عن نوع أداة القياس المستخدمة.

أدوات قياس يُعدها المعلم في مقابل أدوات قياس مقتنة؛ فالمعلم يهتم بقياس مدى تحقيق تلاميذه للأهداف التعليمية للمجالات الدراسية ليس فقط لوضع تقديرات لكل منهم في نهاية مدة دراسية معينة، وإنما لتبيين جوانب القوة والضعف في تحصيلهم لكي يتسمى له اقتراح أساليب علاج مناسبة للطلاب الذين يجدون صعوبة في تحقيق هذه الأهداف.

لذلك فإنه يقوم بإعداد اختبارات يكون لها وظيفة تعليمية مهمة في تحسين دافعية الطلاب وحثّهم على بذل المزيد من الجهد لإثراء تعلمهم، وعادة تكون هذه الاختبارات من نوع الاختبارات المرجعية المحكّ، غير أن هذا يعتمد على تمكّن المعلم من المهارات الأساسية في بناء مثل هذه الاختبارات الصافية.

ولكن المعلم يهتم في بعض الأحيان بمقارنة تحصيل طلاب فصل أو فرقة معينة في مجال أو أكثر من المجالات الدراسية بمجموعة خارجية مماثلة لمجتمع هؤلاء الطلاب ولتكن مجموعة طلاب الفرقة نفسها على المستوى الوطني التي تسمى الجماعة المرجعية Norm Group ، فإنه عندئذ ينبغي أن يعتمد على اختبارات تسمى الاختبارات المقنتة Standardized Tests . ويقوم بناء هذه الاختبارات خبراء في القياس التربوي والمناهج والمعلمين، وتُجرى عملية التقنيّن عن طريق تطبيق الاختبار وتصحيحه في ظروف موحدة ومنظمة على الجماعة المرجعية ، والتوصيل إلى معايير Norms يمكن في ضوئها مقارنة نتائج مختلف فصول فرقة معينة أو مدرسة معينة بأقرانهم على المستوى الوطني . فالدرجة التي يحصل عليها أحد طلاب فرقة معينة في إحدى المدارس في مجال دراسي معين يمكن مقارنتها بمعايير الجماعة المرجعية الذي تم التوصيل إليه لمعرفة المستوى التحصيلي لمجتمع طلاب هذه الفرقة في المجال نفسه ، لذلك فإن مثل هذا الاختبار يكون من نوع الاختبارات المرجعية الجماعة أو المعيار .

ومن الجدير بالذكر أن معظم الاختبارات والمقياسات النفسية تعد أيضاً من هذا النوع؛ نظراً لأنها تعتمد على مقارنة أداء الفرد بأداء أقرانه في الجوانب العقلية والوجدانية والمزاجية. أي أن الاهتمام يكون منصباً على ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم في الاختبار ، ويكون للدرجة تفسيراً فقط في ضوء معيار Group Norm يعتمد على متوسط أداء الجماعة التي يتسمى إليها. وسوف نوضح هذا المفهوم بالتفصيل في الفصل السادس.

خلاصة

تعتمد الدراسة العلمية للسلوك الإنساني على أساليب القياس والتقويم التربوي والنفسى، كما ينبغي أن تستند القرارات المتعلقة بالأفراد، مثل قرارات الانتقاء والتصنيف، والفحص والتشخيص، والتسكين والتوفيق، والتشو吉ه والإرشاد، وكذلك تقويم المدرسة والمناهج والبرامج والمشروعات التربوية المستحدثة إلى هذه الأساليب. ويتضمن مفهوم القياس عملية جمع المعلومات وتنظيمها، وكذلك نتائج هذه العملية. ونظرًا لأن الظواهر السلوكية تميز بالتعقد والتشابك وتعدد المتغيرات، فإن قياسها يتطلب ابتكار طرق وأساليب وأدوات قياس متعددة تميز بخصائص محددة. وقياس السمات الإنسانية يتطلب معرفة طبيعة هذه السمات ومستويات قياسها والقواعد والإجراءات التي تكشف عنها.

لذلك فإن مدخل السمات في القياس التربوي والنفسى يوضح كيفية تعريف السمات المراد قياسها تعريفاً إجرائياً، ويحدد الشروط التي يجب أن تتوافر فيها، وخصائصها الأساسية التي تيسر عمليات قياس السلوك الذى يستدل منه عليها. وبالطبع تبيان أنواع هذه السمات، حيث إن هناك سمات مورفولوجية، وسمات فسيولوجية، وسمات عقلية معرفية، وسمات وجدانية، وسمات مزاجية، وسمات نفسية حركية، وغير ذلك. ويمكن قياس هذه السمات بأدوات متعددة ينبغي أن تفى بشروط معينة، مثل التقنيين، والموضوعية، وتمثيلها للسلوك المراد قياسه، وتحديد معيار جماعة مرجعية، أو محك أداء مقبول. ونظرًا لعدد هذه الأدوات، فإنه يمكن تصنيفها من حيث الغرض من القياس، ونوع السمة المراد قياسها، وطريقة التطبيق، والمحتوى، وأسلوب الأداء، وطريقة التصحيح، وكيفية البناء، وتفسير الدرجات المستمدة منها.

الفصل الثاني

مشكلات القياس التربوى والنفسى
وشروطه العلمية وأخلاقياته

* مقدمة *

- * بعض المشكلات الأساسية فى القياس والتقويم فى الوطن العربى
- * بعض الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس
- * الشروط العلمية لانتقاء الاختبارات والمقاييس
- * الشروط العلمية لتفسير درجات الاختبارات وتقدير نتائجها
- * بعض ضوابط وأخلاقيات استخدام الاختبارات النفسية والتربوية

مقدمة:

اتضح لنا من الفصل الأول أن القياس التربوي والنفسى أكثر صعوبة من القياس الفизيائى، وربما أقل دقة بسبب طبيعة الظواهر السلوكية وتعقدها وتشابكها. إذ تتأثر هذه الظواهر بالعديد من المتغيرات الصريحة والضمنية مما يجعل عزل المتغير أو السمة المراد قياسها عن غيره من المتغيرات من المشكلات التى تتطلب جهوداً بحثية متواصلة.

لذلك فإن استخدام الاختبارات والمقاييس التربوية والنفسية ليس ببساطة استخدام مقاييس الظواهر الطبيعية، وإنما يتطلب دراية تامة بأساسيات علم القياس وأساليبه وأدواته وتطبيقاته، والإفادة الواجبة من التطورات المستمرة التي حدثت فى مختلف مجالاته، ومعرفة مشكلاته وكيفية التغلب عليها بأساليب مناسبة.

كما يتطلب أطراً بشرية مؤهلة تأهيلاً علمياً ومهنياً فاعلاً، ومستدركة تدركاً سليماً على استخدام أدوات القياس المتنوعة، من مثل الاختبارات، والاستبيانات، والقوائم، وبطاقات الملاحظة، والبروتوكولات وغيرها استخداماً واعياً مستبصراً، وتفسير نتائجها تفسيراً مناسباً يتفق وطبيعة أدوات القياس المستخدمة. كذلك يتطلب معرفة ما يتربى على القرارات التى تتخذ فى ضوء نتائج القياس من مشكلات اجتماعية وأخلاقية. لذلك فإن الالتزام بالمعايير والمبادئ الأخلاقية التى تحكم ممارسات استخدام الاختبارات والمقاييس التربوية والنفسية التى أقرتها المؤسسات المعنية المتخصصة يصبح أمراً أساسياً وضرورياً.

لذلك فإننا سنعرض فى هذا الفصل بعض مشكلات القياس التربوى والنفسى، وبعض الأخطاء التى يمكن أن تنتجم عن إساءة استخدام الاختبارات والمقاييس، والشروط والضوابط التى تحكم عمليات انتقاء وتطبيق واستخدام أدوات القياس والتقويم فى مختلف المجالات التطبيقية التى أوضحتها فى الفصل الأول، وكذلك أخلاقيات استخدام هذه الأدوات.

بعض المشكلات الأساسية في القياس والتقويم في الوطن العربي

إن الخدمات التي تقدمها مختلف المؤسسات التربوية والنفسية تستهدف مساعدة الأفراد في الإفادة من قدراتهم وامكاناتهم وطاقاتهم إلى أقصى درجة ممكنة، وتنمية هذه الطاقات، وتمكينهم من وضع أفضل الخطط التي تسهم في تحقيق الأهداف التي يسعون إليها، والتغلب على مشكلاتهم سواء التعليمية أو الحياتية. ويتطلب تقديم هذه الخدمات على أفضل وجه الاستعانة بجميع الوسائل الفنية والاختبارات والمقاييس التربوية والنفسية.

وتُقدَّم هذه الخدمات من خلال المؤسسات التعليمية والتربوية ومراكز العلاج المتخصصة وغيرها. ويشرف على الخدمات النفسية عادة إدارات أو مراقبات لهذه الخدمات تتضمن أقساماً مختلفة من بينها قسم للاختبارات والمقاييس أو القياس والتقويم يختص بإعداد وتقنين أدوات القياس الجماعية والفردية في المجالات النفسية والتربوية والمهنية المتعلقة بعمليات الخدمات النفسية والتربوية، وكذلك إعداد وتقنين أساليب جمع البيانات التي تحتاجها عمليات الإرشاد النفسي والتربوي، وتوفير المقاييس النفسية التي تستخدم في دراسة وتشخيص الحالات.

غير أن هناك - في رأينا - عدداً من المشكلات المتعلقة بالقياس والتقويم التربوي والنفسى تعيق فاعلية هذه الخدمات من أهمها :

إسناد مهام القياس النفسي والتربوي إلى غير الاختصاصيين:

فالأفراد الذين يوكل إليهم القيام بمهام القياس النفسي والعمليات الاختبارية تتباين مؤهلاتهم وأساليب إعدادهم ومدى تدريتهم أو درجة خبرتهم، ويزاولون أحياناً هذه المهام التي تؤثر تأثيراً بالغاً في حياة الأفراد دون تصريح رسمي. إذ أحياناً يقوم بعض أمناء أو مساعدى المختبرات النفسية بانتقاء الاختبارات وتطبيقها، وتقديم نتائجها للمفحوصين، بل وأحياناً يقوم بتطبيق أو استخدام الاختبارات الفردية مثل مقياس بيئي Wechsler Binet أو ويكسنر Rorschach في الذكاء، أو اختبار الرورشاخ، أو اختبار TAT في الشخصية مساعدى باحثين أو أخصائين نفسيين غير مدربين تدريباً كافياً على استخدام هذه الأدوات وتفسير نتائجها في ضوء المحددات المختلفة للاستجابات.

وهذا الأمر يتطلب من المؤسسات المعنية بالخدمات النفسية أن تقوم بإعداد مؤشرات عن مؤهلات العاملين في هذا المجال ونوعية تدريتهم ومدى خبرتهم للاسترشاد بها في تحديد كفايات هؤلاء العاملين. ونحن لا نقترح هنا ضرورة حصول القائمين بأعمال التقويم ومن يسند إليهم مهامه على درجة الدكتوراه في علم النفس،

وإنما يمكن العمل على تطوير برامج للدراسات العليا في الجامعات والمعاهد العليا تمنح الملتحقين بها دبلومات تخصصية في مختلف المجالات التطبيقية لعلم النفس على أن يكون القياس والتقويم النفسي والتربوي أحد هذه المجالات الرئيسية، وأن يكون التركيز فيه على الجوانب التطبيقية للاختبارات والمقاييس كما هو الحال في بعض الكليات. وكذلك يمكن أن يسند الأمر في هذا المجال إلى اختصاصيين نفسيين حاصلين على درجة الماجستير في علم النفس إضافة إلى حصولهم على دورات تدريبية تختلف باختلاف أنواع الاختبارات، وأن يتزموا بالمستويات والمعايير والضوابط المهنية في هذا المجال كما سنعرض فيما بعد.

خبرة الاختصاصيين النفسيين المحدودة بنظر리ات القياس النفسي والتربوي والأسس التي يقوم عليها،

فالقياس والتقويم النفسي والتربوي علم وفن له أسسه ونظرياته وأصوله ومهاراته، ويجب ألا يتصدى له إلا من ملك ناصية علومه وتدرُّب على فنونه، فتفسير نتائج الاختبارات دون دراية كافية بأسسها النظرية يجعل الاختصاصي النفسي كالفنى وليس كالاختصاصى المهني؛ لأنَّه بذلك يقوم بتنفيذ ما يوصى به بدلًا من تسلحه بالمعارف والمهارات التي تمكَّنه من انتقاء أو بناء أدوات قياس وأساليب تقويم تناسب أغراضًا معينة وأفرادًا معينين.

وترتبط هذه المشكلة ارتباطاً وثيقاً بالمشكلة السابقة وهي إسناد هذه المهام إلى أفراد ينقصهم الإعداد والخبرة. فمثلاً يوكل أمر بناء الاختبارات التحصيلية والتربوية في المؤسسات التعليمية إلى معلمين لم يدرسوا علم القياس النفسي والتربوي دراسة وافية، ولم ينالوا قدرًا كافياً من التدرب والخبرة في هذا المجال بما يمكنهم من إعداد هذه الأدوات إعداداً علمياً سليماً، وتفسير نتائجها في ضوء المحکمات والمعايير المحددة. لذلك أوصى اجتماع الخبراء في الحلقة الدراسية الإقليمية لليونسكو حول تنوع التعليم الثانوى وربطه بالعمل المتعجل في الدول العربية في الكويت عام ١٩٨٧ وغيره من المؤتمرات العربية المتخصصة بضرورة تطوير مقررات القياس والتقويم النفسي والتربوي في برامج إعداد المعلمين بحيث تراعى الأساليب والتقييمات الحديثة في هذا المجال.

قلة الاختبارات المقتننة واختبارات الكفايات في البلاد العربية،

في قيام الاختصاصي النفسي بالمهام الموكلة إليه بأسلوب علمي موضوعي يتطلب توافر أدوات قياس مقتننة لمختلف جوانب شخصية الفرد ومقاييس لكتفاليات الأداء في مختلف الأعمال وال المجالات. غير أن المتاح من هذه الأدوات يُعد قليلاً إذا ما قورن

بغيرها من البلاد، وهي في معظمها مقتبسة من ثقافات أخرى تختلف عن ثقافتنا مما يكون له تأثير معين على الأداة بعد تعريبها وتقنيتها على البيئة المحلية.

فالاصل في المقياس أن يتفادى ما يشوهه، ولكن الآثر الشعافي لا يزال واضحاً في الكثير من المقاييس المعتمدة في بلادنا، ومن هذه الزاوية يكون الضعف أو التحيز الذي سبق أن أشرنا إليه. وهذا التحيز لا يظهر في التقليل من شأن الأفراد فحسب بل يظهر كذلك في منح آخرين درجات أكثر مما يستحقون، وكذلك لا يكون التحيز ضمن المجتمع الواحد فحسب، بل يتجاوز ذلك إلى مجتمع آخر يأخذ بالمقاييس دون توفير معايير مناسبة لهذا المجتمع.

فالمقاييس ينبغي أن يجري إعدادها اعتماداً على عينة كافية وضمن شروط علمية محددة. وأن يكون قادراً في البيئة التي وضع لأفرادها أن يقيس ما وضع لقياسه. ومن هذه الزاوية يكون من الخطير اعتماده في بيئه أخرى تختلف كثيراً عن السابقة دون إجراء التعديلات التي تقتضيها عمليات التغيير أو التقنيات ليتناسب مع البيئة الجديدة، وهذا أمر يحتاج إلى عناية فائقة.

كما أن الأصل في المقياس أن تكون لغته مناسبة لمن يفحص به . ولكن وجود لغة يومية، أي لغة دارجة تسود في مجتمع ما وتختلف من جهات عن اللغة التي تظهر في الكتب والصحف اليومية والدوريات تطرح مسألة حادة وخاصة في البلاد العربية .

وقد حاولت بعض هذه البلاد كتابة عبارات أو أسئلة أو بنود الاختبارات والمقاييس النفسية باللغة اليومية «الدارجة» التي يفهمها عامة الناس. غير أن هذا يتطلب وقفة تأمل حول الآثار والمضامين الثقافية الناجمة عن التوسيع في هذا الاتجاه في مجال القياس النفسي وبناء الاختبارات.

ومن المشكلات الهامة الأخرى في هذا المجال استخدام الاختبارات التحصيلية التي يكتبها المعلمون في أغراض متعددة مثل الانتقاء، والفحص، والتوفيق والتوجيه، وعدم إيلاء الاختبارات التشخيصية والبنائية أهمية تذكر. فتحن نعاني في بلادنا من نقص شديد في هذه الأنواع المهمة من الاختبارات التي تحدد أوجه القصور ومواطن القوة في تحصيل الطلاب بالنسبة لمجموعة محددة من المعارف والمهارات. وترشد العملية التعليمية وتصوب مسارها نحو تحقيق أهدافها المرجوة. كما نعاني من قلة الاختبارات التحصيلية المقنة في مختلف المواد الدراسية والتي تفيد في تحديد المركز النسبي للطالب بين زملائه في محتوى أو محتويات دراسية معينة .

وعلى الرغم من الجهدات التي أفادت من التطورات التي حدثت في هذا المجال منذ العقود الماضيين إلا أنه ينبغي تجميع وتنظيم هذه الجهدات وتكييفها في مؤسسات متخصصة.

ويمكن أن يتم ذلك بتحديد الجهات أو المؤسسات التي يُسند إليها بناءً على الاختبارات والمقاييس وتزويدتها بالخبراء والاختصاصيين والتقيين المناسبة. كما ينبغي حصر وتصنيف ما هو متوافر في الهيئة المحلية من اختبارات نفسية وتربيوية ومراجعتها بواسطة خبراء القياس والتقويم لتحديد الصالح منها، وإعادة النظر في الاختبارات التي تتطلب ذلك.

بعض الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربيوية :

على الرغم من أن الاختبارات والمقاييس النفسية والتربيوية تعد أحد مصادر جمع المعلومات؛ إلا أنها بما تميز به من خصائص مهمة، من مثل الموضوعية والتقيين والمعايير والثبات والصدق، تعد من أهم هذه المصادر (Cronbach, 1970). غير أن سوء استخدام هذه الأدوات من قبل أفراد غير متخصصين أو غير متربين تدرّبًا عاليًا، وغير مطلعين على التطورات المتنامية في مجال القياس والتقويم النفسي والتربيوي، يؤدي إلى أضرار بالغة بمن تطبق عليهم هذه الأدوات، ويفقدوها فعاليتها، ويحيد بها عن تحقيق الأغراض الأساسية منها، بل وتصبح مصدرًا من مصادر الشك والنقد من مختلف فئات المجتمع (Aiken, 1985).

وافتقارنا في الوطن العربي إلى أسس ومعايير وضوابط واضحة ومحددة نابعة من ممارساتنا المهنية الصحيحة في مجال استخدام أدوات القياس النفسي والتربيوي، ومسترشدة بواقعنا ومبادئنا الأخلاقية والاجتماعية والثقافية، يؤدي إلى وقوع الاختصاصيين النفسيين في أخطاء ناجمة عن سوء استخدام هذه الأدوات وعن تقديم نتائجها إلى غير المتخصصين أو إلى المفحوصين.

فالمارسات الخاطئة لاستخدام الاختبارات والمقاييس النفسية لا تؤدي فقط إلى قرارات غير صائبة، وإنما لها آثارًا ضارة بالمفحوص وانعكاسات اجتماعية سيئة.

ولا تتعذر الحقيقة إذا قلنا أن هذه الممارسات الخاطئة كانت سببًا أساسياً في الهجوم الشديد والنقد المستمر للختبارات والمقاييس النفسية ذاتها.

ومن أهم هذه الممارسات الخاطئة وما يترتب عليها من مشكلات ما يلى:

تحيز المسؤولين عن البرامج الاختبارية ، فبعض المسؤولين عن البرامج الاختبارية أو التقويمية لا يراعون تباين الأفراد المختبرين في خلفياتهم الثقافية وخصائصهم الشخصية المميزة عند انتقاء الاختبارات والمقاييس النفسية واستخدامها، وهذا يؤثر بلا شك في نتائج الاختبارات ، وبالتالي يقلل من شأن هؤلاء الأفراد.

الثقة التامة في التنبؤ باستخدام الاختبارات ، إن الثقة التامة في التنبؤ باستخدام الاختبارات والمقاييس المقننة يعد من الممارسات الخاطئة . فالتنبؤ بالسلوك المستقبلي باستخدام مختلف أنواع الاختبارات والمقاييس ينبغي أن يكون في حدود معينة من الثقة . إذ إن هذا التنبؤ يرتبط بعوامل متعددة تتعلق بالفرد المختبر ، والسلوك المراد التنبؤ به ، والمدة الزمنية للتنبؤ ، وخصائص المقياس المحك الذي يقدر في ضوئه فاعالية التنبؤ ، وإذا قلت الثقة في التنبؤ بدرجة ملحوظة يصبح استخدام هذه الأدوات مضلاً ، وبخاصة إذا كان القائم باستخدامها شخص غير مدرب أو قليل الخبرة .

كما أن استخدام الاختبارات في أغراض الانتقاء والتصنيف التي لم يتم التتحقق من صدقها أو اتّباع في سبيل التتحقق من ذلك إجراءات واهنة يعد أيضًا من الممارسات الخاطئة . فمشكلة صدق الاختبارات والمقاييس النفسية تتضمن في الواقع الأمر قضياء متعددة ولا تقتصر فقط على مجرد تقدير خاصة من خصائص الاختبار أو المقياس .

فالاستخدام المناسب لل اختبارات يتضمن العديد من المسؤوليات التي يعتمد بعضها على درجة ثبات وصدق الاختبار ، ويعتمد البعض الآخر على الظروف المختلفة المتعلقة بتطبيقات الاختبار .

الاستخدام الروتيني لدرجات الاختبارات :

فالدرجات المستمدة من الاختبارات تزودنا بقاعدة من البيانات تفيد في تصنيف الأفراد تصفيًّا اعتبارياً . وهذا يعني أن هذه الدرجات ليست قياسات ثابتة ولكنها قيم تقديرية لسمات إنسانية ، وبالتالي فهي تحتوى على خطاء معاينات مختلفة (Anastasi, 1982) . غير أن هذه الاختبارات تستخدم عادة بطريقة روتينية لا تسم بالمرونة الكافية التي تأخذ مختلف المتغيرات المؤثرة بعين الاعتبار . وهذه الممازنة الخاطئة تؤدي في بعض الأحيان إلى تصنيف التلاميذ مثلاً في مجموعات بحسب قدراتهم اعتماداً على اختبارات أو مقاييس ربما يصعب تبرير استخدامها ، أو استخدام درجات فاصلة Cut - off Scores في الاختبارات واعتبارها محركات شبه ثابتة لانتقاء الأفراد في مجالات عمل معينة في المؤسسات الصناعية .

افتراض أن الاختبارات العقلية تقيس خصائص موروثة :

فأحياناً يفسر بعض مستخدمي الاختبارات درجاتها على أنها قياسات لخصائص موروثة ، ومثال ذلك سمة الذكاء . وهذا التفسير الخاطئ له آثار غير مرغوبـة ، إذ إنه

يؤدي إلى اعتبار أن القدرات الإنسانية ثابتة أو غير قابلة للنمو، بل وربما يؤدي إلى النظر إلى الإنسان نظرة تصنيفية، لأن نقول أن هذا مختلف عقلياً وهذا عقري وما شابه ذلك، ويرى بعض النقاد أن هذه نظرة غير إنسانية.

أخطاء ناجمة عن تقديم الاختبارات إلى غير المتخصصين أو إلى المفحوصين:

هناك ممارسات خاطئة أيضاً تترجم عن تقديم نتائج الاختبارات والمقاييس النفسية إلى أفراد غير متخصصين أو إلى المفحوصين. ومن بين هذه الأخطاء ما يلى:

مسايرة التوقعات لنتائج الاختبارات: فقد بنت دراسة روزنثال وجاكوبسون (Rosenthal & Jacobson, 1970) أنه إذا أعطى للمعلمين معلومات افتراضية - أي ليست حقيقة - عن قدرات تلاميذهم فإن التحصيل الفعلى لهؤلاء التلاميذ يستفز وتوقعات المعلمين عنهم. فالתלמיד الذي توقع المعلم أن يكون تحصيله منخفضاً كان تحصيلهم كذلك بالفعل. والعكس كان صحيحًا. لذلك فإن الذين يتقدون اختبارات الذكاء يرون أن هذه الاختبارات تؤثر في أداء الأفراد تأثيراً غير مرغوب فيه. إذ يرون أن المعلمين مثلًا ربما ينظرون إلى التلاميذ مرتفعي الذكاء نظرة تختلف عن نظرتهم للتلاميذ الأقل ذكاء في ضوء ما يتوقعونه من كل منهم.

تأثير مفهوم الذات لدى الفرد بنتائج الاختبارات: فالأشخاص النفسي غير المتمرس ربما يقدم نتائج الاختبارات والمقاييس النفسية للمفحوص بطريقة تؤثر سلبيًا في مفهومه عن ذاته. إذ ربما يؤدي تعليم المفحوص لنتائج الاختبار إلى استنتاج أشياء عن نفسه ليس لها ما يبررها، فالشاب الصغير الذي يحاول جاهدًا أن يكون مفهومًا إيجابياً عن ذاته ربما يتأثر هذا المفهوم لديه ويقل مستوى طموحه إذا علم أن درجاته منخفضة في بعض الاختبارات. فالتقارير المقتصبة لنتائج الاختبارات وما يبني عليها من توصيات تقدم للأفراد في بعض مواقف الإرشاد النفسي تؤثر بلا شك تأثيراً سلبياً في نظرتهم إلى أنفسهم ، وفي تشكيل اتجاهاتهم نحو ذاتهم.

والآن ماذا بعد...؟ هل إساءة استخدام الاختبارات والمقاييس النفسية والتربوية تعنى المناداة بعدم استخدامها في مجالات الخدمات النفسية والتربوية؟

إن كرونباك (Cronbach, 1970) يغنينا عن إجابة هذا السؤال بقوله: «إن السياسات الحكيمية لا تدفع عن الاختبارات ولا تهاجمها، وإنما ترشد كيفية استخدامها بما يتفق والأهداف المرجوة منها». وهو بذلك يؤكد أن فاعلية أدوات القياس النفسي والتربوي التي يعني بتصميمها وإعدادها يتوقف تحقيقها لأهدافها على نوعية استخدامها في مختلف مجالات الخدمات النفسية والتربوية وكفاءة وخبرة القائمين بها.

الشروط العلمية لانتقاء الاختبارات والمقاييس النفسية والتربوية :

إن إجراء الاختبارات يعد نوعاً من التواصل والتفاعل الاجتماعي بين طرفين أو أكثر، وهذا يتطلب اتجاهًا أخلاقيًا مسئولاً من جانب الفاخص، والرغبة في التعاون من جانب المفحوص، فجميع أشكال التفاعل الاجتماعي المثمر تتطلب الثقة والاحترام المتبادل بين الأطراف المتفاعلة.

فإذا كان المفحوص واثقاً من النفع الذي سيعود عليه نتيجة للوقت والجهد الذي يبذله في أداء الاختبارات التي يجريها عليه فاخص متخصص، فإنه سوف يحاول جاهدًا أن يتفاعل معه تفاعلاً إيجابياً، ويتعظ تعليماته؛ لذلك تطبق على المفحوص بهدف مساعدته على فهم نفسه وصنع قراراته.

فكثير من الممارسين النفسيين يعتمدون في عملهم على عدد محدود من الاختبارات دون سواها، إما لأنها مألوفة لديهم واعتمدوا على استخدامها، أو لأنها من إعداد أشخاص من ذوى مكانة علمية متميزة. ولكن هذه الأسباب لا يجب أن تكون مانعاً لهم من البحث عن اختبارات ومقاييس جديدة متنوعة. ويمكنهم تبع هذه الأدوات الجديدة في المطبوعات التي تصدرها دور النشر ومؤسساتها المختلفة، وفي الدوريات العلمية المتخصصة.

غير أن هذا يتطلب توافر مهارات معينة لدى الأخصائيين النفسيين تساعدهم في الحكم على هذه الأدوات، ومدى ملاءمتها للغرض الذي سوف تستخدم من أجله. كما يتطلب توفير سبل توثيق الروابط فيما بينهم لتبادل المعلومات والخبرات في هذا المجال وغيره من المجالات المتعلقة بشئون مهنتهم.

وفيما يلى بعض الأسس العلمية والاعتبارات الفنية التي يجب مراعاتها في انتقاء الاختبارات والمقاييس التي تستخدم في المجالات المختلفة للخدمات النفسية:

توافر خصائص الاختبار أو المقاييس الجيد : فعلى الرغم من تنوع أدوات القياس إلا أنه يجب أن يتوافر فيها خصائص معينة يصعب بدونها الاعتماد عليها في اتخاذ قرارات صائبة تتعلق بالأفراد. ومن أهم هذه الخصائص الموضوعية والثبات والصدق وجود معايير مستمددة من البيئة التي يستخدم فيها الاختبار. ويمكن أن يحصل مستخدم الاختبار أو المقاييس على بيانات تتعلق بهذه الخصائص من دليل الاختبار. وهذا يجعلنا نؤكد أهمية تضمين بيانات وافية ومعلومات كافية تتعلق بهذه الخصائص في أدلة الاختبارات، وأن يكون مستخدم الاختبار على درجة عالية من الكفاية في تقييم هذه الخصائص ومعرفة دلالتها لكي يتمكن من انتقاء أداة القياس الصالحة للغرض

المعين. ولا يجوز انتقاء اختبار أو مقياس من مجرد النظر إلى عنوانه أو اسمه بل لابد من فحص محتوى الاختبار فحصاً مستثيراً والاطلاع على خصائصه السيكومترية وتحديد دلالتها، والتأكد من ملاءمة معاييره للفرد أو العينة المستهدفة، والتحقق من مدى صلاحية هذه المعايير، وسوف تناقش هذه الخصائص مناقشة مستفيضة في الفصول الخمسة التالية.

النحو المناسب للمفهوم : فانتقاء الاختبار المناسب للمفهوم يعد أمراً على جانب كبير من الأهمية، فلا يجوز أن تطبق الاختبارات التي تصلح للراشدين على تلاميذ المدارس الابتدائية أو المتوسطة، حيث تكون غير مناسبة للمستوى النمائي لهم، كما أن معاييرها مستمدّة من عينات تختلف عن عينات هؤلاء التلاميذ. لذلك يجب أن تشير أدلة الاختبارات إشارة واضحة إلى طبيعة عينة تقدير الاختبار وخصائصها، من حيث العمر والنوع والمستوى التعليمي والاجتماعي وغير ذلك لكي يسترشد بها الفاحص في انتقاء الاختبار المناسب لفرد أو مجموعة معينة. وينبغي مراعاة أن وجود مشكلات سلوكية، أو عجز بصر أو ضعف سمع، أو حرمان ثقافي، أو صعوبات في اللغة يجعل الاختبار غير مناسب للمفهوم، ويجب عدم استخدام الاختبار في هذه الحالات.

النحو المناسب للغرض الذي يقيسه : فعلى الرغم من تنوع وتباعد استخدامات الاختبارات والمقياسات النفسية والتربوية، إلا أن الهدف الرئيسي منها جمع بيانات ومعلومات تفيد في اتخاذ قرارات معينة تتعلق بالأفراد أو الجماعات بأقل قدر ممكن من الخطأ أو الصدفة أو المخاطرة. غير أن الاختبار أو المقياس الذي يفيد في موقف معين قد لا يفيد أو لا يناسب موقفاً آخر، فاستخدام الاختصاص النفسي أداة قياس مناسبة لموقف معين أو صالحه لغرض معين يساعده بالطبع في اتخاذ قرارات أفضل من تلك التي يستخدمها دون استخدام تلك الأداة.

لذلك يجب تحديد الغرض من عملية القياس أو العملية الاختبارية تحديداً وأصحاً وانتقاء أفضل الأدوات التي تحقق هذا الغرض. وهذا يتطلب بالطبع توافر عدد كثير من الاختبارات والمقياسات الجيدة التي تقيس مختلف السمات الإنسانية حتى يتمكن الفاحص من انتقاء أكثرها ملاءمة للغرض من القياس. وهنا تبرز أهمية تشجيع حركة القياس النفسي والتربوي وبناء الاختبارات بأنواعها المختلفة في المنطقة العربية من خلال مؤسسات متخصصة.

الشروط العلمية لتطبيق الاختبارات وتصحيفها :

بعد انتقاء أدوات القياس التي تحقق بدرجة أفضل أهداف العملية الاختبارية وتناسب المفحوص أو المفحوصين، تجري عملية تطبيق الأدوات.

وقد يظن البعض أن تطبيق الاختبارات عملية آلية بسيطة لا تحتاج إلى توافر مهارات معينة فيمن يقوم بها، غير أننا إذا ذكرنا أن تطبيق الاختبارات يُعد عملية تفاعل اجتماعي كما سبق أن أشرنا، فإننا سرعان ما ندرك تعدد المتغيرات والظروف النفسية والفيزيائية المتعلقة بال موقف الاختباري التي تؤثر في أداء المفحوص في الاختبارات وفي نتائج العملية الاختبارية. وإذا لم يكن القائم بتطبيق الاختبارات شخصاً مؤهلاً أو مدرباً للقيام بهذا العمل فإنه ربما لا يلتفت إلى هذه المتغيرات.

فالقلق والتوتر، وانخفاض مستوى الدافعية، وتشتت الانتباه، والضعف الحسّي، وعدم الالفة بالاختبارات النفسية، والكذب وتزيف الاستجابات، والتخمين العشوائي، والاتجاه نحو الاختبار ونحو الفاحص، واحتمال عدم فهم المفحوص للتوجيهات بسبب ضعف البصر أو انخفاض المستوى اللغوي ، والظروف الفيزيائية غير الملائمة، تؤثر تأثيراً ملحوظاً في أداء المختبرين. ونحن عادة ننصح في أن تطبق الاختبارات في ظروف مقتنة، ولكن هذا لا يتحقق بدرجة تامة في كثير من المواقف الاختبارية.

لذلك يجب أن تشمل كراسة تعليمات الاختبارات أو ملحقاتها على توجيهات وتعليمات واضحة ومحددة عن كيفية ضبط هذه المتغيرات، وأن يقوم الفاحص بدراستها دراسة متأنيّة قبل تطبيق الاختبار، وكذلك الاهتمام بتهيئة المفحوصين لأداء الاختبار من خلال مقابلات فردية أو في مجموعات صغيرة لإزالة أو تخفيف مصادر القلق والتخوف وغيرها من الانفعالات السالبة، ومحاولة رفع مستوى الدافعية بدرجة تسمح للمفحوص ببذل الجهد المطلوب لأداء الاختبار، ومن ثم ينظم الموقف الاختباري في ضوء هذا الإعداد قبل اليوم المحدد لتطبيق الاختبار.

كما يجب أن تشير كراسة تعليمات الاختبار إلى الشروط اللازم توافرها فيمن سيقوم بتطبيق الاختبار.

وسوف نعرض فيما يلى بعض الشروط العلمية والتوجهات الفنية التي يجب مراعاتها عند تطبيق الاختبارات وتصحيفها:

(١) ينبغي على من يقوم بتطبيق اختبار أو مقياس معين أن يتبع الإجراءات المقترنة المذكورة في كراسة تعليمات الاختبار بعناية تامة، فهذه الإجراءات المقترنة ضرورية إذا كان الأخصائى النفسي يود التوصل إلى قرارات تتعلق بالمحظى اعتماداً على

درجات الاختبار. ولا يجوز الحيد عن هذه التعليمات أو اختصارها أو تعديلها. ومن الضروري تدريب القائمين بتطبيق الاختبارات على أداء متطلبات هذا العمل بكفاءة وفعالية. وبالطبع تختلف هذه المتطلبات باختلاف نوع الاختبار، إذ إن بعض الاختبارات وبخاصة الفردية التطبيق تحتاج إلى إشراف مباشر من الأخصائي (Anastasi, 1982).

(٢) ينبغي تهيئة ظروف متسقة لتطبيق الاختبار تساعد على ضبط متغيرات الموقف الاختباري بقدر الإمكان، فمثلا لا يجوز أن يكون هناك اختلاف كبير في درجة الحرارة أو الرطوبة أو الضوضاء، وما يمكن أن يؤدي إلى تشتت انتباه المفحوصين في الغرفة أو القاعة المخصصة لإجراء الاختبار، وأن يجري فحص دورى على الأجهزة والمواد أو الأدوات الاختبارية، من مثل ساعات التوقيت والأجهزة الإلكترونية والميدوية المختلفة للتأكد من سلامتها وتقنيتها.

كما يجب التأكد من أن المفحوص قد فهم تعليمات الاختبار ونوع الاستجابة المطلوبة منه، وأن يهتم الفاحص بيئة إنسانية يسودها المودة والحرص والمسؤولية، وبخاصة إذا كان هناك اختلاف بين الفاحص والمفحوص في النوع أو المركز الاجتماعي حتى يشعر المفحوص بالاحترام والثقة بالنفس.

(٣) إن الأخصائي النفسي مسؤول عن دقة تصحيح الاختبارات ومراجعة عمليات التصحيح وتسجيل النتائج؛ لذلك يجب أن يطمئن على أن القائم بهذه العمليات يتحرى الدقة التامة، ويبذل جهده في التحقق منها.

وإذا كان التصحيح يتطلب تقدير درجات اعتماداً على أحكام خبراء، فإنه يجب على الأخصائي تحديد ثبات هذه التقديرات للتأكد من اتساقها، إذ إنه ليس من العدل اتخاذ قرارات تمس حياة الأفراد اعتماداً على درجات غير صحيحة أو غير متسقة. وإذا أجرى التصحيح باستخدام أجهزة إلكترونية يجب على الأخصائي النفسي التتحقق من سلامة هذه الأجهزة ودقتها، وأن يتتحقق من النتائج بإعادة التصحيح يدوياً لبعض أجزاء الاختبار، أو من خلال نظام معين يضمن الضبط الكيفي لعملية التصحيح.

وعلى الرغم من تطور الأساليب الآلية في تصحيح الاختبارات، وإعداد التقارير، إلا أن العنصر البشري سوف يظل هو العامل الأساسي في جميع مراحل العملية الاختبارية؛ لذلك يجب الاهتمام بانتقاء الأفراد الذين يوكل إليهم تصحيح الاختبارات والمقاييس النفسية بحيث توافق فيهم صفات المثابرة والدقة التامة واللماحية، ويستطيعون إجراء العمليات الحسابية بسرعة ودقة. وأن تُنظم لهم دورات تدريبية تحت إشراف أخصائيين نفسيين من ذوى الكفاءة العالية.

الشروط العلمية لتفسيير درجات الاختبار وتقدير نتائجها :

إن المرحلة التالية لتطبيق الاختبارات وتصحيحها تتعلق بتفسير درجاتها وكتابه تقارير النتائج، وتقديمها للمفحوص أو الجهة المعنية، وتُعد هذه المرحلة من المراحل المهمة في العملية الاختبارية، إذ يترتب عليها اتخاذ قرارات بشأن الأفراد أو المؤسسات.

وعلى الرغم من أن بعض مؤسسات العمل تحيط نتائج الاختبارات النفسية التي تجريها على المتقدمين للوظائف الشاغرة بها سرية تامة بحيث لا يعلم المتقدم عن هذه النتائج سوى قوله أو رفضه بحجة أن هذه المؤسسات ليست بصدق التوجيه المهني، إلا أن هذا يؤدي بالمحفوس إلى الشك في نتائج الاختبارات ووصفها بالتحيز وعدم النفع. فمن حق المفحوص - أو وكيله أو ولی أمره إذا كان المفحوص فاسداً - معرفة نتائج الاختبارات. وتقديم نتائجها يتطلب قدراً كبيراً من الخبرة والسمارة من جانب الاختصاصي النفسي في تطبيق أسس ومقاييس القياس النفسي والإحصاء، وأساليب الإرشاد النفسي في استخلاص المعلومات التي تقدم للمفحوص.

وفيما يلى بعض الشروط التي يجب مراعاتها في هذا الشأن:

(١) ينبغي العناية بتفسير درجات الاختبارات والمقاييس المقتنة في ضوء المعايير الخاصة بها والمدونة في دليل الاختبار أو المقياس، والالتزام بهذه المعايير وعدم الحيد عنها.

(٢) ينبغي تقديم التقارير التي تتضمن درجات الاختبارات لأفراد مؤهلين لتصفيتها واستخدامها استخداماً مناسباً. أما التقارير التي تقدم للأباء أو المعلمين أو المشرفين على العمل، فمن الأفضل أن تتضمن تفسيراً للنتائج بدلاً من تقديم الدرجات ذاتها. فكثير من هؤلاء الأفراد ربما لا يستطيعون تفسير معنى الدرجات المعيارية أو المئويات وغيرها من الدرجات المحوّلة، وأن يشرف الاختصاصي النفسي على هذه التقارير ويكون مستعداً لتقديم العون والمشورة لهؤلاء الأفراد.

(٣) ينبغي أن تباين أشكال تقارير نتائج الاختبارات بتباين الأفراد أو الجهات التي تقدم إليها هذه النتائج، إذ يجب أن تمكّنهم هذه التقارير من فهم تفسير هذه النتائج بيسر وسهولة.

(٤) ينبغي تجنب استخدام كلمات وصفية مثل متخلّف عقلياً أو عدواني عند تفسير درجات الاختبارات والمقاييس النفسية، أو كلمات مثل راسب أو غير كفء في الاختبارات التحصيلية واختبارات الكفايات، فهذه الكلمات أو الأوصاف التي تُعبر عن

أحكام قيمية تكون عرضة لأنخطاء التفسير من جانب الآباء والمعلمين والمفحوص ذاته. لذلك يفضل وصف أو تحديد السلوك الذي يميز الفرد أو الذي يستطيع أو لا يستطيع أداءه.

(٥) ينبغي الحيطة عند تفسير نسب الذكاء والعمر العقلى ومعايير الفرق الدراسية وما شابه ذلك، إذ إن هذه المعايير يشوبها كثيراً من العيوب التي تؤدى إلى عدم دقة التفسير. وإذا اضطر الاختصاصى النفسي إلى تقديم مثل هذه التائج إلى مؤسسة معينة، فإنه يفضل أن يقدم التفسير مستخدماً الدرجات المعيارية أو الميئيات المناظرة في ضوء عينة التقنيين التي استخدمت في اشتاقاق هذه الدرجات، ويجب أن يشار في التقرير أيضاً إلى صورة الاختبار المستخدمة وتاريخ إجراء الاختبار، وطبيعة الموقف الاختبارى.

بعض ضوابط وأخلاقيات استخدام الاختبارات والمقاييس النفسية والتربوية :

يتضح مما سبق عرضه من شروط علمية واعتبارات فنية لمختلف مراحل العملية الاختبارية أن استخدام الاختبارات النفسية علم وفن، فالاختبارات هي أدوات قياس تعين الأخصائيين في مختلف مجالات الخدمات النفسية في أداء واجباتهم الإنسانية بأسلوب علمي موضوعي، واستخدامها في تحقيق الأهداف التي أعدت من أجلها يتوقف على نوعية هذا الاستخدام وكفاءة وخبرة العاملين في هذا المجال.

ولضمان التزام جميع الأطراف المعنية باستخدام الاختبارات والمقاييس النفسية ينبغي أن تكون هناك مجموعة من الضوابط الأخلاقية التي توجه المسار، وتحدد الممارسات في هذا المجال، وتضفي على المهنة الصفة الكاملة التي تستوجبها من المجتمع، وترعى حقوق وواجبات القائمين بهذه الخدمات والمستفيدين منها.

ولذلك حرصت الجمعية الأمريكية لعلم النفس (APA) على إصدار دليل متجدد يتضمن الأسس والمبادئ الأخلاقية التي يجب أن يلتزم بها الأخصائيون النفسيون في أدائهم لمسؤولياتهم، ومن بينها المبادئ الأخلاقية المتعلقة باستخدام الاختبارات النفسية وأدوات القياس التشخيصية.

كما أصدرت دليلاً للمعايير التي ينبغي توافرها في هذه الاختبارات، وإرشادات عن إجراءات استخدامها. وكذلك أصدرت الجمعية الأمريكية لشئون الأفراد والتوجيه دليلاً يتضمن المبادئ الأخلاقية في مجالات الخدمات النفسية متضمنة مجال تطبيق الاختبارات والمقاييس.

وعلى الرغم من أن هذه الضوابط والأخلاقيات يجب أن تتبع من واقع أعضاء المهنة أنفسهم، إلا أنها سوف نقدم بعض التوجيهات التي يمكن الاسترشاد بها في هذا الشأن.

الكفايات اللازم توافرها فيمن يستخدم الاختبارات: إن وضع ضوابط للكفايات اللازم توافرها فيمن يستخدم الاختبارات والمقاييس النفسية تعد خطوة أساسية لتجنب إساءة هذه الاستخدامات، وتختلف بالطبع هذه الكفايات باختلاف نوع الاختبار والهدف منه، ويمكن تصنيف الاختبارات والأدوات التشخيصية والكفايات اللازم توافرها فيمن يستخدمها إلى ثلاثة مستويات (APA, 1973) :

المستوى الأول: اختبارات أو أدوات تشخيصية يمكن تطبيقها وتصحيحها وتفسير نتائجها بأسلوب مناسب بالرجوع إلى كراسة تعليمات الاختبار ومراعاة مطالب المؤسسة التي يعمل فيها مستخدم الاختبار، ومن أمثلة ذلك الاختبارات التحليلية، واختبارات الكفايات المهنية. ويمكن أن يقوم بتطبيق هذه الاختبارات وتفسير نتائجها معلمين ومسئولي عن المؤسسات المختلفة بحيث يكونون ملمين بدرجة معقولة بأسسيات القياس التربوي النفسي.

المستوى الثاني: اختبارات وأدوات تشخيصية يتطلب استخدامها معرفة أساليب بناء الاختبارات وطرق استخدامها ، ودراسة بعض المقررات في الإحصاء النفسي، والفرق الفردية ، وسociولوجية التوافق، وسociولوجية انتقاء الأفراد، والتوجيه المهني والتربوي . ومن أمثلة ذلك اختبارات الذكاء العام، والاستعدادات الخاصة والاستعدادات الفارقة، ومقاييس الميول، واستبيانات الشخصية . وينبغي فيمن يستخدم هذه الاختبارات والأدوات أن يتحقق هذه المتطلبات المذكورة، أو أن يكون مصريحاً له باستخدامها في إطار مؤسسة تربوية أو صناعية أو إدارة حكومية تعمل فيها، أو أن يكون قد أدى بنجاح دورة تدريبية تتعلق بهذه الاختبارات.

المستوى الثالث: اختبارات وأدوات تشخيصية تتطلب معرفة متعمقة في القياس النفسي وبعض المجالات النفسية المصاحبة، وكذلك تدريياً مكثفاً وخبرة عملية تحت إشراف أخصائي إكلينيكي لمدة لا تقل عن عام. ويفضل أن يكون الأفراد الذين يستخدمون هذه الأدوات من الحاصلين على درجة الماجستير في علم النفس . ومن أمثلة هذه الأدوات: اختبارات الذكاء الفردية، والأساليب الإسقاطية، وبعض مقاييس الشخصية الأخرى.

ويجب على من يستخدم الاختبارات والمقاييس النفسية بعامة أن يكون مزوداً بالمهارات التي تمكّنه من انتقاء الاختبارات التي تتفق والهدف من عملية القياس،

وتناسب الفرد الذى تجرى عليه هذه الاختبارات ، وأن يكون مطلعاً على الدراسات والبحوث المتعلقة بالاختبار أو الأداة المستخدمة ، وكذلك يمكنه تقدير الخصائص السيمكومترية للإختبارات ، من مثل المعايير ، والثبات ، والصدق ، وأن يكون على دراية بالعوامل النفسية والظروف المختلفة التى تؤثر فى العملية الاختبارية وفى الأداء ، ويمكنه التوصل إلى استنتاجات ووصيات اعتماداً على درجة أو درجات الاختبار ، آخذ بعين الاعتبار المعلومات الأخرى الهامة المتعلقة بالمفحوص .

حقوق المختبرين : إن من حق الفرد الذى يجرى عليه اختبار أو مقياس معين أن يكون على علم بالهدف من الاختبار وفىم تستخدم نتائجه . فلا يجوز للأخصائى النفسى أن يطلب من المفحوص الإفصاح عن معلومات تخصه أثناء عملية المقابلة الشخصية أو أثناء اختباره أو تقويمه ، أو أن يسعى إلى الحصول على مثل هذه المعلومات إلا بعد تعريفه بالهدف من ذلك وأخذ موافقته كتابياً ، أو أخذ موافقة ولئى أمره إذا كان قاصراً .

عدم تداول الاختبارات بين غير المتخصصين فى علم النفس : إن الحفاظ على سرية الاختبارات تعد من مسئولية مؤلفيها وناشريها ويشاركهم فى ذلك الأخصائى النفسى (Sax, 1974) . فسرية الاختبارات تعد مشكلة إذا أدت إلى تغيير فى الدرجة التى يحصل عليها الفرد فى الاختبارات دون أن يكون هناك تغير مقابل فى درجته الحقيقية ، غير أن سرية الاختبارات لا تعد مشكلة فى بعض الاختبارات . فمثلاً إذا أجرى اختبار على فرد لقياس درجة تمكنه من مهارة معينة ، فإن معرفته السابقة بعينة من أمثلة مشابهة لأسئلة أو تدريبه عليها ربما يوصى به .

ولكن فى كثير من الحالات تؤثر معرفة عينة من الأسئلة أو طريقة تصحيحها تأثيراً بالغاً فى صدق الاختبار؛ لذلك يجب فى مثل هذه الحالات الحفاظ على الاختبارات فى كائن مغلقة ، وتحديد الأشخاص الذين توكل إليهم مهمة استلام هذه الاختبارات .

وإذا أرسلت الاختبارات إلى مكان آخر عن طريق البريد ، فإن الشخص المرسل إليه يجب أن يكون معروفاً وموثوقاً به . والواقع أن آلات التصوير الحديثة والتقنيات المختلفة قد زادت من حدة مشكلة سرية الاختبارات وما يرتبط بها من مخالفات لحقوق الطبع والنشر مما يتطلب اتخاذ الإجراءات التى تضمن منع هذه المخالفات .

نشر وتوزيع الاختبارات : يمكن تلافي كثير من المشكلات الناجمة عن سوء استخدام الاختبارات بتنظيم عملية نشر وتوزيع الاختبارات ، فبعض الناشرين يسمحون ببيع الاختبارات والمقاييس النفسية لكل من يطلبها مادام أنه يدفع ثمن شرائها ، فى حين

أن عدداً محدوداً منهم يقصر البيع على فئة من المتخصصين. وربما يرجع السبب في ذلك إلى أن بعض هؤلاء الناشرين لا يدركون الأضرار التي تنجم عن سوء توزيع الاختبارات سواء للأفراد أو المؤسسات.

لذلك فإن هناك حاجة إلى تقييد توزيع أو بيع الاختبارات والمقياسات النفسية، وقصر ذلك على من توافر لديهم المهارات أو الكفايات الالزمة لاستخدامها الاستخدام المناسب؛ وذلك لضمان سرية الاختبارات وعدم إساءة استخدامها.

ومن مسؤولية القائمين بإعداد الاختبارات أو تأليفها ، ومؤسسات نشرها عدم السماح بتوزيع الاختبارات للاستخدام قبل أن يتم إعدادها في صورتها النهائية. وإذا سمح بنشر اختبار في صورته الأولية لأغراض البحث أو الدراسة يجب أن يوضح ذلك صراحة على غلاف الاختبار وفي كراسة تعليماته، وأن يقتصر توزيعه على تلك الأغراض فقط. ويجب أن تشتمل كراسة تعليمات الاختبار بعامة على بيانات كافية يمكن الاستعانة بها في تقييم الاختبار.

وكذلك تشتمل على معلومات عن كيفية تطبيق الاختبار وتصحيحه وتفسير درجاته في ضوء المعايير الخاصة به، وأن تكون الإرشادات والمعلومات الواردة تعبير حقيقة عما يهدف له الاختبار.

كما يجب على مؤلفي الاختبارات أن يعيدوا النظر في اختباراتهم ومعاييرها كلما دعت الحاجة إلى ذلك حتى لا تقادم هذه المعايير، مع مراعاة أن الاختبارات تختلف اختلافاً ملحوظاً في المدة الزمنية الالزامية لمراجعتها.

ومن الضوابط أيضاً في مجال نشر الاختبارات عدم السماح بالإعلان عن الاختبارات أو نشر أجزاء منها في الصحف أو المجلات أو الكتب العامة حتى لا يساء فهمها أو استخدامها، وأن يقتصر النشر على النوريات العلمية المتخصصة بأسلوب موضوعي بعيد عن الانفعال أو الإقناع أو الوصف.

سرية تقارير نتائج الاختبارات : إن من حق المفحوص الاطلاع على تقرير نتائج الاختبارات التي أدتها، كما يجب أن يسمح له بالتعليق على محتوى التقرير، وتوسيعه أو تعديل بعض المعلومات الخاصة به عند الضرورة. وبهتم المرشدون النفسيون في الوقت الحاضر بأن يشارك المفحوص مشاركة فاعلة في عملية تقويمه؛ لذلك يجب أن تقدم إليه نتائج الاختبارات بشكل يسهل عليه فهمه، وأن يتعد التقرير عن المصطلحات والعبارات الفنية المعقدة.

وإذا كان المفحوص طفلاً يكون من حق والديه الاطلاع على تقرير النتائج، وإن كان هذا يمثل تدخلاً في خصوصياته (Ruebhausen, 1961)، غير أن التوجيهات التي

اقترحتها مؤسسة ريج (Russell Sage Foundation, 1970) تشير إلى أنه إذا بلغ الطفل من العمر ثمانية عشر عاماً، ولم يتمكن من استكمال دراسته الثانوية، أو تزوج، فإن من حقه أن لا يطلع على مثل هذا التقرير سواه.

غير أن هذا التوجيه أعقى تحذير حول مدى إمكانية تنفيذ ذلك ومراعاة القوانين المتعلقة بهذا الشأن. ولكن المشكلة هنا ليست مجرد اطلاع القاصر أو والديه على تقرير نتائجه وإنما كيفية إجراء ذلك. فمن الناحية القانونية للوالدين الحق في الحصول على معلومات تتعلق بأبنائهم، وهو عادة يحرصون على ذلك، إلا أنه في بعض الحالات ربما ترداد حدة المشكلات التعليمية والانفعالية للطفل نتيجة عدم توافقه مع والديه. ففي مثل هذه الحالات ينبغي تدخل المرشد النفسي لكي يحصل على المعلومات الازمة عن الطفل من والديه والعمل على كسب تعاونهم.

أما إذا طُبِقت الاختبارات من خلال مؤسسات عمل أو مؤسسات تربوية أو قضائية فإنه ليس من الضروريأخذ موافقة كتابية من الفرد المختبر، وإنما يلزم إشعاره بموعد الاختبار والهدف منه وكيف ستستخدم نتائجه، ومن هم الذين سوف يستخدمونها. والمشكلة المهمة الأخرى تتعلق باستبقاء المعلومات المتعلقة بتلاق الاختبارات النفسية والاحتفاظ بها، فهذه المعلومات تفيد في الدراسات الطولية وتتبع حالة المختبرين. غير أن هذه المعلومات تتقادم بعد مدة زمنية تختلف باختلاف نوعيتها، وبذلك تؤدي إلى استنتاجات غير واقعية، واستخدامها في غير الأغراض التي جمعت من أجلها. لذلك يجب أن تكون هناك ضوابط تنظم عمليات استبقاء هذه المعلومات والحفظ عليها وعدم النيل منها أو الوصول إليها إلا في الحالات المسموح بها.

ومما يزيد هذه المشكلة تعقيداً التطورات الكبيرة التي حدثت في مجال نظم المعلومات وإمكانية استدعاء المعلومات في سرعة ودقة متناهية، إلا أنه يمكن الإفادة من هذه التطورات في الحفاظ على المعلومات والبيانات المتعلقة بتلاق الاختبارات وابتكار أساليب تضمن سرية هذه التلاق.

خلاصة

تناولنا في الفصل السابق بعض أهم الأسس والمعايير والضوابط التي تحكم وتجه عمليات وممارسات بناء الاختبارات والمقاييس التربوية والنفسية، وتطبيقها وتفسير نتائجها. فهناك مشكلات أساسية تتعلق بالقياس والتقويم في الوطن العربي من أهمها: إسناد مهام القياس النفسي والتربوي إلى غير الاختصاصيين، وخبرة الاختصاصيين النفسيين المحدودة بنظريات القياس النفسي والتربوي والأسس التي يستند إليها، وقلة الاختبارات المدققة واختبارات الكفايات، وإساءة استخدام الاختبارات وبخاصة الاختبارات التحصيلية التي يكتبها المعلمون، وعدم إيلاء الاختبارات الشخصية والبنائية أهمية تذكر.

كما يمكن تصنيف الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربوية في أخطاء تتعلق بتحيز المسؤولين عن البرامج الاختبارية، والثقة التامة في التنبؤ باستخدام الاختبارات، والاستخدام الروتيني لدرجات الاختبارات، وافتراض أن الاختبارات العقلية تقيس خصائص موروثة. أما الأخطاء الناجمة عن تقديم نتائج الاختبارات إلى المفحوصين فإنها تتعلق بمسايرة التوقعات لنتائج الاختبارات، وتأثير مفهوم الذات لدى الفرد بهذه النتائج.

وهناك العديد من الشروط العلمية لانتقاء الاختبارات والمقاييس التربوية والنفسية من حيث توافر خصائص الاختبار أو المقاييس الجيد، ومناسبة الاختبار للمفحوص وللغرض الذي يستخدم من أجله. وكذلك هناك شروط علمية لتطبيق الاختبارات وتصحيحها، وتفسير درجاتها وتقديم نتائجها. ومن بين ضوابط وأخلاقيات استخدام هذه الاختبارات والمقاييس توافر كفايات معينة لدى مستخدمي الاختبارات، وحقوق المختبرين من حيث علمهم بالهدف من الاختبار أو المقاييس، وفيما تستخدمن نتائجه، وعدم تداول الاختبارات والمقاييس بين غير المتخصصين في علم النفس، وضوابط نشر وتوزيع الاختبارات النفسية، وسرية تقارير نتائج الاختبارات.

الفصل الثالث

أسسیات الإحصاء وتوظیفه فی القياس والتقویم التربوی والنفسي

* مقدمة

- * أساليب تنظيم وعرض مجموعات من الدرجات
- * التمثيل البياني لتوزيعات الدرجات
- * المدرج التكراري
- * المنحنى التكراري
- * شكل الأغصان والأوراق
- * استخدام شكل الأغصان والأوراق في المقارنة بين مجموعتين من الدرجات
- * وصف توزيعات درجات الاختبارات والمقاييس
- * مقاييس التزعة المركزية أو الموضع
- * العلاقة بين مقاييس التزعة المركزية أو الموضع
- * مقاييس التشتيت
- * الدرجات المحوّلة
- * تحديد المركز النسبي للأفراد على المنحنى الاعتدالى
- * المئينيات
- * الرتب المئينية
- * استخدام المنحنى الاعتدالى في إيجاد المئينيات
- * مقاييس العلاقة بين مجموعتين من الدرجات
- * معامل ارتباط بيرسون
- * استخدام معامل الارتباط في التنبؤ
- * معادلة خط انحدار المتغير (ص) على المتغير (س)
- * الخطأ المعياري للتنبؤ

مقدمة :

ذكرنا فى الفصل الأول أن القياس التربوى والنفسى يهدف للتحديد الكمى للظواهر السلوكية، ويعتمد ذلك على جمع بيانات عن طريق أدوات القياس المتنوعة، ويتجزء عن تطبيق هذه الأدوات فى معظم الأحيان قيماً عددياً أو درجات تمثل قياسات سمات معينة لدى مجموعة من الأفراد.

وهذه القيم أو الدرجات تتطلب تبويهاً وتصنيفهاً وتحليلهاً لاستخلاص معلومات مفيدة تتعلق بهذه السمات، ويمكن فى ضوء هذه المعلومات تفسير الدرجات المستمدّة من هذه الأدوات تفسيراً مستثيراً، كما يمكن فى ضوئها التتحقق من بعض الخصائص السيكومترية الأساسية لأدوات القياس فى مراحل تصميمها وبنائها والحكم على مدى صلاحيتها فيما تقيسه.

ونظراً لأن علم الإحصاء يتضمن الطرق والأساليب التي يمكن استخدامها في معالجة البيانات الكمية والكيفية، فإن معرفة الباحث والممارس لبعض هذه الطرق والأساليب، واكتساب المهارة في توظيفها توظيفاً مناسباً يمكنه من تفسير الدرجات التي يحصل عليها من أدوات القياس التربوى والنفسى بما يحقق الهدف من استخدام هذه الأدوات.

ونظراً لتعدد أساليب المعالجات الإحصائية ، فسوف نقتصر في هذا الفصل على عرض أهم الأساليب الوصفية التي يمكن توظيفها بسهولة في إعداد أدوات القياس وتفسير نتائجها . وعلى وجه التحديد سوف يشتمل هذا الفصل على الأساليب الإحصائية التي تستخدم في تنظيم ووصف مجموعات من الدرجات عن طريق مقاييس النزعة المركزية والتشتت والعلاقة، وكذلك الأساليب المختلفة لتحويل الدرجات من صورة إلى أخرى لإجراء مقارنات وظيفية فيما بينها، واشتقاق معايير الاختبارات التربوية والنفسية بما يسمح بتفسير الدرجات المستمدّة منها . ونفرد الفصول الأربع التالية لمناقشة أساليب تحديد الخصائص السيكومترية لأدوات القياس .

أساليب تنظيم وعرض مجموعات من الدرجات :

إن تفرد وتميز الأفراد يُعدان من الخصائص الأساسية للحياة، فالأطفال الذين ينشأون في الأسرة نفسها أو في ظروف أسرية متماثلة يختلفون اختلافاً واضحاً في كثير من الصفات، ويزداد الاختلاف بينهم في الاستعدادات والميول والاتجاهات وسمات الشخصية كلما تقدموا في العمر، ويمكننا ملاحظة هذا الاختلاف بوضوح من خلال تعاملنا مع الأفراد. بعض الأفراد يتميزون بقدرات ومهارات ربما لا تتوافق في أقرانهم، مثل القدرة الحسابية أو اللغوية أو الموسيقية أو المهارات اليدوية، وبعضهم يستطيع تكوين صداقات طيبة والبعض الآخر يفضل الوحدة، وهكذا.

وبسبب هذه الاختلافات الواضحة حارت دراسة الفروق الفردية اهتماماً مستمراً من جانب علماء النفس، والدليل على ذلك أن علم النفس الفارق Differential Psychology وهو أحد فروع علم النفس يهتم بالبحث الكمي الموضوعي للفروق الفردية.

ولذلك لا نعجب إذا وجدنا أن معظم المصطلحات المستخدمة في علم النفس لوصف الأفراد تشير إلى خصائص أو سمات تختلف وتتبادر بين الأفراد اختلافاً ملحوظاً. وهذه الفروق والاختلافات لها دلالة عملية مهمة ربما يكون أكثرها وضوحاً ما يتطلبه الأداء الجيد في مختلف المهن والنجاح في الدراسات الأكademie من قدرات ومهارات ربما لا تتوافق بدرجة عالية في جميع الأفراد.

ولذلك فإن المشكلة الرئيسية التي تهم الباحث والممارس في المجالات النفسية والتربية هو كيفية وصف طبيعة ومدى هذه الفروق الفردية وصفاً دقيقاً بالاستعانة بالأساليب الإحصائية التي تساعده في تنظيم وعرض مجموعات الدرجات المستمدة من أدوات القياس وتمثيلها بيانياً بحيث تفصح عن بعض ما يتميز به الأفراد من خصائص أو سمات. كما أن هذا التنظيم يساعد في إلقاء الضوء على إجابة بعض التساؤلات المتعلقة بهذه الفروق الفردية.

ولتوضيح ذلك نفترض أن معلم اللغة العربية طبق اختباراً يقيس فهم المقرء على تلاميذ أحد الصفوف البالغ عددهم ٤٠ تلميذاً ، وكانت الدرجات كالتالي:

٨٠	١٣١	٦٩	١٠٧	١١٤	١٢٣	١٠٢	٨٠	٧٨	٦٧
١٢٣		٨٥	٨٢	٨٦	٩٧	٨٩	٨١	٧٩	٨٤
٨٥			٨٦	٩٧	٨١	٧٩			٩٠
٧٣				٨٨	١٢٦	١١٠	٨١	٨٩	٩١

بالنظر إلى هذه الدرجات نجد أنها تراوح بين ٦٧ ، ١٣١ . وعلى الرغم من أن المعلم يستطيع بالفحص العيني لهذه الدرجات ملاحظة أن هناك فروقاً فردية بين التلاميذ، إلا أن استخلاص معلومات دقيقة عن هذه الفروق من مجرد هذا الفحص العيني يتطلب وقتاً وجهداً ملحوظاً.

لذلك ينبغي أن تنظم هذه الدرجات بطريقة تيسر ذلك، فربما يقوم المعلم بترتيب هذه الدرجات ترتيباً تصاعدياً أو تنازلياً.

غير أن هذا ربما يصعب إجراؤه إذا كان عدد التلاميذ كبيراً، لذلك يفضل تبويب هذه الدرجات في فئات متتالية تشمل كل فئة منها على عدد معين من الدرجات، أي تحديد الفئات ومعرفة عدد التلاميذ الذين يقعون في كل فئة منها، ووضع فئات الدرجات وتكرار أعداد التلاميذ في جدول يسمى جدول التوزيع التكراري Frequency Distribution .

ويمكن اختيار سعة الفئة بحيث ينحصر العدد الكلى للفئات بين ١٠ ، ٢٠ فئة تقريباً، فإذا قل عدد الفئات بحيث أصبح أربع أو خمس فئات مثلاً تزداد سعة الفئة وبالتالي لا تستفيد شيئاً من عملية التبويب، وإذا زاد عدد الفئات عن ٢٠ فئة تصبح سعة الفئة صغيرة وتضيق معالم التوزيع . وبالتالي يعتمد اختيار سعة الفئة على العدد الكلى للدرجات ومدتها والهدف من استخدام التوزيع .

إذا عدنا إلى مجموعة الدرجات السابقة وأردنا تكوين جدول التوزيع التكراري، فإن الخطوات تكون كالتالي :

(١) تحديد مدى الدرجات : وهذا المدى يوضح انتشار الدرجات بين القيمتين العليا وهي ١٣١ والدنيا وهي ٦٧ . أى أن الدرجات تتوزع في مدى قدره (١٣١ - ٦٧) = ٦٤ درجة . وأضفنا الواحد الصحيح لأنه ربما يحصل فرد على الدرجة ٦٧ ويحصل آخر على الدرجة ١٣١ ، فعندها المدى يزيد بقدر الواحد الصحيح .

(٢) تحديد سعة الفئة : ويعنى بذلك تقسيم المدى إلى فئات متساوية لتبويب الدرجات ، فإذا قسمنا المدى وقدره ٦٤ إلى ٩ فئات مثلاً تكون سعة الفئة متساوية $\frac{64}{9} = 7$ ونقربها لتصبح ٧ .

(٣) تحديد الدنيا في التوزيع : وهذا يتطلب البدء بأقل الدرجات بحيث تقبل القسمة على سعة الفئة . ونظراً لأن أقل الدرجات هي ٦٧ لا تقبل القسمة على سعة الفئة وهي ٧ ، فإننا نبحث عن الدرجة الأقل مباشرة من ٦٧ وتقبل القسمة على ٧ فنجد لها ٦٣ . لذلك ينبغي أن تبدأ الفئة الأولى بالدرجة ٣ . وتضيق الفئة الدنيا ٦٣ - ٦٩ ،

والفتة الثانية تبدأ بالدرجة التي تلي ٧٩ أي ٧٠، وتكون هذه الفتة ٧٦ - ٧٧، وهكذا حتى نصل إلى أعلى الدرجات أو بعدها بقليل.

(٤) تكوين جدول التوزيع التكراري : فهذا التوزيع عبارة عن مجسومة من الفئات المرتبة ترتيباً تصاعدياً بحيث تشتمل على جميع الدرجات . والجدول يتكون من ثلاثة أعمدة هي: فئات الدرجات، وعلامات التكرار، والتكرار . ويمكن الحصول على علامات التكرار بأن نمر على الدرجات الأربعين ، ونضع علامة (x) في العمود الثاني أمام الفتة التي تقع فيها كل درجة ، ثم نعد العلامات المناظرة لكل فئة وندون التكرار في العمود الثالث ، ثم نجمع التكرار الكلى للتأكد من أنه يساوى عدد أفراد المجموعة أي ٤٠ .

(٥) إيجاد التكرار المثوى : يفضل عادة إيجاد نسبة التكرار المناظر لكل فئة إلى التكرار الكلى ، وذلك بقسمة كل من التكرارات على العدد الكلى للأفراد وضرب الناتج في ١٠٠ للحصول على التكرار المثوى .

$$\text{أى أن : التكرار المثوى المناظر لفتة معينة} = \frac{\text{التكرار المناظر لفتة معينة}}{\text{التكرار الكلى}} \times 100$$

حيث إن هذا يفيد في مقارنة توزيعات درجات اختبارات مختلفة ، وبخاصة إذا كان التكرار الكلى لدرجات كل من الاختبارين مختلفاً ، وفيما يلى جدول التوزيع التكراري لدرجات السابقة:

التكرار المثوى	التكرار	علامات التكرار	فئات الدرجات
٥,٠	٢	xx	٧٩ - ٦٣
٢,٥	١	x	٧٦ - ٧٠
٢٥,٠	١٠	xxxxxxxxxxxx	٨٣ - ٧٧
٢٢,٥	٩	xxxxxxxxxx	٩٠ - ٨٤
١٠,٠	٤	xxxx	٩٧ - ٩١
١٥,٠	٦	xxxxxx	١٠٤ - ٩٨
٧,٥	٣	xxx	١١١ - ١٠٥
٢,٥	١	x	١١٨ - ١١٢
٥,٠	٢	xx	١٢٥ - ١١٩
٥,٠	٢	xx	١٣٠ - ١٢٦
<hr/> ١٠٠,٠		٤٠ = N	

جدول (١-٣) التوزيع التكراري لمجموعة الدرجات التي عددها ٤٠ درجة

ويلاحظ من جدول (٣ - ١) أن الحدود الدنيا للفئات تقبل القسمة على سعة الفئة ٧، وحصلنا على ١٠ فئات وليس ٩ لأننا اعتبرنا سعة الفئة ٧، وليس ٢، كما بدأنا بالدرجة ٦٣ بدلاً من ٦٧ لأن الدرجة ٦٣ تقبل القسمة على سعة الفئة ٧.

ومن ميزات هذا الجدول أنه يمكن من معرفة توزيع الدرجات وانتشارها بدرجة أفضل من الدرجات غير المبوية، إذ يستطيع المعلم بمجرد النظر إلى علامات التكرار معرفة أن أكبر عدد من الطلاب حصلوا على درجات تتراوح بين ٧٧ ، ٩٠ . وبين ذلك أيضاً التكرار المثوى للفئتين الثالثة والرابعة وهو $25,5 + 22,5 = 47,0 \%$ من المجموع الكلى للتلاميذ.

بينما حصل عدد أقل من التلاميذ على درجات تقع بين ٦٣ ، ٧٦ أو بين ١١٩ ، ١٣٢ ، حيث إن التكرار المثوى في الحالة الأولى ٧,٥٪ وفي الحالة الثانية ١٠٪، ويمكن أيضاً ملاحظة ذلك بوضوح إذا نظرنا إلى علامات التكرار في هذه الفئات.

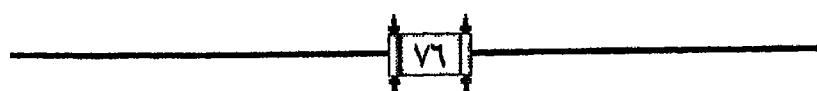
وبالمثل نستطيع تحديد عدد التلاميذ الذين يقعون في مختلف فئات الدرجات.

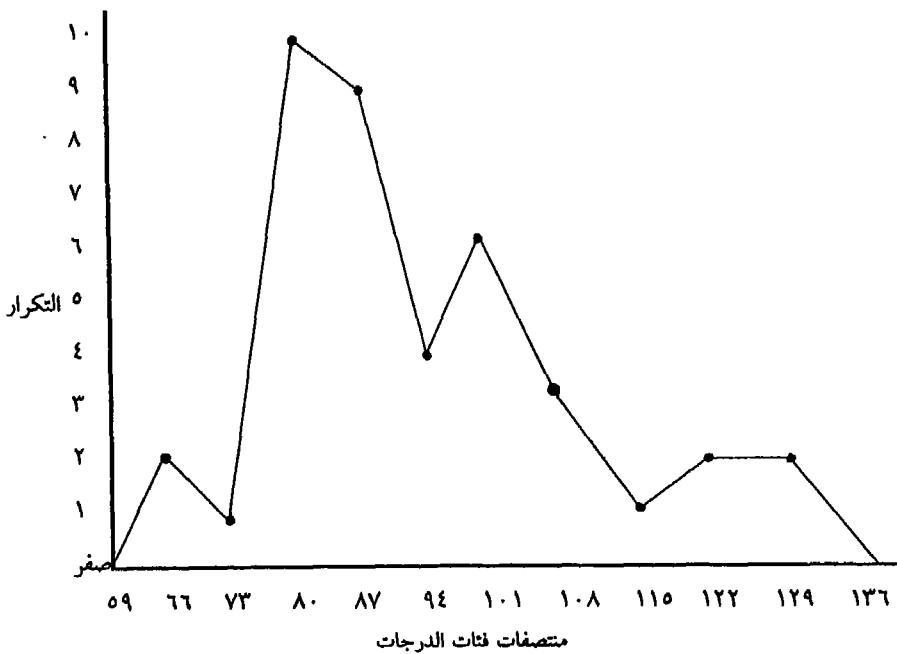
التمثيل البياني للتوزيعات الدرجات :

على الرغم من أن التوزيع التكراري للتوزيعات الدرجات يعد أسلوباً بسيطاً لتنظيم وتلخيص هذه الدرجات، إلا أن التمثيل البياني للتوزيعات يقدم للطلاب والأباء وغيرهم من المعنين صورة أكثر وضوحاً وأيسر فهماً من التوزيع التكراري للدرجات.

وهناك أنواع متعددة من الأشكال البيانية التي يمكن أن يستخدمها المعلم والباحث والممارس في وصفه ودراسته للفروق الفردية ، ومن بين هذه الأشكال البيانية شائعة الاستخدام المنحنى التكراري Frequency Curve ، والمدرج التكراري Histogram .

المنحنى التكراري : نظراً لأن سعة الفئة في المثال السابق أكبر من الواحد الصحيح، فإن النقاط التي نستخدمها لرسم المنحنى التكراري تناظر متصفات الفئات على المحور الأفقي، والتكرارات على المحور الرأسى، ويكون المنحنى التكراري من توسيع متصفات فئات التوزيع بخط منحنى مهد بقدر الإمكان كما هو موضح بشكل (١-٢) التالي :

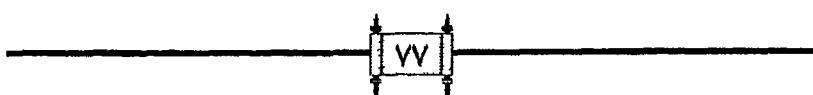


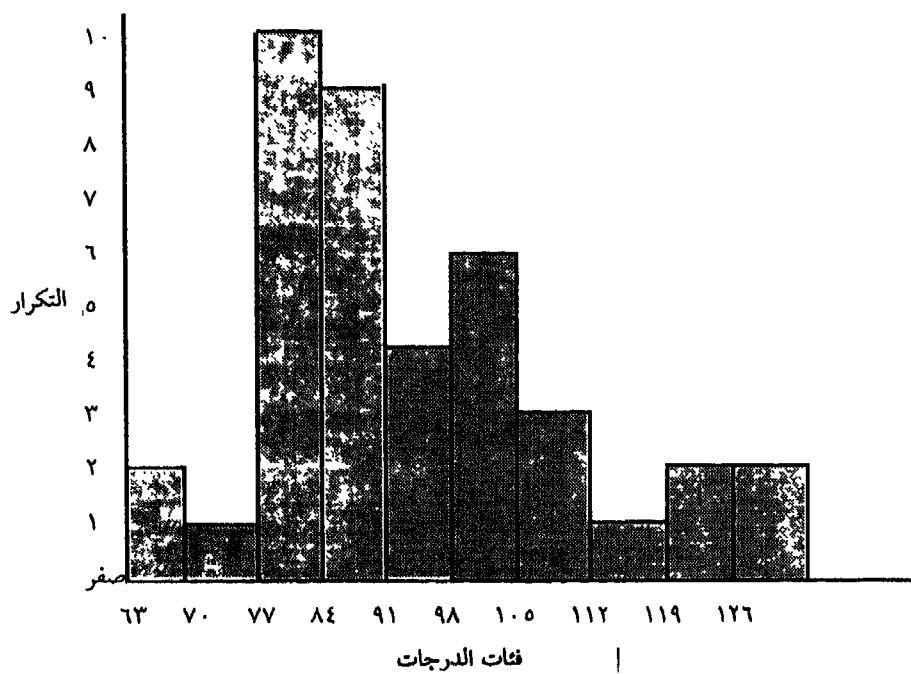


شكل (١-٣) المنحنى التكراري للدرجات ٤٠ تلميذاً

المدرج التكراري: عبارة عن مجموعة من المستطيلات المتلاصقة أو المتباعدة تمثل توزيع درجات الاختبارات والمقاييس، فالمستطيلات تكون متلاصقة إذا كان مستوى قياس المتغير الذي تمثله الدرجات فتراها أو نسبياً، بينما تبتعد المستطيلات عن بعضها البعض إذا كان المتغير كيفياً، مثل النوع أو الديانة أو المهنة، حيث إن الأقسام التي يصنف فيها كل فرد تكون منفصلة وعندئذ تبتعد المستطيلات.

ففي المثال السابق تمثل فئات الدرجات على المحور الأفقي والتكرار على المحور الرأسى كما في المنحنى التكراري، وارتفاع المستطيلات يمثل التكرارات كما هو موضح بشكل (٢ - ٣) التالي:





شكل (٣ - ٢) المدرج التكراري لدرجات ٤٠ تلميذًا

ويتبين من الشكلين (٣ - ١) ، (٣ - ٢) أن كليهما يمثلان مجموعة درجات الاختبار لتلاميذ الفصل ، ولكن المنحنى التكراري يبين انتساقية توزيع الدرجات وشكله العام وحدوده نظرًا لأن التوزيع متصل ومستوى القياس ثابت.

ومن ميزات المنحنى التكراري التي ربما لا يتميز بها المدرج التكراري أنه يمكن استخدامه في تمثيل توزيعين مختلفين للبنين والبنات مثلاً في اختبار ما ، أو توزيع درجات الاختبار نفسه قبل عملية التعليم وبعده ، أو قبل أنشطة علاجية وبعدها ، أو توزيع درجات اختبارين طبقاً على طلاب فصل معين ، وذلك لإجراء مقارنات بين شكل التوزيعين في كل من هذه الحالات . وعندئذ ربما يتقطع المنحنيان في عدة نقاط يمكن تفسيرها ، في حين يصعب إجراء ذلك باستخدام المدرج التكراري .

ويُفضل تمثيل التكرار المثنوي على المحور الرأسى بدلاً من تمثيل التكرار إذا كان عدد أفراد المجموعتين مختلفاً لكي تكون المقارنة دقيقة .

شكل الأغصان والأوراق :

رأينا مما سبق أن التمثيل البياني للدرجات الاختبارات والمقاييس التربوية والنفسية يوضح الشكل العام لتوزيع الدرجات وييسر استخلاص معلومات مفيدة لمصمم الاختبار ومستخدمه، إلا أن المنهج التكراري والمدرج التكراري ليسا كافيين لاعتمادهما على فئات الدرجات بدلاً من الدرجات نفسها، وبذلك تفقد بعض المعلومات المتعلقة بالدرجات، فكل درجة في التوزيع تمثل أحد أفراد المجموعة التي طبق عليها الاختبار أو المقاييس وينبغي الاهتمام بها وتفسيرها.

وشكل الأغصان والأوراق Stem - and - Leaf Display يعد أحد الأساليب الكشفية المستحدثة التي توصل إليها توكي Tukey رائد تحليل البيانات. ويتميز هذا الأسلوب بأنه يعالج بعض أوجه قصور الأساليب التقليدية المستخدمة في التمثيل البياني لتوزيعات الدرجات والتي أشرنا إليها. فهذا الأسلوب يجمع بين ترتيب الدرجات ترتيباً تصاعدياً وتمثيلها تمثيلاً بصرياً، أي يعد بمثابة توزيع تكراري ومدرج تكراري للدرجات في آن واحد، ويحافظ على جميع الدرجات أو القيم الملاحظة مرتبة، وفي الوقت نفسه يساعد في تنظيم الدرجات بحيث يتضح الشكل العام لتوزيع هذه الدرجات أو القيم.

ولتوضيح كيفية إنشاء شكل الأغصان والأوراق للمثال السابق يمكن اتباع الخطوات التالية:

(١) نقسم كل درجة من درجات المجموعة إلى جزءين، فإذا كانت الدرجة تتكون من رقمين نفصل رقم خانة العشرات عن رقم خانة الأحاد، فمثلاً الدرجة الأولى ٧٧ تكون من رقم خانة العشرات ٦ الذي يمثل الغصن Stem، ورقم خانة الأحاد ٧ الذي يمثل الورقة Leaf.

إذا كانت الدرجة تتكون من ثلاثة أرقام نفصل رقمي خاتمة المئات والعشرات عن رقم خانة الأحاد. فمثلاً الدرجة ١٢٣ تتكون من رقم خانة المئات ١، ورقم خانة العشرات ٢، ورقم خانة الأحاد ٣.

لذلك نفصل الرقم ١ ، ٢ عن الرقم ٣ وبذلك يكون العدد ١٢ بمثابة الغصن، والرقم ٣ بمثابة الورقة التي تنتهي إلى هذا الغصن . وهكذا في بقية درجات المجموعة .

(٢) نضع جميع الأغصان في عمود رأسى إلى اليسار بحيث تكون قيمها مرتبة ترتيباً تصاعدياً، ونضع الأوراق التي تنتهي إلى كل غصن في صف أفقي بجوار هذه الأغصان إلى اليمين، ونفصل بين الأغصان والأوراق بخط رأسى . ونظراً لأن عدد

الدرجات أو القيم يعد أساسياً، فإنه ينبغي تسجيل جميع الدرجات حتى لو تكرر بعضها لكي لا فقد أي درجة منها، ويوضح الشكل العام لتوزيع هذه الدرجات. ونظراً لأن الدرجات تتراوح بين ٦٧ ، ١٣١ فإننا نبدأ عمود الأغصان بالقيمة ٦ ونتهي بالقيمة ١٣ . ونظراً لأننا استخدمنا الدرجات كما هي دون تقريب أو تقليص، فإن الوحدة في هذه الحالة تكون درجة واحدة.

إذا اتبعنا هذه الخطوات يمكن أن نحصل على شكل الأغصان والأوراق التالي لمجموعة الدرجات المذكورة في المثال السابق وعدها ٤٠ درجة :

الأغصان	الأوراق
٦	٧٩
٧	٨٩٨٣
٨	٤٩١٠١١٢٩٨٦٢٥٤٠٥
٩	٠١٣٨٧٦
١٠	٢٧٠٧٤٠٣
١١	٠٤
١٢	٣٦٣
١٣	١

شكل (٣ - ٣) يوضح الأغصان والأوراق لدرجات ٤٠ تلميذ

ويوضح من شكل (٣ - ٣) أن معظم الدرجات تتركز في الغصن ٨ حيث يحتوى على أكبر عدد من الأوراق (١٥ ورقة) يليه الغصن ١٠ حيث يحتوى على (٧ أوراق) ثم الغصن ٩ الذي يحتوى على (٦ أوراق).

أى أن التوزيع يميل إلى التراكم في الأغصان ٨، ٩، ١٠ . وهذا يعني أن ٢٨ تلميذاً من بين العدد الكلى للتلاميذ (٤٠ تلميذاً) حصلوا على درجات تتراوح بين ٧٣ ، ١٠٧ . كما يبدو أن التوزيع غير متماثل حول الغصن ٩ الذي يتمى إليه المتوسط العام للدرجات حيث يوجد التواء إلى حد ما ناحية الأغصان ١١، ١٢، ١٣ التي يتمى إليها عدد قليل من الأوراق.

وعلى الرغم من أن شكل الأغصان والأوراق يساعد في استخلاص أكبر قدر من المعلومات من درجات الاختبارات والمقاييس ويوضح النمط العام لتوزيع الدرجات أو

الشكل العام للتوزيع ، إلا أن هناك بعض الملاحظات التي يجب الإشارة إليها لأهميتها:

(١) ينبغي عدم تقليل عدد الأغصان كثيراً : فهذا يؤدي إلى تراكم عدد كبير من الأوراق على كل غصن مما يجعل من الصعب استخلاص معلومات مفيدة تتعلق بتوزيع الدرجات . فمثلاً إذا كانت درجات اختبار ما كالتالي :

٢٧ ، ٢٨ ، ٢٦ ، ٢٥ ، ٢٤ ، ٢٣ ، ٢٢ ، ٢٩ ، ١٩ ، ١٤ ، ١٦ ، ٢١ ، ٩ ، ٧

. ١٥

فإذن نلاحظ أن الدرجات تتراوح بين ٧ ، ٢٩ . فإذا جعلنا الوحدة درجة واحدة فإن عمود الأغصان سوف يشتمل على أرقام خانة العشرات ١ ، ٢ ، صفر فقط ، وعندئذ تراكم الأوراق المناظرة لأوراق خانة الأحاد على هذه الأغصان الثلاثة مما يجعل الشكل قليل الفائدة .

ولذلك يُفضل في هذه الحالة توسيع شكل الأغصان والأوراق وذلك بتقسيم الأغصان . فمثلاً يمكن تقسيم كل غصن إلى غصنين أحدهما للأوراق المناظرة للدرجات ٥ وما فوقها ، والآخر للدرجات ٤ وأقل . وبذلك يصبح شكل الأغصان والأوراق مكوناً من سنتة أغصان بدلاً من ثلاثة ، كما هو موضح بشكل (٣ - ٤) التالي :

الأغصان	الأوراق
صفر	
* صفر	٩ ٧
١	٣ ٤
* ١	٩ ٦ ٥
٢	٢ ١
* ٢	٧ ٩ ٦ ٥ ٨ ٦

* تشير إلى الأوراق المناظرة للدرجات ٥ وما فوقها

شكل (٣ - ٤) يوضح الأغصان والأوراق للدرجات ١٥ تلميذًا بعد تقسيم كل غصن إلى غصنين

ويكن أيضاً تقسيم كل غصن إلى عدد أكبر من الأغصان وليكن أربعة أقسام، وبذلك يصبح الشكل مكوناً من تسعه أغصان كما يوضحه شكل (٣ - ٥) التالي:

الأغصان	الأوراق
صفر ①	
صفر *	
صفر █	٩ ٧
٠ ١	
* ١	٣ ٤ ٥
█ ٢	٩ ٦
① ٢	٢ ١
* ٢	٥
█ ٢	٧ ٩ ٦ ٨ ٦

① تشير إلى الأوراق صفر ، ١ ، ٢ ، .

* تشير إلى الأوراق ٢ ، ٤ ، ٥ ، .

█ تشير إلى الأوراق ٦ وما فرقها ..

شكل (٣ - ٥) يوضح الأغصان والأوراق لدرجات ١٥ تلميذاً بعد تقسيم كل غصن إلى أربعة أغصان وعموماً، فإنه يمكن توسيعة شكل الأغصان والأوراق أو تضييقه إما بتقسيم كل صف إلى صفين أو أكثر أو بضم صفين متتالين أو أكثر، وهذا يعتمد على طبيعة البيانات والمعلومات المراد استخدامها من الشكل.

(٢) **ينبغي عدم الإكثار من الأغصان :** فكثرة الأغصان يؤدى إلى خلوها من الأوراق أو قلتها، ويحدث هذا غالباً عندما تكون القيم مشتملة على أربعة أرقام أو أكثر، ولذلك يفضل في مثل هذه الحالات تقليص الشكل وذلك باختيار وحدة كبيرة ربما تكون ١٠ درجات أو ١٠٠ درجة أو ربما أكثر بحسب قيم الدرجات، وكذلك تقيير هذه القيم تقريراً مناسباً.

وللتوضيح ذلك نفترض أن القيم كانت كالتالى:

١٢١٣٨ ،	٩٣٥٢ ،	١١١٢١ ،	١٢٥٢٧ ،	١٠٢١٢
١١٦٨٧ ،	١٤٢٧٧ ،	١٣٢٧٦ ،	١٠٩٤٤ ،	١١١٣٧
١٢٢٣١ ،	١٥٢٤١ ،	١١٢٢٧ ،	١٠٨٩١ ،	١٦١٢٢

فعندي لا تصلح أن تكون الوحدة درجة واحدة، وإنما يفضل هنا أن تكون الوحدة ١٠٠ مثلاً وتقريب القيم إلى أقرب مائة، فالقيمة ١٠٢١٢ تصبح ١٠٢٠٠ وكذلك القيمة ١٢٥٢٧ تصبح ١٢٥٠٠، وهكذا.

كما يفضل أن يُعبر عن قيم الأغصان بالألف. فمثلاً القيمة ١٠٢٠٠ تقسم إلى جزءين ١٠ ، ١٠ حيث يوضع العدد ١٠ في عمود الأغصان والرقم ٢ في صف الأوراق التي تنتمي إلى هذا الغصن، وبذلك يكون شكل الأغصان والأوراق لهذه المجموعة من الدرجات كالتالى (شكل ٣ - ٦):

الأغصان	الأوراق
٩	٣
١٠	٨ ٩
١١	١ ٢ ١ ٦
١٢	٢ ١ ٥
١٣	٢
١٤	٢
١٥	٢
١٦	١

الوحدة ١٠٠ درجة
الغصن ١٠٠ درجة

شكل (٣ - ٦) يوضح الأغصان والأوراق لمجموعة من الدرجات باستخدام وحدة ١٠٠ درجة

وبالطبع يمكن أيضًا تقسيم كل غصن إلى عدد أكبر من الأغصان إذا تطلب التوزيع التوسيع، أو ضم صفوف أوراق متباورة إذا كانت كثيرة من الأغصان بدون أوراق أو أوراقها قليلة.

ففي الشكل (٣ - ٦) يلاحظ انتشار الأوراق رغم قلة عددها؛ وذلك لأن العدد الكلى للدرجات قليل، ولذلك ربما تحتاج الصفوف المتباورة قليلة الأوراق إلى الضم، مع مراعاة أن تظل الفروق الفردية بين الدرجات واضحة بحيث يمكن أن تستخلص منها أكبر قدر من المعلومات المفيدة عن التوزيع.

استخدام شكل الأغصان والأوراق في المقارنة بين مجموعتين

من الدرجات:

سبق أن أشرنا إلى كيفية المقارنة بين الشكل العام لكل من مجموعتين من الدرجات باستخدام المنحني التكراري، وسوف نوضح كيفية إجراء ذلك باستخدام شكل الأغصان والأوراق. وقد لاحظنا أننا نستخدم النظام العددي في عمود الأغصان بدلاً من سعة الفئة في جدول التوزيع التكراري، ولعل هذا ييسر إنشاء شكل الأغصان والأوراق في وقت قصير، ولكن هذا يؤدي إلى مقارنة غير مناسبة بين مجموعات مختلفة من الدرجات. فإذا كان لدينا مجموعتين من الدرجات وأمكن تمثيلهما بشكلين مختلفين من أشكال الأغصان والأوراق ولكن بعمود أغصان مشترك، فإنه ينبغي أن تكون الوحدة المستخدمة لقييم أو درجات كل من المجموعتين متساوية، فلا يجوز أن تكون الوحدة لأخذ المجموعتين درجة واحدة، وتكون الوحدة ١٠ درجات للمجموعة الأخرى وبخاصة إذا كان التوزيعان لهما نفس النمط أو الشكل العام. إذ يصعب معرفة الارتباط النسبي لكل من التوزيعين في حالة اختلاف الوحدة المستخدمة في عمود الأغصان مما يؤدي إلى مقارنات غير دقيقة. ولعل هذا أيضًا ما جعلنا نشير إلى استخدام التكراري المئوي بدلاً من التكرار عند المقارنة بين مجموعتين من الدرجات باستخدام المنحني التكراري كما سبق أن أوضحتنا.

ولمعرفة كيفية إبراز مقارنة بين مجموعتين من الدرجات في اختبار ما باستخدام شكل الأغصان والأوراق، نفترض أنه أمكن تمثيل مجموعتين من الدرجات بشكل الأغصان والأوراق التالي :

أوراق المجموعة الأولى	الأغصان	أوراق المجموعة الثانية
١٥	٤	
٦٧١٣	٥	٣٢
٨٢٠٤٥٢	٦	٨٧
٩٣٢٤٠	٧	
٧٣١٢٥٦٨	٨	٢٤٥
٣٢٠٥٥٤١٢٣	٩	٣٨
١٢٧٣	١٠	٧
٤٥	١١	٨٠٢١٠٩٧
	١٢	٣٥٣٨٤
٢	١٣	٤٧٢١

الوحدة درجة واحدة لكل من المجموعتين

شكل (٢ - ٧) يوضح الأغصان والأوراق لمجموعتين من الدرجات

وبالنظر إلى شكل (٣ - ٧) يتضح أنه على الرغم من أن درجات المجموعة الأولى تتراوح بين ٤١، ١٣٢ والثانية بين ٥٢، ١٣١ ، أي أن انتشار كل من المجموعتين متقارب، إلا أنها للاحظ أن ذيل توزيع المجموعة الأولى من الدرجات يتوجه بدرجة أكبر نحو الدرجات الدنيا بينما يتوجه ذيل توزيع المجموعة الثانية نحو الدرجات العليا. كما أن درجات المجموعة الأولى تتركز بين ٦٠، ٩٥ (٥٪ من الدرجات) بينما تتركز درجات المجموعة الثانية بين ١١، ١٢٨ (٦٤٪ من الدرجات) مما يدل على اختلاف الشكل العام لكل من التوزيعين، فأخذهما يميل إلى الاتواء تجاه اليمين والآخر تجاه اليسار، وهذا يعني أن مستوى أداء أفراد المجموعة الثانية أفضل من مستوى أداء أفراد المجموعة الأولى.

وبالطبع كلما كان شكل الأغصان والأوراق لكل من المجموعتين أكثر تماثلاً، فإن ذلك يجعل التفسير أيسر. فإذا كان التوزيعان مختلفين اختلافاً ملحوظاً فإنه يمكن إجراء تحويلات معينة على الدرجات، مثل التحويل اللوغاريتمي لمعالجة اختلاف الشكل العام للتوزيعين وجعلهما أكثر تماثلاً. وسوف نتناول بعض أنواع أخرى من التحويلات في هذا الفصل تمكننا من المقارنة بين توزيعات الدرجات بدقة أكبر . ومن

الجدير بالذكر أن هناك برامج للحاسوب يمكن باستخدامها الحصول على أشكال الأغصان والأوراق وإجراء التحويلات المختلفة على الدرجات مثل حزمة برامج MINITAB التي يمكن إدارتها ببيئة النافذ Windows.

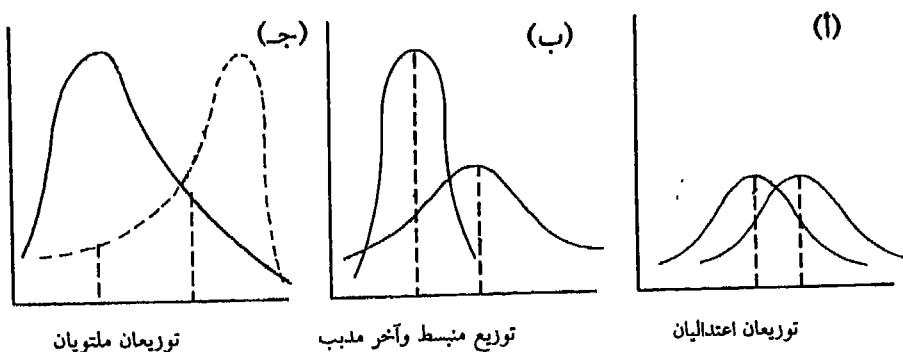
ووصف توزيعات درجات الاختبارات والمقاييس:

عرضنا في الجزء السابق كيفية تنظيم درجات الاختبارات والمقاييس التربوية والنفسية بطرق متعددة كخطوة تمهيدية لوصف توزيعات هذه الدرجات، وإلقاء الضوء على الفروق بين الأفراد في السمة التي يقيسها الاختبار أو المقاييس.

ويعتمد توزيع هذه الفروق الفردية على طبيعة السمة المقاسة في المجتمع وعينة الأفراد ومحض الاختبار أو المقاييس الذي يستخدم في قياس هذه السمة. غير أننا إذا بحثنا ظاهرة نفسية معينة مثل الذكاء أو التوافق أو القلق لدى عينة من طلاب المدارس الثانوية مثلاً، فإننا نجد أن درجات مجموعة كبيرة من هؤلاء الطلاب الذين يتم انتقاوهم عشوائياً من مجتمع طلاب المدارس الثانوية بعامة يتخد شكلًا يشبه الجرس المقلوب ويسمى المنحنى الاعتدالي Normal Curve حيث نجد عدداً قليلاً نسبياً من ذوى الذكاء المنخفض، وعدداً قليلاً نسبياً من ذوى الذكاء المرتفع، بينما نجد أن غالبية الطلاب ذكاؤهم متوسط، أي يقل التكرار كلما اتجهنا نحو طرفى المنحنى بعيداً عن المتوسط.

وهذه الخاصية تنطبق في كثير من الأحيان على معظم الظواهر النفسية والتربوية، أى الميل إلى التمركز أو التزعة المركزية Central Tendency مع مراعاة الشروط التي أشرنا إليها.

وتختلف التوزيعات التكرارية للدرجات بالطبع باختلاف خصائص التوزيع، فمنها ما هو منبسط ومنها ما هو مدبب ومنها ما هو متوسط التواء موجباً أو سالباً، فهذه التوزيعات تختلف في التزعة المركزية وكذلك في التشتت كما هو مبين بالأشكال التالية (٣ - ٨ ، ب، ج).



شكل (٣ - ٨) يوضح بعض أشكال توزيعات الدرجات

مقاييس النزعة المركزية أو الموضع :

إن الوصف الكمي للسمات الإنسانية يتطلب مقاييس إحصائية مناسبة، إذ لا يكفي أن نشخص توزيع درجات الاختبارات والمقاييس التربوية والتفسية من حيث النمط أو الشكل العام وتراكم الدرجات أو تطرفها، وإنما ينبغي أن نحصل على قيم تمثل النزعة إلى التراكم حول درجة معينة. ومن أهم مقاييس النزعة المركزية أو الموضع **Mode**، **Median** ، **Mean** ، والمتوسط **Measures of Location** ولكل من هذه المقاييس مزاياها وعيوبها واستخداماتها. ويعتمد اختيار المقاييس المناسب على الشكل العام لتوزيع الدرجات، ومستوى قياس المتغير الذي يمثل سمة معينة، والتساؤلات التي تود الإجابة عنها. لذلك سوف نوضح هذه المقاييس الثلاثة فيما يلى :

المتوسط الحسابي: يعد من أهم مقاييس النزعة المركزية أو الموضع حيث إنه شائع الاستخدام في تفسير درجات الاختبارات، فكثيراً ما نقول أن درجة فرد ما في أحد الاختبارات تفوق المتوسط أو تقل عن المتوسط. كما يعد أكثر هذه المقاييس استخداماً في دراسة الفروق الفردية والتعبير عنها بقيمة واحدة متوسطة.

ويمكن ببساطة الحصول على قيمة المتوسط الحسابي لمجموعة من الدرجات بقسمة مجموع هذه الدرجات على عددها، وعادة يرمز للمتوسط الحسابي بالحرف سـ.

$$\text{أى أن : } \bar{x} = \frac{\text{مج س}}{ن} \dots \dots \dots (1-3).$$

حيث ترمز (\bar{x}) إلى مجموع الدرجات .

، وترمز (n) إلى عدد الدرجات .

فإذا كان لدينا الدرجات ٢، ٤، ٦، ٨، ١٠ مثلاً،

$$\text{فإن المتوسط الحسابي } \bar{x} = \frac{٣٠}{٦} = \frac{١٠+٨+٦+٤+٢}{٥}$$

وعلى الرغم من أن الصيغة $(\bar{x} = \frac{\text{مج س}}{ن})$ يمكن استخدامها إذا كان عدد الدرجات كبيراً، إلا أنه يفضل تبويب الدرجات في هذه الحالة في جدول يتكون من عمودين أحدهما يحتوى على الدرجات والأخر يحتوى على التكرارات المناظرة للدرجات.

فمثلاً إذا كان لدينا الدرجات :

٢، ٤، ١٠، ٦، ١٠، ٦، ٤، ٨، ٦، ٨، ١٠

فإننا تكون الجدول التالي:

الدرجة (س)	تكرار الأفراد (ت)
٢	١
٤	٢
٦	٣
٨	٢
١٠	٤
<hr/>	
المجموع	١٢

ولإيجاد قيمة المتوسط الحسابي نضرب كل درجة (س) في التكرار المتناظر لها (ت) ونجمع الناتج، ثم نقسم هذا الناتج على عدد الأفراد.

$$\text{أى أن : س} = \frac{\text{مجمـ (س} \times \text{ت)}}{\text{مجـ ت}} = \frac{(٤ \times ١٠) + (٢ \times ٨) + (٣ \times ٦) + (٢ \times ٤) + (١ \times ٢)}{١٢} =$$

$$= \frac{٨٤}{١٢} = \frac{٤٠ + ١٦ + ١٨ + ٨ + ٢}{١٢}$$

ونظراً لأن دراسة الفروق الفردية دراسة علمية والتوصيل إلى نتائج كمية يمكن الاستناد إليها في تفسير هذه الفروق يتطلب تطبيق الاختبارات أو المقاييس المناسبة على عدد كبير نسبياً من الأفراد، وإيجاد المتوسط الحسابي لهذا العدد الكبير من الدرجات باستخدام إحدى الطريقتين السابقتين يتطلب جهداً كبيراً، لذلك ينبغي استخدام برامج الحاسوب في إجراء ذلك.

الوسيلـ، أحياناً يكون توزيع درجات اختبار ما ملتوياً كما في شكل (٩ - ٣) التالي، لأن يكون الاختبار صعباً أو سهلاً لكثير من الأفراد الذين يجري عليهم الاختبار. ونظراً لأن المتوسط الحسابي يتأثر كثيراً بالقيم المتطرفة في التوزيع أي القيم

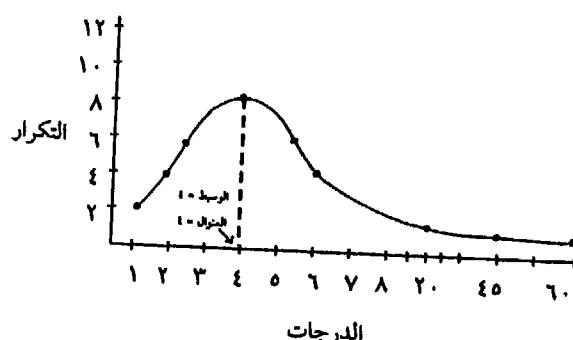
الكبيرة أو الصغيرة نسبياً، فإنه لا يصلح استخدامه في هذه الحالة لأنّه سيؤدي إلى قيمة غير مماثلة للتوزع المركبة، ولذلك يفضل استخدام الوسيط.

ولتوضيح ذلك نفترض أن لدينا توزيع الدرجات التالية الموضح بجدول (٣ - ٢) التالي:

الدرجات	١	٢	٣	٤	٥	٦	٢٠	٤٥	٦٠
التكرار	١	١	٢	٤	٧	٨	٥	٣	٢

| جدول (٣ - ٢) توزيع تكراري لمجموعة من الدرجات

ويمكن تمثيل هذا التوزيع بالمنحنى التكراري الموضح بشكل (٩-٣) التالي:



شكل (٣ - ٣) يوضح المنحنى التكراري لمجموعة الدرجات الموضحة بالجدول (٢-٢)

ويتبّع من شكل (٣ - ٣) الذي يمثل المنحنى التكراري لتوزيع الدرجات المذكورة في جدول (٣ - ٢) أن هناك ذيلاً للمنحنى تجاه الدرجات المرتفعة، ٢٠، ٤٥، ٦٠، أي أن المنحنى ملتوٍ التواء موجياً. قيمة المتوسط الحسابي في هذه الحالة باستخدام المعادلة (٣ - ٢) تساوى ٩,٧ تقريباً، وهي قيمة لا تعبر بدقة عن التوزع المركبة للتوزيع؛ لأنها تأثرت تأثيراً واضحاً بالقيم المتطرفة تجاه ذيل المنحنى، ولذلك يفضل استخدام الوسيط في هذه الحالة. وهذا يتطلب ترتيب الدرجات ترتيباً تصاعدياً أو تناظرياً، وإيجاد القيمة التي تتوسط الدرجات. فالوسيط إذن هو القيمة التي تقسم التوزيع إلى قسمين متساوين.

إذا كان عدد الدرجات فردياً تكون هناك قيمة واحدة، أما إذا كان عددها زوجياً تكون هناك قيمتان نوجد متوسطهما.

ونظراً لأن تكرار الدرجات الموضحة في جدول (٣ - ٢) أى فردياً، فإن هناك قيمة واحدة للوسيط. وفيما يلى الترتيب التصاعدى لهذه الدرجات:

$$\begin{array}{ccccccccc} & & & & 4 & 4 & 4 & 4 & 4 \\ & & & & \boxed{4} & 4 & 4 & 4 & 4 \\ 11 & 222 & 333333 & 6666 & 5000000 & 40 & 20 & 2 & 1 \\ & & & & . & = & \frac{1+33}{2} & & \end{array}$$

قيمة الوسيط = ٤ ، وذلك لأن الدرجة ٤ التي بداخل المربع تقسم التوزيع إلى قسمين متساوين.

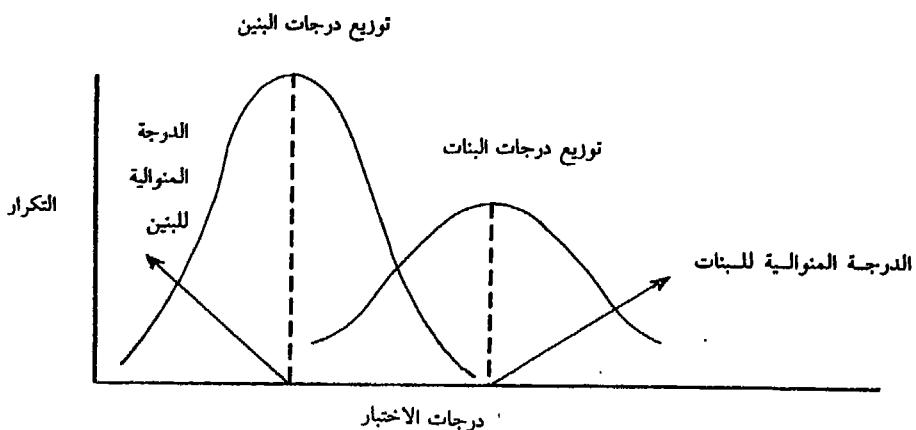
ويلاحظ أن قيمة المتوسط أكبر من قيمة الوسيط بدرجة واضحة وذلك لتأثير المتوسط بالقيم ٢٠، ٤٥، ٦٠ بينما لم يتأثر الوسيط.

لذلك فإن الوسيط في هذه الحالة أكثر تمثيلاً ودقة للتزعة المركزية من المتوسط نظراً للتواه التوزيع. ويلاحظ أن قيمة الوسيط اعتمدت على ترتيب الدرجات فقط دون إجراء أي عمليات حسابية، وذلك لأنه يتطلب أن يكون القياس من المستوى الرتبى على الأقل.

المنوال : إذا قلنا أن ذكاء فرد ما متوسط، فإن هذا ربما يعني أن الدرجة التي حصل عليها في اختبار الذكاء هي الدرجة الأكثر شيوعاً بين أقرانه، وعندئذ تكون درجته هي الدرجة المنوالية.

فالمنوال إذن هو القيمة الأكثر شيوعاً أو تكراراً في توزيع الدرجات، فقيمة منوال الدرجات المذكورة في جدول (٣ - ٢) تساوى ٤ نظراً لأن هذه الدرجة تكررت أكثر من غيرها من الدرجات. وبالنظر إلى التوزيع الملتوي الموضح بشكل (٣ - ٩) نلاحظ أن الدرجة ٤ تناظر قمة التوزيع التي يحددها الخط السريري المتقطع، وتمثل الدرجة المنوالية لأنها تناظر أكبر تكرار وهو ٨.

وأحياناً يكون للتوزيع قيمتين أى منوالين مما يدل على تركيز الدرجات عند أكثر من موضع. فإذا طبقنا اختباراً على مجموعةين أحدهما من البنين والآخر من البنات، فإننا ربما نجد أن قمة المنحنى التكراري للتوزيع درجات البنين تختلف عن قمة المنحنى التكراري للتوزيع درجات البنات كما في شكل (٣ - ١٠) التالي:



شكل (٣ - ١٠) يوضح منحنيان تكراريان لدرجات مجموعتين مختلفتين

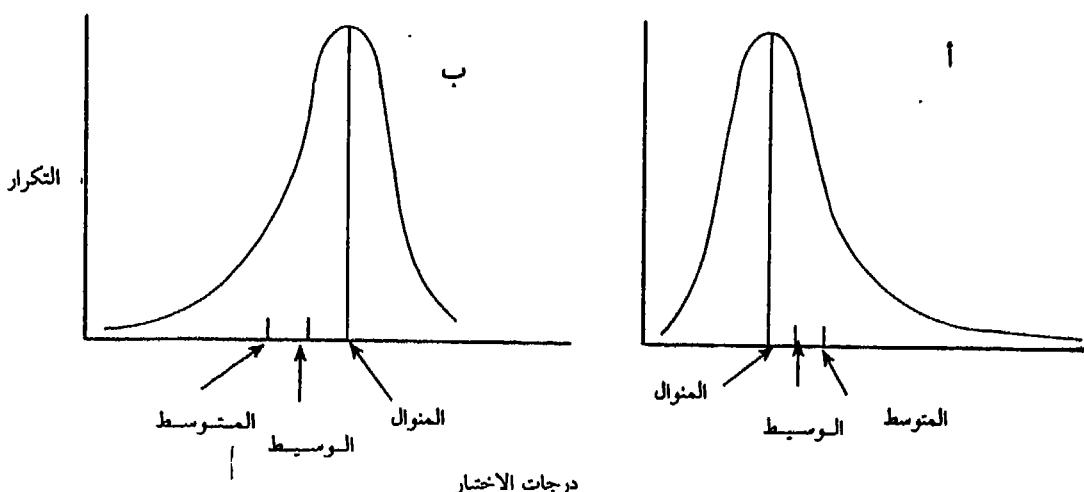
ويتضح من شكل (٣ - ١٠) أنه على الرغم من أن التكرار المناظر للدرجة المنوالية للبنين أكبر من التكرار المناظر للدرجة المنوالية للبنات، إلا أن قيمة الدرجة المنوالية للبنات أكبر من قيمتها للبنين، فالدرجة المنوالية للبنات تفوق الدرجة المنوالية للبنين على الخط الأفقي الذي يمثل درجات الاختبار.

وعلى الرغم من أن المنوال يعد أحد مقاييس الموضع، إلا أن قيمه سريعة التذبذب، حيث إن إضافة أو حذف أي درجة من درجات التوزيع تؤدي إلى تغير قيمته جوهريًا، ولكنه يناسب المتغيرات التصنيفية أي مستوى قياسها يكون أسمياً.

فمثلاً الطالب الذي يتُّخب رائداً للفصل هو ذلك الطالب الذي يحظى بأكبر عدد من الأصوات من أقرانه، كما أنه يعد أبسط مقاييس الموضع من حيث عملياته الحسابية.

العلاقة بين مقاييس النزعة المركزية أو الموضع:

ذكرنا فيما سبق أن التوزيعات التكرارية يمكن أن تكون اعتدالية أي متماثلة أو ملتوية التوء موجباً أو سالباً، فإذا كان التوزيع اعتدالياً أو متماثلاً وأحادي المنوال، فإن قيم المتوسط والوسط والمتوسط تكون متساوية، أما إذا كان التوزيع أحادى المنوال وملتوياً التوء موجباً أو سالباً فإن ذيل التوزيع يتوجه نحو الدرجات العليا أو الدرجات الدنيا على الترتيب كما هو موضح بالشكليين (٣ - ١١)، (٣ - ١١ ب) على الترتيب:



شكل (٣ - ١١) يوضح العلاقة بين مقاييس الموضع إذا كان التوزيع ملتويَا التواه موجياً أو سالباً

فإذا نظرنا إلى شكل (٣ - ١١) نلاحظ أن قيمة المتوسط تأثرت بالقيم العليا المتطرفة على الخط الأفقي الممثل لدرجات الاختبار، بينما نجد قيمة الوسيط تقسم التوزيع إلى قسمين متساوين. وقيمة المنوال هي الدرجة الماناظرة لأكبر تكرار الذي يمثله الخط الرأسى من قمة المحننى. ففى هذه الحالة تكون قيمة المتوسط أكبر من قيمة الوسيط، وقيمة الوسيط أكبر من قيمة المنوال.

وإذا نظرنا أيضاً إلى شكل (٣ - ١١ ب) نلاحظ أن العكس صحيح، حيث إن قيمة المتوسط أصبحت أصغر من قيمة الوسيط لتأثيره بالقيم الدنيا المتطرفة على خط الدرجات، وقيمة الوسيط أكبر من قيمة المنوال.

لذلك ينبغي مراعاة ما يلى عند استخدام مقاييس التزعة المركزية، أو الموضع لوصف توزيعات درجات الاختبارات والمقاييس التربوية والتفسية وصفاً كمياً :

(١) إذا كان مستوى قياس المتغير أو السمة فتررياً أو نسبياً يفضل استخدام المتوسط النحاسى، وإذا كان مستوى رتبياً يستخدم الوسيط، أما إذا كان المتغير اسمياً أو تصنيفياً فالمنوال يكون مناسباً في هذه الحالة.

(٢) إذا كان توزيع الدرجات ملتويَا التواه ملحوظاً، أي يشتمل على درجات

متطرفة، فإنه يفضل استخدام الوسيط كمقاييس للموضع، حيث إن الوسيط لا يتأثر كثيراً بهذه الدرجات، وإنما يتأثر فقط بعدد الدرجات، ويسهل عندئذ إيجاد قيمة الوسيط بترتيب الدرجات ترتيباً تنازلياً أو تصاعدياً واختيار الدرجة الوسطى.

(٣) بعض مقاييس التشتت التي ستناولها بعد قليل مثل الانحراف المعياري والتباين يعتمدان في حسابهما على المتوسط. لذلك إذا طلب الأمر وصف تشتت توزيع الدرجات كمياً، فإنه يفضل استخدام المتوسط كمقاييس للموضع.

وعموماً، فإنه يصعب وصف توزيع الدرجات بقيمة واحدة حيث إننا نفقد بذلك كثيراً من المعلومات المتعلقة بالتوزيع، لذلك فإنه يفضل إيجاد قيم المتوسط والوسيط والمنوال، ويكون لدى مستخدم الاختبار معلومات تفصيلية عن توزيع درجاته، حيث إن كلاً من هذه المقاييس الثلاثة تقدم معلومات مفيدة مختلفة.

مقاييس التشتت :

إن مقاييس النزعة المركزية أو الموضع ليست كافية للمقارنة بين توزيعات درجات الاختبارات المختلفة أو بين توزيعات درجات اختبار ما طبق على مجموعات مختلفة من الأفراد، فلو أن متوسط درجات تلاميذ أحد فصول الصف السادس في اختبار الحساب مثلاً ٣٥، ومتوسط درجات تلاميذ فصل آخر في الاختبار نفسه ٣٥، فإننا لا نستطيع استنتاج أن مستوى تحصيل كل من الفصلين متكافئ، إذ ربما تكون درجات كل تلميذ في الفصل الأول ٣٥، بينما حصل بعض تلاميذ الفصل الثاني على درجات أعلى بكثير من ٣٥، وقلة منهم حصلت على درجات منخفضة، ومع هذا فإن المتوسط في كل من الحالتين يمكن أن تكون قيمته ٣٥.

إذا نظرنا إلى مجموعتي الدرجات التاليتين :

$$\begin{array}{ccccccc} 35 & , & 35 & , & 35 & , & 35 \\ 45 & , & 55 & , & 60 & , & 7 \end{array}$$

نجد أن قيمة المتوسط في كل من الحالتين ٣٥، ومع هذا فإن درجات المجموعة الثانية أكثر تشتتاً من درجات المجموعة الأولى، بل إن المجموعة الأولى تشتتها صفراء.

لذلك ليس من الصواب أن نقارن بين المجموعتين على أساس المتوسط بمفرده، وإنما يلزم أيضاًأخذ قيمة تشتت كل من المجموعتين بعين الاعتبار لكي تكون المقارنة دقيقة وواقعية.

أى أننا يجب أن نهتم عند دراسة الفروق الفردية في إحدى السمات بمتوسطات مجموعات الأفراد في هذه السمة وكذلك تشتت درجات كل مجموعة حول متوسطها. وهناك مقاييس متعددة للتشتت، ولكننا سوف نقتصر على المقاييس التي يمكن الإفادة منها في تفسير درجات الاختبارات والمقارنة بينها. وهذه المقاييس هي: المدى، والانحراف المعياري، والتباين.

المدى : يعد المدى Range أبسط مقاييس التشتت في حسابه، إذ يعتمد على الفروق بين أكبر درجة وأصغر درجة في مجموعة الدرجات، فمثلاً :

$$\text{مدى الدرجات } 15, 12, 11, 9, 6 \text{ يساوى } 15 - 6 = 9$$

$$\text{والمدى الدرجات } 48, 48, 48, 48, 39 \text{ يساوى } 48 - 39 = 9$$

أى أن مدي درجات كل من المجموعتين متساوٍ.

ولكن إذا تأملنا مجموعتي الدرجات نجد أن المجموعة الثانية أكثر تجانساً أو أقل تشتتاً من المجموعة الأولى، وربما يرجع هذا التناقض إلى أن المدى يعتمد على قيمتين فقط، فهو يتاثر تأثيراً بالغاً بإضافة درجة أو قيمة واحدة إلى الدرجات مما يجعله مضللاً كما رأينا ، فالمدى لا يأخذ القيم الوسطى في الحسبان.

لذلك لابد من الاعتماد في حساب التشتت على مقاييس أكثر دقة منها الانحراف المعياري Standard Deviation ، والتباين .

الانحراف المعياري : يعد أكثر مقاييس التشتت شيوعاً واستخداماً في دراسة الفروق الفردية، بل ويعد أكثرها دقة، فهو يعتمد على جميع درجات التوزيع وليس على درجتين فقط كما في المدى.

ونحصل على قيمة الانحراف المعياري لمجموعة من الدرجات بأن نوجد انحراف كل درجة عن المتوسط، ونربع كل من هذه الانحرافات، ونجمع الناتج، ونقسمه على عدد الدرجات، ثم نستخرج الجذر التربيعي لخارج القسمة.

$$\text{أى أن الانحراف المعياري } \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

فإذا كانت الدرجات 8، 8، 9، 9، 9، 9، 10، 10، فإننا يفضل تكون جدول كالتالي (٣ - ٣) :

$(س - س_{\bar{}})^2$	$س - س_{\bar{}} = ح$	الدرجات (س)
١	$1 - 9 - 8$	٨
١	$1 - 9 - 8$	٨
صفر	$= صفر = 9 - 9$	٩
صفر	$= صفر = 9 - 9$	٩
صفر	$= صفر = 9 - 9$	٩
صفر	$= صفر = 9 - 9$	٩
١	$1 + = 9 - 10$	١٠
١	$1 + = 9 - 10$	١٠
<hr/> $\text{مجم } (س - س_{\bar{}})^2 = 4$		
<hr/> $\text{مجم } (س - س_{\bar{}}) = \text{صفر}$		
<hr/> $\text{المجموع مجم س} = 72$		
<hr/> $\text{المتوسط س}_{\bar{}} = 9$		

جدول (٣ - ٣) يوضح خطوات حساب قيمة الانحراف المعياري لمجموعة من الدرجات

يتضح من جدول (٣ - ٣) أن $\text{مجم } (س - س_{\bar{}})^2 = 4$.

$$\sqrt{\frac{\text{مجم } ح^2}{ن}} = \sqrt{\frac{\text{مجم } (س - س_{\bar{}})^2}{ن}}$$

فعدئذ تكون ع =

$$\sqrt{0,5} = \sqrt{0,71} = \sqrt{0,48} =$$

ويمكن تبسيط العمليات الحسابية إذا استخدمنا الصيغة (٤ - ٣) التالية لإيجاد قيمة الانحراف المعياري:

$$\sqrt{\frac{\text{مجم } (س^2) - (\text{مجم س})^2}{ن}} = ع$$

وهذه الصيغة لا تتطلب إيجاد انحرافات الدرجات عن المتوسط، وإنما تتطلب فقط إيجاد مجموع مربعات الدرجات (مجد س^٢).
ففي المثال السابق مجد س^٢ = ٦٥٢.

$$\begin{array}{r} ٢(٧٢) - ٦٥٢ \\ \hline ٨ \\ \hline ٨ \end{array} \quad \boxed{= \sqrt{}} \quad \text{ع}$$

$$\begin{array}{r} ٦٤٨ - ٦٥٢ \\ \hline ٨ \end{array} \quad \boxed{= \sqrt{}}$$

$$٧١,٥ \quad \boxed{= \sqrt{\frac{٤}{٨}}} \quad \boxed{= \sqrt{}}$$

وهي نفس القيمة السابقة.

وكلما اقتربت قيمة الانحراف المعياري من الصفر دل ذلك على قلة التشتت.
ونظراً لأن القيمة ٧١,٥ قريبة من الصفر، فإن معنى ذلك أن مجموعة الدرجات متجانسة، ويمكن ملاحظة ذلك إذا نظرنا إلى الدرجات الثمانية.

$$\text{التبالين : هو مربع الانحراف المعياري أي : } \sigma^2 = \frac{\text{مجد } (س - س_{\bar{}})^2}{ن}$$

فتبالين الدرجات في المثال السابق = (٧١,٥٠)^2 = ٥٠,٧١ . وهي قيمة قريبة أيضاً من الصفر مما يدل على تجانس الدرجات وقلة تشتتها.

طريقة تقريرية لإيجاد قيمة الانحراف المعياري :

يمكن إيجاد قيمة تقريرية للانحراف المعياري والتبالين ببساطة وسرعة إذا كان عدد الدرجات ٢٠ أو أكثر، مثل درجات الاختبارات التحصيلية الصافية التي يختبر المعلم بها تلاميذه في فصل معين؛ وذلك لأن القيم الناتجة تقترب من القيم الفعلية إذا لم يقل

عدد الدرجات عن ذلك، وتصبح القيم متساوية تقريرياً للقيمة الفعلية إذا كان عدد الدرجات ٣٦ أو أكثر. وقد اقترح لاثروب Lathrop الصيغة التالية لإجراء ذلك :

$$\text{مجموع السادس الأعلى من الدرجات} - \text{مجموع السادس الأدنى} = \frac{\text{نصف}}{\text{عدد الدرجات}} \cdot 1$$

ويمكن تطبيق هذه الصيغة على مجموعة الدرجات المذكورة في جدول (٣ - ٣) بأن نرتب الدرجات ترتيباً تناظرياً كالتالي:

١٠ ، ١٠ ، ٩ ، ٩ ، ٩ ، ٩ ، ٨ ، ٨ ، ٨ ، ٨

ثم نأخذ $\frac{1}{6}$ الدرجات أى $\frac{1}{6} \times 8 = 1,3$ درجة.

ولكتنا سنقرها إلى درجة واحدة لكل من السادس الأعلى والسادس الأدنى.

$$\text{ع} = \frac{10 - 8}{\frac{1}{2}} = \frac{2}{\frac{1}{2}} = 4,0 \text{ تقريباً.}$$

وإذا قارئنا هذه القيمة التقريرية بالقيمة التي حصلنا عليها باستخدام الصيغة (٣ - ٣) نجد أن الفرق ١٤،٠ فقط، ويرجع هذا الفرق إلى قلة عدد الدرجات، فكلما زاد العدد قل هذا الفرق بحيث يتحمل ألا يتعدى ١،٠ أو ٢،٠ على الأكثر.

استخدامات الانحراف المعياري:

(١) يمكن استخدام الانحراف المعياري والمتوسط معًا للمقارنة بين مجموعتين أو أكثر من الدرجات، نظراً لأن المتوسط بمفرده لا يصلح لذلك كما سبق أن أشرنا.

ولكن يشترط أن تكون الاختبارات المستمدة منها الدرجات لها نفس وحدات القياس، فالانحراف المعياري يختلف بقدر اختلاف وحدات القياس. لذلك إذا طبق اختباران أحدهما في الاستعداد اللغطي والآخر في القراءة على طلاب فصل واحد، ووجد أن قيمة الانحراف المعياري لدرجات الاختبار الأول أكبر من قيمته للاختبار الثاني، فلا يجوز القول بأن تشتت درجات الطلاب في الاختبار الأول أكبر من تشتت درجاتهم في الاختبار الثاني. فهذا الاستنتاج يتطلب التأكد من أن وحدات ميزان قياس الاستعداد الدراسي هي نفسها وحدات ميزان قياس التحصيل في القراءة. ولذلك يفضل أن تجرى المقارنات إذا طبق اختبار واحد على مجموعتين أو أكثر من الأفراد وذلك لوصف التجانس النسبي للمجموعتين في السمة التي يقيسها الاختبار.

(٢) يمكن أيضًا استخدام الانحراف المعياري في تحديد الأوزان النسبية للدرجات المركبة من مجموع درجات اختبارين أو أكثر: بعض اختبارات الذكاء ومقاييس الشخصية تكون من عدة اختبارات أو مقاييس فرعية مستقلة. والدرجة الكلية عبارة عن مجموع درجات هذه الاختبارات الفرعية، وكذلك درجة التحصيل العام للطالب تكون عادةً مركبة من مجموع درجاته في عدة اختبارات شهرية مستقلة، وفي مثل هذه الحالات لا يجوز جمع الدرجات الفرعية جمجمًا بسيطًا، وإنما يكون الجمع موزونًا. ونقصد بالجمع الموزون مراعاة الأهمية النسبية لكل درجة فرعية من حيث إسهامها في الدرجة المركبة. فأوزان الدرجات الفرعية تعتمد على الانحراف المعياري لهذه الدرجات بالنسبة لغيرها من الدرجات التي تشكل الدرجة المركبة، فكلما زادت مجموعـة من الدرجات زاد وزنها النسبي. لذلك ينبغي أن نعطي وزنًا لكل درجة فرعية يعادل قيمة الانحراف المعياري للدرجات الاختبار الفرعى المناظر. ويجوز أيضًا أن نجعل الوزن مساوياً قيمة الانحراف المعياري مضروباً في أي عامل نراه مناسباً لأهمية الاختبار الفرعى . ولتوسيع ذلك نفترض أن لدينا اختباراً ما يشتمل على ثلاثة اختبارات فرعية مستقلة أي غير مترتبة، وأردنـا الحصول على درجة مركبة من درجات الاختبارات الفرعية الثلاثة، ووجـنا أن الانحراف المعياري لهذه الاختبارات ٧ ، ٦ ، ١٠ على التوالى ، ولكنـا رأينا أنه نظرًا لأهمية الاختبار الثانى فإنـا سوف نضاعف قيمة الانحراف المعياري لهذا الاختبار، وفي الوقت نفسه رأينا تقليـل وزن الاختبار الثالث لقلة أهميته، وذلك بقسمـة قيمة انحرافـه المعياري على ٢ ، وبذلك تصبح أوزان الدرجات الفرعية ٧ ، ١٢ ، ٥ . وفي هذه الحالة تكون الدرجة المركبة مساوية لمجموع الدرجات التي يحصل عليها الفرد في الاختبارات الفرعية بعد ضرب كل درجة منها في الوزن المناظر لها ، أي أن الدرجة المركبة تكون مساوية لمجموع الدرجات الموزونة بقيم الانحراف المعياري أو مضاعفاته أو أجزائه .

(٣) يمكن كذلك استخدام الانحراف المعياري في إجراء تحويلات خطية على الدرجات :

وتفيد هذه التحويلات في تحديد مركز الفرد الذي حصل على درجة معينة في الاختبار بالنسبة لأقرانه الذين طبق عليهم الاختبار نفسه. كما تفيد في إجراء مقارنات أكثر دقة وواقعية بين مجموعـات من الدرجات المستمدـة من اختبارات تختلف في متوسطـاتها وانحرافـاتها المعيارية .

وبذلك تساعد هذه التحويلات في تكوين جداول معايير الاختبارات والمقيـاسـات التربـوية والنـفسـية . ونظرًا لأهمـية ذلك فإنـا سوف نناقشـه فيما يلى بمزيد من التفصـيل .

الدرجات المحوّلة :

تستخدم هذه الدرجات في تحديد المركز النسبي للفرد في توزيع ما، بحيث يمكن وصف أداؤه بالنسبة لأقرانه في اختبار يقيس سمة معينة.

ولعل هذا مرتبط ارتباطاً وثيقاً بما سبق أن ذكرناه في الفصل الأول عن الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests ، فهذه الاختبارات تهدف لتقدير أداء الفرد بالنسبة لأداء الأفراد الآخرين في السمة أو القدرة التي يقيسها الاختبار، ويكون الاهتمام هنا منصباً على ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم في الاختبار، ويكون للدرجة تفسيراً فقط في ضوء متوسط أداء الجماعة أو ما يسمى بمعيار الجماعة Group Norm ، وتصبح الدرجة لا معنى لها ويفسر تفسيرها إذا أغفلنا ردها إلى هذا المعيار.

وتوجد أنواع مختلفة من الدرجات المحوّلة لعل أكثرها استخداماً في القياس التربوي والنفسي الدرجات المعيارية Standard Scores ، والدرجات التائية T-Scores . ويعتمد هذان النوعان من الدرجات على التحويلات الخطية Linear Transformations لدرجات الاختبار باستخدام المتوسطات الحسابية والانحرافات المعيارية معاً.

الدرجات المعيارية

ذكرنا فيما سبق أن إجراء مقارنات بين الدرجات التي يحصل عليها فرد في اختبارات مختلفة يتطلب اشتراكاها في وحدة القياس نفسها. ويمكن إجراء ذلك بتحديد عدد وحدات الانحراف المعياري التي تبعد فيها درجة الفرد في كل من هذه الاختبارات عن متوسطات درجاتها، وتسمى هذه الوحدات «الدرجات المعيارية Z Scores » .

فإذا افترضنا أن الدرجات التي حصل عليها طالب في أربعة اختبارات تختلف في متوسطاتها، وإنحرافاتها المعيارية كانت كالتالي (جدول ٣ - ٤) :

الاختبار	درجة الطالب	المتوسط	الانحراف المعياري (ع)
	س	(س)	(ع)
الأول	٩	١٠	١
الثاني	١٠	١٣	٣
الثالث	٧	٧	٢
الرابع	١٠	١٣	٢

جدول (٣ - ٤) يوضح درجات طالب في أربعة اختبارات ومتوسطات درجات الاختبارات وإنحرافاتها المعيارية

ويتضح من جدول (٤ - ٣) أن درجة الطالب في كل من الاختبارات الأول والثاني والرابع تقل عن متوسط درجات أقرانه بقدر ١ ، ٣ ، ٣ وحدة انحراف معياري، بينما تقع درجته في الاختبار الثالث عند المتوسط.

وفي ضوء هذه الانحرافات عن المتوسطات يبدو أن أداء الطالب في الاختبارين الثاني والرابع منخفضاً بالقدر نفسه مع أن تشتت درجات الاختبار الثاني أكبر، حيث إن الانحراف المعياري لدرجات هذا الاختبار ٣.

فكما تراكمت الدرجات حول المركز النسبي للدرجات المرتفعة، وعلى العكس من ذلك كلما زاد التشتت يقل المركز النسبي لهذه الدرجات بين المجموعة.

لذلك لا ينبغي الاقتصار على الانحرافات عن المتوسط، وإنما يجبأخذ قيم الانحراف المعياري أيضاً بعين الاعتبار عند مقارنة درجات الطالب في الاختبارات الأربع، وذلك للتغلب على مشكلة اختلاف التوزيعات في المتوسطات والانحرافات المعيارية، وهذا يتطلب قسمة انحرافات الدرجات عن متوسطاتها على قيمة الانحراف المعياري المناظرة لها، ويسمى الناتج عندئذ الدرجات المعيارية (د) أي أن :

$$d = \frac{s_n - s_1}{s} \quad \dots \dots \dots \quad (٦ - ٣)$$

ومن الجدير بالذكر أن متوسط الدرجات المعيارية = صفر، وأنحرافها المعياري = الواحد الصحيح.

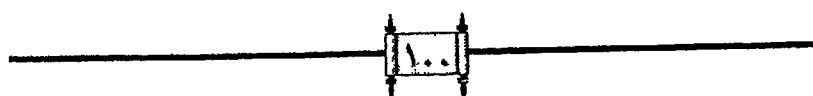
وبذلك تكون الدرجات المعيارية المناظرة لدرجات الطالب في الاختبارات الأربع على الترتيب كالتالي :

$$d_1 = \frac{10 - 9}{1} = 1$$

$$d_2 = \frac{13 - 1}{3} = 4$$

$$d_3 = \frac{7 - 7}{3} = صفر$$

$$d_4 = \frac{13 - 1}{2} = 6$$



ويمكن الآن المقارنة بين درجات الطالب في الاختبارات الأربع بالنسبة لأقرانه، فمركزه النسبي بين أقرانه في الاختبار الثالث أفضل من مركزه النسبي في بقية الاختبارات على الرغم من أنه حصل على الدرجة ٧ في هذا الاختبار وهي أقل من درجاته في الاختبارات الأخرى.

ولذلك فإن الدرجة ٧ التي تعد درجة خام لا تصلح للمقارنة إلا إذا تم تحويلها إلى درجة معيارية لكي يتحدد موقعها بالنسبة لمتوسط الدرجات.

ونظراً لأن الدرجات المعيارية هي درجات مجردة ليس لها وحدة خاصة ، فإنه يمكن جمع الدرجات المعيارية للطالب في الاختبارات الأربع للحصول على درجة معيارية مركبة. كما يمكن تعين أوزان مختلفة لهذه الدرجات قبل حساب الدرجة المركبة. فإذا أردنا أن يسهم الاختبار الأول بنصف ما يسهم به الاختبارين الثاني والثالث، وأن يسهم الاختبار الرابع بقدر مرة ونصف، فإن الدرجة المركبة تكون عندئذ كالتالي :

$$\text{الدرجة المركبة} = ٥٠ \cdot ١٥ + ٣٣ + ٢٢ + ١٥$$

بعض عيوب الدرجات المعيارية :

على الرغم من أهمية الدرجات المعيارية في إجراء المقارنات بين درجات الاختبارات المختلفة، إلا أن هناك بعض المشكلات العملية المتعلقة بها. فالتعلم ومستخدم الاختبار ينبغي عليه تفسير هذه الدرجات للطالب أو المفحوص في ضوء المتوسط والانحراف المعياري، كما أن الدرجات المعيارية تكون أحياناً سالبة أو كسرية مما يزيد من صعوبة تفسيرها.

فإذا نظرنا إلى المثال السابق نجد أن درجة الطالب في كل من الاختبارين الأول والثاني تبعد بقدر انحراف معياري واحد سالب عن متوسط أقرانه، ودرجته في الاختبار الرابع تبعد بقدر ١,٥ انحراف معياري، أي أنه يقل عن متوسط أقرانه في الاختبارات الثلاثة، فهنا يصعب تبرير القيم السالبة والكسرية للطالب أو المفحوص أو ولد الأمر.

ويمكن التغلب على هذه المشكلات بإجراء تحويلات معينة على الدرجات المعيارية والحصول على درجات معيارية معدلة تهدف للتخلص من الإشارات السالبة والكسرية العشرية التي تنتج عن تطبيق الصيغة (٦ - ٣) وبذلك يسهل تفسير هذه الدرجات وفهمها.

وتوجد عدة تحويلات يمكن إجراؤها على الدرجة المعيارية، غير أنها سوف تقتصر على التحويلات التي تستخدم بكثرة في إعداد معايير الاختبارات التربوية والنفسية، مثل الدرجة الثانية T -Scores، والمئويات Percentiles.

الدرجات الثانية:

تعد الدرجات الثانية (t) درجات معيارية معدلة، وتتخرج عن إجراء تحويل خطى للدرجات المعيارية (d). ونقصد بالتحويل الخطى أن نضرب كل قيمة من قيم الدرجات المعيارية (d) في مقدار ثابت وجمع مقدار ثابت آخر على ناتج الجمع. ولعل أفضل مثال يوضح ذلك هو تحويل درجات الحرارة من ميزان مئوى إلى ميزان فهرنهايتى تحويلياً خطياً، فضرب جميع القيم في مقدار ثابت يغير من سعة فترات ميزان القياس، وجمع مقدار ثابت يغير من وضع الصفر الاعتبارى على الميزان بقدر مساو لهذا المقدار الثابت، ولكن عملية الضرب والجمع لا تغيير العلاقات القائمة بين وحدات الميزان.

ولذلك فإن الصيغة العامة للتحويلات الخطية للدرجات المعيارية إلى درجات معدلة تكون كالتالى :

$$\text{الدرجة المعيارية المعدلة} = A + B \times \text{الدرجة المعيارية} \dots \dots \dots (3-7)$$

حيث (A) ، (B) مقداران ثابتان.

وعلى الرغم من أن قيمة كل من (A) ، (B) اختيارية، إلا أن ما كول McCall جعل المتوسط 50 بدلاً من الصفر، والانحراف المعياري 10 بدلاً من الواحد الصحيح لكنه نستطيع التخلص من الإشارات السالبة والقيمة الكسرية للدرجات المعيارية. والصيغة التالية تستخدم في إجراء هذا التحويل باستخدام الصيغة العامة (3-7) :

$$t = 50 + 10 \times d \dots \dots \dots (3-8)$$

ويختصار فإنه يمكن تحويل الدرجة المعيارية (d) إلى درجة تانية (t) بأن نضرب الدرجة المعيارية في 10 ونجمع 50 على حاصل الضرب.

ومن الجدير بالذكر أنه يمكن اختيار أي قيمة أخرى لكل من المتوسط والانحراف المعياري تختلف عن 50 ، 10 بحسب مقدار الدرجات.

إذا كانت الدرجات الخام لاختبار قيمها كبيرة بدرجة واضحة فإنه يمكن تصميم نظام مختلف لتحويل هذه الدرجات إلى درجات معيارية معدلة ، إذ ربما نجعل المتوسط 100 والانحراف المعياري 100 كما في اختبارات الاستعداد الدراسي، أو نجعل المتوسط 100 والانحراف المعياري 16 كما في اختبارات الذكاء وهكذا، وذلك بتطبيق الصيغة العامة (3-7).

ولتوسيع كيفية تطبيق صيغة الدرجة الثانية (ت)، نعود إلى جدول (٤ - ٣)، فنجد أن الدرجات الثانية للطالب المناظرة للدرجات المعيارية للاختبارات الأربع كال التالي (جدول ٣ - ٥) :

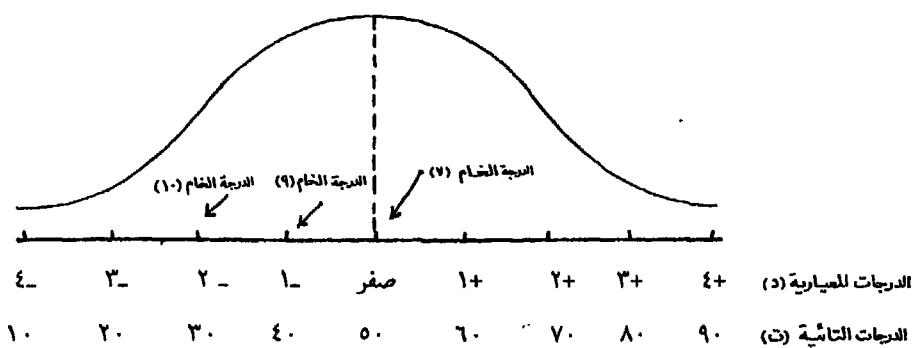
الدرجة الثانية $T = 100 + 5 \times D$	الدرجة المعيارية $D = \frac{S - M}{\sigma}$	الدرجة الخام (س)	الاختبار
٤٠	- ١	٩	الأول
٤٠	- ١	١٠	الثاني
٥٠	صفر	٧	الثالث
٣٥	- ١,٥	١٠	الرابع

جدول (٣ - ٥) يوضح الدرجات المعيارية والثانية المناظرة للدرجات

الخام في أربعة اختبارات

ويتبين من جدول (٣ - ٥) أن الدرجة الخام ٩ تناظر الدرجة المعيارية ١ - ، وهذه بدورها تناظر الدرجة الثانية ٤٠، وهذا يعني أن الدرجة الخام ٩ تنحرف بقدر انحراف معياري واحد عن المتوسط، وهكذا في بقية الدرجات.

ويلاحظ أن الدرجات الثانية قد تخلصت من الإشارات السالبة والقيم الكسرية في الدرجات المعيارية (د) ، وبذلك يصبح تفسيرها معقولاً . ونظرًا لأن الدرجات المعيارية (د) تمثل توزيعاً له معلمين ثابتين هما : المتوسط = صفر ، والانحراف المعياري = ١ ، فإن الدرجات الثانية تمثل التوزيع نفسه بمعلمين ثابتين أيضًا هما المتوسط = ٥٠ ، والانحراف المعياري = ١٠ . والدرجات المعيارية التي تفوق المتوسط (صفر) تنحرف عنه بقدر + ١ ، + ٢ ، + ٣ ، انحراف معياري، بينما تنحرف الدرجات المعيارية التي تقل عن المتوسط بقدر - ١ ، - ٢ ، - ٣ ، انحراف معياري. أما الدرجات الثانية فتنحرف عن المتوسط (٥٠) بقدر ١٠ وحدات معيارية إلى اليمين وإلى اليسار. ويمكن توضيح ذلك بشكل (٣ - ١٢) التالي:



شكل (١٢ - ٣) يوضح الدرجات المعيارية والدرجات التائية

ويتبين من شكل (٣ - ١٢) أن الدرجات المعيارية تتراوح عملياً بين $-3+ \text{ ، } 3-$ وحدة انحراف معياري، بينما تتراوح الدرجات التائية عملياً أيضاً بين $20 \text{ ، } 80$ وجميعها قيم صحيحة موجبة.

كما يتضح من الشكل المركز النسبي للطالب في الاختبارات الأربع حيث إنه لم ينحط المتوسط في أي من الاختبارات، بل إن درجاته في ثلاثة اختبارات تقل عن المتوسط حيث تناظر الدرجات التائية $40 \text{ ، } 40 \text{ ، } 30$ ، مما يدل على انخفاض أداءه فيها بالنسبة لأقرانه.

والخلاصة أن الدرجات المعيارية والدرجات المحوولة تحدد المركز النسبي للفرد في اختبار يقيس سمة معينة بوحدات انحرافات معيارية عن متوسط أقرانه الذين طبق عليهم الاختبار نفسه.

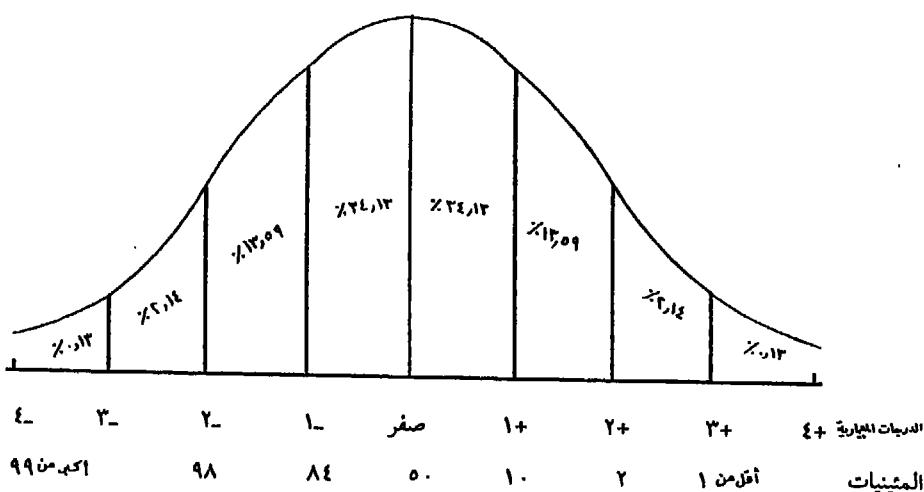
تحديد المركز النسبي للأفراد على المنحنى الاعتدالي:

انصح لنا مما سبق أن الاختبارات مرجعية الجماعة أو المعيار تهدف لتحديد المركز النسبي للفرد بين أقرانه في السمة التي يقيسها الاختبار.

لذلك يعمل مصممو هذا النوع من الاختبارات على إبراز الفروق الفردية والكشف عنها بحيث توزع الدرجات توزيعاً اعتدالياً. فالتوزيع الاعتدالي يمكن تمثيله بيانياً بمنحنى يعرف بالمنحنى الاعتدالي Normal Curve أو منحنى جاوس، ويتميز هذا المنحنى بأنه يشبه الجرس المقلوب، وهو منحنى متصل ومتماطل حول القيمة المئوية التي تناظر أقصى ارتفاع للمنحنى، ويمتد من كلتا الجهات إلى الlanهية، أي يقترب

تدرجياً من المحور الأفقي، ولكنه لا يمسه مهما مددناه من كلتا الجهازين. والمساحة الكلية تحت المنحنى، أي المساحة الممحصورة بين المنحنى والممحور الأفقي تساوي الواحد الصحيح.

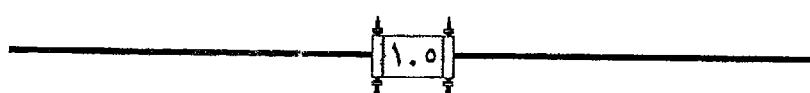
وتميز جميع المنحنيات الاعتدالية بأن نسبة الأفراد التي تتحضر درجاتهم بين درجتين معياريتين مختلفتين ثابتة، إذ يمكن باستخدام مبادئ علم التكامل في الرياضيات التوصل إلى أن حوالي ١٣٪٣٤ من درجات التوزيع الاعتدالي تقع بين المتوسط والدرجة المعيارية $1+$ ، $13,59\%$ من الدرجات تقع بين الدرجتين المعياريتين $1+$ ، $2+$ ، وحوالي $14,2\%$ من الدرجات تقع بين الدرجتين $2+$ ، $3+$. ونظراً لتماثل المنحنى فإن نفس النسب المئوية من الدرجات تقع بين الدرجات المعيارية السالبة. ويوضح شكل (٢ - ١٣) التالي هذه النسب على المنحنى الاعتدالي:



شكل (٢ - ١٣) يوضح النسب المئوية من الحالات تحت المنحنى

الاعتدالي الممحصورة بين الدرجات المعيارية المختلفة

ومن مزايا هذا المنحنى الاعتدالي المعياري أنه يمكن استخدامه في تحويل الدرجات المعيارية إلى مئويات Percentiles ، أي نسبة مئوية من عدد الأفراد الذين تقل درجاتهم عن درجة معيارية معينة كما سيوضح بعد قليل.



المئينيات :

تعد المئينيات Percentiles من أكثر الطرق استخداماً في تفسير درجات الاختبارات والمقياس التربوية والنفسية، فالمئينيات يمكن تفسيرها بسهولة للطالب والمفحوص والمعندين بتائج أدوات القياس. كما يمكن تحويل الدرجات المعيارية والدرجات الثانية وغيرها من الدرجات المعيارية المعدلة إلى مئينيات باستخدام المنهجي الاعتدالي المعياري الموضح بشكل (٣ - ١٣) .

المئيني هو نقطة على توزيع الدرجات تقع عندها أو أقل منها نسبة معينة من الأفراد الذين طبق عليهم اختبار معين.

فالوسيط مثلاً يناظر المئيني ٥ لأنّه يمثل نقطة تقسيم التوزيع إلى نصفين متساوين كما ذكرنا من قبل. فإذا نظرنا إلى شكل (٣ - ١٣) نجد أن الفرد الذي يحصل على درجة معيارية +٢ يتفوق على حوالي ٩٨٪ من أقرانه وبذلك فإن درجته المعيارية تنازلي المئيني ٩٨، والفرد الذي يحصل على درجة معيارية - ١ يتفوق على حوالي ٦٪ من أقرانه، ودرجته المعيارية تنازلي المئيني ٦، وهكذا.

ويمكن أيضاً تقسيم توزيع الدرجات إلى عشرة أقسام متساوية يسمى كل منها إعشاري Decile ، فالإعشاري التاسع مثلاً يناظر المئيني ٩٠. كما يمكن تقسيم التوزيع إلى أربعة أقسام متساوية يناظر كل منها إرباعي Quartile ، فالإرباعي الأول (R₁) يناظر المئيني ٢٥ ، والإرباعي الثاني (R₂) يناظر الوسيط أي المئيني ٥ ، والإرباعي الثالث (R₃) يناظر المئيني ٧٥ .

الإرباعي الأول هو نقطة على التوزيع تقع دونها ٢٥٪ من الحالات، والإرباعي الثالث هو نقطة أخرى على التوزيع تقع دونها ٧٥٪ من الحالات.

أما المدى الربيعي Interquartile Range فهو عبارة عن مدى الدرجات التي تمثل ٥٠٪ الوسطى من الحالات في التوزيع، أي هو الفرق بين الإرباعي الثالث والإرباعي الأول.

وتعد الإعشاريات والإرباعيات حالات خاصة من المئينيات، وتستخدم الإرباعيات أيضاً في تحديد مركز الفرد بالنسبة لأقرانه.

إذا قلنا أن درجة الفرد تنازلي الإرباعي الأول فإنّا نعني بذلك أن درجته تنازلي نقطة معينة على ميزان الدرجات تقع دونها ٢٥٪ من الدرجات، كما تستخدم في معرفة الشكل العام لتوزيع الدرجات، وذلك بإيجاد الفرق بين القيمتين المناظرتين للإرباعي

الثالث والإربعاء الثاني، وكذلك الفرق بين الإربعاء الثاني والإربعاء الأول، أي $(R_3 - R_2), (R_2 - R_1)$.

فإذا كانت القيمتان متساويتان تقريباً فإن التوزيع يكون متماثلاً، أما إذا كانت إحداهما أكبر من الأخرى فإن التوزيع يكون ملتوياً.

ويمكن الإفاده من المدى الربيعي في تعرف مدى تشتت توزيع الدرجات في المنطقة الوسطى المحصورة بين R_1, R_3 وبخاصة إذا كان التوزيع متماثلاً.

طرق حساب المئينيات:

نفترض أن لدينا التوزيع التكراري التالي للدرجات ٤٠ فرداً في أحد الاختبارات النفسية (جدول ٣ - ٦):

الحدود الحقيقية للفئات	التكرار المجتمع المئوي	التكرار المجتمع	التكرار	نوات الدرجات
٦٩,٥ - ٦٤,٥	٧,٥	٣	٣	٦٩ - ٦٥
٧٤,٥ - ٧٩,٥	٢٠,٠	٨	٥	٧٤ - ٧٠
٧٩,٥ - ٨٤,٥	٤٠,٠	١٦	٨	٧٩ - ٧٥
٨٤,٥ - ٨٩,٥	٦٧,٥	٢٧	١١	٨٤ - ٨٠
٨٩,٥ - ٩٤,٥	٨٥,٠	٣٤	٧	٨٩ - ٨٥
٩٤,٥ - ٩٩,٥	٩٥,٠	٣٨	٤	٩٤ - ٩٠
٩٩,٥ - ٩٤,٥	١٠٠,٠	٤٠	٢	٩٩ - ٩٥
				$\sum n = 40$

جدول (٣ - ٦) توزيع تكراري للدرجات اختبار، وكذلك التكرار المجتمع للدرجات

ولحساب قيمة المئيني ٧٠ مثلاً من جدول (٣ - ٦) دون اللجوء إلى استخدام صيغ رياضية معقدة، يمكن اتباع الخطوات التالية :

(١) نوجد عدد الأفراد الذين تقل درجاتهم عن المئيني ٥٠، وذلك بأن نضرب رتبة المئيني \times العدد الكلى للأفراد أي $٥٠ \times ٤٠ = ٢٠$.

(٢) تكون العمود الثالث الذي يشتمل على التكرار المتجمع وذلك بالسده بالتكرار المقابل للفئة الأولى ونقوم بعملية جمع متتال للتكرارات التي في الفئات التي تليها.

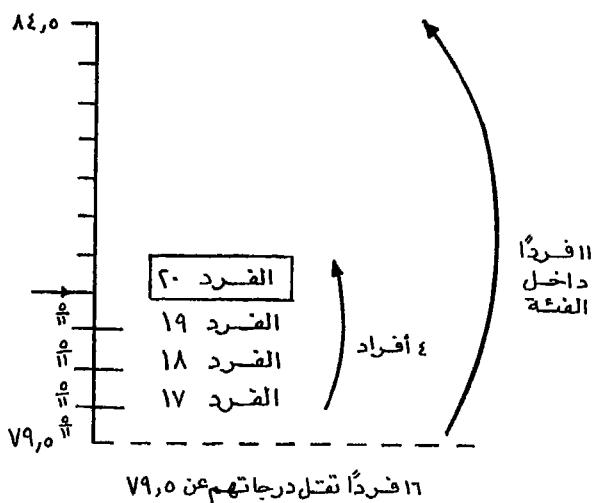
فمثلاً التكرار المتجمع الذي يناظر الحد الأعلى الحقيقي للفئة $٦٩,٥ - ٧٤,٥$ نحصل عليه بجمع تكرار هذه الفئة على تكرار الفئة السابقة عليها أي $٣ + ٥ = ٨$.

والتكرار المتجمع الذي يناظر الحد الأعلى الحقيقي للفئة $٧٤,٥ - ٧٩,٥$ نحصل عليه بجمع تكرار هذه الفئة على تكرار الفئتين السابقتين أي $٣ + ٥ + ٨ = ١٦$. وهكذا في بقية الفئات.

وينبغي التأكد من أن قيمة التكرار المتجمع للفئة الأخيرة $٩٤,٥ - ٩٩,٥$ تساوى العدد الكلى للتكرارات. كما ينبغي ملاحظة أننا استخدمنا الحدود الحقيقية للفئات وذلك بإضافة ٥ إلى الحدود الدنيا للفئات المدونة في العمود الأول، وطرح ٥ من الحدود العليا لهذه الفئات لكي يصبح توزيع الدرجات متصلأً.

(٣) نبحث في عمود التكرار المتجمع (العمود الثالث) عن الحدود الحقيقية للفئة التي يقع فيها التكرار المتجمع ٢٠ المذكور في الخطوة الأولى، فنجد أن هذه الفئة هي : $٧٩,٥ - ٨٤,٥$ ، ونلاحظ أن تكرار هذه الفئة المدون في العمود الثاني ١١. أي أن هناك ١١ فرداً يتبعون إلى هذه الفئة، فإذا افترضنا أن سعة الفئة التي تساوى ٥ درجات مقسمة إلى فترات متساوية على هؤلاء الأفراد، فإن كلاً منهم سيشغل مسافة $\frac{٥}{١١}$ من هذه الفئة.

(٤) نظراً لأننا نبحث عن التكرار المتجمع ٢٠، فإننا عند التكرار المتجمع ١٦ نحتاج إلى ٤ مسافات داخل الفئة $٨٤,٥ - ٨٩,٥$ للوصول إلى هذا التكرار المتجمع المطلوب، أي ٢٠ مطروحاً منها ١٦، ويمكن توضيح ذلك كما يلى :



(٥) نحسب الدرجة المناظرة للمسافات الأربع فنجد لها $\frac{٤}{١١} = ٥ \times ١,٨ = ٢,٠$

(٦) لإيجاد الدرجة المناظرة للمئيني ٥٠ نجمع الحد الأدنى الحقيقي لهذه الفتة أي ٧٩,٥ على الدرجة المناظرة للمسافات الأربع، أي $٧٩,٥ + ٨١,٣ = ٨١,٨$.

$$\text{أي أن الدرجة المناظرة للمئيني } ٥٠ = ٧٩,٥ + \frac{١٦ - ٢,٠}{١١} = ٧٩,٥ + \frac{١٤}{١١} = ٨١,٣$$

$$٥ \times \frac{٤}{١١} + ٧٩,٥ = \\ ٨١,٣ = ١,٨ + ٧٩,٥ \text{ تقريرياً}$$

ومن ذلك يمكن أن نستنتج الصيغة العامة لإيجاد الدرجات المناظرة لرتب مئينية معينة حيث تكون كالتالي :

$$\text{الدرجة المناظرة لرتبة مئينية معينة} =$$

$$\frac{\text{تكرار المجتمع للدرجة المطلوبة} - \text{التكرار المجتمع ت}}{\text{تكرار الفتة التي تحتوى التكرار المجتمع للدرجة}} \times \text{طول الفتة}$$

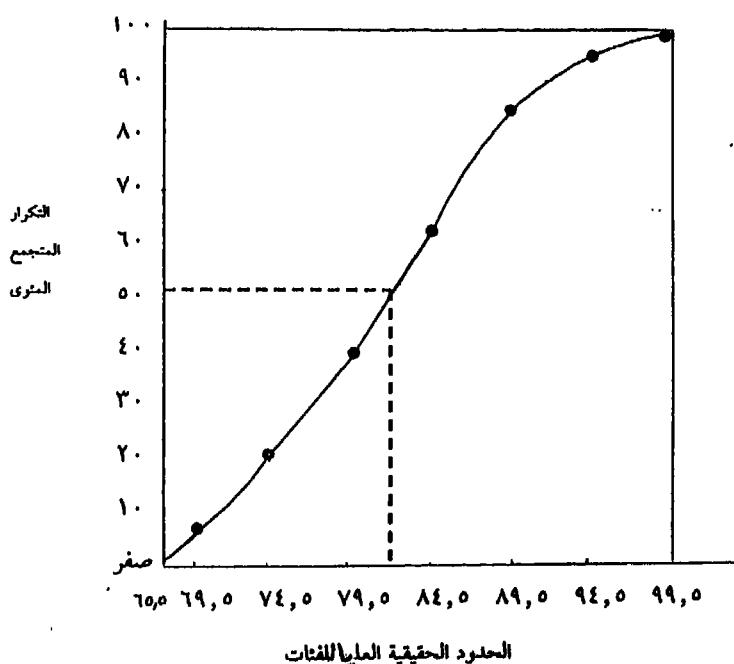
حيث (س) ترمز إلى الحد الأدنى الحقيقي للفترة التي تحتوى على التكرار المجتمع المناظر للرتبة المئينية المطلوبة.



، (ت_م) ترمز إلى التكرار المتجمع للفئة التي تحتوى على التكرار المتجمع المناظر للرتبة المئينة المطلوبة.

إيجاد الدرجات المقابلة للمئينيات بيانياً :

يمكن إيجاد الدرجات المقابلة للمئينيات باستخدام المنهجى التكرارى المتجمع المئوى حيث يتطلب ذلك تحويل التكرارات المتجمعة المقابلة للحدود العليا لفئات توزيع الدرجات إلى تكرارات متجمعة مئوية، وذلك بقسمة كل من هذه التكرارات المتجمعة على التكرار الكلى ونضرب الناتج $\times 100$ كما هو مبين بالعمود الرابع فى جدول (٣ - ٦) ، ثم نمثل هذه الحدود العليا على المحور الأفقي والتكرار المتجمع المئوى على المحور الرأسى، وترسم المنهجى التكرارى المتجمع المئوى كما هو موضح بشكل (٣ - ١٤) التالى:



شكل (٣ - ١٤) يوضح المنهجى التكرارى المتجمع المئوى لدرجات

الاختبار المذكورة في جدول (٣ - ٦)

فلا يجاد الدرجة المقابلة للمئيني ٥٠ مثلاً، نحدد النقطة ٥٠ على المحور الرأسى، ونرسم خطأً أفقياً من هذه النقطة حتى يلاقي المستحقى التكرارى المجتمع المئوى فى نقطة أخرى نسقط منها عموداً على المحور الأفقي، فتكون نقطة الالقاء مناظرة للدرجة المطلوبة وهى ٨١,٣ كما يتضح من شكل (١٤ - ٣)، ويلاحظ أنها متساوية للدرجة التى حصلنا عليها بالطريقة الأخرى.

ونظراً لأن المئيني ٥٠ يناظر الوسيط، فإنه يمكن حساب الدرجات المقابلة لكل من الوسيط والإرباعيات بنفس هاتين الطريقتين، حيث إن الإرباعى الأول يناظر المئيني ٢٥، والإرباعى الثالث يناظر المئيني ٧٥ كما سبق أن أشرنا.

الرتبة المئينية :

تستخدم الرتب المئينية Percentile Ranks بكثرة في تفسير درجات الاختبارات المرجعية الجماعة أو المعيار Referenced Tests - Norm وذلك لسهولة توضيح المقصود منها للطلاب والمفحوصين وأولياء الأمور حيث إنها لا تتطلب منهم إلا معرفة مفهوم النسبة المئوية.

ففي هذا النوع من الاختبارات نود معرفة النسبة المئوية لعدد الأفراد الذين يحصلون على درجات تقل قيمتها عن قيمة درجة معينة في أحد الاختبارات. فإذا كانت الرتبة المئينية المقابلة للدرجة الخام ٥٧ هي ٦٠، فإن هذا يعني أن ٦٠٪ من أفراد مجموعة معينة تقل درجاتهم عن الدرجة ٥٧، بينما تزيد درجة ٤٠٪ منهم عن هذه الدرجة ، فالرتبة المئينية إذن لا يمكن تفسيرها إلا في إطار جماعة مرجعية معينة Reference Group يطبق عليهم الاختبار أو المقياس التربوى أو النفسي. فإذا حصل طالب على درجة خام تقابل رتبة مئينية ٨٥ في اختبار تحصيلي، فربما يبدو لأول وهلة أن أداء الطالب مرتفعاً لأنه يتفوق على ٨٥٪ من أقرانه، ولكن إذا كانت الجماعة المرجعية التي نقارنه بها من ضعاف التحصيل بعامة، فإن أداءه في هذه الحالة يعد منخفضاً. وعلى العكس من ذلك، إذا كانت جماعته المرجعية من المهووبين، وحصل على درجة خام تقابل رتبة مئينية ٣٥ مثلاً، فإن أداءه ربما يبدو جيداً، وهكذا. لذلك ينبغي الحيوطة عند المقارنة بين الرتب المئينية لجماعات مرجعية مختلفة لكي لا يساء تفسير هذه الرتب.

طرق حساب الرتب المئينية :

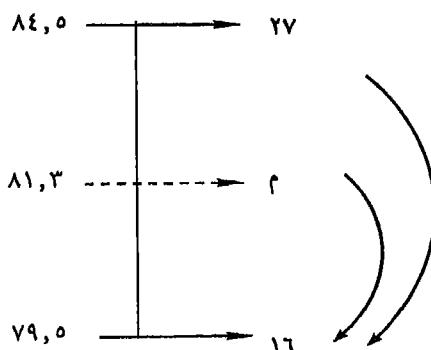
نظرًا لأن الرتب المئينية تشير إلى نسبة من الأفراد الذين يحصلون على درجات أقل من درجة معينة في اختبار ما، فإن هذه الدرجة تمثل نقطة تسمى المئيني، أي أن الرتبة المئينية عكس المئيني، ولذلك فإن طرق حسابها لا تختلف كثيراً عن طرق حساب المئينيات التي ذكرناها.

ولتوضيح هذه الطرق نعود مرة أخرى إلى توزيع الدرجات الموضح في جدول (٣ - ٦).

ونفترض أننا نود إيجاد الرتبة المئينية المقابلة للدرجة الخام $81,3$ ، فما علينا إلا أن نحدد الفتة التي تقع فيها هذه الدرجة حيث نجد أنها الفتة $79,5 - 84,5$ ، ثم نبحث في العمود الرابع عن التكرار المتجمع المثوى لكل من الدرجتين $79,5$ ، $84,5$ فنجد أنه 40 ، $67,5$ على الترتيب.

وهذا يعني أن الرتبة المئينية المقابلة للدرجة $81,3$ تقع بين 40 ، $67,5$ كما هو موضح فيما يلى :

الدرجات	الرتب المئينية
---------	----------------

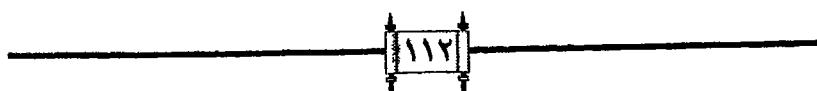


وستستطيع الآن تحديد الرتبة المئينية (m) المقابلة للدرجة $81,3$ باستخدام النسبة والتناسب كالتالي :

$$\frac{79,5 - 81,3}{79,5 - 84,5} = \frac{40 - m}{40 - 67,5}$$

$$\frac{1,8}{5} = \frac{40 - m}{27,5}$$

أي أن :



$$٥ = (٤٠ - ٢٧,٥) \times ١,٨$$

$$٥ = ٢٠٠ - ٤٩,٥$$

$$٥ = ٢٤٩,٥$$

$$\text{أى أن : } م = \frac{٢٤٩,٥}{٥} = ٤٩,٩ \text{ أى تساوى ٥٠ تقريباً.}$$

وهي نفس الرتبة المذكورة في المثال السابق.

ويمكن الحصول على هذه الرتبة بيانياً باستخدام شكل (٣ - ١٤) بطريقة عكسية حيث نبدأ بالمحور الأفقي (المحور الذي يمثل الحدود الحقيقة العليا للفئات الدرجات) ونحدد النقطة ٨١,٣ على هذا المحور ، ونرسم منها خطراً رأسياً يلاقي المنحنى التكراري المجتمع المثوى ، ثم نرسم عموداً على المحور الرأسى (المحور الذي يمثل التكرار المجتمع المثوى أو الرتب المئينية) من نقطة التلاقي ، فنكون نقطة تقاطع العمود الأفقي مع المحور الرأسى هي الرتبة المئينية المطلوبة ، ونلاحظ أنها تساوى ٥٠.

بعض أوجه قصور المئينيات والرتب المئينية :

على الرغم من أهمية المئينيات والرتب المئينية وسهولة استخدامها في تفسير درجات الاختبارات والمقياسين وفي تحديد المركز النسبي للفرد في إطار جماعة مرجعية معينة طبق عليها الاختبار أو المقياس ، إلا أنه يشوبها بعض أوجه القصور في هذا الشأن ونوجزها فيما يلى :

(١) الرتب المئينية تتوزع توزيعاً مستطيلياً في حين أن توزيعات درجات الاختبارات وبخاصة الاختبارات مرجعية الجماعة تتحدد عادة شكل المنحنى الاعتدالي . ولذلك فإن القيم العددية المتساوية للفروق بين الرتب المئينية لا تعنى فروقاً متساوية في الدرجات الخام ، فالفرق بين الرتبة المئينية ٣٢ والرتبة المئينية ٣٨ يساوى ٤ نقاط ، بينما الفرق بين الدرجتين الخام المقابلتين لهما ربما تكون ١٠ درجات ، وربما تكون أقل بكثير . ف مجرد النظر إلى الرتب المئينية لا يكفي لمعرفة الفروق بين الدرجات الخام . كذلك فإن الفروق الضئيلة بين الدرجات الخام بالقرب من مركز التوزيع تناقض رتبة مئينية كبيرة ، بينما الفروق الكبيرة بين الدرجات الخام عند طرف التوزيع تناقض فروقاً صغيرة في هذه الرتب .

(٢) تتوسع الرتب المئينية على ميزان رتبى ، ولذلك لا نستطيع إجراء العمليات الحسابية الأساسية عليها ، غير أن هذا القصور لا يؤثر في تفسير درجات الاختبارات .

وقد سبق أن أوضحنا أن الدرجات المعيارية يمكن ضمها بحيث تشكل درجة معيارية مركبة حيث إنها توزع على ميزان فترى.

ولكن إذا كان توزيع الدرجات الخام اعتدالياً فإنه يمكن الإفاده من خصائص المنهنى الاعتدالى التي سبق توضيحها فى إيجاد المئينيات المناظرة للدرجات المعيارية، وبذلك يجعل تفسير درجات الاختبار أكثر سهولة؛ لأنه كما ذكرنا يصعب على الطالب أو المفحوص العادى تفسير الدرجات المعيارية.

استخدام المنهنى الاعتدالى فى إيجاد المئينيات :

لتوضيح كيفية استخدام المنهنى الاعتدالى فى إيجاد المئينيات دعنا ننظر إلى المنهنى الاعتدالى الموضح بشكل (١٣ - ٢)، ونفترض أن توزيع الدرجات الخام فى أحد الاختبارات كان توزيعاً اعتدالياً متوسطه ١٠٠ وانحرافه المعيارى ٦، وأردنا تحديد الرتبة المئينية المقابلة للدرجة الخام ١٣٢ باستخدام خصائص المنهنى الاعتدالى .

فالخطوة الأولى : نحوال الدرجة الخام إلى درجة معيارية باستخدام الصيغة $(132 - 100) / 6 = 2 + .$

والخطوة الثانية : نحدد المساحة تحت المنهنى الاعتدالى إلى يسار هذه الدرجة ، فنجد أنها تساوى ٩٨ تقريباً، وهذا يعني أن الدرجة ١٣٢ تقابل الرتبة المئينية ٩٨ ، أي أن هذه الدرجة تفوق ٩٨٪ من درجات الجماعة المرجعية .

وعلى العكس من ذلك فإننا نستطيع إيجاد الدرجة الخام المقابلة لرتبة مئينية معينة بأن نتبع الخطوة الثانية أولاً ثم الخطوة الأولى . وينبغي التنويه إلى أنه إذا كانت الدرجة المعيارية كسرية ، فإنه توجد جداول إحصائية تحدد المساحات تحت المنهنى الاعتدالى في هذه الحالة ، ويمكن الرجوع إلى هذه الجداول في أحد المراجع الأولية في الإحصاء .

كما ينبغي التنويه إلى أنه إذا كان توزيع الدرجات الخام في الاختبار يتبع عن التوزيع الاعتدالى ، فإنه يمكن تحويل هذا التوزيع إلى توزيع اعتدالى باستخدام طرق معينة لا مجال لذكرها هنا ، وقد قام المؤلف بتوضيح هاتين الحالتين في مرجع آخر .

إذا أجري هذا التحويل على توزيع الدرجات الخام يصبح شكل توزيع الدرجات اعتدالياً بغض النظر عن شكل التوزيع الأصلي للدرجات الخام . فإذا حوت هذه الدرجات الجديدة إلى درجات معيارية ، فإن الدرجات المعيارية الناتجة تسمى الدرجات

المعيارية الاعتدالية Z ، وكذلك إذا حُولت الدرجات المعيارية الاعتدالية إلى درجات تائية ، فإن الدرجات التائية الناتجة تسمى الدرجات التائية الاعتدالية $Normalized T$.

فإذا علمنا قيمة كل من الدرجة المعيارية الاعتدالية والدرجة التائية الاعتدالية، فإن العلاقة بين هاتين القيمتين وقيمتى المئنيات المقابلة لهما يحددها المنحنى الاعتدالى المعياري الموضع بشكل (٣ - ١٣) بغض النظر عن شكل توزيع الدرجات الخام.

لذلك إذا علمنا أن الدرجة المعيارية الاعتدالية لفرد في اختبار ما (١) مثلاً، فإن هذا يعني أن هذه الدرجة تقابل المئني ١٦ تقريرياً.

وبالطبع لا يجوز مثل هذا التفسير إذا كانت الدرجة المعيارية الخطية المعتادة (١) ما لم يكن توزيع الدرجات الخام اعتدالياً.

مقاييس العلاقة بين مجموعتين من الدرجات :

لاحظنا مما سبق أن مقاييس التزعة المركزية أو الموضع، ومقاييس التشتت، ومقاييس المركز النسبي، مثل الدرجات المعيارية وغيرها من الدرجات المعيارية المعدلة، والمئنيات والرتب المئنية، تستخدم في وصف توزيعات درجات الاختبارات والمقاييس، وتحديد المركز النسبي للفرد في إطار جماعته المرجعية. وجميع هذه المقاييس تصف متغيراً واحداً. ولكننا في القياس التربوى والنفسى نهتم في كثير من الأحيان بوصف العلاقة بين متغيرين أو سنتين من سمات الأفراد. فإذا طبق اختباران يقيسان سنتين مختلفتين فإننا نحصل على مجموعتين من الدرجات. وهنا ربما نهتم بمعرفة العلاقة بين المراكز النسبية للأفراد في إحدى السنتين ومراكزهم النسبية في السنة الأخرى.

فمثلاً ربما نود معرفة العلاقة بين اختبارين أحدهما يقيس الاستعداد الدراسي والأخر يقيس التحصيل، أو بين مقاييس من مقاييس الذكاء أو مقاييس الشخصية، وهكذا، ففى هذه الحالات ربما نجد أن الأفراد الذين حصلوا على درجات مرتفعة فى أحد الاختبارين أو المقاييس يميلون إلى الحصول على درجات مرتفعة فى الاختبار أو المقياس الآخر، فهنا تكون العلاقة بين مجموعتي الدرجات موجبة. وعلى العكس من ذلك، إذا كانت درجاتهم فى أحد الاختبارين منخفضة مثلاً بينما درجاتهم فى الاختبار الآخر تميل إلى الارتفاع فإن العلاقة بين مجموعتي الدرجات تكون سالبة.

ومقاييس العلاقة تحدد قيمة عددية تمثل درجة الارتباط بين مجموعتي الدرجات، أى إلى أى مدى يحافظ الأفراد على مراكزهم النسبية في كل من السنتين موضع القياس.

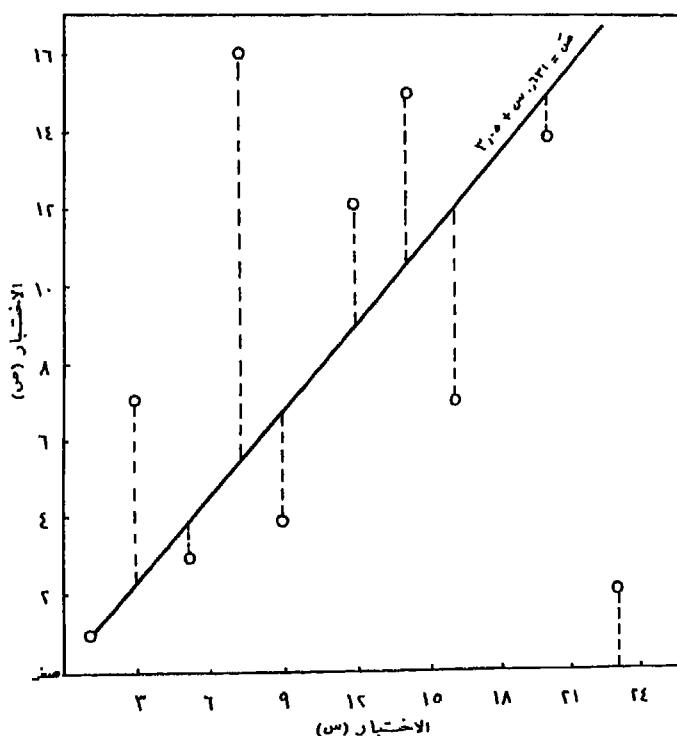
ويمكن أن تتضح مثل هذه العلاقات إذا قمنا بتمثيل قيم المتغيرين أي مجموعتي الدرجات تمثيلاً بيانيًّا في شكل واحد يسمى الشكل الانتشاري Scatter Diagram للمتغيرين.

إذا افترضنا أن لدينا مجموعتي الدرجات التالية لعينة من الطلاب في اختبارين أحدهما يقيس التحصيل في الرياضيات والأخر يقيس التحصيل في الفيزياء (جدول ٣ - ٧) :

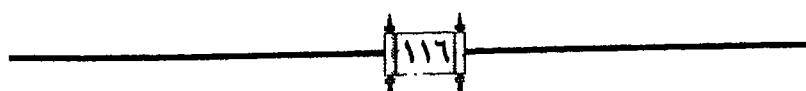
العينة											
الرياضيات						الفيزياء					
الاختبار	الرقم	النوع	القيمة	الاختبار	النوع	القيمة	الاختبار	النوع	القيمة	الاختبار	النوع
١	٢	٥	٧	٩	١٢	١٤	١٦	٢٠	٢٣	٢٨	٣٣
١	٧	٣	١٦	٤	١٣	١٥	٧	١٤	٢	٢١	٣٠

جدول (٣ - ٧) درجات عينة من الطلاب في اختبارين تحصيليين

فإنه يمكن تمثيل قيم (س) على المحور الأفقي ، وقيم (ص) على المحور الرأسى ، ونحدد لكل زوج من القيم المتناظرة لكل من (س) ، (ص) نقطة على مستوى المحورين كما بشكل (٣ - ١٥) التالي :



شكل (٣ - ١٥) شكل انتشاري لأزواج الدرجات المذكورة في جدول (٣ - ٧)

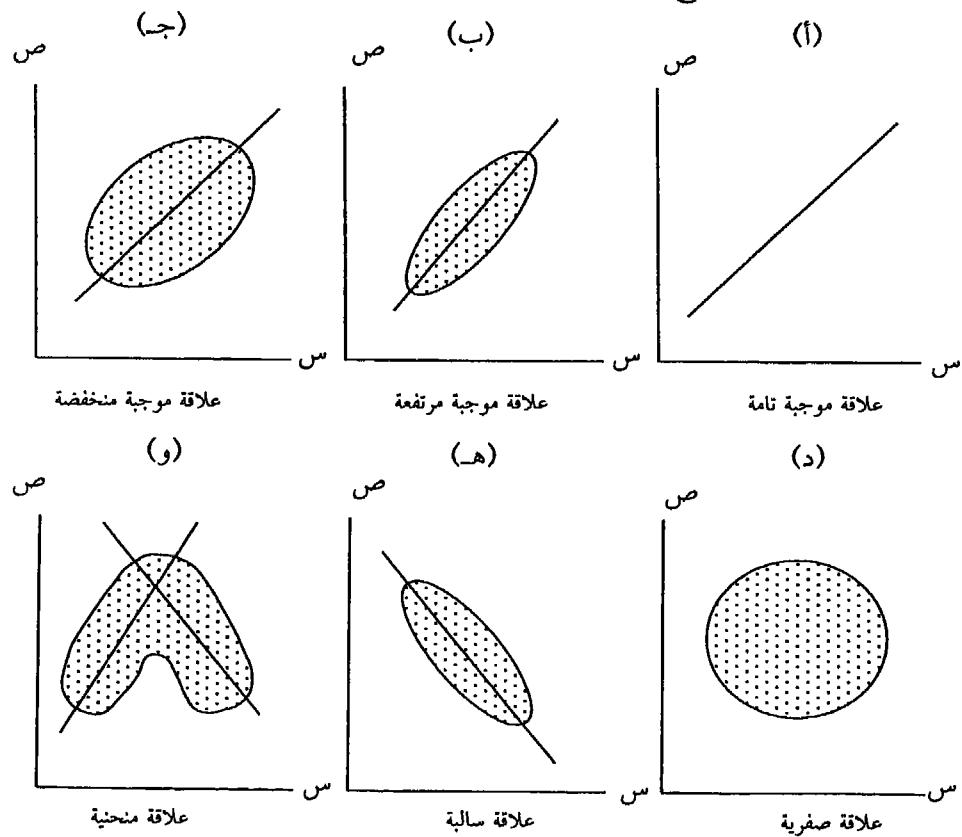


ويتضح من شكل (٣ - ١٥) أنه على الرغم من انتشار النقط التي تمثل أزواج درجات الاختبارين، إلا أنها تقع جميعها داخل شكل بيضاوي، وهناك نزعة عامة لأن تقع هذه النقط على خط مستقيم يمر بمعظمها. لذلك فإن العلاقة بين مجموعتي الدرجات يقال أنها علاقة خطية أو مستقيمة Linear Relationship . ونظرًا لأن الشكل البيضاوي يتجه إلى أعلى جهة المين، فإن العلاقة موجة.

لذلك فإن هذا الشكل الانتشاري يفيد في إلقاء الضوء على شكل العلاقة بين مجموعتي الدرجات لمعرفة ما إذا كانت العلاقة مستقيمة أم منحنية، وموجة أم سالبة.

فالعلاقة المنحنية لا تميل فيها النقط إلى التراكم حول خط مستقيم، بل تبدو هناك نزعة للانحراف، أي يكون هناك نزعة للاتجاه إلى أعلى في جزء معين ثم ينعكس الاتجاه إلى أسفل في جزء آخر من الشكل. أما العلاقة السالبة فتبعد عن اتجاه الشكل البيضاوي إلى أسفل جهة اليمين .

وفيما يلى توضيح أشكال هذه العلاقات (٣ - ١٦) :



٤-١٦) أشكال انتشارية لعلاقات مختلفة

تَدْبِير مُقَدَّارِ الْعَلَاقَة بَيْن مَجْمُوعَتَيْن مِن الدَّرَجَات :

لاحظنا أن الأشكال الانتشارية توضح شكل العلاقة ونوعها ولكن لا تحدد مقدار هذه العلاقة تحديداً كمياً. فلإيجاد مقدار واتجاه العلاقة بين مجموعتين من الدرجات في حالة ما إذا كانت العلاقة خطية أو مستقيمة كما في الأشكال أ ، ب ، ج ، د يمكن استخدام مقاييس إحصائية مناسبة. ولعل أهم هذه المقاييس معامل ارتباط بيرسون Pearson Product Moment Correlation . ولهذا المعامل صيغ رياضية متعددة بعضها يعتمد على الدرجات المعيارية أو الانحرافات عن المتوسط، والبعض الآخر يعتمد على الدرجات الخام مباشرة.

ونظراً لأن الصيغة الأخيرة يمكن تطبيقها بسهولة باستخدام آلة حاسبة صغيرة أو الحاسوب، فإننا سوف نقتصر على هذه الصيغة لأنها شائعة الاستخدام. ونقدم برنامجاً بسيطاً متعلقاً بها يمكن تفريذه بسهولة باستخدام الحاسوب.

معامل ارتباط بيرسون :

لإيجاد العلاقة بين مجموعتين من الدرجات باستخدام معامل ارتباط بيرسون ينبغي التتحقق من استقامة العلاقة من الشكل الانتشاري لمجموعة الدرجات، وكذلك التتحقق من أن السمة مقاسة على ميزان فترى على الأقل.

والصيغة العامة لمعامل ارتباط بيرسون بين متغيرين (س) ، (ص) كالتالي :

$$r = \frac{n \sum sc - \sum s \sum c}{\sqrt{n \sum s^2 - (\sum s)^2} [\sum n \sum c^2 - (\sum c)^2]}$$

(١ - ٣).

ويطلب تطبيق هذه الصيغة إيجاد قيم كل من :

$$(Mg_s \cdot Ch), (Mg_s \times Mg_ch), (Mg_s^2), (Mg_s)^2, (Mg_ch)^2$$

ويوضح جدول (٨ - ٣) كيفية تطبيق هذه الصيغة بمثال لمجموعتي الدرجات (س) ، (ص) المشار إليهما في جدول (٧ - ٣) والشكل الانتشاري (٣ - ١٥) :

س	ص	2	س	ص	س
٤٦٠	٤٠٠		٥٢٩	٢٠	٢٣
٢٨٠	١٩٧		٤٠٠	١٤	٢٠
١١٢	٤٩		٢٥٦	٧	١٦
٢١٠	٢٢٥		١٩٦	١٥	١٤
١٥٦	١٧٩		١٤٤	١٣	١٢
٣٦	١٦		٨١	٤	٩
١١٢	٢٥٦		٤٩	١٦	٧
١٥	٩		٢٥	٣	٥
٢١	٤٩		٩	٧	٣
١	١		١	١	١

جدول (٣ - ٨) يوضح خطوات إيجاد قيمة معامل ارتباط بيرسون

$$\text{مجـ س} = 11.0 \quad , \quad \text{مجـ ص} = 100.0 \quad , \quad \text{مجـ س}^2 = 169.0$$

$$\text{مجـ ص}^2 = 137.0 \quad , \quad \text{مجـ س ص} = 140.3$$

$$r = \frac{100 \times 11.0 - 140.3 \times 1.0}{\sqrt{[2(100) - 137] [2(11.0) - 169]}}$$

$$r = \frac{30.3}{4214.3} = \frac{30.3}{3700 \times 4800} =$$

وينبغي ملاحظة أن معامل ارتباط بيرسون تتراوح قيمه بين (-1) ، (+1) ، فالقيمة (-1) تعنى ارتباط تمام سالب ، والقيمة (+1) تعنى ارتباط تمام موجب ، والقيمة (صفر) تعنى عدم وجود ارتباط بين المتغيرين . وعادة تكون قيم معامل ارتباط

بيرسون بين مجموعتين من درجات الاختبارات والمقاييس التربوية والنفسية التي تستخدم في دراسة الفروق الفردية أقل من الواحد الصحيح.

برنامج الحاسوب الشخصي أو حاسوب الجيب لإيجاد معامل ارتباط بيرسون :

يمكن استخدام برنامج الحاسوب التالي في إجراء العمليات الحسابية التي تتطلبها الصيغة (٣ - ١٠) . ويتميز هذا البرنامج بالبساطة وسرعة الإجراء ، علما بأننا سوف نرمز إلى المتغير (س) بالرمز (x) ، والمتغير (ص) بالرمز (y) ، ومجموع قيم س بالرمز S_x ، ومجموع قيم ص بالرمز S_y ، ومجموع مربعات قيمة س بالرمز S_{xx} ، ومجموع مربعات قيمة ص بالرمز S_{yy} ، ومجموع حواصل ضرب قيمة س في قيمة ص . S_{xy}

```

5   N = 0 : Sx = 0 : Sy = 0 : XY=0.
10  Read X , Y
20  if X < - 999 THEN 110
30  N = N + 1
40  Sx = Sx + X
50  Sy = Sy + Y
60  XX = XX + X * X
70  YY = YY + Y * Y
80  XY = XY + X * Y
100 GO TO 10
110 Rem Calc B And A
120 B = (N * XY - Sx * Sy) / N * XX - Sx * Sx)
130 A = (Sy - B * Sx) / N
140 PRINT A , B
150 R = (N * XY - Sx * Sy) / SQR (N * XX - Sx * Sx) * (N
* YY - Sy * Sy)
160 PRINT R

```



كما أن هذا البرنامج يوجد قيمة كل من الثابتين A , B اللذين يمكن استخدامهما في إيجاد معادلة الانحدار التي سنوضحها بعد قليل.

تفسير قيم معامل ارتباط بيرسون:

إن قيم معامل ارتباط بيرسون لا يجب تفسيرها على أنها نسب مئوية، فلا ينبغي أن تكتب القيمة ٧١٩، ٠، لمعامل الارتباط ٧٢٪، وإنما يمكن تقريرها إلى رقمين عشررين على الأقل، أى يمكن كتابتها ٧٢، ٠، ولا يجوز كتابتها ٧، ٠، تحريا للدقة. كما لا ينبغي إجراء عمليات جمع أو طرح على قيم معاملات الارتباط نظراً لأنها معاملات من المستوى الرتبى، فالفرق بين معامل ارتباط ٣٠، ٤٠، ٠ لا يمثل نفس الفرق بين معامل ارتباط ٨٠، ٩٠، ٠ مثلاً.

ولذلك لا ينبغي تفسير قيم معاملات الارتباط على أنها تشير إلى علاقات علية أو سببية، وإنما تعنى فقط وجود اتجاه ومقدار معين لهذه العلاقات، فالعلاقة بين الذكاء والتحصيل لا تعنى بالضرورة أن الذكاء يسبب التحصيل ، حيث إن هناك عوامل أخرى كثيرة تؤثر في التحصيل، فمعاملات الارتباط لا تتضمن بالضرورة علاقات علية. كما يجب عند تفسير قيم معاملات الارتباط مراعاة عدد أفراد العينة المستخدمة في حساب هذه القيم حيث إن هذه القيم، تتأثر تأثيراً ملحوظاً باختلاف عدد أفراد العينة. فمعامل الارتباط يعتمد على التوزعة المركزية والتشتت لكل من المتغيرين، وكلما زاد التشتت زادت قيمة معامل الارتباط والعكس صحيح. وما لا شك فيه أن عدد أفراد العينة يؤثر في تشتت الدرجات أى تباين هذه الدرجات، مما يؤثر بدوره في قيم معامل ارتباط بيرسون .

ولتفسير قيم معامل الارتباط ينبغي تربيعها للحصول على معامل آخر يسمى معامل التحديد Coefficient of Determination ، وهذا المعامل يحدد نسبة التباين المشترك بين المتغيرين، أى النسبة المئوية للتباين في درجات المتغير الأول التي تعزى إلى تباين درجات المتغير الثاني. فمعامل الارتباط الذى قيمته ٧١٩، ٠، والتي حصلنا عليها في المثال السابق لا نستطيع القول بأنها قيمة مرتفعة ؛ لأن تربيع هذه القيمة يجعلها مساوية ٥٢٪، تقريراً، مما يدل على ٥٢٪ من تباين درجات اختبار الفيزياء تعزى إلى تباين درجات اختبار الرياضيات، أما بقية التباين وهو ٤٨٪ فيعزى إلى عوامل أخرى لا نستطيع تحديدها بدقة، إذ ربما تكون الذكاء أو الدافعية أو عوامل مدرسية معينة، ولذلك فهي قيمة متوسطة .

وعلى الرغم من أن قيم معامل الارتباط لا يجوز المقارنة بينها ، إلا أن القيم النسبية لمعامل التحديد المنشورة لها يمكن المقارنة بينها مقارنة مباشرة ، فمعامل الارتباط $.719$ ، لا نستطيع اعتباره ضعف معامل الارتباط $.360$ ، حيث إن معامل التحديد المنشورة لكل منها $.52$ ، $.13$ ، على الترتيب ، ونظرا لأن القيمة $.52$ تساوى أربعة أضعاف القيمة $.13$ ، فإن قوة معامل الارتباط $.719$ ، تساوى ضعفي قوة معامل الارتباط $.360$ ، وليس كما يبدو أنها الضعف فقط.

والخلاصة أنه ينبغي مراعاة شروط استخدام معامل ارتباط بيرسون من حيث ميزان القياس الفتري لدرجات الاختبارات ، واستقامة العلاقة بين مجموعتي الدرجات . فإذا كان ميزان القياس اسمياً أو رتيباً أو أن العلاقة منحنية فإنه لا يجوز استخدام هذا المعامل ، حيث إن هناك مقاييساً إحصائية أخرى تناسب هذه الحالات يمكن الرجوع إليها في المراجع الإحصائية .

الاختبارات الأخرى عن تفسير قيم معامل الارتباط :

إن إساءة استخدام معامل الارتباط في مجالات القياس التربوي والنفسى سواء في إعداد الاختبارات أو في تفسير درجاتها يؤدي إلى نتائج تفتقر إلى الدقة ، ولذلك ينبغي إلى جانب الملاحظات السابقة مراعاة بعض الاعتبارات الأخرى نوجزها فيما يلى :

(1) **القيم المقبولة لمعامل الارتباط**، فكثيراً ما يسىء كثير من المستغلين بالقياس تفسير قيم معامل الارتباط كما سبق أن ذكرنا . كما أن البعض يتساءل عن القيم المقبولة لمعامل الارتباط ، فهل $.60$ ، $.70$ ، أو $.80$ ، قيمتان مقبولتان أم لا بد أن تكون قيمة معامل الارتباط $.85$ ، أو $.90$ ، مثلاً؟ والحقيقة أن اعتبار قيمة معينة لمعامل الارتباط مرتفعة يحتاج إلىأخذ كثير من العوامل بعين الاعتبار ، إذ ليس هناك نقطة محددة تتصل بين القيم المرتفعة والقيم المنخفضة لهذا المعامل ، وإنما يعتمد ذلك على طبيعة الاختبار أو المقياس المستخدم ، والمتغير أو السمة التي يقيسها ، والعوامل المؤثرة في هذه السمة .

ولذلك فإن القيمة المرتفعة أو المنخفضة لمعامل الارتباط لا تعتمد اعتماداً كلياً على القيمة العددية لهذا المعامل ، وإنما تعتمد أيضاً على طبيعة الاختبار وفيه تستخدم نتائجه ، فهذه القيمة العددية تشير إلى درجة العلاقة وليس إلى كيفية تفسير هذه العلاقة . وربما يمكن ملاحظة هذه المشكلة بوضوح في الاختبارات التي تستخدم في مجال الانتقاء ، حيث ترتبط درجات الاختبارات التنبؤية بمحكمات النجاح في دراسة أو مهنة

معينة، وتبلغ عادة قيمة معامل الارتباط .٦٠ ، أو أكثر. ومع هذا فإن انخفاض قيمة هذا المعامل عن .٦٠ ، في بعض هذه الاختبارات لا يقلل من شأن الاختبار حيث إنه يظل مفيداً في الانتقاء عملاً لو تم الانتقاء عشوائياً .

(٢) **عوامل الخطأ التي تؤثر في قيمة معامل الارتباط :** فقييم معامل الارتباط التي يدونها الباحثون للاختبارات التي يقومون بإعدادها، وكذلك المدونة في أدلة الاختبارات يتم تفسيرها على أنها توضح مقدار العلاقة بين السمات التي تقيسها هذه الاختبارات .

غير أن هناك عوامل خطأ كثيرة ربما تكون قد أسهمت في زيادة هذه القيم أو انخفاضها. ومن بين هذه العوامل المؤثرة : الأخطاء العشوائية، وضيق مدى درجات الاختبار، وشكل توزيع الدرجات، وعدم استقامة العلاقة بين مجموعتي الدرجات. فهذه العوامل تؤثر بلا شك تأثيراً مبايناً في قيم معامل الارتباط ، فالأنخطاء العشوائية في أي من الاختبارين أو اختيار مجموعة متجانسة من الأفراد في متغيرات معينة ، مثل العمر أو الذكاء أو التحصيل المرتفع يؤدى إلى انخفاض قيمة معامل الارتباط .

كما أن انحصار العلاقة أي بعدها عن الاستقامة تؤدي إلى قيم مضللة أو زائفة Spurious Correlation إذا استخدم معامل ارتباط بيرسون .

(٣) **شكل توزيع كل من المتغيرين :** لكي تصل قيمة معامل ارتباط بيرسون إلى أقصى قيمتها وهذا (١+) ، يجب أن يكون توزيعاً المتغيرين لهما نفس الشكل تقريباً. فمثلاً إذا كان أحد المتغيرين متصلةً والآخر ثنائياً (أي يتخذ قيمتين فقط مثل واحد صحيح أو صفر) ، أو أن أحد التوزيعين متلوياً التواه موجباً والآخر متلوياً التواه سالباً ، فإن قيمة معامل الارتباط في كل من هاتين الحالتين تكون أقل من الواحد الصحيح .

استخدام معامل الارتباط في التنبؤ :

لا يقتصر استخدام معامل الارتباط على وصف درجة العلاقة بين متغيرين فحسب بل يمكن استخدامه أيضاً في التنبؤ بقيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر. فمثلاً إذا كان هناك ارتباط موجب بين اختبار يقيس الاستعداد الدراسي لدى الطلاب الذين يودون الالتحاق بالجامعة ومتوسط تقديراتهم في نهاية السنة الجامعية الأولى ، فإن معامل الارتباط يوضح درجة العلاقة بين درجات الاختبار ومتوسط التقديرات ، وكذلك يمكن استخدامه في التنبؤ بمتوسط تقديرات الطلاب إذا علمنا درجاتهم في اختبار الاستعداد الدراسي . وعلى الرغم من أن كثيراً من مقاييس الجوانب

المعرفية ترتبط فيما بينها بعلاقة خطية أى مستقيمة، إلا أنه توجد بعض المقاييس وبخاصة المتعلقة بالمتغيرات الفسيولوجية والدافعية ترتبط فيما بينها بعلاقة منحنية. ولذلك فإننا سوف نقتصر في دراسة التنبؤ على العلاقات المستقيمة.

إذا عدنا إلى شكل (٣ - ١٥) نجد أن معظم النقط تراكم حول خط مستقيم يمثل النزعة العامة للدرجات. فإذا كان هذا الخط يجعل مجموع مربعات المسافات المبينة بالشكل الانتشاري الموازية للمحور الرأسى من النقط إلى الخط المستقيم أقل ما يمكن. فإن هذا الخط عندئذ يسمى خط الانحدار Regression Line ، وهو الخط الذى يستخدم فى التنبؤ بقيم المتغير (ص) من قيم المتغير (س) المعلومة بأقل قدر ممكن من الخطأ الذى يسمى خطأ التنبؤ Prediction Error . وهذا الخطأ عبارة عن الفرق بين الدرجة الملاحظة (ص) والدرجة المتتبأ بها ص ، ونظراً لأن مقدار هذا الخطأ يكون كبيراً نسبياً في القياس التربوى والنفسي، فإنه ينبغي العناية بانتقاء أدوات القياس التى تتبع بالظاهره التى نهم بها بأقل قدر ممكن من أخطاء التنبؤ.

وربما يتسائل البعض عما إذا كان من الضروري أو من المطلوب أن تتبأ بقيمة متغير (ص) من متغير معلوم (س) عندما تكون قيم كل من المتغيرين معلومة. وللإجابة عن ذلك ينبغي أن نعلم أن الاختبارات التى تستخدمن فى التنبؤ تطبق فى بادئ الأمر على عينة عشوائية كافية وممثلة للمجتمع المستهدف وإيجاد الارتباط بين درجاتها وبين قيم المتغير المراد التنبؤ به، والتوصيل إلى خط الانحدار الذى يستخدم مستقبلاً على عينات مماثلة مستمددة من المجتمع نفسه. فاختبار الاستعداد الدراسي مثلاً يطبق على عينة عشوائية مماثلة لمجتمع الطلاب الذين يودون الالتحاق بالكليات، وفي نهاية السنة الجامعية الأولى نوجد متوسط تقديرات هؤلاء الطلاب. ثم نوجد قيمة معامل الارتباط بين درجات الاختبار ومتوسط التقديرات، ونستخدم هذه القيمة فى إيجاد معادلة خط الانحدار الذى تتبأ بواسطتها بمتوسط تقديرات عينات أخرى من الطلاب لم تتحقق بعد بالكليات بمعلومية اختبار الاستعداد الدراسي الذى يطبق عليهم قبل الالتحاق.

المعادلة خط الانحدار المتغير (ص) على المتغير (س) :

يمكن إيجاد معادلة خط الانحدار بطرق متعددة لعل أبسطها المعادلة التالية :

$$\text{ص} = \text{ر} \times \frac{\text{س}}{\text{س}} + \text{ص}_\text{.....} (٣ - ١١)$$

حيث صن ترمز إلى قيمة المتغير ص المتتبأ بها بمعلومية قيمة المتغير (س) .

‘ ر ’ ترمز إلى معامل الارتباط بين المتغيرين (س) ، (ص) .

يعني ، ترمز إلى الانحراف المعياري لكل من المتغير $(س)$ ، $(ص)$ على الترتيب.

، $ص$ ترمز إلى قيمة متوسط المتغير $ص$.

ويتطلب تطبيق هذه المعادلة إيجاد قيمة المتوسط والانحراف المعياري لكل من المتغيرين $(س)$ ، $(ص)$ ، وكذلك قيمة معامل ارتباط بيرسون بينهما. ولتوسيع ذلك نعود إلى جدول (٣ - ٧) لإيجاد معادلة انحدار درجات الاختبار $(ص)$ بمعلومية درجات الاختبار $(س)$ ، ولذلك نوجد أولاً قيمة كل من $س$ ، $ص$ ، $ي$ ، $ر$ فنجد أن :

$$\begin{aligned} س = 11 & , \\ ي = 6,928 & , \end{aligned}$$

$$ص = r \times \frac{ص - س}{ي} + ص$$

$$\begin{aligned} 6,083 & \\ \frac{6,083}{6,928} \times (س - 11) + 10 & = \\ 10 + 631 & , \end{aligned}$$

$$ص = 631,005$$

وهذه هي معادلة انحدار $(ص)$ على $(س)$ التي تستخدم في التنبؤ بقيم $ص$ إذا علمنا أي قيمة من قيم $(س)$ وذلك بالتعويض عن قيمة $س$ في هذه المعادلة. ومن الجدير بالذكر أنه يمكن استخدام برنامج الحاسوب الذي ذكرناه فيما سبق في إيجاد معادلة انحدار المتغير $(ص)$ على المتغير $(س)$. فقيمة كل من A ، B اللذين نحصل عليهما يمكن التعويض بهما في معادلة الخط المستقيم الذي يمثل معادلة الانحدار وهي :

$$ص = B + A \cdot س \quad (12-3)$$

حيث B ترمز إلى ميل خط انحدار $ص / س$

، A ترمز إلى نقطة تقاطع خط الانحدار مع المحور الأفقي.

والمعادلتان (٣ - ١١) ، (١٢ - ٣) هما صيغتان مختلفتان ولكن نتائجهما

واحدة.

الخطأ المعياري للتبؤ:

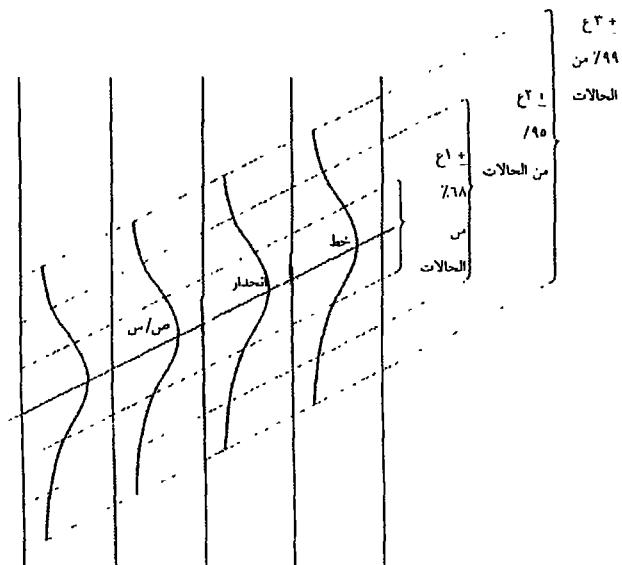
لتقدير أخطاء التبؤ نستخدم الصيغة التالية التي تعتمد على الانحراف المعياري للمتغير المتباين به (ص)، ومعامل الارتباط بين (س)، (ص)، ولذلك يسمى الخطأ المعياري للتبؤ.

$$\text{الخطأ المعياري للتبؤ} = \sqrt{1 - R^2} \quad \dots \dots \quad (12-3)$$

$$\text{فالخطأ المعياري للتبؤ في المثال السابق} = \sqrt{1 - 0.83^2} = 0.4, 0.83$$

$$= 0.4, 0.83 \quad (12-4)$$

ولتفسير مقدار هذا الخطأ نفترض أن أخطاء التبؤ توزع توزيعاً اعتدالياً انحرافه المعياري ± 1 ، فإذا رسمنا خطوطاً موازية لخط الانحدار على كل من جانبيه على مسافات تساوى قيمة الخطأ المعياري ± 1 ، ومضاعفاتها ± 2 ، فإنه بالرجوع إلى المساحات تحت المنحنى الاعتدالي المعياري الموضح بشكل (١٢-٣) نجد أن حوالي ٦٨٪ من الطلاب يقعون بين -1 ، $+1$ خطأ معياري ، ٩٥٪ منهم يقعون بين -2 ، $+2$ خطأ معياري، وهكذا كما هو موضح بشكل (١٢-١٧) التالي :



شكل (١٢-١٧) يوضح توزيع الدرجات المتباينة حول خط الانحدار من على س

أى أن ٦٨٪ منهم تتحضر درجاتهم بين $\pm ٤,٢٣$ حول الدرجة صـ المتبـاً بها، ٩٥٪ منهم تتحضر درجاتهم بين $(+ ٤,٢٣ \times ٢ , - ٤,٢٣ \times ٢)$ ، أى بين $\pm ٨,٤٦$ حول الدرجة صـ.

وبالطبع كلما زاد عدد الأفراد زاد اقتراب عدد القيم التي تتحضر بين الخطين من القيم المتوقعة من التوزيع الاعتدالـى . كما أنه كلما ارتفعت قيمة معامل الارتباط كان التنبـؤ بقيمة المتغير (صـ) أكثر دقة . فمعامل الارتباط $-٠,٩٠$ ، مثلاً أكثر فائدة في التنبـؤ من المعامل $٠,٢٠$ ، ولكن المعامل $-٠,٩٠$ يكون مفيدةً في التنبـؤ تماماً كالمعامل $+٠,٩٠$. أما إذا كانت قيمة معامل الارتباط (-١) أو $(+١)$ فإن جميع النقط تقع على خط الانحدار ، ويكون التنبـؤ عندئذ تاماً.

وفي ختام هذا الفصل نود أن نشير إلى أهمية معامل الارتباط والانحدار الخطـى في تحديد بعض الخصائص السـيـكـوـمـتـرـيـة المهمـة لأدوات القياس وبخـاصـة الاختـبارـات مـرـجـعـيـةـ الجـمـاعـةـ أوـ المـعيـارـ Norm - Referenced Tests ، وهو ما ستتناوله بالتفصـيل في الفـصلـ التـالـىـ .

خلاصة

يعتمد القياس التربوي النفسي على جمع بيانات عن طريق أدوات قياس متنوعة. وهذه البيانات إما أن تكون كمية أو كيفية، وتتطلب البيانات الكمية تسويفاً وتصنيفاً وتحليلًا لاستخلاص معلومات مفيدة تتعلق بسمات معينة. ويبدأ هذا التحليل للدرجات الاختبار بتكونين جدول توزيع تكراري وتمثيله بيانياً بطرق متعددة، مثل المدرج التكراري والمضلعل التكراري والمنحنى التكراري، وكذلك يمكن استخدام شكل الأغصان والأوراق الذي يعد بمثابة توزيع تكراري ومدرج تكراري للدرجات في آن واحد، ويوضح الشكل العام لهذا التوزيع، ويحافظ على جميع الدرجات مرتبة ويمكن حساب ثلاثة مقاييس للتوزعة المركزية، هي: المتوسط، والوسيط، والمنوال. ويعتمد اختيار المقاييس المناسب على مستوى قياس السمة وطبيعتها، وشكل توزيعها.

ويمكن تحديد مقدار تشتت توزيع الدرجات باستخدام الانحراف المعياري ونصف المدى الريعي، كما يمكن استخدام الانحراف المعياري في المقارنة بين توزيعين أو أكثر من الدرجات إذا علمنا قيم متوسط هذه التوزيعات، وكذلك يمكن تحديد المركز النسبي للأفراد في سمة معينة بتحويل الدرجات الخام إلى درجات معيارية أو معيارية مُعدلَّة ، أو درجات ثانية أو تائية معدلَّة أو مثنائيات. وجميع هذه الدرجات المحولة تعتمد على الانحراف المعياري والمتوسط. ويستخدم المنحنى الاعتدالى أيضًا في إيجاد المئويات والرتب المئوية المناظرة لدرجات خام معينة.

أما معامل ارتباط بيرسون فيستخدم في إيجاد العلاقة بين مجموعتين من الدرجات، وتتراوح قيمه بين - ١ ، ٠ ، + ١ ويُفيد معامل الارتباط في وصف درجة العلاقة واتجاهها، كما يُفيد في التنبؤ بدرجة الفرد في أحد المتغيرين إذا علمنا درجته في المتغير الآخر، وذلك باستخدام معادلة الانحدار الخطى البسيط، وتحديد الخطأ المعياري للتنبؤ الذي يدل على مدى دقة التنبؤ.

الفصل الرابع

الذئائص السيميكوتيرية للاختبارات

أولاً : ثبات درجات الاختبارات

* مقدمة

* ثبات درجات الاختبارات

* الثبات والدرجات الحقيقية

* المفهوم الإحصائي لمعامل الثبات

* مؤشر الثبات

* مصادر أخطاء القياس

* أنواع معاملات ثبات درجات الاختبارات

* كيفية تقدير قيم أنواع معاملات الثبات

* معامل التكافؤ

* معامل الاستقرار

* معامل الاستقرار والتكافؤ

* معامل الاتساق الداخلى

* معامل التجانس لكيودر وريتشاردسون

* معامل (α) لكرونياك

* ثبات تقديرات المحكمين

* بعض الطرق الأخرى لتقدير الثبات

* ثبات الفرق بين درجتي اختبارين

* معامل إمكانية التعميم

* تعقيب عام على طرق تقدير الثبات وتقدير نتائجها

* الخطأ المعياري للقياس

* تفسير الخطأ المعياري للقياس

* عوامل أخرى تؤثر في قيم معامل الثبات

مقدمة :

أوضحنا في الفصل السابق كيفية تنظيم ووصف توزيعات درجات الاختبارات والمقاييس التربوية والنفسية باستخدام أساليب إحصائية متعددة لكي يتسع تفسير الدرجات تفسيراً ينفعه عما تتضمنه من معلومات مفيدة تحقق الهدف من عملية القياس. غير أن الاختبارات والمقاييس تعد أدوات قياس وينبغى أن تتوافر فيها شروط معينة لكي تكون صالحة لقياس سمة معينة.

بعض أدوات القياس والتقويم التربوي والنفسى ربما لا تكون معدة إعداداً جيداً مما يتطلب الحيطة أو الحذر عند استخدامها. وهذا الأمر لا يقتصر على العلوم السلوكية، وإنما يمتد ليشمل أيضاً أدوات القياس في العلوم الطبيعية، فما بالك بميزان أو ترمومتر أو أميتر أو ميكرومتر تعطى قراءات مختلفة إذا تكرر استخدامها في قياس الشيء نفسه في أوقات مختلفة. أليس هذا ممكناً؟ بل وربما يحدث كثيراً في حياتنا اليومية.

وما بالك إذا كانت لا تقيس الظاهرة التي هدفت لقياسها، وإنما تقيس ظاهرة مختلفة، فعندئذ تكون أداة القياس غير موثوق بها، ويجعل الاعتماد عليها في القياس مضلاً. غير أن مشكلة صلاحية وجودة أدوات القياس تكون أكثر إلحاحاً وتعقيداً في العلوم السلوكية. فالدرجات المستمدة من هذه الأدوات ينبغي أن تتميز بالاتساق ولا تتغير تغيراً جوهرياً من تطبيق إلى آخر على الفرد نفسه ما دامت الظروف لم تتغير تغيراً ملحوظاً. كما ينبغي أن نطمئن إلى أن أداة القياس تقيس بالفعل السمة المحددة لكي تؤدي من نتائجها في اتخاذ قرارات صائبة تتعلق بالفرد أو بالجماعة، سواء في الانتقاء أو التصنيف أو التشخيص أو العلاج.

وتشير هذه الخصائص إلى مفهومين من المفاهيم الأساسية التي تتعلق بالاختبارات والمقاييس التربوية والنفسية هما مفهوم ثبات درجات الاختبار Test Reliability، وصدق الاختبار Test Validity. ونظرًا لأهمية هذين المفهومين، فإننا سنفرد هذا الفصل لمناقشة مفهوم الثبات في الاختبارات مرجعية الجماعة أو المعيار، ونعرض مفهوم الصدق في الفصل الخامس.

وسوف نتناول الطرق المختلفة لتقدير كل من المفهومين، وكيفية تفسيرهما والعوامل المؤثرة في كليهما ، وعلاقة كل منهما بالآخر .

ثبات درجات الاختبارات :

يُقصد بمفهوم ثبات درجات الاختبارات مدى خلوها من الأخطاء غير المتناظمة التي تسبب التباس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة التي يهدف لقياسها، فدرجات الاختبار تكون ثابتة Reliable إذا كان الاختبار يقيس سمة معينة قياساً متسقاً في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس. فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس.

والأخطاء غير المتناظمة التي تؤثر في درجات الاختبارات تكون أخطاء عشوائية يصعب التنبؤ بها من موقف إلى آخر ولذلك تعمل على خفض ثبات الدرجات. وهذه الأخطاء العشوائية ترجع إلى عوامل بعضها يتعلق بالاختبار، مثل عدم وضوح مفرداته، وغموض تعليماته، وعدم تحديد محكّات تصحيح مفرداته، والبعض الآخر يتعلق بالظروف البيئية مثل الإضاءة والتهوية والضوضاء وملاءمة غرفة الاختبار، وكذلك عوامل تتعلق بخصائص الأفراد المختبرين، مثل قلة دافعيتهم، وشعورهم بالتعب أو الملل أو القلق، وحالاتهم النفسية والصحية بعامة وقت إجراء الاختبار. وكذلك الأخطاء المتناظمة ربما تؤدي إلى تضخيم أو تقليل ثبات درجات الاختبار ولكن بطريقة متسقة، وبالتالي لا تؤثر في ثبات الدرجات.

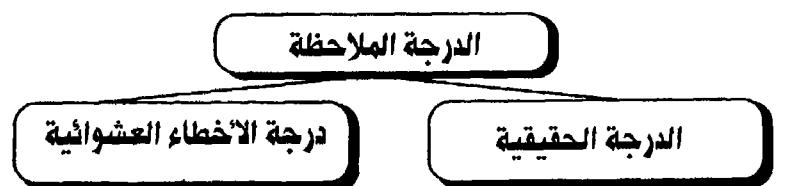
الثبات والدرجات الحقيقة :

إن الدرجة التي يحصل عليها الفرد في مقياس معين تسمى الدرجة الملاحظة Obtained Score ، غير أن هذه الدرجة تكون في كثير من الأحيان مشوبة بأخطاء القياس التي إذا أمكن تحديد مقدارها فإننا نحصل على درجة الخطأ Error Score ، وإذا طرحنا درجة الخطأ من الدرجة الملاحظة فإننا نحصل على درجة الفرد خالية من الأخطاء العشوائية، وهذه تسمى الدرجة الحقيقة True Score ، أي أن :

$$\text{الدرجة الحقيقة} = \text{الدرجة الملاحظة} - \text{درجة الأخطاء العشوائية}$$

$$\text{أو أن الدرجة الملاحظة} = \text{الدرجة الحقيقة} + \text{درجة الأخطاء العشوائية}$$

ويمكن توضيح ذلك بالشكل التخطيطي (٤ - ١) التالي:



شكل (٤ - ١) يوضح مكونات الدرجة الملاحظة

فالدرجة التي يحصل عليها الفرد في الاختبار إذن ليست هي الدرجة الحقيقية، وإنما هي الدرجة التي تتكون من الدرجة الحقيقة مضافاً إليها درجة الأخطاء العشوائية كما هو مبين بشكل (٤ - ١). أى أن الدرجة الملاحظة للفرد في الاختبار لا تساوى درجته الحقيقة إلا إذا كانت درجة الخطأ صفرًا، ولكنها تزيد عن درجته الحقيقة إذا كانت درجة الخطأ موجبة، وتقل عنها إذا كانت درجة الخطأ سالبة.

إذا لجأ الفرد إلى التخمين في إجابتة عن مفردات الاختبار وتوصل إلى الإجابة الصحيحة فإن درجته الملاحظة تزيد عن درجته الحقيقة، أما إذا كان الفرد متعباً أو أن مفردات الاختبار أو تعليماته غامضة فإن درجته الملاحظة سوف تقل عن درجته الحقيقة.

غير أن هذه الأخطاء الموجبة والسلبية يلاشى بعضها البعض الآخر على المدى البعيد استناداً إلى قوانين الصدفة، وبذلك يصبح متوسط درجات الخطأ صفرًا. ولذلك فإن متوسط الدرجات الملاحظة لمجموعة الأفراد الذين يطبق عليهم الاختبار يعد أفضل تقدير لمتوسط درجاتهم الحقيقة في الاختبار.

كما أن تباين الدرجات الملاحظة يساوى مجموع تباين كل من الدرجات الحقيقة وتباین الخطأ (لاحظ أن مقدار التباين = مربع مقدار الانحراف المعياري) (كما سبق أن أوضحنا في الفصل الثالث)، فإذا كانت الدرجات الملاحظة خالية من الخطأ فإن مقدار تباين الخطأ يساوى صفر، ويصبح مقدار تباين الدرجات الملاحظة مساوياً مقدار تباين الدرجات الحقيقة، أى تكون الدرجات متسقة اتساقاً تاماً، وتكون قيمة معامل الثبات واحداً صحيحاً .

أما إذا كانت درجات الاختبار غير متسقة على الإطلاق، فمعنى ذلك أن الدرجات لا تمثل سوى أخطاء القياس، أى أن الفروق الملاحظة بين الأفراد في هذه الحالة لا يكون لها أي علاقة بالفروق الحقيقة بينهم، وتكون قيمة معامل الثبات صفرًا.

لذلك فإن الانحراف المعياري لدرجات الخطأ يسمى الخطأ المعياري للقياس Standard Error of Measurement . فإذا كان مقدار هذا الخطأ المعياري صفرًا تكون درجات الاختبار متسقة اتساقاً تاماً. ولكننا نلاحظ عادة في القياس التربوي والنفسى أن هناك قدرًا معيناً من الخطأ المعياري مما يؤدي إلى خفض ثبات درجات الاختبارات بقيمة متساوية بحسب مقدار هذا الخطأ. وسوف نناقش مفهوم الخطأ المعياري للقياس فيما بعد .

المفهوم الاحصائي لمعامل الثبات :

على الرغم من أن الدرجات الحقيقة ودرجات الأخطاء العشوائية تعد من المفاهيم الافتراضية، إلا أنها تفيد بدرجة كبيرة في التوصل إلى المفهوم الإحصائي للثبات، فقد تبين لنا أنه إذا كانت درجات الخطأ غير مرتبطة بالدرجات الحقيقة فإن تباين الدرجات الملاحظة = تباين الدرجات الحقيقة + تباين الخطأ.

$$\text{أي أن } \frac{\sigma^2}{\text{تح}} = \frac{\sigma^2}{\text{تح}} + \frac{\sigma^2}{\text{خط}} \quad \dots \quad (1 - 4)$$

وإذا قسمنا كلا من طرفي المعادلة (4 - 1) على $\frac{\sigma^2}{\text{تح}}$ نحصل على :

$$\frac{\frac{\sigma^2}{\text{تح}}}{\frac{\sigma^2}{\text{تح}}} + \frac{\frac{\sigma^2}{\text{خط}}}{\frac{\sigma^2}{\text{تح}}} = 1$$

$$\text{أي أن : } \frac{\frac{\sigma^2}{\text{تح}}}{\frac{\sigma^2}{\text{تح}}} - 1 = \frac{\frac{\sigma^2}{\text{خط}}}{\frac{\sigma^2}{\text{تح}}} \dots \quad (2 - 4)$$

يدل على نسبة التباين الحقيقي إلى التباين الملاحظ.

$$\frac{\sigma^2}{\text{تح}}$$

أما الكسر $\frac{\sigma^2}{\text{خط}}$ فيدل على نسبة تباين الخطأ إلى التباين الملاحظ.

$$\frac{\sigma^2}{\text{خط}}$$

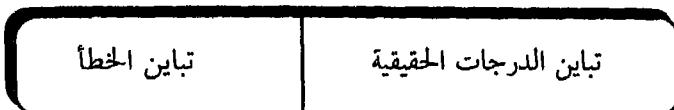
ولذلك فإن الكسر الأول يُعبر عن معامل ثبات درجات الاختبار Coefficient of Reliability . أي أن معامل الثبات من المنظور الإحصائي يمكن اعتباره نسبة تباين الدرجات الحقيقة إلى تباين الدرجات الملاحظة كالتالي :

$$\text{معامل الثبات } R_{11} = \frac{\text{تباین الدرجات الحقيقة}}{\text{تباین الدرجات الملاحظة}} \dots \quad (3-4)$$

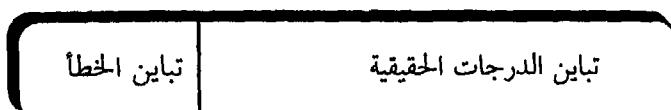
وهذا يعني أن معامل الثبات هو ذلك الجزء من تباين الدرجات الملاحظة الذي يُعزى إلى تباين الدرجات الحقيقة. وتتراوح قيم هذا المعامل بين صفر ، ١ ، ٠ ، ولذلك

يعبر عن قيمته بكسر عشري مثل معامل الارتباط. فكلما اقتربت قيمة من الواحد الصحيح دل ذلك على قدر أكبر من ثبات الدرجات، وبالتالي تزداد الثقة في دقة هذه الدرجات واتساقها والاعتماد عليها. ويمكن توضيح ذلك بشكل (٤ - ٢) التالي:

معامل ثبات منخفض



معامل ثبات مرتفع



بيان الدرجات الملاحظة

شكل (٤ - ٢) يوضح زيادة قيمة الثبات بزيادة القدر النسبي

لبيان الدرجات الحقيقة

العلاقة بين الدرجات الحقيقة والدرجات الملاحظة (مؤشر الثبات) :

لعل من المفيد أن نوضح العلاقة بين الدرجات الحقيقة لمجموعة من الأفراد ودرجاتهم الملاحظة في اختبار ما في ضوء المفهوم الإحصائي لمعامل الثبات. وهذا يتطلب العودة إلى الفصل الثالث لمراجعة مفهوم معامل الارتباط، مع ملاحظة أنه يمكن كتابة الصيغة (٣ - ١٠) بطريقة مختصرة كالتالي:

$$\text{معامل الارتباط } R = \frac{\text{مج س ص}}{ن \times ع \times ع من} \quad (٤ - ٤)$$

حيث ترمز (س)، (ص) إلى انحراف كل منها عن متوسطها.

وترمز (ع)، (ع من) إلى الانحراف المعياري لكل من (س)، (ص).

وترمز (ن) إلى عدد أزواج الدرجات.

إذا رمزنا لانحرافات كل من الدرجات الحقيقة، والدرجات الملاحظة، ودرجات الخطأ عن متوسط توزيعاتها الممناظرة بالرموز (ح)، (م)، (خ) على الترتيب، وتطبيق الصيغة (٤-٤)، فإن معامل الارتباط بين الدرجات الحقيقة ودرجات الخطأ يكون كالتالي :

$$\frac{\text{مجمـع} \times ح}{ن \times ع} = ح$$

$$\frac{\text{مجمـ ح} (ح + خ)}{\text{ن} \times \text{ع} \times \text{ع}} =$$

(وذلك لأن $m = ح + خ$)

$$(0-4) \dots - \frac{\text{مجم ح} \times \text{ع}}{\text{ن} \times \text{ع} \times \text{ع}} + \frac{\text{مجم ح} \times \text{ع}}{\text{ن} \times \text{ع} \times \text{ع}} = \frac{\text{مجم (ح} \times \text{ح خ)}}{\text{ن} \times \text{ع} \times \text{ع}} =$$

ولكن $\frac{\text{موج} \hat{H}}{n}$ = تباین الدرجات الحقيقة \hat{U}

مج_ج خ = مجموع حواصل ضرب الدرجات الحقيقية في درجة الخطأ،
ونظراً لأنهما غير مرتبطين كما سبق أن ذكرنا ، فإن هذا المجموع = صفر.
أي أن : مج_ج خ = صفر ، وبالتالي ينعدم كسر النسبة الثانية .
وبالتعويض في (٤ - ٥) نجد أن :

$$(7 - \xi) \dots \dots \dots \quad \frac{\sum_{\text{C}}}{\sum_{\text{C}}} = \frac{\sum_{\text{C}}}{\sum_{\text{C}} \times \sum_{\text{C}}} = \text{PC}$$

ولكن سبق توضيح أن : معامل الثبات R_m =

في التعويض، في، (٤ - ٦) نجد أن :

$$R^2 = \frac{R^2}{R^2 + R^2} \text{ أى } R^2 = \frac{R^2}{2R^2} = \frac{1}{2}$$

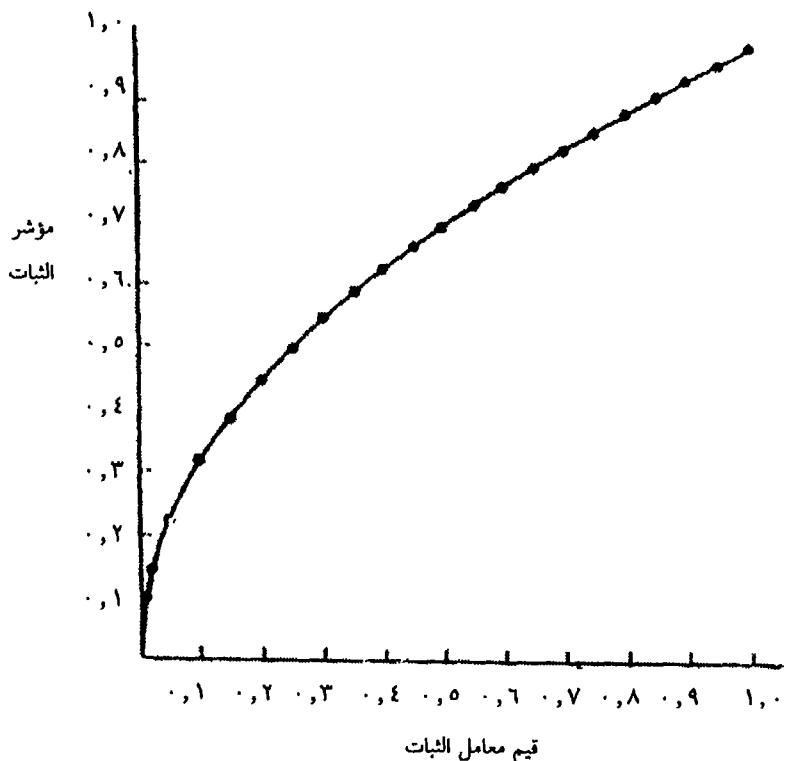
وهذا يعني أنه يمكن تقدير معامل الارتباط بين الدرجات الحقيقة والدرجات الملاحظة لمجموعة من الأفراد في اختبار ما بإيجاد الجذر التربيعي لمعامل ثبات درجات هذا الاختبار. ويسمى معامل الارتباط عندئذ «مؤشر الثبات» Reliability Index، فإذا كانت قيمة معامل ثبات درجات اختبار .٨٠، مثلاً فإن القيمة التقديرية لمعامل الارتباط بين الدرجات الحقيقة والدرجات الملاحظة في الاختبار تساوى $\sqrt{.80} = .89$ ، وهذه هي قيمة مؤشر الثبات.

ونظراً لأن معامل الثبات يساوى مربع مؤشر الثبات أى مربع معامل الارتباط بين الدرجات الحقيقة والدرجات الملاحظة، فإن معامل الثبات يعد بمثابة معامل التحديد الذى أوضحناه عند مناقشتنا لكيفية تفسير معامل الارتباط. فمعامل الثبات إذن يحدد نسبة تباين الدرجات الملاحظة للفرد التي تعزى إلى التباين الحقيقى للدرجات فيما يقيسه الاختبار. وإذا طرحتنا قيمة من الواحد الصحيح، فإننا نحصل على نسبة تباين في الدرجات الملاحظة التي تعزى إلى الأخطاء العشوائية.

وتجدر الإشارة هنا إلى أن قيمة معامل الثبات هي القيمة التي يذكرها الباحثون ومعدو الاختبارات والمقياس عادة في أدلة الاختبارات. فمعامل الثبات يدل على العلاقة بين مجتمعتين من الدرجات الملاحظة (المعلومة)، مثل درجات اختبارين متكافئين يقيسان السمة نفسها، بينما يدل مؤشر الثبات على العلاقة بين الدرجات الحقيقة (غير المعلومة) والدرجات الملاحظة (المعلومة). لذلك فإن مؤشر الثبات يعطى قيمة تقديرية لهذه العلاقة ، وعلى الرغم من ذلك فإن كليهما يقدم نفس المعلومات، ولكن بطريقتين مختلفتين.

وقد أوضح جاليسان Gulliksan أن مقدار صدق الاختبار ينبغي أن يكون دائمًا أقل من قيمة مؤشر الثبات، أى أقل من R^2 كما سنرى عند مناقشتنا لفهم الصدق. فمن الناحية النظرية لا يعني أن ترتبط درجات اختبار معين بأى متغير آخر بدرجة أكبر من ارتباطها بالمتغير الذى يقيسه الاختبار، أى ارتباط الاختبار بنفسه.

ويكفي توضيح العلاقة بين قيم كل من معامل الثبات، ومؤشر الثبات بيانياً بتمثيل قيم معامل الثبات التي تتراوح بين صفر ، .١، على المحور الأفقي ، وقيم مؤشر الثبات والتي تتراوح أيضًا بين صفر ، .١، على المحور الرأسى . كما هو موضح بالشكل (٤ - ٣) التالي :



شكل (٤ - ٣) يوضح العلاقة المنحنية بين معامل الثبات ومؤشر الثبات

وبالنظر إلى شكل (٤ - ٣) يتضح أن قيمة مؤشر الثبات تكون أكبر من قيمة معامل الثبات المناظرة لها فيما عدا عندما تكون قيمة معامل الثبات صفرًا أو واحدًا صحيحةً. فمؤشر الثبات يساوي ٠,٨٩ ، إذا كان معامل الثبات ٠,٨ ، كما سبق أن رأينا ، ويصبح مؤشر الثبات مساوياً ٠,٩٥ ، إذا كان معامل الثبات ٠,٩ ، وهكذا.

ويمكن الحصول على قيمة مؤشر الثبات برسم خط رأسى من النقطة الممثلة لقيمة معامل الثبات ، وعند تقائه هذا الخط بالمنحنى نرسم خطًا أفقياً يلتقى المحور الرأسى عند نقطة تمثل قيمة مؤشر الثبات.

مصادر الأخطاء القياسية :

ذكرنا أن هناك نوعين من أخطاء القياس أحدهما الأخطاء المتناظمة Systematic Errors، والأخرى الأخطاء غير المتناظمة أو العشوائية Unsystematic Errors التي ترجع إلى عوامل الصدفة Chance Factors.

وهذه الأخطاء العشوائية هي تلك الأخطاء التي لا ترتبط بأداء الفرد المستقبلي الذي نود الاستدلال عليه، وبالتالي فهي لا ترتبط بدرجة الفرد الحقيقة أو بدرجة الأخطاء العشوائية إذا ما أعيد تطبيق الاختبار نفسه أو صيغة مكافأة له على الفرد نفسه.

فدرجة الفرد الحقيقة في الاختبار هي تلك الدرجة الناجمة عن جميع العوامل المتناظمة التي نقررها بما في ذلك الأخطاء المتناظمة التي ربما تسهم بدرجة متناظمة في تباين درجات الاختبار وتكون جزءاً من تباين الدرجة الحقيقة. ولتوسيع ذلك نفترض أننا اختبرنا مجموعة من الأفراد بمفردات اختبارية مختلفة فعندئذ تسهم أخطاء الصدفة في تباين درجاتهم الحقيقة نظراً لاختلاف عينة المفردات، أما إذا اختبرنا جميع الأفراد بعينة واحدة من المفردات، فإننا بذلك نستبعد مثل هذه الأخطاء على الرغم من أن اختلاف طبيعة مفردات هذه العينة بالنسبة لكل فرد يمكن أن تؤدي إلى تباين إجاباتهم عنها، ولكن بطريقة متناظمة. وكذلك إذا أضفنا لكل طالب درجتين على درجته في الاختبار، فإن هذه الإضافة تعد من الأخطاء المتناظمة التي لا تؤثر في قيمة معامل ثبات الاختبار.

ولذلك فإن هذه الأخطاء المتناظمة أو الثابتة لا تهمنا كثيراً في هذا الشأن وإن كانت تؤثر بلا شك في دقة البحث العلمي.

ونظراً لأن هذا النوع من الأخطاء له علاقة بصدق الاختبار فسوف نرجو مناقشته للجزء الخاص بالصدق في الفصل الخامس.

غير أن الذي يجب مراعاته جيداً هو الأخطاء غير المتناظمة أو العشوائية، حيث إنها تؤثر بدرجات متفاوتة في معامل الثبات كما سبق أن أوضحنا.

وهنا ربما يتساءل القارئ عن مصادر هذه الأخطاء العشوائية من أجل التحكم فيها أثناء إعداد الاختبار وعند تطبيقه وتصحيح مفرداته، وذلك لجعل تأثيرها في الدرجات الحقيقة أقل ما يمكن، وبذلك تقترب الدرجات الملاحظة للأفراد في الاختبار من درجاتهم الحقيقة ويزيد تباين الدرجات الحقيقة إلى أقصى حد ممكن، وبالتالي تكون قيم معامل الثبات أكبر ما يمكن.

ونظراً لأن الأخطاء العشوائية التي تؤثر في موقف اختباري معين تختلف باختلاف مجموعة الأفراد المختبرين، ونوع الاختبار وعينة مفرداته، وظروف تطبيقه

وتصحيح مفرداته، فإن مصادر هذه الأخطاء تكون كثيرة ومتنوعة. غير أننا سوف نقتصر فيما يلى على مناقشة بعض أهم هذه المصادر التي تؤثر في القياس التربوى والنفسى، ويمكن تصنيفها في ثلاثة مصادر أساسية كالتالى :

أولاً : مصادر تتعلق بآداة القياس،

تعد أدوات القياس فى بعض الأحيان مصدراً من مصادر الأخطاء العشوائية. فعلى الرغم من أنه ينبغي أن يراعى فى تصميم هذه الأدوات جعل هذه الأخطاء أقل ما يمكن، إلا أننا نلاحظ أن مفردات الاختبار التربوى أو النفسى قد تكون غاية فى الصعوبة، أو تحتمل التخمين أو أن تكون صياغتها غامضة أو مربكة، أو أن تكون تعليمات الإجابة عن الاختبار غير محددة، أو تشجع على التخمين كما فى مفردات الصواب أو الخطأ ، أو الاختيار من متعدد، وما إلى ذلك من العوامل التى تقلل من ثبات درجات الاختبار.

غير أن المصدر الأساسى للأخطاء العشوائية المتعلقة بالاختبار تتعلق بأسلوب معاینة المفردات للنطاق السلوکي الذى يقيسه الاختبار مما يتربّ عليه أن يكون عدد المفردات قليلاً أو أن المفردات لا تمثل مكونات السمة المقاسة تمثيلاً كافياً.

ولكى يتضح الإطار النظري لهذا النوع من الأخطاء ينبغي أن نلقي بعض الضوء على أهم النماذج التى تتناولها النظرية الكلاسيكية للاختبارات Classical Test Theory ، فهذه النظرية تتعلق بالاختبارات والمقياسات مرجعية الجماعة أو المعيار Norm - Referenced Tests التى تهتم بقياس الفروق الفردية والكشف عنها، لذلك تفترض أن الأفراد المختبرين تتوزع درجاتهم توزيعاً اعتدالياً على متصل السمة أو القدرة التى يقيسها الاختبار. ويتحقق ذلك باستخدام أساليب انتقاء المفردات الاختبارية التى تجعل توزيع الدرجات خاضعاً لهذا التوزيع الاعتدالى Normal Distribution. ويندرج تحت هذه النظرية نموذجين أحدهما يسمى نموذج معاینة النطاق The Domain Sampling Model ، والآخر نموذج الاختبارات المتوازية The Parallel Tests Model ، وكلا النموذجين يهتمان بالأخطاء العشوائية للقياس.

(١) نموذج معاینة النطاق : يعتبر هذا النموذج أن أي اختبار أو قياس يشتمل على عينة عشوائية من المفردات المسحوبة من نطاق أو مجتمع افتراضي من المفردات، مثل اختبار عمليات الجمع البسيط الذى يفترض أنه يشتمل على عينة عشوائية من الأعداد ذات الرقمان التى تتحصر بين ١٠ ، ١٠٠ ، ١٠٠٠ مثلاً. وعلى الرغم من أن هذا الافتراض ربما لا يكون واقعياً، إلا أن النموذج يمكن أن يؤدى إلى تقديرات دقيقة للدرجات الحقيقية نظراً لأن مصمم الاختبار يكتب عادة عدداً كبيراً من مختلف أنواع المفردات فى اختبار معين ويعتبرها كما لو كانت عينة عشوائية من المفردات. فالهدف الرئيسي

من بناء الاختبارات والمقاييس في إطار هذا النموذج هو تقدير القياسات التي يمكن الحصول عليها إذا اختبر الفرد بجميع المفردات التي يشتمل عليها النطاق أو المجتمع (أى جميع الأعداد ذات الرقمين مثلاً). والدرجة التي يمكن أن يحصل عليها الفرد في النطاق الشامل للمفردات Items Universe تسمى الدرجة الحقيقية True Score أو درجة النطاق الشامل Universe Score ، فإذا ارتبطت عينة المفردات التي يُختبر بها الأفراد ارتباطاً مرتفعاً بالدرجة الحقيقية، يقال: إن عينة مفردات الاختبار تميز درجاتها بالثبات. لذلك يختص هذا النموذج بالاختبارات المتوازية عشوائياً Randomly Parallel Tests حيث تختلف درجاتها الكلية إلى حد ما في المتوسط والانحراف المعياري والارتباط فيما بينها نظراً إلى الأخطاء العشوائية في معاينات المفردات التي يشتمل عليها الاختبار (Lord & Novick , 1950 ; Gulliksen , 1950).

(٢) نموذج الاختبارات المتوازية :

يفترض هذا النموذج التوازي الفعلى للاختبارات وليس تواليها على أساس المعاينات العشوائية. أي يفترض تساوى متosteات درجات اختبارات معينة وتساوى انحرافاتها المعيارية، وارتباط درجاتها الملاحظة بالدرجات الحقيقية ارتباطاً متساوياً في قيمة. وتبين درجات أي من هذه الاختبارات التي لا يعزى إلى الدرجات الحقيقية إنما يرجع إلى محض الصدفة أو الأخطاء العشوائية.

وعلى الرغم من هذا الاختلاف في فرضيات كل من النموذجين، إلا أنهما يتوصلان إلى نتائج متماثلة فيما يتعلق بالأخطاء العشوائية للقياس.

غير أن ثالثي Nunnally يرى أنه يصعب التتحقق الإمبريقي المباشر من إحدى فرضيات النموذج الثاني المتعلقة بقيم الارتباط بين الدرجات الملاحظة والدرجات الحقيقية.

وفي ضوء هذين النموذجين نلاحظ أن المصدر الأساسي لعدم اتساق درجات الاختبارات والمقاييس الذي يرجع إلى الاختبار أو المقياس نفسه يتعلق بأسلوب معاينة نطاق المفردات التي يتكون منها الاختبار.

فإذا كان للاختبار صيغ متوازية أو متكافئة فإنه ينبغي أن تتحقق فيها شروط أى من النموذجين. ونظراً لصعوبة الانتقاء العشوائي لمفردات الاختبار من نطاق شامل لهذه المفردات، فإنه يمكن المزاوجة بين المفردات في الصيغ المتعددة للاختبار من حيث مكونات السمة التي تقيسها، ومحض المفردات ودرجة صعوبتها، وتماثل توزيع درجاتها.

وهذا بالطبع يتطلب العناية بتصميم الاختبار وبناء مفرداته وتحليل نتائجها ، وإذا لم يحظ الاختبار بمثل هذه العناية ، فإن درجات الأفراد سوف تباين من صيغة إلى أخرى للاختبار نفسه ، مما يجعل الدرجات مشوبة بأخطاء القياس التي تعزى إلى أداة القياس

ثانياً ، مصادر تتعلق بإجراءات تطبيق الاختبار وتصحيحه

إن الاختبارات والمقاييس التربوية والنفسية يجب أن تطبق على الأفراد في ظروف مقتنة . فقد سبق أن أوضحتنا في الفصل الأول أن التقنيات مكونة أساسية من مكونات تعريف أداة القياس ، وناقشتنا عناصر هذه المكونة . وإجراءات تطبيق الاختبار وتصحيحه تعد مصدراً من مصادر الأخطاء العشوائية في القياس ، غير أن هذا المصدر يمكن ضبطه أو التحكم فيه لأنه يتعلق بالبيئة الفيزيائية المحيطة بالفرد أثناء تطبيق الاختبار . فالإضافة الجيدة لغرفة الاختبار ، وحسن تهيئتها ، ومنع الضوضاء حولها ، وتوزيع المقاعد بها وغير ذلك من العوامل الفيزيائية التي ربما تؤثر في إجابات الأفراد وبالتالي في درجاتهم الملاحظة ينبغي أن يتم ضبطها قبل بدء العملية الاختبارية .

كما ينبغي أن تكون تعليمات الاختبار واضحة ومفهومة وبخاصة في الاختبارات الفردية أو التي تطبق على الأطفال أو الأمينين ، حيث يفضل أن تكون التعليمات شفوية ومناسبة للمستوى العمري للأطفال . كما يجب أن يقاس زمن الإجابة عن الاختبارات التي تتطلب السرعة قياساً دقيقاً ، وأن تكون طباعة الاختبار واضحة ، وأوراق الإجابة منظمة بطريقة تسمح للفرد بتحديد إجابته بيسر وسهولة . وهذا يتطلب أن تشتمل أدلة الاختبارات على إشارات تفصيلية لمن يستخدم الاختبار بحيث يمكنه اتباعها بدقة ، على أن تتضمن هذه الإشارات كيفية التعامل في الظروف غير المتوقعة التي يمكن أن تطرأ أثناء تطبيق الاختبار .

وكذلك تشتمل على كيفية تصحيح الاختبار بطريقة موضوعية لتقليل أثر الأخطاء الناجمة عن اختلاف التقديرات وبخاصة إذا كانت مفردات الاختبار تتطلب إجابات مفتوحة كما في أسئلة المقال أو بعض مقاييس الشخصية والاتجاهات ، وكذلك كيفية رصد الدرجات ، والدقة في العمليات الحسابية التي ربما يقوم بها الفاحص للتوصيل إلى الدرجة الكلية في الاختبار وبخاصة إذا كانت تجري بطريقة يدوية .

وي ينبغي ملاحظة أن عوامل التحيز في أي من هذه الإجراءات التي ربما تكون متسقة تؤثر في التباين المحقق للدرجات ولكنها تؤدي إلى خفض صدق الاختبار . فالمعلم الذي يقوم بتقدير درجات اختبار مقالى في مادة دراسية معينة ربما يتأثر بعوامل لا تتعلق بإجابات الأسئلة ، مثل أخطاء النحو والهجاء ، أو جودة الخط ، أو فكرته

الذاتية عن الطالب مما يؤثر في الدرجة التي يحصل عليها الطالب، وبالتالي تؤثر في ثبات درجات الاختبار. ولعل الاتفاق على محكمات واضحة قبل تقدير درجات الاختبار يسهم في استبعاد مثل هذه العوامل الداخلية.

ثالثاً : مصادر تتعلق بالأفراد المختبرين :

على الرغم من أن هناك خصائص معينة للأفراد المختبرين تكون متتسقة أثناء أدائهم الاختبارات، مثل الخبرة السابقة لدى الفرد ومستوى قدرته، حيث تسهم في التباين الحقيقي للدرجات، إلا أن هناك خصائص أخرى قد تبدو متتسقة وتسهم أيضاً في التباين الحقيقي للدرجات ولكنها غير مرغوبة، مثل تذكر الإجابات عن مفردات الاختبار نتيجة الخبرة السابقة أو إعادة تطبيق صيغة مكافئة للاختبار، أو نتيجة مرض مزمن لدى الفرد أو انخفاض دافعيته، أو اتجاهه السلبي، أو تشتيت انتباذه بصورة دائمة. وهذا المصدر من مصادر الخطأ يصعب التحكم فيه؛ لأنه يتعلق بالفرد نفسه. فالاختبارات التحصيلية أو اختبارات الذكاء تتطلب من الفرد بذل أقصى جهده للحصول على درجة مرتفعة، غير أن هناك أفراداً يتسمون في معظم الأحيان بانخفاض دافعيتهم للتحصيل أو اتجاههم السلبي نحو الاختبارات، فعندئذ تنخفض درجاتهم في الاختبار نتيجة لذلك. وكذلك في مقاييس الشخصية التي تستخدم بغرض انتقاء أصلح الأفراد للالتحاق بوظيفة أو مهنة معينة، ربما يحاول الفرد إعطاء انطباع جيد للمهتمين بنتائج القياس، وذلك بأن يجيب عن القياس كما لو كان يتطلب بذلك أقصى جهده للحصول على درجة مرتفعة، وبذلك لا يعد مقياساً للأداء المميز الذي يجب فيه الفرد كما يسلك عادة في حياته اليومية، وإنما قد يجيب إجابات تبدو مرغوبة اجتماعياً . ومع هذا فإن هذه الخصائص المتتسقة ربما تكون مرتبطة بقدرات الفرد أو سماته.

غير أن ما يهمنا هو تلك الخصائص التي تسهم في أخطاء القياس وينبغي العمل على تقليل أثراها بقدر الإمكان. وتعلق هذه الخصائص بالظروف الطارئة أو الوقتية للأفراد مثل التعب ، والانخفاض النسبي للدافعية ، وتقلب الحالة المزاجية ، وقلة الاكتراث بالاختبار ، وما شابه ذلك. وعلى الرغم من صعوبة التحكم في هذه التغيرات أيضاً إلا أن معامل الثبات يجب أن يعكس الأخطاء الناجمة عنها فقط.

والخلاصة أنه ينبغي أخذ جميع هذه المصادر بعين الاعتبار عند تقدير ثبات درجات الاختبار وفي تفسير نتائجه، كما ينبغي مراعاة المصادر الأخرى المتعددة التي قد تؤثر في دقة الدرجات واتساقها.

أنواع معاملات ثبات درجات الاختبارات :

أوضحنا في الجزء السابق المفهوم الإحصائي لمعامل الثبات ومصادر الأخطاء التي يمكن أن تؤثر في قيم هذا المعامل، وتبين لنا أن هناك مصادر متعددة للأخطاء غير المنتظمة أو العشوائية التي ينبغي التقليل من أثرها بقدر الإمكان لكن تعبير الدرجات الملاحظة التي يحصل عليها الأفراد في الاختبارات عن السمات التي تقيسها هذه الاختبارات مماثلة في درجاتهم الحقيقية. ونظرًا لأنه يصعب إلغاء تأثير هذه الأخطاء العشوائية في القياس التربوي والنفسى، فإنه يمكن تقدير أثرها في الدرجات الملاحظة بحيث نستطيع معرفة مدى إسهامها في اختلاف هذه الدرجات عن الدرجات الحقيقية، وبالتالي يمكن تقدير قيم معامل ثبات الدرجات الملاحظة.

ولذلك فإن جميع الطرق التي تستخدم في تقدير هذه القيم تعتمد على الدرجات الملاحظة، حيث إن الدرجات الحقيقة تكون غير معلومة.

وتحتفل هذه الطرق باختلاف مصدر الأخطاء العشوائية التي نود تقدير تأثيرها في الدرجات الملاحظة، وهذا يعتمد على طبيعة الاختبار وفيما تستخدم نتائجه .

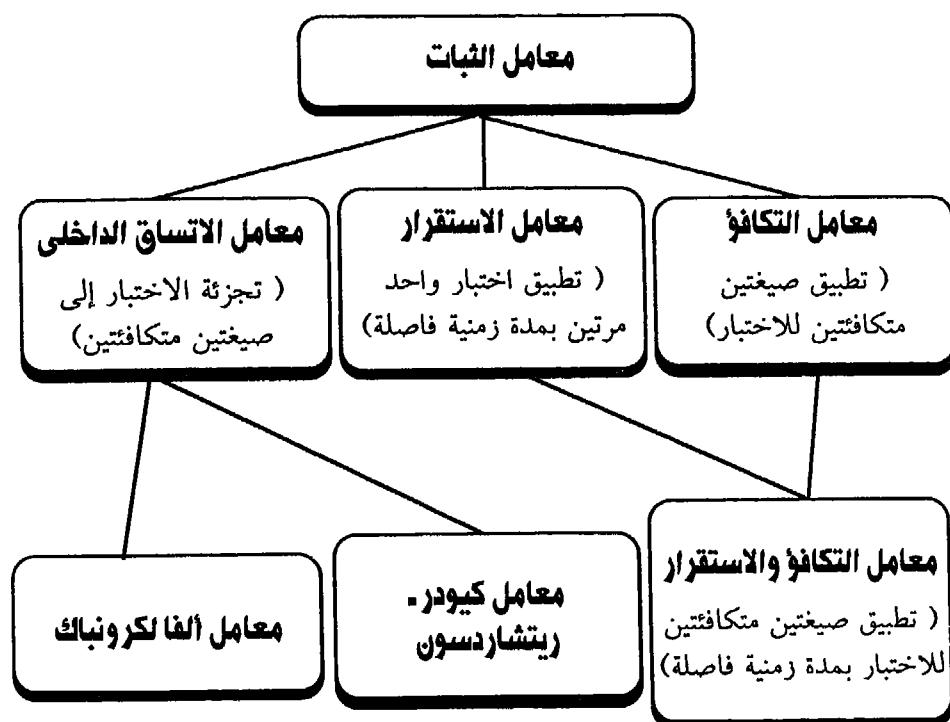
ويتعلق هذا التقدير بمعرفة مدى اتساق درجات الاختبار في قياسات متكررة لنفس الفرد أو مجموعة من الأفراد. ويحدث ذلك عادة في القياس الفيزيائى لتقدير دقة أداة القياس أو عمليات القياس حيث يعاد قياس الظاهرة عدة مرات للتحقق من ذلك. غير أنها في القياس التربوي والنفسى لا نستطيع تكرار عملية القياس على الفرد أو مجموعة الأفراد عدة مرات نظرًا لأن هذه العملية تؤدى إلى تغيرات في سلوك الفرد المختبر .

ولذلك فإننا نعتمد على تكرار القياس مرتين على الأكثر لفرد نفسه أو مجموعة الأفراد، وعندئذ نحصل على درجتين لكل فرد. ويكون لدينا مجموعتين من أزواج الدرجات المعاشرة لمجموعة الأفراد. ويمكن الحصول على تقدير لمتوسط اتساق درجات المجموعة في الاختبار بإيجاد قيمة معامل ارتباط بين أزواج الدرجات، وهذا يعد تقديرًا لمعامل ثبات درجات الاختبار.

وقد ذكرنا فيما سبق أن معامل الثبات هو النسبة بين تباين الدرجات الحقيقة وتباين الدرجات الملاحظة، وعند تكرار القياس باستخدام صيغتين متكافتين للاختبار نفسه على مجموعة واحدة من الأفراد فإن هذه النسبة تساوى قيمة معامل الارتباط بين درجات الاختبارين المتكافئين، وذلك للداخل التام بين تباين الدرجات الحقيقة لكل منهما، ويساوى النسبة بين تباين الدرجات الحقيقة والتباين الكلى للدرجات كل من الاختبارين .

وتشير أدبيات القياس والتقويم التربوي والنفسى إلى طرق متعددة لتقدير قيم معامل الثبات، بعضها يعتمد على تطبيق صيغتين متكافئتين للاختبار لتقدير تكافؤ Consistency أو اتساق Equivalence درجات الصيغتين، وبعضها يعتمد على المدة الزمنية الفاصلة بين تطبيق اختبار واحد مرتين على مجموعة الأفراد نفسها لتقدير استقرار Stability درجات الاختبار، وبعضها يعتمد على ضم الطريقيتين السابقتين معاً بحيث تطبق صيغتان متكافئتان للاختبار بمدة زمنية فاصلة لتقدير معامل الاستقرار والتكافؤ Stability and Equivalence، في حين أن بعضها يعتمد على تجزئة الاختبار إلى مجموعتين متكافئتين من المفردات لتقدير الاتساق الداخلى لمفردات الاختبار Internal Consistency، والبعض الآخر يعتمد على تحليل البنية الداخلية للاختبار Tests Items Internal Structure مثل معامل كيودر وريتشاردسون Homogeneity Kuder - Richardson ومعامل ألفا Cronbach لكرونباك.

والشكل التخطيطي (٤ - ٤) التالي يوضح هذه الأنواع من معاملات الثبات:



شكل (٤ - ٤) يوضح الأنواع المختلفة من معاملات الثبات
(من تصميم المؤلف)

ويتضح من شكل (٤ - ٤) أن بعض أنواع معاملات الثبات يتطلب تطبيقين للاختبار أو صيغ مكافئة له، والبعض الآخر يتطلب تطبيقاً واحداً.

وسوف نتناول هذه الأنواع المختلفة من معاملات الثبات فيما يلى بالمناقشة والتحليل ، ونبين كيفية تقدير قيم كل منها إمبريقيا.

كيفية تقدير قيمة أنواع معاملات الثبات:

سوف نتناول أولاً كيفية تقدير قيمة معاملات الثبات التي تعتمد على تطبيقين مختلفين لصيغ مكافئة من الاختبار أو للاختبار نفسه، ثم ننتقل إلى معاملات الثبات التي تعتمد على تطبيق الاختبار مرة واحدة.

معامل التكافؤ:

يعتمد تقدير الثبات اعتماداً أساسياً كما سبق أن أوضحنا على الارتباط بين مجموعتين من القياسات ، ولذلك فإن تقدير قيمة معامل الثبات باستخدام صيغتين متكاففتين للاختبار يتفق منطقياً مع هذا المفهوم.

غير أن المشكلة المتعلقة بذلك هي كيفية إعداد صيغ متكافئة أو متوازية لاختبار معين استناداً إلى أساليب إجرائية محددة، فإحدى الصيغتين يمكن أن تختلف في محتواها وأنواع مفرداتها ودرجة صعوبتها عن الصيغة الأخرى مما يؤدي إلى اختلاف واضح في تباين الدرجات الحقيقية للفرد في كل منها.

ولعل نموذج معاينة النطاق Domain Sampling Model الذي سبق أن أشرنا إليه يقدم إجابة نظرية لهذه المشكلة، حيث يفترض نطاقاً شاملًا للمفردات Items تنتهي منه مفردات كل من صيغتي الاختبار انتقاءً عشوائياً.

ونظراً لصعوبة إجراء ذلك عملياً فإنه يمكن إعداد مواصفات تفصيلية دقيقة لنطاق المفردات بحيث تشمل حدود محتوى الاختبار، وتوزيعه على مفرداته، ونوع المفردات وعددتها ودرجة صعوبتها، وأساليب انتقاءها وفحصها ، وكذلك تعليمات الإجابة، وحدود زمنها، وأمثلة توضيحية، وجميع الجوانب الأخرى المتعلقة بتطبيق الاختبار وتصحيح الإجابات.

إذا اتبعت هذه المواصفات في بناء صيغتي الاختبار مع مراعاة عدم تكرار المفردات في كل منها، وكذلك عدم تداخل محتواها بدرجة كبيرة، وانتقاء مفردات كل منها وفقاً لهذه المواصفات، فإن ذلك يكون كافياً لضمان تكافؤ الصيغتين، ويقال: إن الاختبارين في هذه الحالة متوازيان Parallel Tests. كما يمكن في مرحلة إعداد الاختبار تجميع عدد كبير من المفردات المتعلقة بالظاهرة المراد قياسها وتجريبيها

ميدانياً. فإذا كان محتوى الاختبار متجانساً أو يشتمل على محتويات فرعية مت詹سة، فإنه يمكن انتقاء المفردات بحيث يكون توزيع درجات صعوبتها في كل من صيغتي الاختبار له نفس الشكل، وكذلك توزيع قيم معامل الارتباط بين درجات المفردات والدرجة الكلية في كل منها.

وربما يسهل إجراء أي من هاتين الطريقتين في اختبارات التحصيل التي تشمل على مفردات تتطلب معرفة الكلمات أو الهجاء أو العمليات الحسابية، ولكن ربما يصعب إجراؤها في بعض اختبارات القدرات وكثير من مقاييس الشخصية، حيث يكون إعداد صيغ متكافئة أو متوازية للاختبار أو المقاييس أمراً يتطلب جهداً كبيراً.

وعلى الرغم من هذه الصعوبات، إلا أن استخدام صيغة متكافئة أو متوازية للاختبار يكون ضرورياً في مواقف كثيرة. فعند تقويم المعلم لطلابه أو للمنهج الدراسي مثلاً فإنه يقوم بتطبيق اختبار في محتوى دراسي معين قبل بدء عملية التعليم، وعند انتهاء عملية التعليم يطبق اختبار آخر مكافئاً للاختبار الأول وذلك لتعرف النمو التحصيلي للطلاب.

فتواتر صيغتين للاختبار يتطلب معلومات تتعلق بمعامل تكافؤهما، وعلاوة على ذلك فإن تطبيق هاتين الصيغتين يجعل أثر الذاكرة والاستذكار من أجل الاختبار أقل مما لو اختر الطلاب بالاختبار الأول نفسه قبل عملية التعليم وبعدها.

وكذلك في مواقف العلاج النفسي حيث يجرى تطبيق صيغتين متكاففتين لقياس معين أحدهما قبل العلاج والآخر بعده، لتعرف فاعلية العلاج، فإن هذا يتطلب مؤشرات تتعلق بتكافؤ درجات صيغتي المقاييس.

كذلك إذا كان اختبار معين يقيس أداء يتعلق بحل مشكلة محددة وتوصيل الفرد إلى حلها في هذا الاختبار، فإن إعادة تطبيق الاختبار نفسه على الفرد سوف يجعله يتذكر خطوات الحل السابقة بطريقة آلية، وبذلك يكون الاختبار مقياساً للتذكر وليس للقدرة على حل المشكلات.

أى أنه إذا توافرت صيغتان متكاففتان لاختبار أو مقاييس معين (أ) ، (ب) فإن مشكلة الثبات في هذه الحالة تتعلق باتساق درجات كل من الصيغتين، ويصبح من الضروري تقدير قيمة معامل التكافؤ . *Equivalent Coefficient*

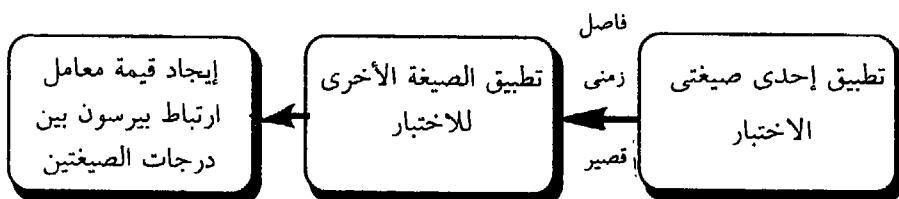
وهذا يتطلب تطبيق إحدى صيغتي الاختبار أو المقاييس (الصيغة أ مثلاً) في وقت ما ، وبعدها بقليل تطبق الصيغة الأخرى (الصيغة ب) ، أى أن الصيغتين تطبقان على مجموعة الأفراد واحدة تلو الأخرى أو بفواصل زمني قصير ، لكنى لا تؤثر التغيرات التي ربما تحدث للأفراد المختبرين خلال هذا الفاصل الزمني في ثبات الدرجات باعتبار

أن هذه التغيرات غير متسقة، أى تعد من الأخطاء العشوائية. كما أن عوامل التذكر والمران والتعب لا تؤثر تأثيراً غير منتظم في قيمة معامل التكافؤ.

وهذا الإجراء يشبه إلى حد ما وزن فرد مرتين متاليتين بميزانين مختلفين في اليوم نفسه، فهنا لا نتوقع أن يكون هناك اختلاف في القيمتين يرجع إلى زيادة وزن الفرد، فإذا وجد أى قدر من الاختلاف فإنما يرجع إلى اختلاف أداة القياس أى الميزان أو عملية القياس. ولإبطال التأثير الناجم عن ترتيب تطبيق كل من صيغتي الاختبار يمكن تقسيم مجموعة الأفراد إلى قسمين إحداهما تخبر بالصيغة (أ) تلتها الصيغة (ب)، والأخرى تخبر بالصيغة (ب) تلتها الصيغة (أ).

فال المصدر الأساسي للأخطاء العشوائية في هذا النوع من الثبات يرجع إلى الفروق بين مفردات كل من صيغتي الاختبار أو المقياس التي ربما تكون نتيجة عدم تمثيل مفردات كل من الصيغتين للنطاق الشامل تمثيلاً كافياً، أو عدم الالتزام بمواصفات هذا النطاق وبخاصة إذا كان هذا النطاق الشامل غير محدد تحديداً جيداً، أو أن السمة التي يقيسها الاختبار سمة عريضة أو مركبة، مثل القدرة على التفكير الناقد أو الإبداع وكذلك ربما ترجع هذه الفروق إلى اختلاف إجراءات أو ظروف تطبيق كل من الصيغتين على الرغم من عدم وجود فاصل زمني بين مرتب التطبيق.

ويمكن تمثيل إجراءات هذه الطريقة بالشكل التخطيطي (٤ - ٥) التالي:



شكل (٤ - ٥) يوضح إجراءات تقدير قيمة معامل التكافؤ

واللحصول على قيمة تقديرية لمعامل التكافؤ توجد معامل ارتباط يرسون بين مجموعة درجات صيغتي الاختبار أو المقياس. وتعد هذه القيمة تقديرية؛ لأننا اعتمدنا على عيتيترين من المفردات ممثلتين لنطاقها الشامل كما سبق أن ذكرنا.

ولذلك فإن تفسيرها يتعلق بدقة قياس الاختبار للسمة المرجوة بغض النظر عن صيغة الاختبار المستخدمة أو ظروف تطبيقها. فكثير من القرارات التربوية والنفسية تعتمد على ذلك، فمثلاً إذا أراد مسئول عن العمل تقدير مهارة أحد العمال، فإنه لا

يقتصر على تقدير أداءه في موقف واحد أو بواسطة مشرف واحد، وإنما يعتمد على تقدير الأداء في أكثر من موقف وأكثر من مشرف لكي يطمئن على اتساق التقديرات.

ومن الجدير بالذكر أن استخدام صيغتين متكافئتين يعد من أكثر الطرق دقة في تقدير ثبات الاختبارات والمقاييس التربوية والنفسية. غير أن كلفة بناء هاتين الصيغتين وما يتطلبه من جهد ووقت يقلل أحياناً من استخدام هذه الطريقة ما لم تكن هناك صيغ متكافئة متوافرة. لذلك يفضل كثير من المهتمين ببناء أدوات القياس استخدام طريقة تعتمد على إعادة تطبيق الاختبار نفسه لتقدير قيمة معامل الثبات الذي يعرف بمعامل الاستقرار Stability Coefficient.

معامل الاستقرار:

يُستخدم معامل الاستقرار في تقسيم الخطأ الناجم عن تطبيق الاختبار مرتين متاليتين، أي يعاد تطبيق الاختبار نفسه بحيث يكون هناك فاصل زمني بين مرتب التطبيق. ولذلك يصلح هذا المعامل إذا كانت السمة التي يقيسها الاختبار لا تتغير كثيراً بمرور الزمن ، أي تكون مستقرة نسبياً.

فمثلاً لا يجوز استخدام هذا المعامل في تقدير ثبات درجات اختبار يقيس ذكاء الأطفال الصغار، أو درجات مقاييس الحالة المزاجية، أو حالة القلق حيث إن هذه الخصائص تكون متذبذبة. أما إذا كانت السمة تميز بالثبات النسبي أو الاستقرار ، فإن اختلاف درجات الفرد في مرتب تطبيق الاختبار، ربما يعزى إلى الأخطاء العشوائية للقياس. كما يصلح إذا أردنا اتخاذ قرارات لوضع خطط بعيدة الأمد مثل التنبؤ بالنجاح في إحدى الكليات باستخدام اختبار الاستعداد الدراسي .

وقد أوضحنا أن معامل التكافؤ يتعلق بتقييم التأثير المحتمل لاختلاف مفردات كل من صيغتي الاختبار في درجات كل منها.

أما هذا المعامل فيتعلق بتقييم التأثير المحتمل لاختلاف الظروف المتعلقة بالفرد والبيئة المحيطة عند إعادة تطبيق الاختبار عليه.

ومن ميزات هذا النوع من الثبات أنه يسهل تقديره، وذلك بتطبيق الاختبار على مجموعة من الأفراد وإعادة تطبيقه عليهم بعد مرور مدة زمنية معينة في ظروف مناسبة ومحددة، وإيجاد قيمة معامل ارتباط بيرسون بين مجموعة الدوائر التي تحصل عليها في مرتب التطبيق، وبذلك تحصل على قيمة تقديرية لمعامل الاستقرار Stability Coefficient، ويسمى هذا الإجراء طريقة إعادة الاختبار Test - Retest . والشكل التخطيطي (٤ - ٦) التالي يوضح ذلك :



شكل (٤ - ٦) يوضح إجراءات تقدير قيمة معامل الاستقرار

غير أن هناك بعض الاعتبارات التي ينبغي مراعاتها عند استخدام هذه الطريقة في تقدير ثبات درجات الاختبارات والمقياس. فمعامل الاستقرار لا يعكس الخطأ الناجم عن معاينة المحتوى في الاختبار، كما أن هذا المعامل يتأثر تأثيراً ملحوظاً بالمددة الزمنية الفاصلة بين مرتب تطبيق الاختبار. فإذا كانت هذه المدة قصيرة جداً فإن عوامل الذاكرة أو انتقال الأثر والتمرن ربما تؤدي إلى اتساق الدرجات في مرتب التطبيق إلى حد كبير مما يؤدي إلى تضخيم قيمة معامل الاستقرار. فإذا إعادة تطبيق الاختبار نفسه تؤدي في كثير من الأحيان إلى تغيرات معينة في الفرد ربما تسهم في تحسين أدائه فيما يقيسه الاختبار كما هو الحال إذا طبق اختبار لقياس المهارة اليدوية لدى مجموعة من الأفراد، فإنهم يحصلون على درجات أفضل إذا ما أعيد تطبيق الاختبار نفسه عليهم نتيجة لاختبارهم في المرة الأولى.

أما إذا كانت المدة الزمنية الفاصلة كبيرة بحيث تبلغ عدة أشهر مثلاً، فإننا نتوقع تدخل كثير من العوامل غير المرغوبية التي ربما تفسر اختلاف الدرجات في مرتب التطبيق وبخاصة في قياس التحصيل الدراسي، أو إذا كان المختبرون من الأطفال الصغار.

واختلاف هذه المدة الزمنية يؤدي إلى اختلاف قيم معامل الاستقرار حيث تعتمد هذه القيم على زيادة أو نقص هذه المدة. ويُوضّح ذلك إذا حاولنا تمثيل العلاقة بين قيمة معامل الاستقرار والمدة الزمنية الفاصلة بين مرتب التطبيق، فإننا سوف نلاحظ أن قيمة معامل الاستقرار تقل بزيادة هذه المدة. غير أن المدة الزمنية في ذاتها ليست سبباً في اختلاف الدرجات في مرتب التطبيق، وإنما عامل المران أو الخبرة التي يكتسبها الفرد خلال هذه المدة.

لذلك ينبغي عند تقرير القيمة التقديرية لمعامل استقرار اختبار أو مقياس معين أن يذكر بالإضافة إلى هذه القيمة المدة الزمنية الفاصلة بين مرتب التطبيق، كما يجب أن تذكر بعض المعلومات أو البيانات من عينات مختلفة من الأفراد طبق عليهم الاختبار أو المقياس مررتين تفصل بينهم فترات زمنية مختلفة كلما أمكن ذلك.

وكذلك ينبغي عند الاطلاع على أدلة الاختبارات Test Manuals توجيه الاهتمام إلى المدة الزمنية الفاصلة التي في ضوئها تم تقدير قيمة معامل الاستساق، وذلك قبل استخدام اختبار أو مقياس معين، وذلك للتأكد من أن درجات الاختبار سوف تكون مستقرة أى ثابتة خلال المدة الزمنية التي يتطلبها مستخدم الاختبار في دراسته. كما ينبغي تعرف الظروف أو المتغيرات التي ربما تكون قد لعبت دوراً خلال المدة الفاصلة مثل تدرب المختبرين في برامج معينة أو مشاهدة برامج التلفاز أو التحاقدتهم ببرامج تعليمي معين مما يؤثر تأثيراً ملحوظاً في القيمة التقديرية لمعامل الاستقرار.

ولكن انخفاض قيمة هذا المعامل في بعض الأحيان لا تعني أن درجات الاختبار غير مستقرة، إذ ربما تكون السمة المقاسة قد تغيرت على الرغم من افتراضنا أنها مستقرة نسبياً .

لذلك يرى بعض علماء النفس المعاصرين مثل أتكينسون (Atkinson, 1985) أن السمات الإنسانية الجوهرية مثل الدافعية تتذبذب بمرور الوقت. وعلى الرغم من أنه في إطار طريقة إعادة الاختبار يعد هذا التذبذب من الأنخطاء العشوائية، إلا أن هؤلاء العلماء يرون أنه يمكن التنبؤ بالتغييرات التي تحدث في دافعية الأفراد باستخدام نماذج نظرية معينة على اعتبار أن هذه التغييرات ليست عشوائية ، وإنما تغييرات متتظمة .

وعموماً فإن قلة مصادر الأنخطاء العشوائية في هذه الطريقة يمكن أن تؤدي إلى قيم تقديرية مرتفعة نسبياً لمعامل الاستقرار.

ولذلك فإنه يمكن الإفادة من ميزات كل من معامل الاستقرار ومعامل التكافؤ في تقدير ثبات درجات الاختبارات، ويسمى المعامل عندئذ معامل الاستقرار والتكافؤ . Stability and Equivalence

معامل الاستقرار والتكافؤ :

ذكرنا فيما سبق أن معامل التكافؤ يتطلب تطبيق صيغتي الاختبار في الوقت نفسه تقريباً أو بفواصل زمني قصير من أجل اتخاذ قرارات أو التوصل إلى تنبؤات قصيرة الأمد، مثل تقديرات المشرفين على العمل أثناء أداء العاملين لمهامهم. أو تبع التقدم الدراسي للطلاب خلال مدة دراسية محدودة.

ولكتنا أحياناً نون استخدام نتائج الاختبارات والمقياس في التوصل إلى تنبؤات بعيدة الأمد، مثل التنبؤ بنجاح الطلاب في الكليات باستخدام درجات اختبارات القبول، أو درجات الاستعداد الدراسي أو المعدل التراكمي في الثانوية العامة، حيث يفترض أن هذه الدرجات مستقرة، أى أننا نهتم هنا بمعامل استقرار الاختبار.

ولكنا في أحيان أخرى ربما لا يقتصر اهتمامنا على التبؤات بعيدة الأمد فقط، وإنما أيضاً بالاستدلال على نطاق شامل لمحنتي مجال مهني أو دراسي معين. فمثلاً إذا اعتمدنا في قياس كثير من المفاهيم السيكلوجية مثل الذكاء، والميول، والاتجاهات على مجموعة محددة من المفردات الاختبارية، فإن هذه المفاهيم تصبح قليلة الجدوى.

ولكنا نود التتحقق من إمكانية قياس هذه المفاهيم بمجتمعات مختلفة من المفردات المستمدة من نطاق شامل Items Domain بحيث تكون درجات الأفراد الذين يُخبرون بكل مجموعة منها متطابقة أو متقاربة. لذلك يفضل في هذه الحالة استخدام معامل الاستقرار والتكافؤ Stability and Equivalence.

ويمكن تقدير قيمة هذا المعامل بطريقة مماثلة لطريقة تقدير قيمة معامل التكافؤ، غير أنه يجب أن تكون هناك مدة زمنية أكبر تفصل بين مرتب التطبيق، ونوجد معامل ارتباط بيرسون بين مجموعة الدرجات التي نحصل عليها.

والشكل التخطيطي (٤ - ٧) التالي يوضح ذلك:



شكل (٤ - ٧) يوضح إجراءات تقدير قيمة معامل الاستقرار والتكافؤ

ونظرًا لأن هذه الطريقة تجمع بين الطريقتين السابقتين فإنها تؤدي إلى قيم تقديرية لمعامل الثبات أقل من نظيرتها في أي من الطريقتين السابقتين؛ وذلك لأنها تجمع بين الأخطاء العشوائية التي تؤثر في كل منها.

فمعامل الاستقرار والتكافؤ يعكس الأخطاء العشوائية الناجمة عن اختلاف مفردات صيغتي الاختبار، وكذلك اختلاف الدرجات نتيجة التغيرات التي تحدث للأفراد المختبرين أو تذبذب السمة التي يقيسها الاختبار. ولذلك يعد هذا المعامل أكثر المعاملات تدقيقاً، وقيمه تعد بمثابة الحد الأدنى لتقدير معامل الثبات.

وسوف نقدم فيما يلى مثالاً يوضح كيفية إيجاد القيمة التقديرية لكل من معامل التكافؤ ، ومعامل الاستقرار ، ومعامل التكافؤ والاستقرار :

نفترض أننا طبقنا صيغتين متكافئتين (أ) ، (ب) من اختبار نفسى على مجموعة تتكون من ٢٠ فرداً في أوقات مختلفة ، فالصورة (أ) طبقت في يوم ما تلتها الصورة

(ب) في اليوم نفسه، أما الصورة (أ) فقد أعيد تطبيقها بعد انتهاء أسبوعين، وسوف نرمز لهذه التطبيقات الثلاثة بالرموز: أ_١ ، ب_١ ، أ_٢ على الترتيب ، وفيما يلى الدرجات المناظرة لها :

٢٠ ٢٠ ٢٢ ٢٤ ٢٨ ٣٠ ١٢ ٢٤ ١٤ ٢٠ ١٨ ١٦ ١٦ ١٦	١٩
٢٠ ٢٢ ٢٦ ٢٠ ١٨ ١٨	
٣٢ ١٦ ٢٨ ٢٦ ١٨ ١٤ ٢٠ ١٨ ٢٠ ١٨ ٢٢ ٢٤ ٢٨ ٢٠ ١٤	ب _١
٣٠ ٢٢ ١٨ ١٦ ٢٤ ٢٢ ٣٠ ٢٢ ٢٠ ١٤ ٢٠ ٢٢ ٢٦ ٢٠ ٢٤ ٢٨	٢٩

فإيجاد القيمة التقديرية لمعامل التكافؤ نستخدم نتائج الاختبارين (أ_١ ، ب_١) ونوجد قيمة معامل ارتباط بين درجاتها باستخدام الصيغة (٣ - ١٠)، وهذه تتطلب إيجاد قيم مج س ، مج س^٢ ، مج س ص لكل من الاختبارين كالتالى:

		الاختبار
١		١
ب _١	أ _١	
٤٠٠	٣٩٨	مج س
٨٥٧٦	٨٤٠٤	مج س ^٢
	٨٤٧٢	مج س ص

$$R = \frac{400 \times 398 - 8472 \times 20}{\sqrt{[400 \times 398 - 8404 \times 20] [8576 \times 20 - 8472 \times 20]}}$$

$$= .979$$

ولإيجاد القيمة التقديرية لمعامل الاستقرار نستخدم نتائج الاختبارين أ_١ ، أ_٢ ونكرر ما سبق كالتالى:

الاختبار ١ ٢

٤٥٦	٣٩٨	مجـ س
١٠٨٥٦	٨٤٠٤	مجـ س ٢
٩٠٠		مجـ س ص

$$\frac{456 \times 398 - 900 \times 20}{[456 \times 20 - (456 - 10856) \times 8404]} = r \\ .903 = r$$

ولإيجاد القيمة التقديرية لعامل الاستقرار والتكافؤ نستخدم نتائج الاختبارين بـ ١ و ٢ ونكرر الخطوات نفسها كالتالي :

الاختبار ٢ ١ ب

٤٥٦	٤٠٠	مجـ س
١٠٨٥٦	٨٥٧٦	مجـ س ٢
٩٥٨٨		مجـ س ص

$$\frac{456 \times 400 - 9588 \times 20}{[456 \times 20 - (456 - 10856) \times 8576]} = r \\ .909 = r$$

من هذا يتضح أن ترتيب القيم التقديرية لمعامل الثبات هو : ٩٦٩ ، ٠ ، ٩٠٩ ، ٠ ، ٩٠٣ ، ٠ . حيث يلاحظ أن قيمة معامل التكافؤ ٩٦٩ ، ٠ تعد أعلى القيم الثلاث نظراً لتكافؤ درجات الاختبارين (أ) ، (ب) ، فمتوسط درجات كل منهما: ٢٤٦ ، ٢٠ ، ١٩ ، ٩ ، والانحراف المعياري للدرجات ٢،٦٨ ، ٢،٦٨ على الترتيب.

أى أن هناك تقارباً كبيراً بين متوسطي درجات كل منها وانحرافهما المعياري وبذلك قلت الأخطاء العشوائية التي تعزى إلى اختلاف عينة مفردات كل من الاختبارين ، حيث إن نسبة التباين الحقيقي (٩٦٩ ، ٠) ، ٩٤٪ تقريباً مما يدل على توازي الاختبارين .

كما يلاحظ أن قيمة كل من معامل الاستقرار ، ومعامل الاستقرار والتكافؤ ٩٠٣ ، ٠ ، ٩٠٩ ، ٠ على الترتيب ، وهما قيمتان متقاربتان مما يدل على اتساق درجات الاختبار (أ) في مرتب التطبيق ، وكذلك تكافؤ درجات الاختبار (أ) والاختبار (ب) مما يدل على توازيهما كما سبق أن أشرنا .

وهذا أيضاً يدل على أن السمة التي تقيسها كل من صيغتي الاختبار (أ) مستقرة نسبياً خلال المدة الزمنية الفاصلة على الأقل . ونظراً لاختلاف متوسطي كل من الاختبارين (أ) ، (ب) حيث إن هذين المتوسطين ١٩،٩ ، ٢٢،٨ على الترتيب ، فإن هذا يدل على أن الأخطاء العشوائية الناجمة عن تأثير المران والتدريب والتى ظهرت فى تحسن الدرجة فى الاختبار (أ) (أى بعد إعادة تطبيقه) قد انعكست فى معامل الاستقرار حيث بلغت قيمته التقديرية ٩٠٣ ، ٠ ، ومع هذا فإن نسبة التباين الحقيقي فى الدرجات (٩٠٣) ، ٠ ، ٩٠٣٪ تقريباً وهى نسبة مرتفعة، حيث إن نسبة تباين الخطأ فى الدرجات تبلغ ١٨٪ فقط .

معامل الاتساق الداخلى:

تطلب الطرق الثلاث السابقة تطبيق اختبار أو صيغة متكافئة منه مرتين بفواصل زمنية متباينة ، غير أنه فى بعض الأحيان يصعب بناء صيغتين متكاففتين أو الحصول عليهما ، أو قد يصعب تطبيق الاختبار مرتين ، فعندئذ يفضل تقدير ثبات باستخدام اختبار واحد وتطبيقه مرة واحدة .

فالتعلم لا يمكنه إعادة تطبيق الاختبار التحصيلي نفسه مرتين متتاليتين لتقويم طلابه ، كما لا يمكنه بناء صيغة متعددة من الاختبار ، ولذلك يمكنه تقدير ثبات الاختبار باستخدام معامل الاتساق الداخلى Internal Consistency .

وهذه الطريقة تماثل طريقة الصيغتين المتكاففتين ، غير أنها لا تهدف لتقسيم تكافؤ الصيغتين نظراً لأنها تعتمد على تطبيق اختبار واحد ثم تجزئه إلى نصفين متكاففين ،

وإيجاد معامل ارتباط يرسوسون بين درجات كل من النصفين بعد الانتهاء من تطبيق الاختبار، ولذلك فهي تهتم بتقييم الاتساق الداخلى لمفردات الاختبار.

ويتم تقدير درجات كل من نصفى الاختبار كما لو كان كل منها اختباراً منفصلاً. غير أن المشكلة الأساسية تتعلق بكيفية تجزئة الاختبار إلى نصفين متكافئين أو متوازيين وبخاصة إذا كان الاختبار يشتمل على مفردات غير متجانسة في محتواها أو يقيس سمة مركبة أو عدلة سمات، مثل بطاريات اختبارات الاستعداد العقلى أو المهارات الأساسية مثلاً، وكذلك إذا كان الاختبار موقوتاً، أو تباين مفرداته في درجة صعوبتها أو متزايدة الصعوبة، أو متداخلة فيما بينها.

لذلك فإن هذه الطريقة تتطلب الدقة والحرص عند تجزئة الاختبار إلى نصفين متكافئين ، حيث إنه يمكن إجراء هذا التقسيم بطرق متعددة.

ومع هذا فإنه إذا كان الاختبار غير متجانس في محتواه، فإنه يمكن تجميع المفردات المتجانسة معًا أو تقسيمه إلى اختبارات فرعية متجانسة في محتواها وتجزئتها كل اختبار فرعى إلى نصفين متكافئين ، وضم أجزاء كل نصف على حدة.

أما إذا كانت مفردات الاختبار متزايدة الصعوبة فإنه يمكن تجزئة الاختبار إلى نصفين متكافئين يشتمل أحدهما على المفردات الفردية والآخر على المفردات الزوجية .

ولكن من الضروري في جميع الحالات التأكد من تشابه مضمون المفردات التي تكون النصفين وفق خطة منتظمة، وكذلك التتحقق من تساوى متوسط درجات كل من النصفين والانحراف المعيارى.

كذلك لا يجوز استخدام هذه الطريقة إذا كان الاختبار من نوع اختبارات السرعة Speed Tests حيث يتطلب أن يجيب الفرد عن عدد كبير من المفردات السهلة نسبياً في زمن محدد، مثل اختبار القدرة العددية الذى يشتمل على عدد كبير من المفردات التي تتطلب عمليات جمع بسيط في زمن وجيز. واختلاف الأفراد في معدل سرعة إجابتهم يجعل تجزئة الاختبار إلى نصفين متكافئين غير مناسب بل وغير مطلوب. لذلك يفضل استخدام معامل الاتساق الداخلى في تقدير ثبات اختبارات القوة Power Tests.

ولكن إذا كان من الضروري تقدير ثبات درجات اختبار موقوت،أى يعتمد على السرعة فإنه يمكن إجراء ذلك بإيجاد معامل ارتباط بين اختبارين لكل منها زمنه المحدد ويقيسان السمة نفسها بدلاً من تجزئة الاختبار إلى نصفين متكافئين.

ويقترح ستانلى Stanley فيما يتعلق باختبارات القدرات الموقوتة التى لا تهدف أساساً إلى جعل سرعة الأفراد المستجيبين تؤثر في درجاتهم تأثيراً كبيراً ترتيب مفردات

الاختبار من حيث درجة صعوبتها أي من السهل إلى الصعب، والسماح بوقت كاف يمكن جميع الأفراد من محاولة إجابة كل مفردة قدر استطاعته بالترتيب التي ترد به في الاختبار إلى نصفين متكافئين بالنسبة لزمن الإجابة وليس بالنسبة للمفردات، أي إعطاء زمن محدد لكل من نصفي الاختبار.

ومن المشكلات الأخرى لهذه الطريقة أن الدرجات تعتمد على نصف عدد مفردات الاختبار. ونظرًا لأن القيمة التقديرية لمعامل الثبات تتأثر بطول الاختبار أي عدد مفرداته، فإن قيمة معامل الاتساق الداخلي الناتجة عن إيجاد معامل ارتباط بيرسون بين نصفي الاختبار تكون أقل مما لو استخدم العدد الكلى لمفردات الاختبار، أي أنها تدل على ثبات درجات نصف الاختبار.

ولتقدير ثبات الاختبار ككل ينبغي افتراض إطالة الاختبار إلى الضعف ، وإجراء تعديل على قيمة معامل ارتباط بيرسون التي حصلنا عليها في ضوء هذه الإطالة.

وقد اشتقت كل من سبيرمان وبراون Spearman - Brown صيغة رياضية عامة لتقدير معامل ثبات الاختبار إذا زاد طوله أو نقص (ك) من المرات ، وذلك في إطار مبادئ النظرية الكلاسيكية للختبارات ، وهذه الصيغة هي :

$$\text{معامل الثبات التقديرى } R_{\text{est}} = \frac{k \times r}{1 + (k - 1) r} \quad \dots \dots \dots \quad (4 - 7)$$

حيث r ترمز إلى معامل ثبات درجات الاختبار الأصلى قبل الإطالة.
 r_{est} ترمز إلى معامل الثبات التقديرى لدرجات الاختبار بعد إطالته (ك) من المرات.

وفي حالة التجزئة النصفية ، فإن $k = 2$ ، وتصبح صيغة سبيرمان وبراون كالتالى :

$$r_{\text{est}} = \frac{2r}{1 + r} \quad \dots \dots \dots \quad (4 - 8)$$

حيث r_{est} ترمز إلى تقدير معامل ثبات درجات الاختبار ككل .
 r ترمز إلى قيمة معامل ثبات درجات نصف الاختبار .
وتفترض هذه الصيغة تساوى تباين درجات كل من نصفي الاختبار .

إذا كان معامل الارتباط بين نصف الاختبار = ٠,٨٠ ، فإن قيمة معامل ثبات درجات نصف الاختبار = ٠,٨٠ ، وتصبح القيمة التقديرية لمعامل ثبات درجات الاختبار ككل أى إذا زاد طوله إلى النصف كالتالى :

$$r = \frac{0,80 \times 2}{0,80 + 1} = 0,89$$

ويلاحظ أن هذه القيمة التقديرية لمعامل ثبات درجات الاختبار ككل قد زادت عن القيمة المناظرة لمعامل ثبات درجات نصف الاختبار.

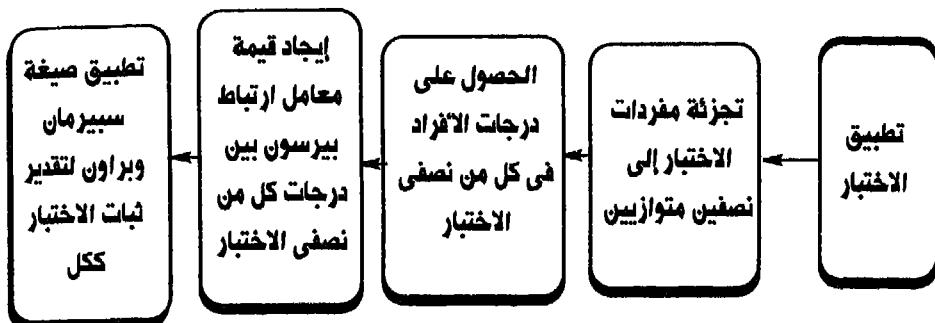
والجدول التالي (٤ - ١) يوضح تأثير استخدام تصحيح سيرمان ويراون في القيمة التقديرية لمعامل الاتساق الداخلي لقيم تتراوح بين ٠,٩٠ ، ٠,٩٥ ، ٠,٨٥ ، ٠,٨٩ ، ٠,٨٢ ، ٠,٧٥ ، ٠,٧٧ ، ٠,٥٧ ، ٠,٤٦ ، ٠,٣٣ ، ٠,٢٠ ، ٠,١٠ ، ٠,٠٨ ، ٠,٠٥ ، ٠,٠٩ ، ٠,١٢ ، ٠,١٥ ، ٠,١٧ ، ٠,١٧ ، ٠,١٦ ، ٠,١٢ ، ٠,٠٨ .

قيم المعامل قبل التصحيح	قيم المعامل بعد التصحيح	مقدار التغيير في قيم المعامل
٠,٩٠	٠,٨٠	٠,٠١
٠,٩٥	٠,٨٩	٠,٠٥
٠,٨٥	٠,٨٢	٠,٠٣
٠,٨٩	٠,٧٥	٠,١٤
٠,٨٢	٠,٧٧	٠,٠٥
٠,٧٥	٠,٥٧	٠,١٨
٠,٧٧	٠,٤٦	٠,٣٣
٠,٥٧	٠,٣٣	٠,١٦
٠,٤٦	٠,٢٠	٠,٠٨
٠,٣٣	٠,١٠	٠,٠٨
٠,٢٠	٠,٠٨	٠,٠٠
٠,١٠	٠,٠٠	٠,٠٠
٠,٠٨	٠,٠٠	٠,٠٠
٠,٠٥	٠,٠٠	٠,٠٠
٠,٠٩	٠,٠٠	٠,٠٠
٠,١٢	٠,٠٠	٠,٠٠
٠,١٥	٠,٠٠	٠,٠٠
٠,١٧	٠,٠٠	٠,٠٠
٠,١٧	٠,٠٠	٠,٠٠
٠,١٦	٠,٠٠	٠,٠٠
٠,١٢	٠,٠٠	٠,٠٠
٠,٠٨	٠,٠٠	٠,٠٠
٠,٠٥	٠,٠٠	٠,٠٠

جدول (٤ - ١) يوضح القيم التقديرية لمعامل الاتساق الداخلي قبل وبعد التصحيح باستخدام صيغة سيرمان ويراون

ويتبين من جدول (٤ - ١) أن مقدار التغيير في قيم معامل الاتساق الداخلي تزداد بعد إجراء تصحيح سيرمان ويراون، غير أن الزيادة تكون أكبر لقيم المعامل الممحضورة بين ٠,٣٠ ، ٠,٦٠ ، أي القيم الوسطى.

ويمكن توضيح إجراءات تدبير قيمة معامل الاتساق الداخلي لاختبار أو مقياس معين باستخدام طريقة التجزئة النصفية Split-Half Method بالشكل التخطيطي (٤-٨) التالي :



شكل (٤ - ٨) يوضح إجراءات تقدير قيمة معامل الانساق الداخلي لل اختبار

وقد توصل فلانجان Flanagan إلى صيغة أخرى تختلف عن صيغة سبيرمان وبراون ولا تتطلب توافر شرط تساوى تباين درجات كل من نصفى الاختبار ، وهذه الصيغة هي :

$$r_s = \frac{1 - \frac{U_1^2 + U_2^2}{2U}}{2} \quad (4 - 9)$$

حيث (r_s) ترمز إلى معامل ثبات الاختبار ككل .

، (U_1^2) ترمز إلى تباين درجات أحد نصفى الاختبار .

، (U_2^2) ترمز إلى تباين درجات النصف الآخر .

، (U_{total}^2) ترمز إلى تباين الدرجات الكلية في الاختبار .

وتطبيق هذه الصيغة يتطلب إيجاد تباين درجات كل من نصفى الاختبار ، وكذلك تباين درجاته الكلية باستخدام الصيغة (٣ - ٥) التي سبق أن أوضحتها فى الفصل الثالث .

والقيم التى نحصل عليها باستخدام هذه الصيغة تقترب من القيم الناتجة عن صيغة سبيرمان وبراون وبخاصة إذا تساوى تباين كل من نصفى الاختبار أى :

$$U_1^2 = U_2^2$$

ولتوضيح كيفية استخدام هذه الصيغة فى إيجاد القيمة التقديرية لدرجات اختبار نفترض أننا قمنا بتجزئه إلى نصفين أحدهما يشتمل على المفردات الفردية والأخر يشتمل على المفردات الزوجية . ثم حسبنا قيمة تباين درجات كل من نصفى الاختبار

(البيان يساوى مربع الانحراف المعياري) ودرجات الاختبار ككل فوجدنا أنها تساوى ٢,١٦ ، ١,٨٥ ، ١,١٨ على الترتيب. وبالتعويض عن هذه القيم في صيغة فلانجان (٤ - ٩) نجد أن :

$$\frac{1,85 + 2,16}{7,18} - 1] 2 = \text{رس}$$

$$[\frac{4,1}{7,18} - 1] 2 =$$

$$= ., 882$$

ويمكن أن نحصل على هذه القيمة نفسها بإيجاد معامل الارتباط بين درجات نصفى الاختبار وتصحيح القيمة الناتجة باستخدام صيغة سيرمان وبراؤن كما سبق.

ومن مزايا الصيغة العامة لسيرمان وبراؤن أنه يمكن باستخدامها تقدير طول الاختبار أي عدد مفرداته اللازم لجعل ثبات درجاته قيمة معينة إذا علمنا قيمة معامل ثبات درجات الاختبار الأصلي.

فمثلاً إذا اشتمل اختبار على ١٥ مفردة، وقيمة معامل ثبات درجاته = ٧٠،٧٠ فإنه يمكن تقدير طول الاختبار وعدد مفرداته التي تجعل قيمة الثبات ٨٠،٨٠ وذلك بتطبيق الصيغة العامة (٧-٤) والتعويض فيها عن قيمة $\rho = 80,80$ ، $R = 70,70$ ، $k = 15$ ، وحل المعادلة الناتجة لإيجاد قيمة (ك) كالتالي:

$$\frac{k \times 0,50}{(k - 1)(0,50) + 1} = 0,80$$

$$\frac{0,50 \times k}{0,50 + k} = 0,80$$

$$0,40 \times k + 0,40 = 0,50 \times k$$

$$0,10 \times k = 0,10$$

$$k = 4$$

لهذا فإنه ينبغي إطالة الاختبار إلى أربعة أمثال طوله الأصلي ، أي أن عدد المفردات يجب أن يزداد بحيث يصبح $4 \times 10 = 60$ مفردة.

ولكن ينبغي ملاحظة أن هذه الصيغة العامة تتطلب أن تكون المفردات التي ستضاف إلى الاختبار الأصلي مستمدّة من نطاق المفردات الذي تمثله مفردات هذا الاختبار ، كما ينبغي مراعاة الواقعية في اختيار القيمة المتوقعة (ك) لإطالة الاختبار من حيث الزمن اللازم للإجابة ، وطبيعة عينة الأفراد المختبرين.

وتفسير القيم التقديرية لمعامل الاتساق الداخلي يماثل تفسير قيم معامل التكافؤ ، حيث إن كل من نصفى الاختبار يطبقان في آن واحد كما هو الحال في الصيغتين المتكافئتين للاختبار .

والأخطاء العشوائية التي تدخل في تقدير معامل الاتساق الداخلي تتعلق جزئياً بأخطاء معاينة محتوى كل من نصفى الاختبار ، ولكن الخطأ الناجم عن عدم اتساق الدرجات عبر مدة زمنية فاصلة ليس له أي تأثير نظراً لأن نصفى الاختبار يطبقان معاً . ولذلك فإن هذه الطريقة تؤدي عادة إلى قيم تقديرية لمعامل الثبات أكبر من حقيقتها .

معامل التجانس لكيودر وريتشاردسون :

ذكرنا فيما سبق أن مشكلة تقدير قيمة معامل الاتساق الداخلي للاختبار بطريقة التجزئة النصفية تتعلق بكيفية تجزئة الاختبار إلى نصفين متكافئين ، فانختلف طريقة التجزئة يؤدي إلى اختلاف في القيمة التقديرية لمعامل الثبات ، كما أن هذه الطريقة أيضاً تتطلب تحقيق شرط تساوى تباين درجات كل من نصفى الاختبار . غير أن هناك طرقاً أخرى هدفت لمعالجة أوجه قصور طريقة التجزئة النصفية ، ومن بين هذه الطرق شائعة الاستخدام طريقة كيودر وريتشاردسون Kuder - Richardson ، وطريقة كرونباك Cronbach .

وتهدف طريقة كيودر وريتشاردسون Kuder - Richardson للتوصيل إلى قيمة تقديرية لمعامل ثبات الاختبارات غير الموقوتة أي اختبارات القوة Power Tests ، والتي تكون درجات مفرداتها ثنائية أي إما واحد صحيح أو صفر ، مثل مفردات الاختيار من متعدد ، أو مفردات الصواب والمخطأ .

وتتغلب هذه الطريقة على مشكلة تعدد طرق التجزئة النصفية للاختبار بأن توجد قيمة سقدرية لمتوسط قيم معامل الارتباط بين كل من نصفى الاختبار لجميع طرق التجزئة الممكنة دون أن تقوم بهذه التجزئة فعلاً . فعلى الرغم من أن جميع الطرق السابقة استخدمت الأساليب الإحصائية في تقدير قيمة معامل الثبات إلا أن طريقة كيودر وريتشاردسون تعد اسلوباً إحصائياً في ذاتها ، فتقدير متوسط قيم معامل

الارتباط بين كل من نصفى الاختبار اعتمد على المبادئ الإحصائية للنظرية الكلاسيكية لل اختبارات .

و تؤكد هذه الطريقة العلاقات القائمة بين المفردات التي يشتمل عليها الاختبار أكثر من تأكيدها اتساق درجات الاختبار بمرور الزمن أو تكافؤها على الرغم من اختلاف صيغ الاختبار .

و تعد طريقة التجزئة النصفية التي سبق ذكرها حالة خاصة من هذه الطريقة . فالاتساق الداخلى لمفردات الاختبار يعني درجة الارتباط الموجب بين هذه المفردات . ولذلك يفضل كيودر وريتشاردسون تسمية ذلك تجانس المفردات Item Homogeneity ، أي اتساق الأداء عبر جميع مفردات الاختبار .

ولذلك يرى براون Brown أن تجانس الاختبار يمكننا من التنبؤ بأداء الفرد في مفردات الاختبار إذا علمنا أداءه في إحدى هذه المفردات ، وهذا يتطلب أن يقيس الاختبار سمة أحادية البعد Unidimensional ، أي تشتمل على متغير واحد .

وتجانس مفردات الاختبار يعد شرطاً أساسياً للاختبار الذي يقيس سمة أحادية البعد . ولذلك يُنصح القائم ببناء أداة قياس أن يراعى عند استخدام طريقة كيودر وريتشاردسون هذا الشرط إلى جانب الشروط الأخرى التي ذكرناها . فإذا اشتملت أداة القياس على اختبارات فرعية غير متجانسة فإنه يمكن تقدير ثبات درجات كل منها على حدة . ولذلك فإن معامل الثبات الذي نحصل عليه باستخدام هذه الطريقة يسمى معامل التجانس Homogeneity Coefficient .

ونظراً لعدد الحالات المتعلقة بالخصائص الإحصائية لمفردات الاختبارات التي اهتم كيودر وريتشاردسون بدراستها ، فقد تعددت الصيغ التي توصلنا إليها . غير أن أكثر هذه الصيغ شيوعاً واستخداماً في تقدير قيمة معامل تجانس درجات مفردات الاختبارات الصيغة ٢٠ Kuder - Richardson Formula 20 (KR - 20) ، والصيغة ٢١ (KR - 21) Kuder - Richardson Formula 21 ، ولذلك سوف نقتصر على توضيح كل من هاتين الصيغتين .

الصيغة (٢٠) لـ كيوهدر وريتشارتسون :

تستخدم الصيغة (٢٠) إذا كانت درجات المفردات ثنائية (صفر ، ١) فقط، وتوافرت الشروط سالفة الذكر، وهذه الصيغة كما يلى :

$$\text{معامل التجانس} = \frac{n}{n-1} \times \frac{\sum_{i=1}^n \text{مجس}_i}{\sum_{i=1}^n}$$

حيث (ن) ترمز إلى عدد مفردات الاختبار.

، (ع^٢) ترمز إلى تباين الدرجات الكلية في الاختبار (مربع الانحراف المعياري).
، (س) ترمز إلى نسبة عدد الأفراد الذين أجابوا عن أي مفردة إجابة صحيحة (درجة صعوبة المفردة).

، (ص) ترمز إلى نسبة عدد الأفراد الذين أجابوا عن أي مفردة إجابة خطأ.
أى أن ص = ١ - س ، (مجس ص) ترمز إلى مجموع تباين درجات
مفردات الاختبار.

وبالطبع لن نناقش البرهان الرياضي لهذه الصيغة حيث إنه يتطلب مراجعة لمبادئ وفرضيات النظرية الكلاسيكية لل اختبارات ، ولكن ما يهم القائمين ببناء الاختبارات والمقياسes ومستخدميها هو كيفية استخدام هذه الصيغة استخداماً صحيحاً مع تحقق الشروط التي تتطلبها.

وإذا نظرنا إلى الصيغة (٤ - ١٠) نجد أن كسر النسبة الثانية

$$\frac{\sum_{i=1}^n \text{مجس}_i}{\sum_{i=1}^n} = \frac{\text{مجس}_1 + \text{مجس}_2 + \dots + \text{مجس}_n}{\text{مجس}_1 + \text{مجس}_2 + \dots + \text{مجس}_n}$$

فإذا كانت مجس ص = ع^٢ ، أى أن مجموع تباين درجات مفردات الاختبار يساوى تباين الدرجات الكلية في الاختبار ، فإن هذا الكسر يصبح مساوياً صفرًا ، وبذلك تندم القيمة التقديرية لمعامل التجانس .

ولذلك إذا أردنا أن تكون قيمة معامل ثبات الاختبار مرتفعة ينبغي أن تكون (مجس ص) أكبر من (ع^٢) ، أى ينبغي أن يكون تباين الدرجات الكلية في الاختبار أكبر من مجموع تباين درجات مفردات الاختبار. وهذا لا يتأتى إلا إذا كان هناك ارتباط مرفق بين المفردات ، أى تقيس جميعها سمة واحدة ، وهو أحد الشروط الأساسية لهذا المعامل .

ولتوسيع ذلك نفترض أن لدينا مقياساً يشتمل على ٢٠ عبارة تتعلق بسمة الخجل ، وتنطلب هذه العبارة من الفرد المستجيب أن يذكر ما إذا كانت العبارة تنطبق عليه (يحصل على الدرجة واحد صحيح) ، أم لا تنطبق (يحصل على الدرجة صفر) وطبق المقياس على ٤ فرداً.

فإيجاد القيمة التقديرية لمعامل التجانس لهذا المقياس باستخدام الصيغة (٢٠) لكيدر وريتشاردسون ، فإنه ينبغي أولاً التتحقق من توافر شروط هذه الصيغة.

ونظراً لأن جميع العبارات تقيس سمة يفترض أنها أحادية البعد، والمقياس لا يعتمد على السرعة ، ومفرداته ثنائية الدرجة ، فإننا نستطيع استخدام هذه الصيغة وفقاً للخطوات التالية :

الخطوة الأولى : تكون مصفوفة درجات المفردات لمجموعة الأفراد ، حيث تحتوي خلايا هذه المصفوفة على الدرجات صفر ، ١ . فإذا وجد الفرد الأول مثلاً أن العبارة الأولى لا تنطبق عليه ، فإنه يحصل على الدرجة صفر ، وتدون في الخلية المقابلة للفرد الأول والمفردة الأولى في المصفوفة ، وهكذا في بقية المفردات والأفراد.

الخطوة الثانية : نجمع الدرجات الكلية لكل فرد على حدة ، وكذلك المجموع الكلي للدرجات الأفراد ، ونحسب قيمة كل من المتوسط والانحراف المعياري لهذه الدرجات الكلية ، فيكون تباين هذه الدرجات (ع^٢) .

الخطوة الثالثة : نوجد قيمة (س) وذلك بقسمة عدد الأفراد الذين حصلوا على الدرجة واحد صحيح في كل عبارة من عبارات المقياس على العدد الكلي لأفراد المجموعة .

الخطوة الرابعة : نوجد قيمة (ص) وذلك بطرح قيمة (س) التي حصلنا عليها في الخطوة الثالثة من الواحد الصحيح .

الخطوة الخامسة : نوجد حاصل ضرب قيم (س) ، (ص) المتقابلة ، ونجمع نواتج حاصل الضرب لجميع العبارات لحصول على مجـ (س ص) .

الخطوة السادسة : نعرض عن قيمة كل من (ع^٢) (مجـ س ص)، (ن) في الصيغة (٤ - ١٠) السابقة للحصول على القيمة التقديرية لمعامل التجانس .

فمثلاً إذا وجدنا أن ع^٢ = ٩ ، مجـ س ص = ١,٤

$$\frac{1,4 - ٣,٩}{٣,٩} \times \frac{٢٠}{١ - ٢٠} = \text{فإن معامل التجانس} =$$

$$= \frac{٢,٥}{٣,٩} \times \frac{٢٠}{١٩} =$$

وتفسير هذه القيمة لا يختلف عن تفسير القيم التقديرية لمعامل الثبات.

الصيغة (٢١) لـ **لـكـيـوـدـ وـرـيـتـشـارـدـ سـوـجـ** :

يتطلب استخدام الصيغة (٢١) توافر شرط آخر إلى جانب الشروط الأخرى التي تتطلبها الصيغة (٢٠)، وهذا الشرط هو تساوى جميع مفردات الاختبارات والمقاييس في درجة صعوبتها أو على الأقل يتراوح متوسط درجة صعوبة جميع المفردات .، ولكن نظراً لصعوبة تحقق هذا الشرط في كثير من الاختبارات أو المقاييس ، فإن الصيغة (٢٠) ربما تكون أكثر ملاءمة لهؤلاء الاختبارات.

وهذه الصيغة مشتقة جبرياً من الصيغة (٢٠)، وهي كالتالي:

$$\text{معامل التجانس} = \frac{\frac{n}{n-s}}{\frac{n-1}{n-1} - \frac{s(n-s)}{n \times ٤^2}}$$

حيث (س) ترمز إلى متوسط الدرجات الكلية في الاختبار.

أما بقية الرموز فقد سبق تعريفها في الصيغة (٢٠).

ويتطلب استخدام هذه الصيغة إيجاد المتوسط والانحراف المعياري للدرجات الكلية فقط، والتعریض في الصيغة (١١-٤).

ومن الجدير بالذكر أن الصيغة (٢٠) هي الصيغة الأكثر شيوعاً واستخداماً في القياس التربوى والنفسي؛ وذلك لأن الصيغة (٢١) تعطى قيمة تقديرية لمعامل التجانس أقل من حقيقتها، وبخاصة إذا استُخدمت دون مراعاة للشروط التي تتطلبها، كما أن الصيغة (٢٠) تتميز بسهولة عملياتها الحسابية.

وعموماً، فإن معامل التجانس يتعلق بتقييم مدى قياس المفردات المختلفة التي يشتمل عليها الاختبار لسعة أحاديثه البعض. فإذا كان الاختبار يقيس أكثر من سمة واحدة، فإن معامل التجانس يقلل من القيمة التقديرية لثبات درجات الاختبار.

معامل (α) الكرونباخ :

يتضمن مما سبق أن كلًا من صيغتي كيدور وريتشاردسون تستخدم في الاختبارات التي تشتمل على مفردات ثنائية الدرجة، غير أنه إذا كان الاهتمام منصبًا على بناء مقاييس متدرج الميزان مثل موازين التقدير، أو استبيانات قياس الاتجاهات واستطلاع الرأي أو بعض مقاييس الشخصية التي يستجيب الفرد لعبارات المقاييس على ميزان ثالثي أو خماسي التدريج مثل (موافق جداً = ٥ ، موافق = ٤ ، غير متأكد = ٣ ، غير موافق = ٢ ، غير موافق على الإطلاق = ١) فهنا لا نستطيع اعتبار إحدى الاستجابات صحيحة والأخرى خطأ وإنما تقع الاستجابات على متصل يتراوح بين موافق جداً وغير موافق على الإطلاق.

وكذلك في اختبارات التحصيل التي تشتمل على أسئلة مقال يتم تقييم درجاتها بحسب نوعية الإجابة التي ربما تكون من مستويات متعددة، فعندها لا تصلح أي من صيغتي كيدور وريتشاردسون في تقييم ثبات درجات مثل هذه المقاييس والاختبارات.

وقد تمكنت Cronbach من اشتقاء صيغة عامة من الصيغة (٢٠) السابقة لتقييم ثبات درجات أنواع الاختبارات والمقاييس المختلفة. وتؤدي هذه الطريقة إلى معامل اتساق داخلي لبنية الاختبار أو المقاييس ويسمى أيضًا معامل التجانس ، غير أنه أطلق على هذه الصيغة معامل (α) ، وهي كالتالي :

$$\text{معامل } (\alpha) = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \text{مفردات}}{\sum_{i=1}^n \text{مفردات}} \right]$$

حيث (Σ) ترمز إلى تباعين درجات كل مفردة من مفردات الاختبار.

(Σ مفردات) ترمز إلى مجموع تباعين درجات جميع المفردات.

(ن) ترمز إلى العدد الكلى لمفردات الاختبار.

والصيغة (٢٠) تختلف عن الصيغة (٤ - ١٠) فى بسط الكسر الثانى فقط حيث حلت (Σ مفردات) محل (Σ مفردات)، وكلاهما يرمز إلى تباعين الدرجات الكلية للاختبار بصيغتين مختلفتين .

ومن الجدير بالذكر أن قيمة معامل (α) تساوى متوسط القيم التقديرية لمعامل ثبات كل من نصفى الاختبار لجميع طرق التجزئة النصفية الممكنة، كما هو الحال فى معامل التجانس لكيودر وريتشاردسون . وبعبارة أخرى فإن معامل (α) يعطى قيمة تقديرية للارتباط بين عينتين عشوائيتين من المفردات المستمدة من نطاق شامل للمفردات التى تمثل مفردات كل من العينتين.

وقد وجد كرونباك Cronbach أن هذا المعامل يعد مؤشراً للتكافؤ أى يعطى قيمة تقديرية جيدة لمعامل التكافؤ، إلى جانب الاساق الداخلى أو التجانس. ومن الملاحظ أن معامل (α) ومعامل التجانس لكيودر وريتشاردسون يتاثران بطول الاختبار أى عدد مفرداته ومدى تجانس هذه المفردات. ويعطى معامل (α) الحد الأدنى للقيمة التقديرية لمعامل ثبات درجات الاختبارات ، أى أن قيمة معامل الثبات بعامة لا تقل عن قيمة معامل (α)، فإذا كانت قيمة معامل (α) مرتفعة ، فإن هذا يدل بالفعل على ثبات درجات الاختبار، أما إذا كانت منخفضة فربما يدل ذلك على أن الثبات يمكن أن تكون قيمته أكبر من ذلك باستخدام طرق أخرى.

وللتوسيع كيفية تطبيق معامل (α) عندما يكون المقياس متدرج الميزان ، نفترض أن لدينا مقياساً للاتجاهات يشتمل على ٦ عبارات على ميزان ثلاثي التدرج (موافق = ٣ ، محاید = ٢ ، غير موافق = ١) طبق على خمسة أفراد ، وكانت درجاتهم كالتالى (جدول ٤ - ٢) :

	٦	٥	٤	٣	٢	١	الفرد
١٢	١	٣	٢	٣	١	٢	١
١٤	٣	٢	٣	٣	٢	١	٢
١١	٢	١	٣	٢	١	٢	٣
١١	١	١	١	٣	٣	٢	٤
١٣	٢	٣	٢	٣	٢	١	٥
٦١	٩	١٠	١١	١٤	٩	٨	المجموع
١٢,٢٠	١,٨٠	٢,٠٠	٢,٢٠	٢,٨٠	١,٨٠	١,٧٠	المتوسط س
٦,٨٠	٠,٥٦	٠,٨٠	٠,٥٦	٠,١٦	٠,٥٦	٠,٢٤	التبان ع

جدول (٤ - ٢) يوضح توزيع درجات خمسة أفراد في مقياس للاتجاهات

الخطوة الأولى : نوجد قيمة تباين درجات كل عبارة، وذلك بطرح الدرجة من متوسط درجات العبارة، وتربيع الناتج وقسمته على عدد الأفراد (٥) ، وذلك بعد تكوين الجدول المبين .

الخطوة الثانية : نوجد قيمة تباين الدرجات الكلية في المقاييس وذلك بطرح الدرجة الكلية لكل فرد من المتوسط العام للدرجات، وتربيع الناتج وقسمته على عدد الأفراد .

الخطوة الثالثة : نطبق صيغة معامل (α) لكرونباك (٤ - ١٢) كالتالي :

$$\text{القيمة التقديرية لمعامل } (\alpha) =$$

$$\frac{6}{[\frac{2,4 - 1]{\frac{6}{1-6}}]} = \frac{2,88}{6,8} = \frac{1}{5}$$

$$= 0,58 \times 1,2$$

= ٠,٧٠ . ويلاحظ أنها قيمة متوسطة .

ثبات تقديرات المحكمين :

يحتاج كثير من الباحثين إلى مؤشر ثبات تقديرات المحكمين وبخاصة فيما يتعلق بأساليب قياس الشخصية الإنسانية التي تعتمد على الملاحظة أو المقابلة الشخصية أو موارين التقدير أو الطرق الإسقاطية التي سوف نوضحها في الفصل الثالث عشر وكذلك في قياس التحصيل الدراسي باستخدام اختبارات تشتمل على أسئلة المقال حيث تعتمد جميع هذه الأساليب على تقديرات مجموعة من المحكمين .

وهنا يبرز التساؤل حول مدى الثقة في هذه التقديرات، أي ما مقدار الخطأ الذي يعزى إلى فردية المحكم أو المحكمين، و يؤثر في الدرجات التي يقدرونها للأفراد .

ولعل أبسط الطرق هو إيجاد قيمة معامل الارتباط بين تقديرى محكمين مختلفين فمثلاً يقوم أحد المحكمين بتقدير درجة سؤال مقال معين أو يقوم بمشاهدة فرد في موقف معين، ويقوم محكم مستقل آخر بتقدير درجة السؤال نفسه أو يقوم بمشاهدة

الفرد نفسه في موقف آخر في ظروف مشابهة، فعندئذ تدل قيمة معامل الارتباط على مدى اتساق تقدير كل منهما أو اتفاقيهما.

أما إذا أردنا معرفة مدى اتساق متوسط تقديرات اثنين من المحكمين لمجموعة من الأفراد، فإنه يمكن إيجاد قيمة معامل الارتباط بين تقديرات كل منهما، وتطبيق صيغة سبيرمان وبراؤن (٤ - ٨) التي ذكرناها فيما سبق، حيث نعرض عن ر_{٢١} بقيمة معامل الارتباط لكي نحصل على القيمة التقديرية لثبات متوسط التقديرات (١١).

وفي حالة تعدد المحكمين يمكن إيجاد قيم معامل الارتباط بين تقديراتهم لمجموعة الأفراد واستخدام الصيغة العامة لسبيرمان وبراؤن (٤ - ٧) للحصول على القيمة التقديرية لثبات متوسط تقديراتهم، حيث نعرض عن قيمة (ك) في هذه الصيغة بعد المحكمين.

ويمكن أيضاً تقدير الثبات عن طريق تحديد النسبة المئوية للاتفاق بين المحكمين أو القائمين بال مقابلة الشخصية إذا كان الهدف مقصوراً على تصنيف الأفراد إلى مجموعتين مثل راسب أو ناجح، متمنك أو غير متمنك، منطوى أو منبسط ، وهكذا.

بعض الطرق الأخرى لتقدير الثبات :

هناك بعض الطرق الأخرى التي تستخدم في تحديد تجانس مفردات الاختبار، وتعتمد على أساليب إحصائية متقدمة مثل التحليل العاملى Factor Analysis ، وتحليل التباين Analysis of Variance . فالتحليل العاملى هو أسلوب إحصائى لتحديد أقل عدد من العوامل أو التكوينات الفرضية التي تفسر قيم معاملات الارتباط بين مفردات الاختبار ، وبذلك نستطيع باستخدام هذه العوامل تصنيف هذه المفردات في مجموعات متجانسة تقيس كل مجموعة منها سمة معينة. أما إذا بینت نتائج التحليل العاملى وجود عامل مشترك يجمع المفردات فإن هذا يعني أن مفردات الاختبار متجانسة فيما تقيسه .

أما تحليل التباين فهو أسلوب إحصائي آخر يعتمد على تجزئة التباين الكلى للدرجات الاختبار إلى ثلاثة مصادر للتباین ترجع إلى الأفراد والمفردات وتفاعل الأفراد مع المفردات .

ويمكن تقدير تباين الدرجات باستخدام مجموع مربعات انحرافات درجات الأفراد عن المتوسط العام لدرجاتهم ، أما تباين الخطأ فيمكن تقديره باستخدام مجموع

المربعات الخاص بتفاعل الأفراد مع المفردات، وبذلك يمكن تقدير ثبات درجات الاختبار باستخدام الصيغة التالية لتحليل التباين الذى تنسب إلى هويت Hoyt :

$$\text{معامل الثبات} = \frac{\text{التباین المتعلق بالافراد} - \text{التباین المتعلق بتفاعل الافراد مع المفردات}}{\text{التباین المتعلق بالافراد}} \quad (13-4) \dots \dots \dots$$

والقيمة التقديرية لهذا المعامل تدل على مدى تجانس مفردات الاختبار أو اتساقه الداخلى.

ويمكن للقارئ المهتم باستخدام هاتين الطريقتين الرجوع إلى أحد المصادر الإحصائية التى تناولت التحليل العاملى وتحليل التباين، وقد تناول المؤلف هذان الأسلوبان فى مرجع آخر.

ثبات الفرق بين درجتي اختبارين :

يهتم الباحثون ومستخدمو الاختبارات فى بعض الأحيان بمقارنة درجة فرد فى اختبارين أو أكثر ، فمثلاً ربما نود مقارنة درجة فرد فى مقياس المسئولية الاجتماعية بدرجته فى مقياس تقدير الذات ، أو درجه فى اختبار التحصيل فى الرياضيات بدرجته فى اختبار الذكاء ، أو درجه فى اختبارين طبق أحدهما قبل عملية التعليم وطبق الآخر بعد انتهائهما ، فعندئذ يجب أن نتحقق من ثبات الفرق بين كل من الدرجتين لمعرفة تأثير الخطأ العشوائى فى هذا الفرق الذى ربما يؤدي إلى اختلاف الدرجات الملاحظة عن الدرجات الحقيقية . فدرجات كل من الاختبارين تتأثر بعوامل خطأ متباينة ، وتتجمع هذه الأخطاء فى درجة الفرق بين الدرجتين .

ولذلك فإن القيمة التقديرية لمعامل ثبات فرق الدرجتين تكون عادة أقل من قيمة ثبات كل درجة على حدة . وهذه القيمة التقديرية تساوى نسبة تباين الفرق بين الدرجتين الملاحظتين إلى نسبة تباين الفرق بين الدرجتين الحقيقتين ، ويمكن إيجاد هذه القيمة باستخدام الصيغة التالية :

$$\frac{\frac{R}{21} + \frac{R}{11}}{2} - \frac{1}{21} - \frac{1}{11} = \frac{R}{\text{فرق بين الدرجتين}} \quad (14-4) \dots \dots \dots$$

حيث R ترمز إلى معامل ثبات درجات أحد الاختبارين .

، ر ترمز إلى معامل ثبات درجات الاختبار الآخر .

، ر ترمز إلى معامل الارتباط بين درجات الاختبارين .

أى أن قيمة ثبات درجات كل من الاختبارين على حدة ، وكذلك قيمة معامل الارتباط بينهما تؤثر في قيمة ثبات الفرق بين الدرجتين .

فإذا كانت قيمة معامل ثبات درجات كل من الاختبارين ٧٥ ، ٨٠ ، ٠٠ ،

وقيمة معامل الارتباط بينهما ٢٥ ، ٠ ، فباستخدام الصيغة (٤ - ١٤) نجد أن معامل ثبات الفرق بين درجتي الاختبار

$$\frac{. , ٨٠ + . , ٧٥}{. , ٢٥} = \frac{٢}{١ , ٢٥}$$

$$\frac{. , ٧٠}{. , ٧٥} = \frac{. , ٥٢٥}{. , ٧٥} =$$

ويلاحظ أن القيمة الناتجة أقل من قيمة معامل ثبات درجات كل اختبار على حدة . وكلما قل الارتباط بين درجات كل من الاختبارين وزاد ثبات كل منهما تزداد قيمة معامل ثبات الفرق بين الدرجتين ، والعكس يكون صحيحاً .

فإذا كانت قيمة متوسط ثبات درجات كل من الاختبارين مساوية لقيمة معامل الارتباط بينهما ، فإن معامل ثبات الفرق بين درجتيهما يساوى صفرًا ، أما إذا كانت قيمة معامل ثبات كل منهما واحداً صحيحاً فإن قيمة معامل ثبات الفرق بين الدرجتين سوف يكون مساوياً واحداً صحيحاً أيضاً بغض النظر عن قيمة معامل الارتباط بين درجات كل من الاختبارين . ويمكنك إثبات ذلك بالتعريض في الصيغة (٤ - ١٤) عن :

$$R = 1_{\frac{1}{2}} = 1_{\frac{1}{2}} \text{ مهما كانت قيمة } R_{\frac{1}{2}}$$

وعلى الرغم من أنه يمكننا التتحقق من ارتفاع قيمة معامل ثبات درجات الاختبارين قبل اختيارهما لمقارنة درجة فرد في كل منهما ، إلا أننا لا نستطيع أن نقلل قيمة معامل الارتباط بين درجات الاختبارين ، وبخاصة إذا كان الاختباران يقيسان نفس المحتوى ويطبق أحدهما قبل عملية التعليم والآخر بعدها .

فهنا يفترض أن درجات الاختبارين تكونان مرتبتين ارتباطاً مرتفعاً . ولذلك فإن القيمة التقديرية لمعامل ثبات درجات التحسن تكون عادة منخفضة . ولذلك يفضل

استخدام الصيغة (٤ - ١٤) في التحقق من ثبات الفرق بين درجتي اختبارين يقيسان سمتين مختلفتين وكلاهما يتميز بثبات مرتفع نسبياً.

معامل إمكانية التعميم:

يرى كرونباك Cronbach وراجاراتنام Rajaratnam وجليزر Glaser أن درجات الاختبار أو أي درجات ملاحظة تعد عينة من نطاق شامل للدرجات الممكنة. وتبين أن ظروف جمع الملاحظات وفقاً لأبعاد متعددة.

ومثال ذلك تطبيق اختبار معين مرة واحدة، أو تطبيقه وإعادة تطبيقه، وأشتماله على عينة ممثلة من المفردات دون بقية العينات المستمدة من نطاق شامل، وتطبيقه في ظروف معينة دون باقي الظروف.

وهذا التباعين في الأبعاد المتعلقة بأداة القياس وما تشتمل عليه من مفردات، وظروف تطبيقها، وخصائص مجموعة الأفراد التي تطبق عليها تؤدي إلى مشكلات في عمليات القياس وإمكانية تعميم نتائجها. وقد اقترح كرونباك وزملائه عام ١٩٧٢ نظاماً لتقدير تأثير كل من هذه الأبعاد في تباين الدرجات الملاحظة ، وكذلك تقدير دقة الدرجة الحقيقية من الدرجات الملاحظة ، وقد أصبح هذا النظام معروفاً باسم نظرية إمكانية التعميم Generalizability Theory. ونظراً لأن الدرجات الملاحظة تتأثر بجميع هذه الأبعاد، فإن الدرجة الحقيقية تسمى درجة النطاق الشامل Universe Score. والمشكلة الرئيسية في هذه الحالة تتعلق بإمكانية تعميم ما تدل عليه الدرجات الملاحظة في ضوء هذه الأبعاد المتعددة.

وبعبارة أخرى ، فإن معامل الثبات في إطار النظرية الكلاسيكية للاختبارات لا يأخذ بعين الاعتبار تعدد أبعاد عملية القياس، وإنما يركز على أحد هذه الأبعاد. ولكن يكون تعميم القياس ممكناً لابد من مراعاة هذه الأبعاد وانعكاس أثرها في قيمة معامل إمكانية التعميم Generalizability Coefficient .

ونظراً لأهمية هذه النظرية وإسهاماتها في توسيع نطاق استخدامات كل من مفهومي الصدق والثبات وتكاملهما، فسوف نلقى عليها مزيداً من الضوء في الباب الرابع (الفصل الرابع عشر).

تحقيق عام على طرق تقييم الثبات وتفسير نتائجها :

لعل تعدد طرق تقييم ثبات درجات الاختبارات كما أوضحتنا يجعل القارئ يتساءل عن كيفية اختيار إحدى هذه الطرق عند تحقيقه من ثبات درجات الاختبار أو المقياس الذي يُعده أو يستخدمه.

والحقيقة أنه لا توجد إجابة مباشرة أو محددة لهذا التساؤل، إذ إن الاختيار ينبغي أن يستند إلى عوامل متعددة يتعلق بعضها بطبيعة السمة المقاسة، ونوع الاختبار وفيما تستخدمن تنتائجها، وخصائص عينة المختبرين، والبعض الآخر يتعلق بمصادر الأخطاء العشوائية التي يود أن تؤخذ بعين الاعتبار في تقييم معامل الثبات، وكذلك الأخطاء التي يرى أنها متسقة أو منتظمة.

فكما رأينا هناك طرق تعتمد على تطبيق الاختبار مرة واحدة، وطرق أخرى تعتمد على تطبيق الاختبار أو صيغ مكافئة له مرتين، واختيار أي من هذه الطرق يعتمد على إمكانات الباحث أو مستخدم الاختبار من حيث توافر صيغتين متكافتين للاختبار أو تمكنه من إعدادهما، وسهولة تطبيقهما على مجموعة من الأفراد إما في آن واحد أو بفواصل زمني معين.

وهذا يتطلب مراعاة ظروف الأفراد وقوتهم إجراء الاختبار عليهم دون تعب أو ملل، كما يتطلب مراعاة ما إذا كان الاختبار أو كل من صيغتيه من نوع اختبارات السرعة أو اختبارات القوة، وكذلك نوع القرار الذي سيتخد بالنسبة للفرد في ضوء نتائج الاختبار. فإذا كان القرار يتعلق بالفرد فإن قيمة معامل ثبات درجات الاختبار ينبغي أن تكون مرتفعة بدرجة أكبر مما لو كان القرار متعلقاً بمجموعة من الأفراد.

إذ يرى خبراء القياس أنه ينبغي في الحالة الأولى إلا تقل هذه القيمة عن ٨٥٪ . كما في حالة استخدام الاختبارات في مجالات التشخيص الكlinيكي والتصنيف، ولا تقل في الحالة الثانية عن ٦٥٪ ، ومع هذا فإن هاتين القيمتين ليستا مؤكدين، وإنما تعدان بمثابة مرشد لمستخدم الاختبار. فقرارات الانتقاء تختلف عن قرارات التصنيف أو التشخيص في تأثيرها على مستقبل الفرد مما يتطلب توافر معلومات أخرى تستند إليها هذه القرارات إلى جانب المعلومات المتعلقة بثبات درجات الاختبارات والمقياسين.

ولذلك فإن قيم معامل الثبات التي يتم تقاديرها باستخدام أي من هذه الطرق تختلف في معناها ودلالتها وذلك لاختلاف مصادر الأخطاء العشوائية التي ينعكس أثرها في هذه القيم.

فالطريقة التي تعتمد على صيغتين متكاففتين من الاختبار يطبقان في مرتين متباينتين تعد أكثر الطرق مراعاة لمصادر الأخطاء العشوائية الناجمة عن التغيرات التي تحدث للأفراد خلال المدة الزمنية الفاصلة بين مرتب التطبيق، والتغيرات الناجمة عن عدم كفاية تمثيل عينة المفردات للنطاق الشامل الذي تقسيه، وكذلك الأخطاء الناجمة عن ظروف تطبيق الاختبار. وتنعكس جميع هذه الأخطاء في معامل الاستقرار والتكافؤ مما يجعل قيمته التقديرية أقل من قيمة المعاملات الأخرى.

أما طريقة إعادة تطبيق الاختبار نفسه فإنها تراعي فقط الأخطاء العشوائية الناجمة عن التغيرات اليومية التي تطرأ على الأفراد المختبرين، والأخطاء الناجمة عن ظروف التطبيق، مما يجعل القيمة التقديرية لمعامل الاستقرار أقل من قيمة معامل الاستقرار والتكافؤ لعدم تأثر معامل الاستقرار بأخطاء معينة مفردات الاختبار.

أما طرق الأساق الداخلي مثل طريقة التجزئة النصفية، وطريقة كيدور وريتشاردسون، ومعامل (α) فإنها تختلف أيضاً فيما بينها من حيث مصادر الأخطاء العشوائية التي تؤخذ بعين الاعتبار في تقييم معامل الثبات.

فالأخطاء الناجمة عن اختلاف عينة مفردات كل من نصف الاختبار يؤدي إلى انخفاض قيمة معامل الأساق الداخلي، في حين أن هذا المعامل لا يعكس الأخطاء الناجمة عن التغيرات التي تحدث للأفراد أو لظروف التطبيق.

غير أن اعتماد هذه الطرق الثلاث على تطبيق الاختبار مرة واحدة يجعلها مناسبة من الناحية العملية فيما يتعلق بالوقت المخصص لتطبيق الاختبار، وعدم ضرورة توافر صيغ مكافئة للاختبار. ولكن هذه الطرق لا تناسب الاختبارات التي تعتمد على السرعة أو الاختبارات غير متجانسة المحتوى، فكثير من اختبارات الاستعدادات تكون موقوتة وتستخدم في اتخاذ قرارات متعددة بعيدة الأمد في المؤسسات التربوية والمهنية.

لذلك ربما يكون من المناسب استخدام طريقة إعادة تطبيق الاختبار في تقييم ثبات درجات هذه الاختبارات بدلاً من التتحقق من اتساق درجات كل من نصف الاختبار. ولكن إذا كان الاختبار يقيس استعدادات متعددة، فإنه ينبغي تقييم ثبات درجات الاختبارات الفرعية التي يشتمل عليها الاختبار.

وإذا كان من الضروري استخدام طريقة التجزئة النصفية لهذه الاختبارات فيمكن إجراء هذه التجزئة بالنسبة لزمن إجابة مفردات كل من نصف الاختبار بعد ترتيب المفردات بحسب درجة صعوبتها. أما اختبارات التحصيل الدراسي فإنه يستدل من درجات الطلاب في عينة من المفردات التي تشتمل عليها على درجة تمكّنهم من نطاق سلوكي لمعارف ومهارات معينة. لذلك فإنه من المناسب في هذه الحالة تقييم ثبات الدرجات باستخدام أي من طرق الأساق الداخلي. أما إذا اشتمل الاختبار على أسئلة مقال فإن معامل (α) يكون هو المعامل المناسب، والقيم التي تحصل عليها من هذا

المعامل تعد أدنى قيم معاملات الثبات التي يمكن الحصول عليها من المعاملات الأخرى . ونظراً لأننا لا نتخذ قرارات بعيدة الأمد عن تحصيل الطلاب في منهج دراسي معين ، فإن معامل الاستقرار لا يناسب هذه الاختبارات .

في حين أن هذا المعامل يكون أكثر ملاءمة لمقاييس الجوانب الوجدانية والشخصية نظراً لأن السمات غير المعرفية أكثر استقراراً على المدى البعيد .

كما يفضل استخدام معامل (α) إذا كان الهدف تقدير الاتساق الداخلي لمفردات الاختبارات التي تقيس هذه السمات نظراً لأنها تشتمل على موازين متدرجة . وفي جميع الأحوال ينبغي ألا تقتصر أدلة الاختبارات على تقديم مؤشرات الثبات فقط ، وإنما يجب أن تقدم المزيد من المعلومات حول خصائص مجتمع المختبرين ، والعينة أو العينات المستعملة منه وعدد أفرادها ومدى تمثيلها للمجتمع ، وكذلك المتوسط والانحراف المعياري لدرجات عينة الأفراد ، التي استندت إليها قيمة معامل الثبات . كما ينبغي أن تزود مستخدم الاختبار بقيم الأخطاء المعيارية للفياس التي سوف نوضحها بعد قليل .

والخلاصة أنه يمكن ترتيب طرق حساب الثبات ترتيباً تناظرياً بحسب حجم القيم التقديرية التي نحصل عليها باستخدامها ، وهي كالتالي :

التجزئة النصفية ، كيودر وريتشاردسون ، إعادة تطبيق الاختبار ، الصيغ المتكافئة ، معامل (α) .

لذلك ينبغي توقع احتلاف هذه القيم باختلاف الطريقة المستخدمة في موقف معين ولعينة أفراد معينة . فهذه القيم ليست مطلقة ، وإنما هي قيم نسبية حيث إنها مرتبطة بكل من الموقف وطبيعة العينة المستخدمة . كما أن هذه القيم تزودنا بتقدير Estimate لدرجة الاتساق في درجات الاختبار ، ولكنها لا توضح أسباب عدم اتساق الدرجات .

الخطأ المعياري للفياس:

تبين لنا مما سبق أن درجة الفرد الملاحظة في الاختبار تختلف في غالبية الأحيان عن درجته الحقيقة نظراً لتأثير الدرجة الملاحظة بمصادر أخطاء متعددة .

لذلك ذكرنا في مستهل هذا الفصل أن الدرجة الملاحظة تنقسم إلى جزئين هما: الدرجة الحقيقية ودرجة الأخطاء العشوائية . فإذا افترضنا أننا استطعنا تحديد درجة الأخطاء العشوائية التي أثرت في الدرجة الملاحظة لكل فرد من الأفراد المختبرين ، فإنه يمكن إيجاد الانحراف المعياري لدرجات الخطأ .

والقيمة الناتجة تسمى الخطأ المعياري للقياس Standard Error of Measurement، ولكننا لا نستطيع في واقع الأمر إيجاد درجة الخطأ لكل فرد من أفراد المجموعة إلا إذا أعيد تطبيق الاختبار على الفرد نفسه عدداً كبيراً من المرات ، وهو أمر غير ممكن ؛ لذلك فإننا لا نستطيع إيجاد قيمة الخطأ المعياري للقياس ، ولكننا نستطيع تقدير هذه القيمة إذا علمنا قيمة الانحراف المعياري للدرجات الملاحظة، وكذلك قيمة معامل ثبات درجات الاختبار، وذلك باستخدام صيغة رياضية يمكن اشتقاقها مباشرة من الصيغة (٤ - ١). كالتالي:

$$\frac{\sigma^2}{\rho} = \frac{\sigma^2_x - \sigma^2_u}{\sigma^2_x}$$

لكن الكسر الأول يدل على معامل ثبات درجات الاختبار (ρ_{xx}) كما أوضحنا من قبل. أي أن:

$$\rho_{xx} = 1 - \frac{\sigma^2_u}{\sigma^2_x}$$

$$\text{أي أن } \rho_{xx} \times \sigma^2_x = \sigma^2_x - \sigma^2_u$$

$$\sigma^2_u = \sigma^2_x - \rho_{xx} \times \sigma^2_x$$

$$\sigma_u = \sqrt{\sigma^2_x (1 - \rho_{xx})}$$

وباستخراج الجذر التربيعي لكل من الطرفين نجد أن :

$$\sigma_u = \sqrt{\sigma^2_x (1 - \rho_{xx})} \quad (4 - 15)$$

أي أن الخطأ المعياري للقياس = حاصل ضرب الانحراف المعياري للدرجات الملاحظة في الجذر التربيعي لمعامل الثبات مطروحاً منه الواحد الصحيح.

إذا افترضنا أن قيمة معامل ثبات درجات الاختبار (ρ_{xx}) تساوى ٠,٨٠ ، وقيمة الانحراف المعياري للدرجات الملاحظة ٦,٣ ، فإن الخطأ المعياري للقياس يساوى

$$\sqrt{0,80 \times 6,3} = 2,82$$

ويلاحظ أن الخطأ المعياري للقياس يساوى صفرًا، إذا كانت قيمة معامل الثبات متساوية للواحد الصحيح، ولكنه يصبح متساوية للانحراف المعياري للدرجات الملاحظة إذا كانت قيمة معامل الثبات متساوية صفرًا.

وهذا يعني أنه إذا كانت درجات الاختبار غير متسقة على الإطلاق، فإن تشتت الدرجات الملاحظة يعزى إلى محض الصدفة.

ونظرًا لأننا لا نستطيع كما ذكرنا إعادة تطبيق الاختبار على الفرد عدًّا كبيرًا من المرات، فإن الصيغة (٤ - ١٥) تعطى قيمة تقديرية للخطأ المعياري للقياس مستمدة من أفراد المجموعة ككل. غير أن هذه الصيغة يمكن أن تدل أيضًا على الانحراف المعياري للدرجات الملاحظة للفرد عن درجة الحقيقة لو أن الاختبار أعيد تطبيقه عليه عدًّا كبيرًا من المرات نظرًا لتبين درجات الفرد من تطبيق إلى آخر نتيجة عوامل الخطأ، وبذلك تتذبذب درجاته حول درجة الحقيقة. فإذا افترضنا أن الأخطاء التي تؤثر في هذه الدرجات كانت عشوائية، فإنها تعمل أحياناً على زيادة درجة الملاحظة، أى تكون درجة الخطأ موجبة، وأحياناً أخرى تعمل على خفض هذه الدرجة، أى تكون درجة الخطأ سالبة.

وبذلك يكون متوسط درجات الخطأ صفرًا، ودرجة الملاحظة تساوى درجة الحقيقة في الاختبار.

وقد سبق أن أوضحنا في الفصل الثالث أن المتغير العشوائي يخضع للتوزيع الاعتدالي؛ لذلك فإن الدرجات الملاحظة في الاختبار تخضع أيضًا للتوزيع الاعتدالي حول الدرجة الحقيقة وذلك لتأثيرها بالأخطاء العشوائية. ونظرًا لأن درجة الفرد الحقيقة في الاختبار لا تتغير مهما أعددنا تطبيق الاختبار عليه، فإن تباين هذه الدرجة = صفر. وبالعودة إلى الصيغة (٤ - ١) وهى:

$$\text{ع}^2 = \text{ع}^2 + \text{ع}^2$$

والتعریف عن قيمة تباين الدرجة الحقيقة $\text{ع}^2 = \text{صفر}$.

$$\text{فإن } \text{ع}^2 = \text{ع}^2 \quad \dots \quad (4 - 16).$$

أى أن تباين درجات الفرد الملاحظة في التطبيقات المتكررة للاختبار يساوى تباين درجات الخطأ. أما بالنسبة للمجموعة ككل إذا اختبرت مرة واحدة، يكون تباين درجات الخطأ أقل من تباين الدرجات الملاحظة نظرًا لأن درجاتهم الحقيقة سوف

تكون مختلفة على الرغم من أن تباين الدرجات الحقيقة لكل منهم على حدة تساوى صفر.

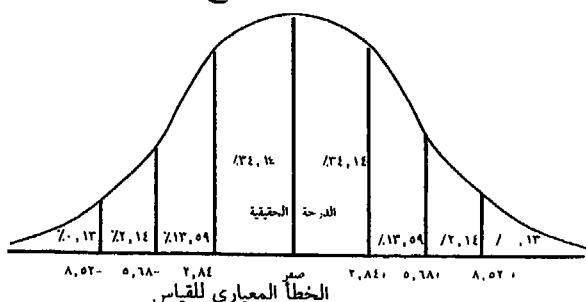
لذلك فإن الخطأ المعياري للقياس يتم تقدير قيمته من درجات مجموعة من المختبرين، مع افتراض أن هذا الخطأ المعياري متساوٍ لكل فرد. غير أن هذا الافتراض يبدو غير واقعٍ وبخاصة إذا كانت مجموعة الأفراد غير متGANسة. ولذلك فإن بعض الاختبارات الجيدة المنشورة تُقدر أحطاء معيارية مختلفة للمجموعات المتباينة من الأفراد سواء في النوع أو العمر أو الصف الدراسي على متصل السمة التي يقيسها الاختبار. كما أن التطورات المعاصرة التي حدثت في نظرية الاختبارات أكدت ضرورة تقدير الخطأ المعياري لكل مستوى معين من مستويات السمة المقاسة، ولعل أهم هذه التطورات ما أصبح معروفاً في أوساط القياس التربوي والنفسى بنظرية الاستجابة للفرد الاختبارية (IRT) Item Response Theory التي سوف نلقى الضوء عليها في الباب الرابع (الفصل الرابع عشر).

والخطأ المعياري للقياس يفيد في تقدير مدى تشتت الدرجات الملاحظة حول الدرجة الحقيقة الافتراضية في اختبار ما، كما يفيد في تقدير قيمة الخطأ العشوائي الذي يؤدي إلى تذبذب الدرجة الملاحظة للفرد حول درجة الحقيقة.

تفسير الخطأ المعياري للقياس :

لتوضيح كيفية تفسير قيمة الخطأ المعياري للقياس نعود للمثال السابق المتعلق بالاختبار الذي قيمة معامل ثبات درجاته .٨٠ ، والانحراف المعياري للدرجات الملاحظة للفرد .٦٣ حيث وجدنا أن القيمة التقديرية للخطأ المعياري للقياس باستخدام الصيغة (٤ - ١٥) متساوية .٢٨٢

فالخطوة الأولى : نرسم المنهجي الاعتدالى ونحدد على المحور الأفقي قيم الخطأ المعياري للقياس ومضاعفاتها كما هو موضح بشكل (٤ - ٩) التالي:



شكل (٤-٩) يوضح المنهجي الاعتدالى لتوزيع درجات الخطأ الذى متواسطها صفر ، وانحرافها

٢,٨٤ المعياري

الخطوة الثانية: نحدد مدى الدرجات Score Bands التي تتحضر بين وحدة خطأ معياري واحد إلى يمين المتوسط وإلى يساره أي $+ 1 \sigma$ ، وكذلك بين وحدتين وثلاث وحدات أي بين $+ 2 \sigma$ ، وبين $+ 3 \sigma$.

وبذلك يكون مدى الدرجات كالتالي : $+ 2,84$ ، $+ 4,68$ ، $+ 8,52$.

الخطوة الثالثة : نحدد النسبة المئوية من المساحة تحت المنحنى الاعتدال المحسورة بين وحدات الخطأ المعياري المذكورة في الخطوة الثانية. وهذه النسب موضحة في شكل (٤ - ٩) فنجد مثلاً أن :

68% تقريباً من المساحة تتحضر بين $+ 1 \sigma$ أي بين $+ 2,84$.

95% تقريباً من المساحة تقع صر بين $+ 2 \sigma$ أي بين $+ 4,68$ وهكذا.

وهذا يعني أن الدرجة التي يحصل عليها الفرد في الاختبار سوف تقع بين $+ 2,84$ حول درجته الحقيقة 68% من المرات تقريباً.

فالدرجة الحقيقة ثابتة، والدرجة الملاحظة تتبدل حولها، فإذا حصل فرد في الاختبار على الدرجة ١٢ مثلاً فإنه يمكننا أن نستدل على أن درجته الحقيقة سوف تقع بين $12 - 2,84$ ، $12 + 2,84$ ، أي بين $9,16$ ، $14,84$ ، وثقتنا في هذا الاستدلال 68% تقريباً.

كذلك يمكننا أن نستدل بدرجة ثقة 95% تقريباً على أن درجته الحقيقة سوف تقع بين $12 - 5,68$ ، $12 + 5,68$ أي بين $6,32$ ، $17,68$.

ويلاحظ أن المدى الذي يحتوي الدرجات الحقيقة يزداد بزيادة درجة ثقتنا، غير أنه يكتفى عادة بالمدى $+ 1 \sigma$ في تفسير القيم التقديرية للخطأ المعياري للقياس. ولكن إذا أردنا مقارنة الدرجات التي يحصل عليها فرد في اختبارات متعددة مثل الاختبارات الفرعية التي تشتمل عليها مقاييس الشخصية أو بسطاريات الاستعداد العقلي والتي تمثل درجاتها بيانياً في صفحة نفسية Profile ، فإنه يفضل في هذه الحالة رسم شريط عريض (مدى) Band حول الدرجة الملاحظة في كل اختبار بحيث تكون سعة $+ 2 \sigma$ فلكي يكون الفرق بين درجتي الفرد في اختبارين مختلفين دالاً يجب أن يكون أكبر من ضعف الخطأ المعياري للقياس لأي من الاختبارين.

ويرجع ذلك إلى أن الخطأ المعياري للفرق بين درجتي الاختبارين أكبر من الخطأ المعياري لكل اختبار على حدة كما أوضحنا فيما سبق. ولكن ينبغي ملاحظة أن إجراء هذه المقارنة يتطلب أن تكون وحدات القياس المستخدمة في هذه الاختبارات موحدة نظراً لأن الخطأ المعياري للقياس يعبر عنه بنفس وحدات الانحراف المعياري. فإذا

كانت قيمة الانحراف المعياري تم حسابها من درجات خام، فإن القيمة التقديرية للخطأ المعياري للقياس ستكون درجات خام أيضاً، وإذا تم حساب قيمة الانحراف المعياري من نسب ذكاء فإن قيمة الخطأ المعياري للقياس سوف يعبر عنها بحسب ذكاء أيضاً، وهكذا.

وهذا يدل على أننا لا نستطيع مقارنة القيم التقديرية للخطأ المعياري للقياس لاختبارات يختلف نوع معاييرها، لذلك يجب في هذه الحالة إجراء المقارنة باستخدام القيم التقديرية لمعامل الثبات.

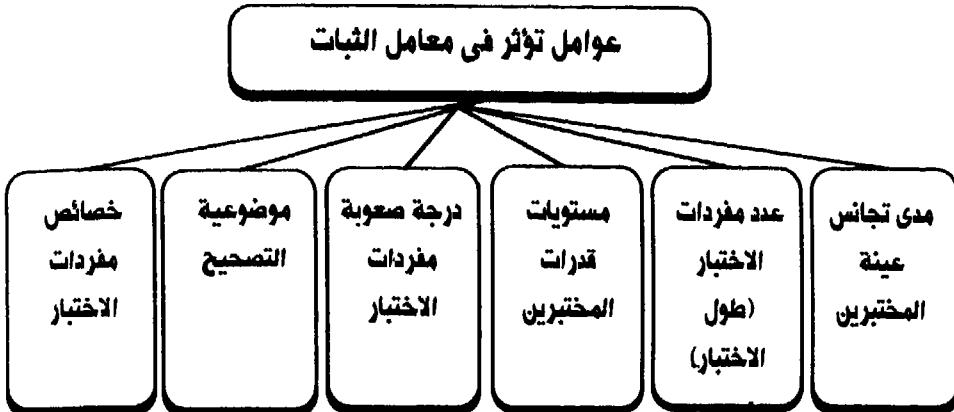
فالقيم التقديرية للخطأ المعياري للقياس تُعد قيمًا مطلقة وليس نسية كما هو الحال في القيم المناظرة لمعامل الثبات، وتُعبر عن مدى تكافؤ الدرجات الملاحظة والدرجات الحقيقة لفرد في الاختبار.

والخلاصة أن الخطأ المعياري للقياس يعد عاملًا أساسياً في تقرير نتائج الاختبارات والمقاييس وتفسيرها، وهو مرتبطة ارتباطاً وثيقاً بمفهوم الثبات ولكنه أيسر فهماً لمستخدمي الاختبارات والباحثين، ولا يتأثر كثيراً بتشتت الدرجات على عكس الثبات الذي يعتمد على التشتت.

لذلك تهتم أدلة الاختبارات بتحديد القيم التقديرية لهذا الخطأ لكي يتمكن مستخدم الاختبار من تفسير الدرجات تفسيراً احتمالياً. فإذا حصل فرد في اختبار ما على الدرجة ٦٥ فإن هذه الدرجة ليست نقطة محددة وإنما يمكن أن تذبذب في مدى يتراوح بين قيم معينة على متصل الدرجات تحدها القيمة التقديرية للخطأ المعياري للقياس ودرجة الثقة المطلوبة. بخلاف من القول بأن الفرد حصل على الدرجة ٦٥ نقول مثلاً: بأننا نثق ٦٨٪ أن درجته الحقيقة تقع بين ٦٣ ، ٦٦ إذا ما أعيد تطبيق الاختبار عليه عدداً كبيراً من المرات.

عوامل أخرى تؤثر في قيم معامل الثبات:

سبق أن ذكرنا أن القيم التقديرية لمعامل الثبات تتأثر بمصادر متعددة للأخطاء العشوائية، غير أن هناك العديد من العوامل الأخرى التي تؤثر في هذه القيم، وهذه العوامل تهم القائمين ببناء الاختبارات والمقاييس أكثر مما تهم مستخدميها، ومع هذا فإن معرفة مستخدمي الاختبارات لهذه العوامل ومراعاتها يساعدهم في انتقاء اختبارات تتميز درجاتها بالثبات، كما يساعدهم في تفسير قيم معامل الثبات تفسيراً مستنيراً. والشكل التخطيطي (٤ - ١) التالي يوضح بعض هذه العوامل:



شكل (٤ - ١٠) يوضح بعض العوامل المؤثرة في قيم معامل الثبات

وسوف نوضح فيما يلى العوامل الموضحة فى شكل (٤ - ١٠)

مدى تجانس عينة المختبرين :

تعتمد القيمة التقديرية لمعامل الثبات اعتماداً كبيراً على مدى الفروق بين الأفراد المختبرين، فكلما زادت هذه الفروق ازداد تباين الدرجات الحقيقية للأفراد وبالتالي تزداد قيمة معامل الثبات. فالثبات كما أوضحنا يشير إلى اتساق قياس الفروق الفردية الحقيقية، فكلما زادت هذه الفروق يسهل الحصول على قياسات متسقة إذا تكررت عملية القياس على مجموعة الأفراد، فإذا كانت مجموعة الأفراد متتجانسة في القدرة أو السمة التي يقيسها الاختبار فإن تباين الدرجات الحقيقية يقل وبالتالي تنخفض قيمة معامل الثبات أو ربما تصبح مساوية صفرأً . ولهذه الخاصية تطبيقاتها الميدانية عند استخدام درجات الاختبارات والمقاييس في اتخاذ قرارات تتعلق بالطلاب أو العاملين الذين تكون قدراتهم متتجانسة . فعندئذ ينبغي مراعاة انتقاء اختبارات اعتمد تقدير ثبات درجاتها على عينة من الأفراد مستوى قدراتها في نطاق مستوى المجموعة التي نهم بدراستها .

إذا كانت قيمة معامل الثبات المذكورة في دليل الاختبار تم تقديرها من مجموعة من تلاميذ الصفوف الثلاثة الأخيرة من مرحلة التعليم الأساسي ، ولكننا طبقنا الاختبار بعد ذلك على تلاميذ صف واحد فقط ، فإننا تتوقع أن تقل القيمة التقديرية لمعامل ثبات الدرجات بما هو مذكور في دليل الاختبار نظراً لأن تلاميذ هذا الصف أكثر تجانساً من تلاميذ الصفوف الثلاثة .

عدد مفردات الاختبار،

سبق أن أوضحنا عند مناقشتنا معامل الاتساق الداخلي أنه كلما زاد عدد مفردات الاختبار أي طول الاختبار كلما زادت قيمة معامل ثبات درجاته . ويرجع ذلك إلى أن زيادة عدد المفردات يسمح للأخطاء العشوائية الموجبة والسلبية أن تتلاشى بعضها بعضاً مما يجعل الدرجة الملاحظة للفرد في الاختبار تقترب من درجته الحقيقة . وعند مناقشتنا صيغة سبيرمان وبراؤن Spearman - Brown Formula أوضحنا أنه يمكن تقدير القيمة المتوقعة لمعامل ثبات درجات الاختبار إذا أطلنا الاختبار أو جعلناه أقل طولاً . ولعل جدول (٤ - ١) يوضح كيف أن زيادة عدد مفردات الاختبار تؤدي إلى زيادة القيمة المتوقعة لمعامل الثبات .

مستويات قدرات المختبرين :

يؤثر مستوى قدرات المختبرين أيضاً في قيم معامل الثبات الذي نحصل عليه ، فتبين درجات الخطأ يزيد بالنسبة للمجموعة منخفضة الدرجات حيث تلعب عوامل التخمين والصدفة دوراً أكبر في درجاتهم .

لذلك لا يجوز استخدام درجات مجموعة من مستوى قدرة معين للتبؤ بثبات درجات الاختبار إذا طبق على مجموعة من مستوى آخر . فمثلاً إذا أردنا استخدام اختبار معين لانتقاء الأفراد في إحدى المهن الدنيا ، أو أفراد مبتدئة في إحدى المهن يجب إلا يتأثر القرار بالبيانات المتعلقة بالثبات المستمدة من أفراد مهنة عليا أو أفراد من ذوى مهارات متقدمة . فالاختبار عندئذ تباعين درجة دقته في قياس هذه المستويات المختلفة . وهذا يتطلب مراجعة دليل الاختبار بعناية لتعرف مدى تأثير درجات الاختبار بتباين مستوى متوسط قدرات المختبرين .

درجة صعوبة مفردات الاختبار،

تؤثر درجة صعوبة مفردات الاختبار في قيم معامل الثبات ، فإذا كانت المفردات غالية في السهولة أو الصعوبة ، فإنه لا يستطيع باستخدامها قياس الفروق الفردية . ففي الحالة الأولى يستطيع كل فرد أن يجيب إجابة صحيحة عن جميع المفردات ، والعكس في الحالة الثانية ، وبذلك يكون توزيع الدرجات منتظمًا في الحالتين . فدرجة صعوبة الاختبار ككل هي متوسط توزيع درجات الاختبار مقسوماً على عدد مفرداته ، فكلما

اقربت قيمة هذا المتوسط من عدد مفردات الاختبار كان الاختبار أكثر سهولة، وإذا تساوى المتوسط مع عدد المفردات كان الاختبار غاية في السهولة، وعندئذ يستطيع كل فرد أن يجيب إجابة صحيحة عن جميع مفرداته ويصبح تباين الدرجات صفرًا، وبالتالي تكون قيمة معامل الثبات صفرًا أيضًا. وبالطبع تختلف الاختبارات في مدى سهولة أو صعوبة مفرداتها، فالاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests التي تصمم لقياس الفروق الفردية تكون مفرداتها عادةً متوسطة الصعوبة لإبراز هذه الفروق، أما الاختبارات مرجعية المحك Criterion - Referenced Tests التي تصمم للأغراض التعليمية مثل الاختبارات التحصيلية الصفية، فإنها تهدف لقياس تمكن جميع الطلاب من مجال دراسي معين. لذلك فإن مفرداتها ربما تميز بالسهولة النسبية لكي تقلل من عوامل التخمين الذي يؤدي إلى الأخطاء العشوائية في الدرجات التي تسهم بدورها في خفض قيمة معامل الثبات. ولذلك فإن تقدير ثبات درجات هذا النوع من الاختبارات باستخدام الطرق السابقة التي تعتمد على معامل ارتباط بيرسون لا تكون مناسبة، وسوف نشير إلى طرق أخرى تناسب الاختبارات مرجعية المحك في الفصل الثامن.

موضوعية التصحيح:

سبق أن ناقشنا مفهوم الموضوعية Objectivity ، وتبيّن لنا أنه كلما تأثر تصحيح الاختبار أو تقدير درجاته بعوامل ذاتية أو عوامل التحيز انخفضت قيمة معامل ثبات الدرجات .

فتتصحيح الاختبارات التي تشمل على مفردات مثل الاختيار من متعدد، أو الصواب أو الخطأ، أو الإكمال يكون عادةً موضوعياً سواءً أجري يدوياً أو آلياً. ولكن المشكلة تبدو واضحة في تقدير درجة اختبارات المقال وبعض مقاييس الأداء والشخصية، حيث يتضمن التصحيح أحکاماً فردية تتعلق بنوعية الاستجابات، وهذا بدوره يؤثر تأثيراً بالغاً في ثبات التقديرات. وقد سبق أن أوضحتنا في هذا الفصل إمكانية التغلب على هذه المشكلة عند مناقشتنا ثبات تقديرات المحكمين .

خصائص مفردات الاختبار:

تؤثر مفردات الاختبار في ثبات درجات الاختبار ككل، فخلو المفردات من الخطأ يعتمد على كيفية بناء هذه المفردات، وبعض المفردات ربما تشمل على مؤشرات لاجابة مفردات أخرى في الاختبار مما يساعد بعض الأفراد على التخمين بدرجة جيدة مما يعمل على خفض قيمة معامل الثبات.

وكذلك المفردات الغامضة ، أو غير محددة الهدف ، أو التي تكون صياغتها أو تعليمات إجابتها غير دقيقة، أو غایة في الصعوبة، تؤثر تأثيراً بالغاً في ثبات درجات الاختبار.

ومن هذا يتضح أن قيم معامل الثبات ليست قيمًا مطلقة وإنما تعد قيمًا تقديرية تؤثر فيها عوامل متعددة ينبغي مراعاتها في تصميم أدوات القياس وعند انتقاء هذه الأدوات واستخدامها وتفسير نتائجها. فالفرض الذى تستند إليه طرق تقدير الثبات يصعب تحقيقها جمیعاً في الواقع الميداني، مما يتطلب الحیطة في تفسير قيم معامل الثبات. إذ لا يوجد أسلوب إحصائي يُعني عن الحكم المنطقى والتفكير المستنير في تفسير هذه القيم وبخاصة في القياس التربوى والنفسى.

خلاصة

إن الاختبار أو المقياس ينبغي أن تتوافر فيه بعض الخصائص السيكومترية الأساسية من أهمها ثبات درجاته وصدقه. ويشير مفهوم الثبات إلى عدم تأثير الدرجات بالخطاء غير المنظمة. وتتعدد مصادر هذه الأخطاء، وبعضاً منها يتعلّق بأداة القياس، أو إجراءات تطبيق الاختبار وتصحيحه، وبعضاً آخر يتعلّق بالأفراد المختبرين. فإذا انعدمت هذه الأخطاء تصبح الدرجة الملاحظة للفرد في الاختبار متساوية لدرجته الحقيقية التي تعد بمثابة درجة افتراضية. وإذا كان تباين الدرجات الملاحظة متساوياً لتباين الدرجات الحقيقة، فإن تباين الخطأ يكون متساوياً صفرًا، وبهذا يصبح الشات تاماً.

ونظراً لأننا لا نستطيع قياس التباين الحقيقى بطريقة مباشرة، فإنه يمكننا تقدير قيمة أخطاء الصدفة التى تؤثر فى الدرجات باستخدام أساليب مختلفة نحصل منها على قيم تقديرية لمعامل التكافؤ، أو معامل الاستقرار، أو معامل الاستقرار والتكافؤ، أو معامل الاتساق الداخلى، أو معامل التجانس، أو معامل إمكانية التعجم. كما يمكن تقدير ثبات الفرق بين درجتي اختبارين، وثبات تقديرات المحكمين.

كما يمكن إيجاد قيمة الخطأ المعياري للقياس التي تتناسب عكسياً مع قيمة معامل الثبات، حيث تُحدد مدى معين يحتوى الدرجة الحقيقية للفرد في الاختبار. ويمكن تقليل مصادر الأخطاء العشوائية التي تؤثر في قيم معامل الثبات، وذلك بإطالة الاختبار بقدر معين، ومراعاة الموضوعية في تصحيحه، وكذلك مراعاة تباين الأفراد المختبرين في السمة التي يقيسها الاختبار، والعناية بصياغة وبناء مفرداته، والتحقق من درجة صعوبتها وفقاً لما إذا كان الهدف من الاختبار إبراز الفروق الفردية، أو قياس درجة التمكّن في مجال دراسى معين.

الفصل الخامس

الخطائين السيفيكومترية للإختبارات ثانياً: صدق الاختبارات

* مقدمة *

- * جوانب صدق الاختبارات
- * صدق المحتوى
- * طرق تقدير صدق المحتوى
- * الصدق المرتبط بمحك
- * الصدق التنبؤى
- * طرق تقدير الصدق التنبؤى
- * بعض المشكلات المتعلقة بالمحك
- * الصدق التلازمى وطرق تقديره
- * بعض العوامل التى تؤثر فى الصدق المرتبط بمحك
- * صدق التكوين الفرضى
- * طبيعة التكوينات الفرضية
- * منهجية التحقق من صدق التكوين الفرضى
- * أساليب جمع أدلة تتعلق بصدق التكوين الفرضى
- * منظور تكاملى لطرق تقدير الصدق والثبات

مقدمة :

تعد جوانب الصدق من أهم خصائص الاختبارات والمقاييس التربوية والنفسية. فصدق الاختبار Test Validity يتعلق بالهدف الذي يبني الاختبار من أجله، وبالقرار الذي يتخذ استناداً إلى درجاته، فدرجات الاختبار تستخدم عادة في التوصل إلى استدلالات معينة، وهنا يبرز التساؤل حول ما يمكن الاستدلال عليه بدرجة عالية من الدقة أو الثقة.

فعلى الرغم من أن ثبات درجات الاختبار يعني اتساق الدرجات، فإن الصدق يتعلق بدرجة تتحققنا من تفسير درجات الاختبار واستخدامها، ولذلك يمكن أن تميز درجات الاختبار بالثبات أو الاتساق، ولكن لا يعني ذلك أن الاختبار يكون صادقاً. إذ يمكن أن يُبني اختبار لقياس المهارات الحسابية ويهصل كل تلميذ على درجات متسقة إذا أعيد تطبيق الاختبار عليه عدة مرات، ومع هذا لا نستطيع الحكم على درجة صدق الاختبار، إذ إن هذا يتطلب معلومات كثيرة تتعلق بمحتوى الاختبار، وعدد مفرداته ونوعها، ومدى تمثيلها للنطاق الشامل للأعداد والعمليات، وما إذا كانت المفردات تشتمل على مسائل كلامية، وغير ذلك مما يؤثر في أداء التلاميذ في الاختبار وبالتالي في تفسير درجاته.

ثبات درجات الاختبار يتأثر فقط بمصادر الأخطاء العشوائية للقياس كما سبق أن أوضحنا، بينما يتأثر صدق الاختبار بكل من مصادر الأخطاء العشوائية والأخطاء المنتظمة أو الثابتة. وهذا يؤكد ما ذكرناه من أن درجات الاختبار يمكن أن تكون على درجة عالية من الثبات دون أن يكون الاختبار صادقاً. ولكن الاختبار لا يمكن أن يتسم بالصدق دون أن تكون درجاته متسقة، أي أن الثبات شرط ضروري للصدق ولكنه غير كاف، ومع هذا لا نستطيع الفصل بينهما أو اعتبارهما ثنائية. فقد أكد Cronbach أن هذين المفهومين مترابطان ويمكن أن يندرجَا تحت اسم «مقاييس قابلية التعميم Measures of Generalizability» التي أشرنا إليها في الجزء المتعلق بالثبات.

فالفرق الرئيسي بينهما يمكن في الأبعاد التي يمكن أن تعمم في ضوئها نتائج الاختبار أو القياس. كما أكد أن الصدق ليس خاصية تتعلق بالاختبار في ذاته، وإنما تتعلق بتفسير الدرجات المستمدّة من الاختبار. فالاختبار الواحد يمكن أن يستخدم في أغراض متعددة، فاختبار القراءة يمكن أن يستخدم في فحص المتقدمين للدورة تدريبية معينة، وفي تشخيص صعوبات القراءة، وفي قياس فاعلية برنامج تعليمي في اللغة العربية، وغير ذلك . ونظراً لأن كلاً من هذه الاستخدامات يعتمد على تفسيرات مختلفة، فإن الأدلة التي تبرر استخداماً معيناً ربما لا تصلح لتبرير استخدام آخر .

فجميع هذه التفسيرات تتميز بقدر معين من الصدق، لذلك يصعب أن نستنتج ببساطة أن الاختبار «صادق Valid». لذلك ليس لكل اختبار مؤشر صدق واحد، وإنما يمكن أن يكون له عدة مؤشرات للصدق بحسب تعدد أغراض استخداماته، ونوع القرارات التي يمكن اتخاذها. فالاختبار الذي يتميز بدرجة معينة من الصدق بالنسبة لغرض معين أو لمجموعة معينة من الأفراد، ربما لا يكون كذلك بالنسبة لغرض آخر أو لمجموعة أخرى.

فالصدق إذن يتعلق بمدى فائدة أداة القياس في اتخاذ قرارات تتعلق بغرض أو أغراض معينة. فقرارات الانتقاء مثلاً يمكن أن تكون أكثر صواباً إذا استطاع المعلم تقدير احتمال نجاح الطالب في مقررات مختلفة.

فإذا لم تكن لنتائج أدوات القياس فائدة في اتخاذ قرارات صائبة ربما لأنها لا تقدم معلومات غير دقيقة أو غير متعلقة بمجال القرار، فإن نتائج هذه الأدوات لا تسم بالصدق مهما بلغت قيمة معامل الثبات.

ولكي نحدد مدى ملاءمة الاختبار لاستخدامات معينة ينبغي جمع معلومات مناسبة تتعلق بصدق الاختبار، وتعتمد هذه المعلومات على هدف أو أهداف عملية القياس وليس على نوع الاختبار أو المقاييس.

وقد سبق أن ذكرنا في الفصل الأول أن الاختبار يمكن أن يكون بمثابة عينة Sample أو بمثابة مؤشر Sign ، وهذا التمييز يعتمد على درجة تمثيل المفردات للنطاق السلوكي الشامل للسمة المراد قياسها.

فإذا كان الاختبار يشتمل على عينة من المفردات تمثل النطاق السلوكي الشامل تمثيلاً جيداً كما في الاختبارات التحصيلية مثلاً، فإن الاختبار يكون بمثابة عينة Sample، أما إذا كان النطاق السلوكي الشامل غير محدد تحديداً دقيقاً وكاملاً كما في بعض مقاييس الشخصية والقدرات، فإن الاختبار أو المقاييس يكون بمثابة مؤشر Sign . إذ يمكن أن يستخدم اختبار في الذكاء كمؤشر لبعض الاضطرابات العقلية ، ومثال ذلك اختبار ويكسنر Wechsler للذكاء.

فالذى يطلع على الأدبيات السيكلوجية والتربوية يلاحظ تعدد أنواع الصدق ونمطية استخداماتها بحيث لم يعد ممكناً التتحقق مما يقصد بهذه الأنواع، وما إذا كانت تختلف بالفعل فيما بينها ، وربما يرجع ذلك إلى استخدام الباحثين والممارسين لمفهوم الصدق استخداماً ضيقاً ومحدوداً.

لذلك أصدرت لجنة فنية مشتركة من الجمعية الأمريكية لعلم النفس APA، والجمعية الأمريكية للبحث التربوى AERA ، والمجلس القومى الأمريكى للقياس

التربوي NCME كتيباً متعددًا يتضمن معايير الاختبارات التربوية والنفسية وأدلتها سبق أن أشرنا إليه في الفصل الثاني، ويؤكد هذا الكتيب أن الاختبارات والمقاييس تستخدم لاتخاذ قرارات وإصدار أحكام متعددة، وكل قرار أو حكم يتطلب أنواعاً مختلفة من الدراسات للتحقق من الصدق.

ومن بين هذه القرارات والأحكام المتعددة يميز الكتيب بين ثلاثة رئيسة منها : كال التالي :

(١) تحديد كيفية أداء الفرد في الوقت الحاضر في نطاق شامل لموافقات سلوكية تمثلها مفردات الاختبار تمثيلاً جيداً. فمثلاً ربما يود مستخدم اختبار معرفة أداء تلميذ في الصف الثاني الابتدائي في الحساب في منتصف العام الدراسي ، فالاختبار يكون مستهتملاً على عينة مماثلة من المفردات أو الأسئلة المتعلقة بالأهداف السلوكية المرجوة من هذه المادة الدراسية.

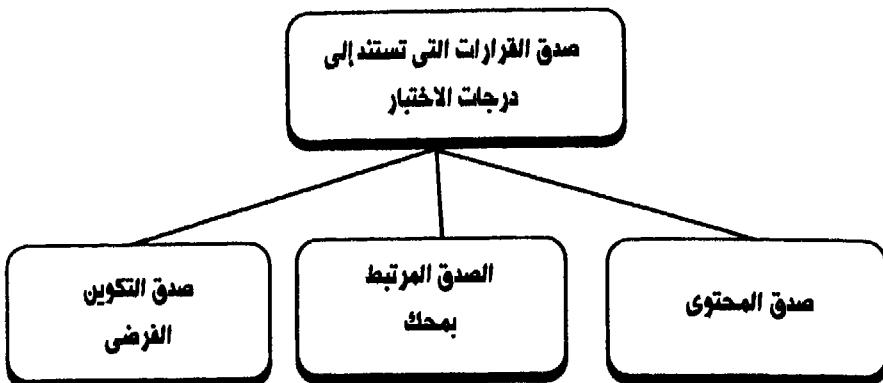
(٢) التنبؤ بالأداء المستقبلي أو تقدير مكانة الفرد في أحد المتغيرات ذات الأهمية ، ولكنها تختلف عما يقيسه الاختبار . فمثلاً يمكن استخدام اختبار الاستعداد الأكاديمي في التنبؤ بالمعدل العام للطالب في نهاية السنة الأولى بإحدى الكليات ، فعندئذ يكون المعدل العام هو المتغير موضع الاهتمام ، ويلاحظ أنه يختلف عن متغير الاستعداد الأكاديمي . وكذلك ربما يستخدم استبيان التوافق في تقدير ما يمكن أن يكشف عنه الفحص السيكولوجي .

(٣) الاستدلال على درجة تملك الفرد سمة أو خاصية معينة تعد بمثابة تكوين فرضي Construct يbedo أثره في سلوك الفرد أو أدائه . فمثلاً ربما نود معرفة ما إذا كان الفرد يتسم بقدر كبير من سمة معينة مثل الذكاء أو الابتكارية أو الاكتتاب ، علما بأن هذه السمات مجردة أي غير منظورة ولا نستطيع ملاحظتها ملاحظة مباشرة .

وربما يكون الهدف من هذا الاستدلال هو تقدير درجة السمة لدى الفرد ، أو دراسة الاختبار وتعرُّف مكوناته ، أو دراسة علاقة الاختبار بالاختبارات الأخرى التي تقيس السمة نفسها أو سمات مشابهة ، أو بناء نظرية سيكولوجية .

جوانب صدق الاختبارات :

حددت اللجنة المشتركة ثلاثة جوانب من الصدق تناولت أنواع القرارات الأساسية الثلاثة السابقة ، وهى : صدق المحتوى ، والصدق المرتبط بمحك ، وصدق التكوين الفرضي أو المفهوم كما يوضحه الشكل التخطيطي (١٥) التالي :



شكل (٥ - ١) يوضح الجوانب المختلفة لصدق القرارات التي تستند إلى درجات الاختبارات ونما هو جدير بالذكر أن هذه الجوانب الثلاثة متربطة إجرائياً ومنطقياً . وقد أكدت اللجنة ذلك ، حيث أشارت إلى أنه نادراً ما يكون أحد هذه الجوانب ذو أهمية بمعزل عن الجانبيين الآخرين ، فدراسة الاختبار دراسة متعمقة تتضمن عادة معلومات حول الجوانب الثلاثة .

لذلك يرى كثير من علماء القياس مثل كرونباك (Cronbach, 1970) وأنساتاري (Anastasi, 1976)، وستانلى (Stanley , 1981) وغيرهم أن الأدلة التي نترشد بها للتحقق من الصدق وليس الاختبار هي ما ينبغي أن يطلق عليها مفهوم «الصدق» .

كما يرون أنه لا ينبغي التفكير في الجوانب الثلاثة كأنواع مختلفة من الصدق ، وإنما كأدلة متعددة عن مدى صلاحية الاختبار في ترشيد القرارات . فيقدر ما تكون المعلومات التي يمدنا بها الاختبار مضللة أو غير كافية لإصدار قرارات معينة تكون البيانات المستمدة من هذا الاختبار غير صادقة .

لذلك يؤكّد ساكس Sax أنه لا يمكننا أن نبرهن على صدق الاختبار وإنما نستطيع أن نقدم أدلة تتعلق بصلاحيته في غرض أو أغراض معينة .

وعلى الرغم من تكامل الجوانب الثلاثة للصدق إلا أننا سوف نتناول كل منها على حدة بالمناقشة والتحليل لكي يتيسر على القارئ استيعابها .

طرق المحتوى :

يدل صدق المحتوى على مدى تمثيل محتوى الاختبار للنطاق السلوكي الشامل Universe للسمة المراد الاستدلال عليها، إذ يجب أن يكون المحتوى ممثلاً تمثيلاً جيداً لنطاق المفردات الذي يتم تحديده مسبقاً.

ونقصد بنطاق المفردات: المعرف والمهارات والعمليات التي يتم معainتها بواسطة مفردات الاختبار. ويمكن إجراء ذلك عن طريق تحليل المحتوى أو النطاق السلوكي بحيث يمثل الاختبار جميع مكونات هذا المحتوى أو هذا النطاق تمثيلاً جيداً. لذلك فإن صدق المحتوى يناسب بدرجة أكبر الاختبارات والمقياس التربوية، وبخاصة الاختبارات التحصيلية. فإذا أجرى المعلم اختباراً على تلاميذه لقياس مهاراتهم في استخدام الحاسوب، فإنه يهتم بالتحقق من أن درجات الطلاب في هذا الاختبار تمثل بالفعل المهارات المرجوة. غير أن كثيراً من العوامل المتعلقة بخصائص مفردات الاختبار ومدى تمثيلها للنطاق السلوكي لهذه المهارات ربما تؤثر في درجاتهم.

وهذه العوامل تؤدي إلى انخفاض صدق الاستدلال على المهارات التي يحددها هذا النطاق. ونظراً لأن الاختبارات التحصيلية واختبارات الكفايات تعد بمثابة عينة Sample ، فإن صدق المحتوى في هذه الحالة يتعلق بأسلوب معainات المفردات وكفاية العينة الممتدة التي يشتمل عليها الاختبار.

وهذا يتطلب أن يكون النطاق الشامل للمفردات Items Universe معرفاً تعرifa دقیقاً؛ لأنـه يعد بمثابة تعريف إجرائي للسمة المراد قياسها وهـى التحصيل في مجال معين. ولعل تحديد الأهداف السلوكية التي تقيسها الاختبارات التحصيلية يعد خطوة ضرورية في هذا الشأن كما سيتضح في الفصل الثامن عند تناولنا أساليب قياس التحصيل .

وقد أشار كتيب اللجنة المشتركة إلى إمكانية التحقق من صدق المحتوى بالنسبة لاختبارات الاستعدادات واستبيانات الشخصية والميول وكذلك مقاييس التوافق والسلوك الاجتماعي، حيث يمكن الاعتماد على عينات من المهارات التي يتضمنها الاستعداد الميكانيكي أو الموسيقى مثلاً، أو عينات من الملاحظات في مواقف متعددة تتتعلق بالميول أو التوافق أو غير ذلك من السمات، ولكن لا ينبغي في هذه الحالات الاقتصار على صدق المحتوى وإنما يجب الاستناد إلى أدلة من جوانب الصدق الأخرى. غير أن بعض علماء النفس لا يتفقون مع هذه اللجنة في هذا الشأن.

إذ ترى Anastasi عدم ملاءمة صدق المحتوى لاستخدامات هذه المقاييس، وحجتها في ذلك أن مقاييساً معيناً ربما يقيس وظائف مختلفة لدى

المجموعات المتباعدة من الأفراد، مما يجعل تحديد هذه الوظائف من مجرد تأمل أو فحص محتوى المقياس أمراً غير ممكن.

ويوجه عام فإن مدى ملاءمة استخدام صدق المحتوى يعتمد على نوع الاستدلالات أو التفسيرات التي نود التوصل إليها استناداً إلى نتائج الاختبارات والمقاييس.

طرق تقييم صدق المحتوى :

اتضح لنا مما سبق أن صدق المحتوى Content validity يتعلق بمحاولة تحديد مدى كفاية بناء الاختبار بحيث تمثل مفرداته النطاق السلوكي لمهارة أو مجال دراسي معين تمثيلاً جيداً.

لذلك يتطلب هذا الجانب من الصدق دون غيره من الجوانب أدلة منطقية وليس إحصائية. فتقدير صدق المحتوى يستند إلى منطق مستثير ومهارات تحليلية ومثابرة، حيث إن أدلة هذا الصدق تحتاج إلى مراجعة مستمرة في ضوء التغيرات التي يمكن أن تطرأ على النطاق السلوكي وما يتضمنه من محتوى وعمليات، أو على المناهج الدراسية أو أساليب التعليم.

ولذلك فإن معظم أساليب تقدير صدق المحتوى تعتمد على الأحكام التقيسية لخبراء المواد الدراسية أو المهتمين بتنمية المهارات والكفايات التعليمية والمهنية والفنية، وتعمل هذه الأحكام بتقدير مدى التمازج بين مفردات الاختبار والنطاق السلوكي الذي تمثله هذه المفردات.

وهذا يتطلب صياغة مكونات النطاق السلوكي من معارف ومهارات وعمليات ومصادر صياغة واضحة ومحددة، وكذلك المكونات الفرعية لهذا النطاق والأهمية النسبية لكل منها، ومحتوى كل مفردة وما تتضمنه من عمليات وإجراءات.

ويمكن إعداد استماراة تشتمل على ميزان تقدير تتضمن الأبعاد الرئيسية المتعلقة بالنطاق السلوكي، مثل محتوى المفردات، ونوع المهارات و مجالها، ومصادر أو مواد ذات أهمية، ونوع المفردات وملاءمتها لكل من المحتوى والمهارة المرجوة.

ويقوم المحكم بتقييم كل مفردة من مفردات الاختبار في ضوء هذه الأبعاد على ميزان التقدير الموضوعي في الاستماراة. ويفضل الاعتماد على أكثر من محكم للحصول على تقييمات متسبة بدرجة أفضل.

ويمكن التتحقق من ذلك بتحليل التقديرات إحصائياً باستخدام أساليب تحليل التباين أو غيرها من الأساليب التي ترودنا بمؤشرات إحصائية تتعلق بدرجة اتساق التقديرات.

كما يمكن التتحقق من صدق المحتوى إمبيريقياً، وذلك بتطبيق الاختبار على مجموعة من الطلاب قبل بدء عملية التعليم، وإعادة تطبيقه عليهم عقب الانتهاء منها، وفحص نتائج الاختبار في المرتبين لمعرفة ما إذا كان الاختبار يقياس بالفعل المجال الذي اهتمت به عملية التعليم. وهذا يتطلب أن يكون لدى الطلاب معرفة محددة بال المجال الذي سوف تتناوله عملية التعليم قبل البدء.

وينبغي التنويه إلى أن بعض الباحثين والمهتمين ببناء الاختبارات ومستخدميها يفترضون أن صدق المحتوى يناظر الصدق الظاهري Face Validity الذي يتحققون منه بالقاء نظرة سريعة على محتوى الاختبار لتعرف ماذا يقياس، غير أن هذا ربما يكون مفيداً إلى حد ما في المواقف التي تتطلب انتقاء العاملين للالتحاق بهم أو وظائف معينة، حيث ربما يكون من المناسب أن يشعر المختبرون بأن الاختبار يبدو أنه يقياس خصائص تتعلق بالمهنة أو الوظيفة المتقدمين لشغلها.

ولكن هناك مواقف أخرى ربما يؤدي الاعتماد فيها على الصدق الظاهري للاختبار أو المقياس إلى نتائج غير مرغوبة وبخاصة في قياس الشخصية باستخدام الاستبيانات، إذ ربما يلجم الفرد الذي يتجنب عن مفردات الاستبيان إلى تزييف استجاباته لشعوره بالتدخل في خصوصياته أو إظهاره بمظهر معين.

لذلك فإن دليل معايير الاختبارات والمقياس الذي أشرنا إليه استبعد الصدق الظاهري واكتفى بصدق المحتوى.

والخلاصة أن تقدير صدق المحتوى يستند إلى فروض ثلاثة صاغها لينون Lennon كالتالي :

(١) ينبغي أن يكون المجال الذي يختبر فيه الأفراد محدوداً بنطاق شامل للمفردات الذي تبدو أهميته لهم، ويمكن تعريفها دقيقاً.

(٢) يمكن انتقاء عينة من المفردات من هذا النطاق بطريقة هادفة ومناسبة.

(٣) يمكن تحديد عينة المفردات وأسلوب المعاينات المستخدم وتعريفها بدقة كافية لكي يتمكن مستخدم الاختبار من الحكم على مدى تمثيل عينة المفردات للنطاق السلوكي الشامل الذي تقيسه.

ويلاحظ أن جانب صدق المحتوى شأنه شأن جوانب الصدق الأخرى يتعلق بالغرض من استخدام الاختبار ومجموعة الأفراد المختبرين.

فأدلة صدق المحتوى لمجموعة معينة ربما لا تُنطبق بالضرورة على مجموعات أخرى نظراً لاختلاف العمليات التي تعتمد عليها كل مجموعة في إجابتهم عن مفردات الاختبار.

لذلك نؤكد مرة أخرى أنه ليس هناك صدق محتوى للاختبار، وإنما صدق محتوى الغرض أو الاستخدام المناسب لمجموعة معينة من الأفراد. وقد وضح كثير من علماء القياس أهمية محاولة تطوير أساليب وإجراءات يمكن باستخدامها تحديد مؤشرات كمية أو مجموعة من المحکات يستند إليها في تلخيص درجة صدق المحتوى لكي تُسر عملية التواصل بالمعلومات المتعلقة بهذا الجانب من الصدق.

الصدق المرتبط بمحک:

كثيراً ما يهتم مستخدمو الاختبارات ومتخذلو القرارات بالتبؤ بنجاح أحد الطلاب في دراسته الجامعية، أو بفاعلية أداء المعلم بعد تخرجه من كلية التربية، أو بمستوى أداء فرد في إحدى المهن بعد انتهائه من دورة تدريبية معينة، فالاهتمام في مثل هذه الحالات ينصب على التنبؤ بالأداء المستقبلي باستخدام درجات اختبارات مناسبة تطبق في الوقت الحاضر، فتقسيم القرارات التنبؤية يعتمد على تقدير الصدق المرتبط بمحک - Criterion Related Validity ، أي مدى صلاحية الاختبارات والمقاييس في أغراض التنبؤ من أجل اتخاذ قرارات عملية.

فالصدق المرتبط بمحک يناسب المواقف التي نود فيها استخدام أداة قياس في تقيير سلوك معين ذي دلالة، وهذا السلوك خارج عن نطاق الاختبار ذاته ويعود بشائبة المحک Criterion ، ففي الأمثلة السابقة ربما يكون معدل تقديرات الطالب في نهاية السنة الأولى بإحدى الكليات هو المحک، أو أن نتائج بطاقة تقييم المعلم بعد مرور عام من التحاقه بمهنة التدريس هي المحک، وهكذا.

فالاختبارات التي تُستخدم في اتخاذ قرارات التنبؤ ليس بالضرورة أن يكون محتواها مثلاً لمحک المقاييس المعلم، إذ ربما نستطيع أن نتنبأ ببعض الاضطرابات النفسية باستخدام اختبار للذكاء، أو أن نتنبأ بنجاح الطالب في دراستهم الجامعية باستخدام اختبارات قبول معينة، أو نتنبأ بالأفراد الذين ينجحون في المهن التبغية الدعائية باستخدام استبيانات الشخصية.

ففي مثل هذه الحالات كلما كانت التنبؤات دقيقة كان الاختبار أكثر فائدة. ونظراً لتأثير القرارات المترتبة على نتائج الاختبارات والمقاييس في مجالات انتقاء الأفراد، فإنه تبدو أهمية توافر الصدق المرتبط بمحک في هذه القرارات إذا استخدمت الاختبارات والمقاييس لترشيد عمليات الانتقاء.

، ومن هذا يتضح أن الصدق المرتبط بمحك يستند إلى الأسلوب الإمبريقي في دراسة العلاقة بين درجات اختبارات أو مقاييس معينة تعد بمثابة منبئات Predictors ودرجات مقاييس خارجية مستقلة تعد بمثابة محکات Criteria .

غير أن المشكلة الأساسية في تقدير الصدق المرتبط بمحك تتركز في إمكانية الحصول على محکات مناسبة ، فجمع بيانات تتعلق بالمحکات تعد من مشكلات القياس التربوي والنفسى التي تتطلب دراسات متعددة . فكما أن الاختبار التنبؤى ينبغي أن يتميز بخصائص معينة ، كذلك ينبغي أن يتميز المقياس المحک بخصائص يمكننا من الثقة فيما يترتب عليه من بيانات .

إلى جانب الأحكام التقييمية للمقياس المحک ينبغي أيضاً التحقق من ثبات درجاته وخلوها من التحيز ومواءمة المقياس للأداء المطلوب . فإذا اعتبرنا تقدیرات المشرفين على العمل محکا لأداء العاملين ، فإن هذا يتطلب التتحقق من دقة هذه التقدیرات وعدم تحيزها ، ومدى مواءمتها لطبيعة العمل وخصائص العاملين ؛ لكن تكون ذات فائدة في الحكم على صدق القرار الذي يتخذ في ضوء درجات الاختبار المنبئ .

أنواع الصدق المرتبط بمحک:

تميز أدبيات القياس والتقويم بين نوعين من الصدق المرتبط بمحك أحدهما يسمى الصدق التنبؤى Predictive Validity والأخر يسمى الصدق التلازمي Concurrent Validity . ويعتمد هذا التمييز على إمكانية توافر محکات أداء لاحقة أو مستقبلية ، ومحکات أداء راهنة .

وترجع أهمية هذا التمييز إلى أن المدة الزمنية التي تفصل بين تطبيق الاختبار التنبؤى والاختبار أو المقياس المحک تؤثر في درجة العلاقة بين الاختبارين . فالاختبار الذي يمكن باستخدامه تقدير الكفاية الحالية لدى الطلاب أو العاملين ربما لا يستطيع باستخدامه التنبؤ بالأداء المستقبلي أو اللاحق والعكس أيضاً يكون صحيحاً .

والشكل التخطيطي (٢-٥) التالي يوضح هذين النوعين من الصدق :



شكل (٢-٥) يوضح نوعي الصدق المرتبط بمحك

النحوين التنبؤى :

تعملق أدلة الصدق التنبؤى Predictive Validity بتقدير مدى صلاحية الاختبار في التنبؤ بأداء الفرد المستقبلى الذى يقاس باختبار محك باستخدام درجات اختبار يطبق عليه فى الوقت الحاضر ، وبعبارة أخرى تتعلق هذه الأدلة بمدى اقتران تباين درجات اختبار منبئ بتباين درجات اختبار محك ، أى درجة العلاقة بين درجات اختبار ونمط معين من السلوك المستقبلى مما يجعل من الممكن التنبؤ بهذا السلوك . وهذا يتطلب مرور مدة زمنية بين الحصول على الدرجات فى الاختبار التنبؤى ودرجات أو تقديرات الاختبار المحك الذى نقدر فى ضوء صدق القرار .

فإذا وجدنا ارتباطاً مرتفعاً بين اختبار الاستعدادات الفنية الذى يطبق على الطلاب المتقدمين للقبول بكليات الفنون الجميلة والمعدل العام لتقديراتهم فى نهاية السنة الأولى بهذه الكليات ، فإن هذا يمكن أن يقدم دليلاً على الصدق التنبؤى للقرارات المتعلقة بقبول الطلاب للالتحاق بهذه الكليات أو رفضهم .

فالمحك فى هذه الحالة هو المعدل العام للتقديرات كمؤشر لنجاح الطلاب . وربما لا يتفق الجميع على هذا المحك ، فاختبار الاستعدادات الفنية ربما يرتبط ارتباطاً مختلفاً بمحكمات مختلفة .

ولذلك يمكن أن يكون هناك معاملات صدق تنبؤى مختلفة لاختبار واحد ، أى أنه لا يوجد معامل صدق تنبؤى وحيد يمثل صدق مجموعة من المقاييس أو البيانات المستمدة منها .

طرق تقدير الصدق التنبؤى :

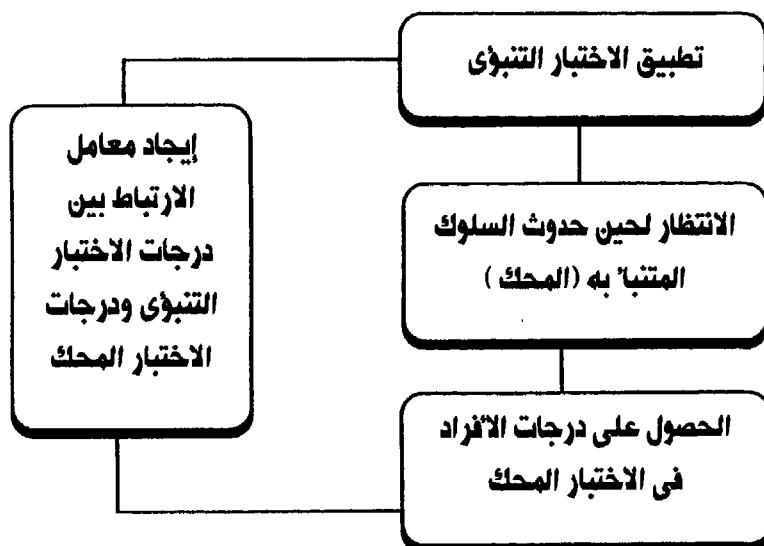
تعدد طرق تقدير الصدق التنبؤى اعتماداً على مجالات استخدام الاختبارات والمقاييس ونوع القرارات التي تسترشد بالبيانات المستمدة من دراسات الصدق، غير أننا سوف نناقش بإيجاز بعض الطرق شائعة الاستخدام في هذا الشأن.

طريقة تعتمد على الارتباط بين الاختبار التنبؤى واختبار المحك،

يمكن تقدير الصدق التنبؤى في المجالات التربوية ومجالات الأعمال والوظائف عن طريق متابعة تطبيق اختبار معين على جميع المتقدمين في شهور مختلفة دون أن يستخدم الاختبار في أغراض الانتقاء وإنما يستخدم في أغراض البحث، بل وأحياناً لا يجري تصحيح الاختبار إلا بعد إلتحق بعض هؤلاء المتقدمين بالعمل.

فالذين تم إلتحقهم يستمرون في العمل مدة زمنية معينة ثم تجمع بيانات باستخدام الاختبار أو المقياس المحك، أو أى أداة لجمع معلومات مناسبة عن هؤلاء المقبولين، وتحسب قيمة معامل الارتباط بين درجات كل من الاختبار التنبؤى والاختبار أو المقياس المحك.

ويمكن توضيح خطوات هذه الطريقة بالشكل التخطيطي (٥ - ٣) التالي:



شكل (٥ - ٣) يوضح خطوات تقدير الصدق التنبؤى للاختبار

فإذا كانت هذه القيمة مرتفعة بدرجة كافية يمكن اعتبار قرارات القبول أو الرفض صادقة، ويصبح الاختبار التنبؤي جزءاً من عملية الانتقاء لشغل وظائف أو أعمال معينة.

وهنا يفترض أن درجة صلاحية الاختبار في الواقع الميداني لا تختلف عن درجة صلاحيته عندما تم تقديم الصدق التنبؤي لأغراض البحث. غير أن هذا الافتراض يصبح غير صحيح إذا حدث تغير ملحوظ في الظروف المحيطة بالعمل، حيث يصبح الصدق التنبؤي عندئذ موضع تساؤل.

ويطلق على معامل الارتباط بين درجات الاختبار التنبؤي ودرجات محك الأداء أو محك النجاح في وقت لاحق معامل الصدق التنبؤي. وينبغي أن تشتمل اختبارات المحك أو أي أدوات أخرى على عينة مماثلة للسمهارات أو الصفات المطلوبة للعمل بحيث يتعلق محتوى الاختبار أو الأداة المحكبة بالسلوك المتوقع.

وينبغي مراعاة أن العوامل التي تؤثر في معامل الارتباط والتي ناقشناها في الفصل الثالث تؤثر أيضاً في معامل الصدق التنبؤي، كما أنه يمكن تفسير قيم معامل الصدق التنبؤي بنفس الطريقة التي فسرنا بها قيم معامل الارتباط في الفصل نفسه، وذلك باستخدام مفهوم التباين المشترك للدرجات كل من الاختبار التنبؤي واختبار المحك الذي نحصل عليه بتربيع قيم معامل الصدق التنبؤي. فإذا كانت قيمة هذا المعامل مثلًا فإن هذا يعني أن ٥٠٪ من تباين درجات الاختبار المحك (أي نصف مقدار التباين) يعزى إلى تباين درجات الاختبار التنبؤي.

وينبغي الإشارة هنا إلى أنه لا ينبع أن نتوقع الحصول على قيم مرتفعة لمعامل الارتباط بين درجات كل من الاختبارين التنبؤي والمحك في دراسات الصدق التنبؤي، وذلك لتعقد وتشابك السلوك الإنساني مما يجعل التنبؤ الدقيق بأداء الأفراد في مختلف المواقف التربوية والعملية باستخدام اختبار تنبؤي واحد أو أكثر أمراً غير واقعي.

ولذلك يرى نانلى Nunnally أنه من الأفضل تفسير قيم معامل الصدق في ضوء التحسن المحتمل في متوسط أداء الأفراد الذي يمكن الحصول عليه باستخدام الاختبار التنبؤي، فالاختبار التنبؤي الذي يرتبط بالمحك ارتباطاً متوسطاً نسبياً مثل ٤٠٪ يمكن أن يفيد في التنبؤ بهذا التحسن المحتمل في بعض المواقف التربوية والعملية، فالتنبؤ بأداء فرد معين يشوبه قدر من الخطأ أكبر من التنبؤ بمتوسط أداء مجموعة من الأفراد.

طريقة تعتمد على استخدام معادلة الانحدار في التنبؤ بدرجات الاختبار المحك

سبق أن أشرنا في الفصل الثالث إلى أنه يمكننا التنبؤ بقيمة أحد المتغيرات إذا علمنا قيمة متغير آخر باستخدام صيغة معادلة خط الانحدار Regression Equation التي تعتمد على قيمة معامل الارتباط بين المتغيرين.

لذلك يمكن استخدام هذه المعادلة في التنبؤ بدرجة الفرد في الاختبار المحك بعمومية درجته في الاختبار التنبؤ اعتماداً على قيمة معامل الصدق التنبؤى. غير أن هذا يتطلب أن تكون العلاقة بين درجات الاختبارين مستقيمة. ويمكن تقدير أخطاء التنبؤ باستخدام صيغة الخطأ المعياري للتنبؤ (١٢ - ٣) التي أشرنا إليها أيضاً في الفصل نفسه.

ويمكن التوصل إلى معادلة خط الانحدار في مرحلة دراسة صدق الاختبار المستخدم في التنبؤ، إذ يمكن استخدام قيمة معامل الارتباط بين هذا الاختبار والاختبار المحك، وكذلك قيمة المتوسط والانحراف المعياري لدرجات كل منهما في التوصل إلى معادلة الانحدار

$$\bar{X}_n = \bar{X}_c + r \times \frac{\bar{X}_c - \bar{X}_n}{S_n}$$

وتشتمل هذه المعادلة على متغيرين أحدهما المتغير المتنبؤ (\bar{X}_n) والأخر المتغير المتنبأ به (\bar{X}_c) ، أما بقية رموز المعادلة وهي (\bar{X}_n) ، (r) ، (\bar{X}_c) ، (S_n) ، ($\bar{X}_c - \bar{X}_n$) ، فإنها يتم حساب قيمة كل منها في مرحلة الدراسة الإمبريقية للصدق.

ويلاحظ أن (\bar{X}_c) ترمز لمتوسط درجات عينة الأفراد المستخدمة في دراسة الصدق ، (S_n) ترمز لمتوسط درجات اختبار المحك المتنبأ بها من هذه العينة التي تكون درجات أفرادها في الاختبار التنبؤى (\bar{X}_n) .

ونظراً لأننا نتوقع عند استخدام الاختبار التنبؤى فيما بعد على عينات أخرى مماثلة أن تختلف درجاتهم المتنبأ بها (\bar{X}_n) عن الدرجات المتنبأ بها لأفراد عينة دراسة الصدق، لذلك فإنه يمكن تقدير هذه الأخطاء باستخدام صيغة الخطأ المعياري للتنبؤ $\bar{X}_n - \bar{X}_c$. وقد سبق أن أوضحتنا كيفية تفسير القيمة التقديرية لهذا الخطأ في الفصل الثالث.

ويمكن استخدام معادلة خط الانحدار بعد ذلك في التنبؤ بدرجات أفراد عينات أخرى مماثلة للعينة التي استمدت منها بيانات هذه المعادلة في الاختبار المحك ، بعمومية درجاتهم في الاختبار التنبؤى .

وكما سبق أن ذكرنا فإننا نتوقع أن تكون القيم التقديرية للخطأ المعياري للتنبؤ كبيرة وذلك لاعتمادنا على اختبار تنبؤ واحد. ولذلك ربما يفضل استخدام أكثر من اختبار تنبؤ واحد، وضم درجاتها معاً بأوزان مختلفة تناسب وأهميتها في التنبؤ بدرجات اختبار المحك.

ولإجراء ذلك نوجد معامل الارتباط المتعدد **Multiple Correlation** بين التركيب الخطى لدرجات الاختبارات التنبؤية ودرجات اختبار المحك، أو استخدام أسلوب الانحدار الخطى المتعدد **Multiple Linear Regression** فى تحديد هذه الأوزان. وقد وضع المؤلف هذه الأساليب الإحصائية وكيفية استخدامها فى مرجع آخر.

غير أنه ينبغي فى هذه الحالة مراعاة عدم استخدام عدد كبير من الاختبارات التنبؤية، والاقتصار على اختبارين أو ثلاثة، نظراً لأن كثرة الاختبارات المرتبطة بالاختبار المحك لا تسهم إسهاماً ملحوظاً فى تفسير تباين درجات هذا الاختبار أكثر مما تسهم به ثلاثة اختبارات مثلاً. ولكن ينبغي مراعاة أن يكون الارتباط بين الاختبارات التنبؤية منخفضاً، وارتباطها باختبار المحك مرتفعاً لكي تسهم فى التنبؤ إسهاماً جيداً، وبذلك تكون أكثر جدوى فى التنبؤ ولا تتطلب وقتاً أو جهداً كبيراً فى التطبيق.

وعلى الرغم من أن الاختبارات التنبؤية تشتمل على مفردات يفترض أنها تميز بين المجموعات أو تتبأ بالنتائج، إلا أن المفردات المستقلة التى تميز بين أفراد مجموعة معينة ربما لا تميز بين أفراد مجموعة أخرى، فالارتباط بين درجات اختبار تنبؤ معين ودرجات اختبار محك فى المجموعة الأصلية التى استخدمت فى بناء الاختبار التنبؤى لا تؤدى إلى معاملات صدق مرتبطة بمحلك معقولة ولكن تذبذب من مجموعة إلى أخرى. لذلك يفضل إجراء دراسة صدق مستعرض **Cross Validation** على عينة جديدة مستقلة من الأفراد تكون مستمدة من المجتمع الذى سيطبق عليه الاختبار فيما بعد. وهذا الأسلوب يهدف لإعادة فحص بيانات الصدق على المجموعة الجديدة، فالمفردات التى كانت مميزة نتيجة للصدفة سوف تجد أنها أصبحت غير مُميزة عند إعادة تطبيق الاختبار.

أما إذا بینت دراسات الصدق أن النتائج متماثلة في المجموعتين، فإن هذا يعد دليلاً على العلاقة بين الاختبار المبني والمحلك، ولا ترجع إلى عوامل الصدفة. ولكن إذا كانت النتائج مختلفة فإنه ينبغي بناء مفردات جديدة ترتبط بالمحلك.

بعض المشكلات المتعلقة بالمحك :

تعتمد دقة تقدير الصدق المرتبط بمحك على نوع المحك الذى يستخدم في مجالات معينة. و اختيار المحك المناسب والاختبار أو المقياس الذى يزودنا بدرجات أو تقديرات أو معلومات عن سلوك الفرد أو أدائه فى موقف معين يعد من المشكلات الأساسية التى تواجه القائم بإعداد الاختبارات التنبؤية والمسئولين عن اتخاذ قرارات الانتقاء والتصنيف والتسكين وربما التشخيص أيضاً. فالمحاكمات تختلف باختلاف المجال موضوع الاهتمام، إذ ربما تستخدم درجات الاختبارات التحصيلية التى تطبق في نهاية مدة دراسية معينة، أو تقديرات المعلمين حول سلوكيات تلاميذهم وإنجازاتهم، أو المعدل العام كمحاكمات لتقدير الصدق التنبؤى للختبارات فى مجالات انتقاء الطلاب أو تسكينهم فى برامج تعليمية مناسبة. وربما تستخدم تقديرات المشرفين على العمل، والتقييم الذانى، ومؤشرات الإنتاج وسجل الخدمة، وتقديرات الدورات التدريبية فى مجالات انتقاء العاملين أو إلحاقيهم بوظائف أو مهن أو أقسام مناسبة، أو ترقيتهم إلى وظائف أو مهن عليا.

واستخدام محك مناسب يتطلب توافر بعض الشروط الأساسية، من أهمها كما سبق أن ذكرنا مواعيده لنوع القرار المراد اتخاذه ولطبيعة السلوك المراد التنبؤ به. فالسلوك الذى يتضمنه المحك مثل تقديرات المعلمين فيما يتعلق بتحصيل تلاميذهم لا تكون مقياساً موائماً لنجاح التلاميذ إلا إذا عكست هذه التقديرات ما يتضمنه النجاح من خصائص معينة، إذ ربما يختلف مفهوم النجاح من زمان إلى آخر، فالطالب الذى يعد ناجحاً الآن ربما لا يكون كذلك مستقبلاً، وهذا يعتمد على ما فصله بمفهوم النجاح وما يتضمنه من خصائص.

فعندما يستند تقدير النجاح في هذه الحالة على تقديرات المعلمين، فإن هذا المحك سوف يكون متحيزاً. وللتغلب على مشكلة تحيز المحكمات وبخاصة التي تعتمد على الأحكام القيمية والتقديرات ينبغي تحديد العناصر والمكونات الأساسية التي ينبغي أن تسترشد بها هذه الأحكام تحديداً إجرائياً للتقليل من الذاتية وتأثير الهالة Halo Effect أى الأفكار المسبقة عن الفرد أو عن درجته في الاختبار التنبؤى، وجميع عوامل التحيز قدر الإمكان.

ومن المشكلات الأساسية أيضاً المتعلقة بالمحاكمات ثبات درجات المحك، فالاختبار أو المقياس الذى يستخدم فى تقدير الصدق التنبؤى يفترض أن درجاته تتسم بالثبات أو الاتساق، غير أن واقع الأمر ليس كذلك. فكثير من مقاييس المحك تعتمد

على التقارير الذاتية، والتقديرات، والمقابلات الشخصية، وجميعها تعتمد على الأحكام الشخصية بدرجات متفاوتة، وهذه المقاييس تؤثر فيها كما ذكرنا عوامل تحيّز متعددة مما يقلل من ثبات درجاتها.

ويكن استخدام بعض الأساليب الإحصائية التي تساعد في تصحيح الأخطاء التي ربما تكون قد أدت إلى خفض معامل ثبات درجات كل من الاختبار التنبؤي واختبار المحكّم. ومن بين هذه الأساليب الصيغة الإحصائية التالية:

أقصى، معامل، ارتباط بين الاختيار التئوي واختيار المحك =

معامل الصدق التنبؤ
معامل، ثبات الاختبار التنبؤي \times معامل، ثبات الاختبار المحك
 $(1-5) \dots$

فإذا كان معامل الصدق التنبؤى للاختبار التنبؤى .٣٠ ، مثلاً، ومعامل ثبات درجاته .٨٥ ، ومعامل ثبات درجات المحك .٥٠ ، فإن أقصى قيمة لمعامل الارتباط بين درجات كل من الاختبارين =

$$\therefore V1 = \frac{\cdot , V \cdot}{\therefore o : x : A o}$$

وهذه القيمة تساوى قيمة معامل الارتباط بين درجات الاختبار التنبؤى ودرجات اختبار المحك إذا افترضنا أن درجات كل منها غير مشوبة بأى نوع من الخطأ، أي أن ثبات درجات كل منها تام. ولكن ينبغي أن يميز دليل الاختبار بين القيمة الأصلية لمعامل الصدق وقيمة المتوقعة بعد إجراء عملية التصحح.

والصيغة (٥ - ١) تبين القيمة المتوقعة لعامل الارتباط بين درجات الاختبارين إذا ارتفعت قيم معامل ثبات درجات الاختبار التنبؤي والاختبار المحك. فإذا لم نجد ارتفاعاً ملحوظاً في القيمة المتوقعة مما كانت عليه بالفعل، فإنه ينبغي إعادة النظر في الاختبار التنبؤي أو في طريقة بنائه، نظراً لأنه في هذه الحالة لا يعزى إليه نسبة كبيرة من تباين الدرجات الحقيقة للاختبار المحك.

طريقة تعتمد على جداول التوقع :

تستخدم هذه الطريقة في وصف العلاقة بين درجات اختبار تنبؤي ودرجات اختبار المحك. وتعتمد على جداول يمكن تكوينها تسمى جداول التوقع Expectancy Tables ، وتنظم بشكل يسمح بتقدير أرجحية حصول فرد على درجة أو تحقيقه لمستوى معين من الكفاية في اختبار محك إذا علمنا درجته في اختبار تنبؤي. أي أن هذه الجداول تفيد في تقدير احتمالات نجاح الأفراد الذين حصلوا على درجات معينة في الاختبار التنبؤي. فمشكلة انتقاء الأفراد تتعلق بالتبؤ بمستوى معين من مستويات الكفاية أو النجاح في دراسة أو مهنة معينة، أي أن التنبؤ هنا يكون تصنيفياً. فالملعلم أو المسؤول عن العمل لا يستطيع على وجه التحديد معرفة ما إذا كان الطالب أو العامل سيؤدي أداء جيداً في دراسته أو مهنته، ولكن البيانات المتعلقة بالصدق التنبؤي يمكن أن تساعد في تقدير أرجحية ذلك استرشاداً بالنسب المئوية أو التكرارات النسبية لعدد الحالات التي حققت مستويات نجاح معينة والمدونة في جداول التوقع الناتجة عن دراسات الصدق.

ولتكوين جدول التوقع ينبغي الحصول على توزيع درجات كل من الاختبار التنبؤي واختبار المحك كما سبق أن أوضحنا، وتقسيم درجات كل من الاختبارين إلى عدد من الأقسام . ويفضل أن تكون هذه الأقسام متساوية في عددها، وليس في مدى درجاتها .

ونوجد عدد الأفراد الذين يتسمون إلى كل قسم من هذه الأقسام، وكذلك التكرارات النسبية أو النسب المئوية من الأفراد في كل قسم. ثم تكون جدول التوقع الذي يوضح العلاقة الملاحظة بين النسب المئوية من الأفراد في كل قسم من أقسام اختبار المحك لكل فئة من فئات درجات الاختبار التنبؤي .

ويمكن إجراء ذلك بتمثيل فئات درجات الاختبار التنبؤي على المحور الأفقي ، وأقسام درجات اختبار المحك على المحور الرأسى، وتدوين النسب المئوية من الأفراد الذين يتسمون إلى فئة معينة من فئات درجات الاختبار التنبؤي ويتمون في الوقت نفسه إلى قسم معين من أقسام درجات اختبار المحك. ويمكن توسيع ذلك بالمثال الافتراضي التالي :

نفترض أن الاختبار التنبؤي هو اختبار القبول بالكلليات واختبار المحك هو المعدل التراكمي للطالب في نهاية السنة الأولى الجامعية، ونفترض أنه أجريت دراسة الصدق التنبؤي للبيانات المستمدة من هذين الاختبارين ، فعندئذ يمكن تكوين جدول

التوقع وفقاً للمخطوطة السابقة، وذلك بتمثيل فئات درجات اختبار القبول على المحور الأفقي ، وأقسام أو فئات المعدل التراكمي على المحور الرأسى ، وتكون جدول توزيع تكرارى لعدد الطلاب فى كل خلية من خلايا الجدول (٥ - ١) التالى :

نسبة القبول (اختبار القبول)	درجات الاختبار	درجات اختبار المحك (المعدل التراكمي)				المجموع
		١,٠	٢,٠	٣,٠	٤,٠	
٦٤-٦٠	٢		٥	١	٨	
٥٩-٥٥	١	١	٦	١	٩	
٥٤-٥٠	٢	٥	٧	٢	١٦	
٤٩-٤٥	٥	٤	٤	٢	١٥	
٤٤-٤٠	٢	٤	٧	٣	١٦	
٣٩-٣٥	٢	٧	٦	١	١٦	
٣٤-٣٠	١	٢	٣	٣	٩	
٢٩-٢٥	٢	٤	١		٧	
٢٤-٢٠	١	٣			٤	
المجموع		٨	٢٨	٢٩	٢٩	٦
						١٠٠

جدول (٥ - ١) جدول توزيع تكرارى للمعدل التراكمي لكل فئة من فئات درجات اختبار القبول

والخطوة الثالثية : نحوال عدد الطلاب فى كل خلية إلى نسب مئوية من العدد الكلى لطلاب كل فئة من فئات درجات اختبار القبول كما فى جدول (٥ - ٢) التالى :

درجات الاختبار التبؤى (اختبار القبول)	درجات اختبار المحك (المعدل التراكمي)	المجموع					
		صفر	١,٠	٢,٠	٣,٠	٤,٠	
٦٤-٦٠			٢٥		٦٣	١٢	١٠٠
٥٩-٥٥		١١	١١		٦٧	١١	١٠٠
٥٤-٥٠		١٣	٣١		٤٤	١٢	١٠٠
٤٩-٤٥		٣٣	٢٧		٢٧	١٣	١٠٠
٤٤-٤٠	١٢	٢٥	٤٤		١٩		١٠٠
٣٩-٣٥	١٢	٤٤	٣٨		٦		١٠٠
٣٤-٣٠	١٢	٢٢	٣٣		٣٣		١٠٠
٢٩-٢٥	٢٩	٥٧	١٤				١٠٠
٢٤-٢٠	٢٥	٧٥					١٠٠

جدول (٥ - ٢) جدول توقع للمعدل التراكمي لكل فئة من فئات

درجات اختبار القبول

وبالنظر إلى جدول (٥ - ١) يتضح أن ١٦ طالبًا حصلوا على درجات في اختبار القبول تتراوح بين ٤٠ ، ٤٤ ، والمعدل التراكمي لطالبين منهم (صفر)، ولأربعة طلاب (١،٠) ولسبة طلاب (٢،٠) ولثلاثة منهم (٣،٠).

ويوضح جدول (٥ - ٢) النسب المئوية لعدد طلاب كل فئة من فئات درجات اختبار القبول الم対اظرة لمستويات المعدل التراكمي. وإذا نظرنا إلى الخلايا الم対اظرة للأعداد السابقة نجد أنها تنازل النسب المئوية ١٢ ، ٢٥ ، ٤٤ ، ١٩ على الترتيب.

ويمكن تفسير هذه النسب على أنها احتمالات حصول طلاب على معدل تراكمي معين إذا علمتنا درجاتهم في اختبار القبول.

لذلك يمكننا القول بأن ١٢ % من الطلاب الذين تتراوح درجاتهم في اختبار القبول بالكليات بين ٤٠ ، ٤٤ يتوقع أن يحصلوا في نهاية السنة الجامعية الأولى على معدل تراكمي (صفر)، ٢٥ % منهم يتوقع أن يحصلوا على معدل تراكمي (١)، وهكذا.

وإذا اعتبرنا الطلاب الناجحين هم الذين يزيد معدلهم التراكمي عن (٠،٢)، بينما يُعد راسباً من يقل معدله التراكمي عن (٠،٢)، فالنظر إلى جدول (١ - ٥) يلاحظ أن: $٣٥ = ٢٨ + ٧$ طالباً من بين المجموع الكلي للطلاب (١٠٠ طالب) يُعدون ناجحين، فإذا تم قبول العدد الكلي للطلاب فإنه يتوقع أن ينجح منهم ٣٥ طالباً، أي ٣٥٪، أما النسبة الباقيّة وهي ٦٥٪ فيتُوقع رسوبيهم. ويوضح ذلك الخط الرأسى الفاصل عند المعدل التراكمي (٠،٢) في الجدول.

وإذا تم قبول جميع الطلاب الذين حصلوا على الدرجة ٤٥ وما فوقها في اختبار القبول، فإن عدد المقبولين يكون $٤٨ = ٢٠ + ٢٨$ طالباً، ويُتوقع أن ينجح منهم ٢٨ طالباً. ويوضح ذلك المربعين العلوين الأيمن والأيسر، ونسبة هؤلاء الطلاب الناجحين إلى عدد المقبولين تسمى نسبة النجاح Success Ratio، أي $\frac{٢٨}{٤٨} = ٥٨\%$. استناداً إلى درجتي القطع المختارتين، أي ٤٥، ٠٢.

وإذا قسمنا عدد الطلاب المقبولين الذين تفوق درجاتهم الدرجة ٤٥ في اختبار القبول (٤٨) على العدد الكلي للطلاب (١٠٠)، أي ٤٨٪، فإن هذه النسبة تسمى نسبة القبول Success Ratio. وينبغي ملاحظة أنه كلما زاد عدد المقبولين فإن نسبة القبول تنخفض، كما ينبغي ملاحظة أنه إذا كان معامل الصدق تماماً أي مساوياً الواحد الصحيح فإن جميع القيم في خلايا الجدول تتراكم بحيث تقع على خط مستقيم، وعندئذ يمكن التنبؤ بدرجة المحك بمعلومية درجة اختبار القبول تنبؤاً تماماً بغض النظر عن موقع درجة القطع، وبذلك تكون نسبة النجاح واحداً صحيحاً، أي لا يحتمل رسوبي أي طالب تم قبوله، أما من رفض قبوله فإننا تتوقع أنه كان سيرسب إذا تم قبوله عن طريق الخطأ.

أي أن هناك علاقة تفاعلية بين صدق التنبؤ ونسبة النجاح ونسبة القبول. وقد وضع هذه العلاقة تايلر وراسل Tayler - Russell في جداول تتضمن قيمًا مختلفة لكل من الصدق ونسبة النجاح ونسبة القبول. ويمكن باستخدام هذه الجداول تقدير نسبة الأفراد الناجحين الذين يقبلون استناداً إلى اختبارات تنبؤية متباينة في درجة صدقها، ونسب قبول مختلفة.

ونظراً لأن كلاً من نسبة النجاح ونسبة القبول تعتمد على درجة القطع في الاختبار التنبؤى ، فإنه ينبغي مراعاة التصنيف الخاطئ الذي يترتب على هذه الدرجة عند تقييم فاعلية اختبار كأدلة فحص. ونعني بالتصنيف الخاطئ رفض قبول فرد كان من الممكن نجاحه إذا التحق بالدراسة أو بالعمل ويسمى الخطأ الموجب False Positive، وكذلك نسب النجاح والرسوب لمجموعة الأفراد.

وتحتختلف هذه النسب باختلاف طبيعة الدراسة أو العمل الذي سيوكىء إلى الأفراد المقبولين. بعض أنواع العمل تتطلب كفايات مرتفعة وبالتالي تكون درجة القطع مرتفعة ارتفاعاً نسبياً مما يؤدي إلى الخطأ الموجب، وهنا يمكن أن نقلل من هذا الخطأ بخفض درجة القطع. ولكن هناك أعمال ربما تستدعي قبول عدد كبير من الأفراد على الرغم من احتمال فشل بعضهم وهو ما يسمى بالخطأ السالب . False Negative

وعموماً، فإن موقع درجة القطع يتحدد بناء على حاجة العمل من الأفراد العاملين، وعدد الأماكن الشاغرة ومدى الحاجة الملحة لشغلها.

ويمكن للمرشدين التربويين في المدرسة الثانوية تكوين جداول توقع مثل جدولى (١ - ٥) ، (٢ - ٥)، لمساعدة الطلاب في اتخاذ قرارات تتعلق بالكلبات التي يزمعون الالتحاق بها مع مراعاة الأخطاء التي تنتجه عن التنبؤ باستخدام هذه الجداول كما هو الحال في أخطاء الانحدار.

ولتفسير درجة أحد الطلاب نحدد موقعها في العمود الأول للفئات في جدول التوقع (١ - ٥) ، ونقرأ النسبة المئوية المدونة في الخلية الناتجة من تقاطع الصف الذي تقع فيه درجته في اختبار القبول مع العمود الذي يناظر المعدل التراكمي المتوقع، وذلك لتقدير احتمال حصوله على هذا المعدل في نهاية السنة الجامعية الأولى .

ومن الجدير بالذكر أنه يمكن تكوين جداول توقع لاختبارين من اختبارات التنبؤ بدلاً من اختبار واحد، فمثلاً يمكن استخدام اختبار القبول ودرجات نهاية العام في الثانوية العامة للتنبؤ بالمعدل التراكمي للطالب في نهاية السنة الجامعية الأولى ، واتباع الخطوات السابقة في تكوين جدول التوقع .

غير أنه في هذه الحالة ينبغي تحديد درجة قطع لكل اختبار على حدة بحيث يمكن استبعاد الأفراد من ذوى الدرجات المنخفضة في أي اختبار؛ نظراً لأن الدرجة المركبة من درجات الاختبارات تعوض بعضها البعض الآخر .

والخلاصة أن جداول التوقع تتضمن بيانات تتعلق بالصدق التنبؤى ، ويستطيع المسؤولون عن انتقاء الأفراد استخدامها في تفسير درجات الاختبارات التنبؤية بعبارات احتمالية في ضوء نواتج معينة متوقعة .

ولكن يعاب على هذه الجداول استخدام فئات درجات الاختبار التنبؤى لكل قسم من أقسام درجات المحك مما يقلل من الدقة فى التفسير ، كما أن بيانات هذه الجداول تستمد من عينات من الأفراد فى الماضى . وربما تختلف خبراتهم عن خبرات الأفراد الذين نتبأ بدرجاتهم فى المحك بعد ذلك . وهذا يتطلب الحيطة والنظرية الفاحصة عند

تفسير البيانات المدونة في هذه الجداول ؛ نظراً لأن الاختبارات التنبؤية لا تقيس خصائص مهمة أخرى لدى الأفراد مثل المثابرة ، والمبادرة ، والدافعية ، مما قد يؤثر في نتائج التوقع .

كذلك تُعتمد هذه الجداول عادة من عينة عدد أفرادها قليل نسبياً مما يجعل هناك خلايا فارغة أو قليلة العدد كما في الجدول (٥ - ١) ، مما يؤدي إلى عدم اتساق قيم الاحتمالات .

ونظراً لأن هذه الجداول استمدت من مجموعة من الأفراد ، فإن تفسير هذه القيم ينطبق على المجموعة ككل . ففي الجدول السابق وجدنا أن ١٩٪ من طلاب المجموعة الذين تراوح درجاتهم بين ٤٠ ، ٤٤ يتوقع حصولهم على معدل تراكمي (٣) ، ومعنى ذلك أنه إذا تم قبول الطلاب الذين تراوح درجاتهم في اختبار القبول بين ٤٠ ، ٤٤ فإننا نتوقع أن ١٩٪ من هؤلاء الطلاب سوف ينجحون في دراستهم بالكلية ، فإذا حصل أحد الطلاب على درجة في اختبار القبول تراوح أيضاً بين ٤٠ ، ٤٤ فهو إما سيكون ناجحاً أو راسباً ، ويصبح احتمال نجاحه ١٩٪ غير ذي معنى . فالاحتمال ١٩٪ يشير إلى توقع النجاح لمجموعة الأفراد الذين حصلوا على درجات تنحصر بين درجتين معيتين في اختبار القبول ، أي أن التوقع أو التنبؤ هنا لا يتعلق بكل فرد على حدة ، وإنما يتعلق بمتوسط المجموعة ؛ إذ ربما لا يؤدي الأفراد أداء يتفق وتوقعاتنا ، وإنما ربما نجد بعضهم يؤدي أداء أفضل مما توقعنا ، لذلك لا يجب أن نخبرهم بتوقعنا عن أدائهم المنخفض .

ومع هذا فإن استخدام جداول التوقع يسهم في ترشيد قرارات الانتقاء بدرجة أفضل مما لو تم الانتقاء دون استخدام أساليب تنبؤ علمية .

طريقة تعتمد على نظرية المنفعة :

اتضح لنا أن جداول التوقع تفيض في ترشيد قرارات انتقاء الأفراد ، كما اتضح كيف تتأثر هذه القرارات بدرجة الصدق التنبؤى لأدوات القياس المستخدمة وبنسبة القبول ونسبة النجاح ، غير أن هذه القرارات تتأثر أيضاً بالمنفعة النسبية Utility للنواتج المتوقعة وما يتعلق بهذه النواتج من كلفة وعائد . فنظرية القرارات تستند إلى نظام لتعيين قيم لمختلف النواتج من حيث المنفعة النسبية ، وهذا يتطلب تحليلًا منظماً لعلاقة النواتج بكلفتها وعائدها ، وهذا التحليل يعد ذى فائدة كبيرة في مجال انتقاء الأفراد باستخدام الاختبارات والمقاييس التنبؤية .

فكثير من المؤسسات تُعنى بانتقاء الأفراد من خلال برامج تختص بعمليات الفحص والانتقاء، وتتطلب هذه البرامج تمويلاً لإعدادها وتطويرها مستمراً. وعندما يُقبل فرد ويتحقق بالعمل فإنه يؤدي منفعة للمؤسسة إذا كان ناجحاً مما يعد بمثابة عائد للمؤسسة، أما إذا لم يتبيّن نجاحه فإنه بذلك يعد بمثابة فاقد للمؤسسة. وفي جميع هذه الأحوال تبدو أهمية تحليل كلفة وعائد القرار الذي يتخذ وما يترتب عليه.

وهذا يتطلب تقدير قيم المنفعة لكل من هذه القرارات متضمنة الكلفة المالية والعائد. فمثلاً عائد العامل الناجح على المؤسسة ربما يتم تقديره بطرح المرتب الذي يتلقاه وكلفة الأجهزة أو الآلات التي يستخدمها في العمل من جملة عائد إنتاجه. وربما تدخل أيضاً بعض المتغيرات الأخرى مثل تأثيره في إنتاجية زملائه، أو مدة انقطاعه عن العمل بسبب ظروف طارئة، وهكذا. أما كلفة العامل الذي يتبيّن عدم فاعليته فإنه يمكن تقديرها من كلفة برنامج الانتقاء، وكلفة إلحاد فرد آخر للقيام بعمله، وفاقد الإنتاجية التي ترتب على ذلك.

والخلاصة أن تقدير الصدق التنبؤى لقرارات الانتقاء ينبغي أن يأخذ بعين الاعتبار كلفة وعائد البرنامج الاختباري وما يترتب عليه من منفعة للمؤسسة، ومحاولة ترجمة ذلك اقتصادياً، مما يجعل مفهوم الصدق التنبؤى له دلالة عملية لصانعى القرار.

الصدق التلازمى وطرق تقديره :

سبق أن ميزنا بين نوعين من الصدق المرتبط بمحك أحدهما الصدق التنبؤى Predictive Validity الذى ناقشناه فى الجزء السابق ، والصدق التلازمى Concurrent Validity ، وذكرنا أن الصدق التنبؤى يتعلق بدرجة اقتران تباين درجات اختبار تنبؤى بتباين درجات اختبار محك يطبق بعد مرور مدة زمنية معينة، تفصّل بين تطبيق كل منها.

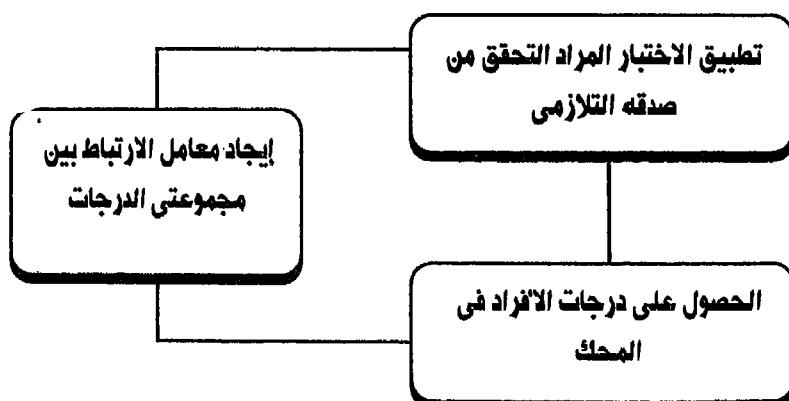
أما الصدق التلازمى فيتعلق بدرجة اقتران تباين درجات اختبار بتباين درجات اختبار آخر يطبق فى الوقت نفسه تقريباً، أى أن الصدق التنبؤى يهتم بالتنبؤ، أما الصدق التلازمى فيهتم بالوصف. فمثلاً يمكن مقارنة درجات الاختبارات أو المقاييس التى تطبق بهدف التشخيص الكلى بى مع أحکام الأطباء النفسيين الذين يجرؤون مقابلات مع المرضى، أو يمكن مقارنة الكفاية فى مهنة معينة بتقديرات المشرفين على العمل، كذلك يمكن للمعلم أن يقارن درجات اختبار تحصيلي مقتنن فى مجال دراسى معين بدرجات اختبار تحصيلي يُعده بنفسه لتلاميذه فى المجال المعين ، أو يمكن مقارنة

درجات اختبار جديد تم بناؤه لقياس سمة معينة بدرجات اختبار معترف به، أى تم التتحقق من صدقه بأدلة متعددة. ففى هذه الحالات يتركز الاهتمام على الصدق التلازمى الذى يهدف لتحديد ما إذا كان من الممكن الاستعاضة عن درجات أحد الاختبارات بدرجات اختبار آخر.

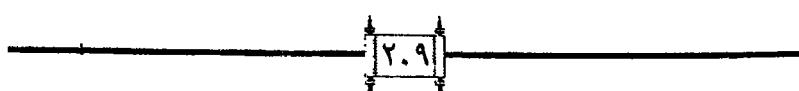
وهذا يتطلب أن تكون هناك علاقة بين الاختبارين، وأن يكون الاختبار البديل متاحاً، ويتميز باتساق درجاته وسهولة تطبيقه بكلفة منخفضة. فالصدق التلازمى إذن يفيد في إحلال اختبار أو مقياس محل آخر، أما الصدق التبؤى فيستخدم درجات اختبار أو مقياس معين في التنبؤ بدرجات اختبار محك؛ ولذلك فإنهما يمثلان جانباً متكملاً للصدق المرتبط بمحيث.

ويستخدم الصدق التلازمى بدرجة أكبر في مجالات الصناعة وانتقاء الأفراد والتشخيص الكلينيكي، فمثلاً يمكن تحليل العمل Job Analysis لتحديد المهارات أو السلوك الذي يتطلبه، وتقدير درجة تملك الفرد الذي يود الالتحاق بهذا العمل لعيوب من السلوك مماثلة لهذا المهارات.

وقد نال هذا الأسلوب اهتمام المشغلين بعلم النفس الصناعي في مجال انتقاء العاملين في المهن المختلفة، حيث بيّنت دراساتهم ارتباط مثل هذه العينيات من السلوك بالأداء الفعلى في العمل ارتباطاً مرتفعاً. لذلك تستخدم هذه العينيات السلوکية في فحص وانتقاء العاملين في مهن معينة استناداً إلى دراسات الصدق التلازمى.
ويوضح الشكل التخطيطي (٥ - ٤) التالى خطوات تقدير الصدق التلازمى:



شكل (٥ - ٤) يوضح خطوات تقدير الصدق التلازمى للاختبار



إذ يمكن التتحقق من الصدق التلازمي بتطبيق اختبار مناسب مثل اختبار كفاية، أو اختبار تحصيلي، أو اختبار معلومات متعلقة بعمل معين على مجموعة من الأفراد العاملين بالفعل، وفي الوقت نفسه تقريرياً نطبق اختبار محك يقيس مهاراتهم الفعلية في العمل، ونقدر الصدق التلازمي بإيجاد قيمة معامل الارتباط بين درجات كل من الاختبارين.

كذلك يستخدم الصدق التلازمي عندما نطبق اختبار معين على أفراد تم بالفعل تشخيصهم تشخيصاً كلينيكياً؛ وذلك لمعرفة ما إذا كانت متوسطات درجات المجموعات المتباعدة تختلف فيما بينها اختلافاً دالاً. فإذا بینت النتائج اختلاف قيم متوسطات الدرجات اختلافاً كبيراً من مجموعة إلى أخرى، فإن هذا الاختبار يمكن أن يكون بدليلاً مناسباً لأغراض تصنيف الأفراد.

فمثلاً استبيان الشخصية متعدد الأوجه (MMPI) يُعد مقياساً جيداً لتصنيف الأضطرابات العقلية، حيث بینت الدراسات المتعلقة بهذا المقياس أن كثيراً من حالات الأضطرابات العقلية التي يتم تشخيصها بواسطة الأخصائيين النفسيين تتميز درجاتهم التي يحصلون عليها في مجموعات معينة من مفردات المقياس (بعض المقياس الفرعية) بنطء معين.

كذلك يمكن أن يقدم الصدق التلازمي معلومات مفيدة للمعلم تتعلق بالاختبارات الصافية التي يقوم بإعدادها، أو الاختبارات التحصيلية المقنتة، ومدى مواءتها للأهداف التعليمية التي يهتم بتحقيقها لدى تلاميذه. إذ يمكن أن يتحقق المعلم مما إذا كانت الاختبارات التحصيلية التي يُعدها في مجال دراسي معين تميز بالفعل بين الطلاب الأقوية والطلاب الضعاف، وذلك بإيجاد الارتباط بين درجات كل اختبارين على حدة، أو بين الدرجات الكلية للختبارات معاً وبين درجات اختبار تحصيلي مقتن يفي بخصائص الاختبارات الجيدة ويتعلق بال المجال الدراسي نفسه. كما يمكنه أن يقدر العلاقة بين مجموعة الدرجات باستخدام الشكل الاستشاري الذي أوضحتنا عند مناقشتنا لمفهوم الارتباط في الفصل الثالث. فعندئذ يمكنه تمثيل درجات الاختبار التحصيلي المقتن على المحور الأفقي، ودرجات الاختبار الذي أعده بنفسه على المحور الرأسى، ووضع علامات تقابل أزواج درجات كل من الاختبارين لمجموعة الطلاب.

ويمكن توضيح ذلك للدرجات افتراضياً بشكل (٥ - ٥) التالي:

			XXX X	XXX XX	
درجات					
اختبار التحصيل		XX	XXXX XX	XXX	
المقى	XX	XXXX XXXX			
٢٤-٢٠					
١٩-١٥	X	X	XX		
١٤-١٠	XX	XXX			
٩-٥	X				
	٩٤	١٥-١٠	٢١-١٦	٢٧-٢٢	٣٣-٢٨

درجات الاختبار الذى أعده المعلم
شكل (٥ - ٥) يوضح انتشار درجات الاختبارين تفصيلياً

فمعظم العلامات في الشكل ينبغي أن تقع في الربع الأدنى الأيسر الذي يحتوي درجات الأفراد الصعاف في كل من الاختبارين، وكذلك في الربع الأعلى الأيمن الذي يحتوى درجات الأفراد الأقوباء في كل من الاختبارين، أما بقية الدرجات فتقع في الربعين الباقيين.

ولكن النزعة العامة للعلاقات ينبغي أن تتجه من الربع الأدنى الأيسر إلى الربع الأعلى الأيمن لكي يدل على نزعة موجبة مما يقدم مؤشراً جيداً للصدق التلازمي. وإذا لم يحدث ذلك فإن الصدق التلازمي يكون منخفضاً، وعندئذ ينبغي على المعلم أن يعيد النظر في الاختبار الذي أعده، أو يتحقق من أن الاختبار التحصيلي المقى مناسباً ومرتبطاً بالأهداف التعليمية التي يود قياسها.

كما أن الصدق التلازمي يمكن أن يفيد المعلم نظراً لصعوبة تقدير الصدق التنبؤ لل اختبارات الصافية . فالمعلم يهتم عادة بالسلوك المستقبلي تلاميذه ، وبالطبع لا يستطيع إجراء دراسات تتبعية بعيدة الأمد لتقدير مدى صدق الأساليب التعليمية التي يتبعها ؛ لذلك يمكنه تقدير صدق هذه الأساليب من خلال سلوك تلاميذه في الوقت الراهن. فمن المعلوم أن درجة العلاقة بين متغيرين تقل بمرور الزمن ؛ لذلك ترتبط

اختبارات الاستعدادات بالتحصيل الدراسي في الوقت الراهن بدرجة أكبر من ارتباطها به بعد مرور عدة سنوات.

لذلك تتوقع انخفاض قيمة الصدق التنبؤي عن قيمة الصدق التلازمي، فقيم معامل الصدق التلازمي تُعد قيمًا قصوى لمعامل الصدق التنبؤي، وهذا يعني أنه إذا كانت قيمة معامل الصدق التلازمي ٦٥٪، فإننا تتوقع أن يكون الصدق التنبؤي أقل من ذلك إذا استخدم اختبار المحك نفسه في الحالتين.

بعض العوامل التي تؤثر في الصدق المرتبط بمحك :

تتأثر قيم معامل الصدق المرتبط بمحك بعوامل متعددة أشرنا إلى بعض منها في سياق مناقشاتنا السابقة. وسوف نوضح فيما يلي مزيداً من هذه العوامل التي ينبغي على مستخدمي الاختبارات والمقاييس والمسئولين عن اتخاذ قرارات الانتقاء والتسكين مراعاتها عند تفسير البيانات المتعلقة بهذا النوع من الصدق.

مدى تجانس عينة الأفراد: نظرًا لأن الصدق المرتبط بمحك يعتمد على معامل الارتباط بين درجات الاختبار التنبؤي واختبار المحك، فإن جميع العوامل التي تؤثر في قيم معامل الارتباط ومعامل الثبات التي أوضحتناها في الفصلين الثالث والرابع تؤثر أيضًا في قيم معامل الصدق. فكلما كانت المجموعة متتجانسة انخفضت قيمة معامل الصدق المرتبط بمحك، وكلما كانت غير متتجانسة ازدادت بين درجات الاختبار وبالتالي تزداد قيمة معامل الصدق. وتتجانس المجموعة بعد من المشكلات المتعلقة بالصدق عند اتخاذ قرارات انتقاء الأفراد. فقيمة معامل الصدق تستمد عادة من تطبيق الاختبار التنبؤي على مجموعة من الأفراد الذين تم انتقاهم بالفعل، وهذه المجموعة تكون أكثر تجانساً من المجموعة التي سيتخذ قرار الانتقاء بشأنها في المستقبل مما يؤدي إلى انخفاض هذه القيمة مما قد يثير الشك في فاعلية الاختبار التنبؤي.

لذلك ينبغي عند تقدير قيمة معامل الصدق التنبؤي لاختبار معين أن تُستخدم الدرجات المستمدة من مجموعة الأفراد التي تقدم للالتحاق بدراسة أو بعمل معين قبل أن تُجرى عمليات الفحص والانتقاء.

فمثلاً نوجد العلاقة بين درجات اختبار الاستعداد الدراسي لجميع المتقدمين للالتحاق بالكليات ودرجاتهم أو تقديراتهم في نهاية السنة الأولى الجامعية، ولكن

الكليات لا تسمح عادة باللحاق جميع الطلاب، وكذلك لا تسمح مؤسسات العمل باللحاق جميع المتقدمين للمهن أو الوظائف.

لذلك تعتمد دراسات الصدق التنبؤى على الأفراد المقبولين فقط، مما يؤدى إلى ضيق مدى درجات الاختبارات التنبؤية؛ نظراً لأن هؤلاء الأفراد يمثلون مجموعة أكثر تجانساً من المجموعة التى ستطبق عليها هذه الاختبارات فى المستقبل بغرض انتقاء أفضلهم. وهذا التناقض فيما يتعلق بالصدق يتلخص فى أن الاختبارات التى تستخدمن فى اتخاذ قرارات الانتقاء ينبغى أن تكون قيمة معامل صدقها التنبؤى مرتفعة بالنسبة للمجموعة التى ستطبق عليها هذه الاختبارات مستقبلاً قبل بدء عملية الانتقاء، ولكن نظراً لأن هذه القيم يتم تقديرها فى واقع الأمر استناداً إلى المجموعة المتناثرة، فإننا متوقع أن تكون منخفضة.

لذلك ينبغي عند تفسير قيم معامل الصدق المرتبط بمحك عدم التسرع فى الحكم بعدم صلاحية الاختبار التنبؤى، أو عدم الثقة فى فاعليته فى التنبؤ بالنجاح فى الدراسة أو الأداء فى العمل، إذ ربما يرجع انخفاض قيمة معامل الصدق إلى قلة تباين الدرجات فى الاختبار نتيجة الاعتماد على المجموعة المتجانسة التى تم انتقاوها، أو ربما يرجع ذلك إلى الثبات المنخفض لدرجات الاختبار التنبؤى أو اختبار المحك، أو إلى عدم استقامة العلاقة بين درجات الاختبارين، أو غير ذلك من العوامل التى تؤثر فى قيمة هذا المعامل.

ثبات درجات المحك: سبق أن أوضحنا أن مشكلة ثبات درجات الاختبار التنبؤى ودرجات الاختبار المحك تعد من المشكلات الأساسية، وذكرنا كيفية تصحيح الأخطاء التى ربما تسهم فى انخفاض قيمة معامل ثبات كل منها.

فالمحاذفات تكون عادة تقديرات المشرفين على العمل، أو نتائج ملاحظات المعلم أو المعدل التراكمى للطالب، وجميعها يشوبها الخطأ مما يؤثر فى قيمة معامل ثبات درجات المحك. كما سبق أن ذكرنا عند مناقشتنا لمفهوم الثبات أن قيمة معامل صدق الاختبار تكون دائمًا أقل من قيمة مؤشر الثبات، أي أقل من \sqrt{r} حيث r ترمز إلى ثبات درجات المحك.

فمثلاً إذا كانت قيمة معامل الثبات $r = 0.4$ فإن معامل صدق المحك لا يمكن أن يزيد عن $\sqrt{0.4} = 0.63$ ، حيث إن درجات المحك لا ترتبط عادة بدرجات اختبار آخر أكثر من ارتباطها بذاتها.

تأثير المحك بمتغيرات أخرى: يتأثر معامل الصدق المرتبط بمحك بالمتغيرات الأخرى التى تتدخل مع المحك، فعندما يوجد معامل الارتباط بين درجات اختبار

الاستعداد الميكانيكي مثلًا، وتقديرات المشرفين على العمل لمجموعة من الأفراد، فإن معرفة المشرف على العمل بدرجات الأفراد في الاختبار ربما تؤثر في تقديراته لهم، مما يؤثر في قيمة هذا المعامل؛ لذلك لا ينبغي على المشرف الاطلاع على درجات اختبار الاستعداد عند تقديره للعاملين.

عدد مفردات الاختبار التنبؤي؛ نظرًا لأن القيمة القصوى للصدق المرتبط

بمحك تعتمد على مؤشر الثبات R^2 ، فإن هذه القيمة تزداد بزيادة قيمة معامل الثبات. وزيادة قيمة معامل الثبات تعتمد على طول الاختبار، أي عدد مفرداته، فكلما زاد هذا العدد زادت قيمة معامل الثبات إذا طبق الاختبار على مجموعة الأفراد نفسها. وقد سبق أن أوضحنا كيفية تقدير قيمة معامل الثبات إذا أطلنا الاختبار، أي أضفنا إليه بعض المفردات التي تقيس السمة نفسها، وذلك باستخدام الصيغة العامة لسييرمان وبراؤن Spearman - Brown . ويمكن الرجوع إلى جدول (٤ - ١) لمعرفة تأثير هذه الإطالة على القيم المتوقعة لمعامل الثبات وبخاصة القيم المتوسطة، غير أن هذا التأثير يكون أقل فيما يتعلق بقيم الصدق.

المدة الزمنية الفاصلة بين تطبيق الاختبار التنبؤي والاختبار المحك؛ فمما

لا شك فيه أننا نتوقع انخفاض قيمة معامل الصدق المرتبط بمحك بزيادة المدة الزمنية التي تفصل بين تطبيق كل من الاختبارين، نظرًا لتأثير الأخطاء العشوائية في درجات كل من الاختبارين، ولذلك يفضل أن تكون هذه المدة الزمنية قصيرة لكي يتيسر إيجاد معامل الارتباط بين درجاتها، ويتحقق هذا بوضوح في الصدق التلارمى.

عدد أفراد مجموعة المختبرين : فكلما قل عدد الأفراد الذين يطبق عليهم الاختبار التنبؤي نتوقع عند تطبيق المحك بعد مرور مدة زمنية معينة أن يقل العدد أكثر مما يؤثر كما ذكرنا في قيمة معامل الصدق التنبؤي. لذلك ينبغي البدء بعينة كبيرة من الأفراد حيث يترتب على طول المدة الزمنية فقدان بعضهم نتيجة الانقطاع أو الرسوب أو أى ظروف طارئة أخرى.

صدىق التكوين الغريغى (صدق المفهوم) :

يعد مصطلح الصدق Validity من أكثر مصطلحات القياس التربوى والنفسى إثارة للجدل وتبادر وجهات النظر، والدليل على ذلك تعدد المقالات المتعلقة به والمنشورة في الدوريات المتخصصة التي تصدر عن الجمعية الأمريكية لعلم النفس (APA)، والجمعية الأمريكية للبحث التربوى (AERA) ، والجمعيات والمجالس القومية المتخصصة في القياس التربوى والنفسى .

وتهتم هذه المقالات بمراجعة مصطلح الصدق ومحاولة توسيع مضمونه لكي يأخذ في الاعتبار ما يتربّب على استخدامات أساليب القياس في المجالات النفسية والتربوية، وبخاصة في مجال قياس السمات الإنسانية. فصدق التكوين الفرضي أو صدق المفهوم Construct Validity نشأ وتطور نتيجة النقد المستمر للمفهوم التقليدي للصدق. فالذى يطلع على البحوث التربوية والنفسية يجد كثيراً من التسميات المتعلقة بصدق أدوات القياس، مثل الصدق الظاهري، والصدق المنطقي ، والصدق التجريبى ، وصدق المضمون ، والصدق المنهجى ، والصدق الذاتى ، وصدق التعميم ، والصدق العاملى ، والصدق الإحصائى ، والصدق التمايزى ، والصدق التطابقى ، والصدق التقاربى ، وغيرها مما يشير العديد من التساؤلات حول مصطلح الصدق وفيه يستخدم.

ولعل هذا ما دعا هذه الجمعيات إلى مراجعة هذا المصطلح، وأوصت بتقسيم دراسات الصدق إلى الصدق المرتبط بمحك بجانبيه التنبؤي والتلازمي، وصدق المحتوى، وصدق التكوين الفرضي، أو صدق المفهوم.

وتتعلق هذه الدراسات بأنواع القرارات الثلاثة التي أشرنا إليها في مستهل الجزء الخاص بالصدق. وعلى الرغم من أن مصطلح صدق التكوين الفرضي قد اقترحه كل من كرونباك Cronbach وميهل Meehl عام ١٩٥٠، إلا أنه قد نال اهتماماً ملحوظاً وقبولاً متزايداً من جانب علماء القياس وبخاصة في الآونة الأخيرة نتيجة غموض كثير من التكوينات الفرضية أو المفاهيم السيكولوجية، الذي ترتب عليه إعادة تطوير اختبارات ومقاييس أكثر صدقاً.

ويتناول صدق التكوين الفرضي العلاقة بين نتائج الاختبارات والمقاييس وبين المفهوم النظري الذي يهدف الاختبار لقياسه، مثل مفاهيم الذكاء، والقلق، وقوية الأنما، والتصلب، والانطواء، ودافعية الانجاح، والابتكارية، والاستعداد المكاني، والقدرة التحليلية، ومستوى الحافظ ، والمناخ المدرسي ، والإدارة الذاتية ، والقيادة المدرسية ، وهكذا .

وبعبارة أخرى فإن صدق التكوين الفرضي يهدف لتحديد التكوينات الفرضية التي يُعزَّى إليها تباين الأداء في الاختبارات، أي أن هذه التكوينات الفرضية هي التي يتركز عليها الاهتمام وليس درجات اختبار المحك أو سلوك الفرد.

ولكن ما المقصود بالتكتونات الفرضية، وهل تختلف عن السمات الإنسانية؟ فعلى الرغم من أن كرونباك Cronbach استخدم المعنى العام أو القاموسى لمصطلح السمات Traits من حيث إنها خصائص موجودة داخل الفرد، وكذلك المعنى العام لمصطلح التكتونات الفرضية Hypothetical Constructs من حيث إنها تتواجد في

فکر علماء النفس، إلا أن لوفنجر (Loevinger, 1967) قارنت بين المصطلحين بما يشبه المقارنة التي تشير إليها المراجع الإحصائية بين بارامتر أو معلم المجتمع Parameter وإحصاء العينة Statistic. فبارامتر المجتمع هو الذي يهدف علم الإحصاء الاستدلالي إلى تقدير قيمة، أما إحصاء العينة الذي يناظر البارامتر فإنه يمثل أفضل تقدير متاح لهذا البارامتر.

ولذلك فإن السمات هي التي تهدف لفهم طبيعتها كما أوضحنا في الفصل الأول، أما التكوينات الفرضية المناظرة لهذه السمات، فإنها تمثل أفضل ما يتوصل إليه العلماء في وقت ما لتوسيع طبيعة هذه السمات. ولذلك يركز صدق التكوين الفرضي على ثلاثة عناصر هي : الاختبار، والسمات المراد قياسها، وماذا يقيس الاختبار من وجهة نظر القائم بإعداده، أي أن هذه العناصر تتعلق بالتكوين الفرضي، والتفسير، والنظرية التي يستند إليها التفسير. ويرى كرونباك أن هناك تساؤلين مستقلين يتضمنهما صدق التكوين الفرضي هما: إلى أي حد يقيس الاختبار ما يتضمنه تكوين فرضي أو مفهوم معين؟ وإلى أي حد يتضمن المفهوم فرضياً صادقاً؟

غير أن لوفنجر Loevinger تصوّغ التساؤلين صياغة بديلة كالتالي:
إلى أي حد يقيس الاختبار سمة لها وجود فعل؟ وإلى أي حد يكون هناك تناظر بين التفسير المقترح وما يقيسه الاختبار؟

فالتساؤلان اللذان اقترحاهما كرونباك يقسمان صدق التكوين الفرضي إلى : صدق الاختبار فيما يتعلق بتكوين فرضي معين، وصدق التكوين الفرضي ذاته، ويرى أن هذين النوعين من الصدق لا ينفصلان عادة إلا إذا كان الاختبار أو التكوين الفرضي قد نال قبولاً واسعاً على مدى زمني طويل.

أما التساؤلان الآخرين فيقسمان صدق التكوين الفرضي إلى الصدق الذاتي للاختبار، وصدق التفسير. وترى لوفنجر أن هذه الصياغة البديلة تسمح بتقسيم التساؤلين كل على حدة، فقيم الارتباط الداخلي بين مفردات الاختبار، والقيم المرتفعة للارتباط بين درجات الاختبار ومتغيرات خارجية تقدم أدلة على أن الاختبار يقيس شيئاً ما بطريقة منظمة، أما محتوى المفردات وطبيعة العلاقات القائمة بينها، وطبيعة المتغيرات الخارجية التي تباين في درجة ارتباطها بالاختبار فتقدم أدلة على التفسير المقترح. أي أن الأدلة الأولى تفيد في إضفاء معنى سيكومترى على الاختبار، أما

الأدلة الثانية تُفضي على الاختبار معنى سيكولوجي أو تربوي، وكلاهما أمراً ضرورياً. فالعلم يجب أن يقدم إطاراً نظرياً يسمح بتكامل وتفصيل البيانات المتعلقة بمفهوم معين، وتوجيه البحث المستقبلي المتعلقة به.

طبيعة التكوينات الفرضية:

إن التكوينات الفرضية Constructs تعد بمثابة خصائص يفترض أنها تميز الأفراد، وينعكس أثراها في سلوكهم. ويمكن أن تكون هذه التكوينات كيفية مثل فقدان الذاكرة أو كمية مثل الذكاء، ويفترض أن الفرد الذي يمتلك خاصية معينة يتحمل أن يسلك في موقف معين سلوكاً معيناً.

وعادة يجري تعريف هذه التكوينات الفرضية تعريفاً إجرائياً بحيث يمكن قياس أثراها بواسطة أدوات القياس. ولكن يمكن أيضاً أن يتخطى التعريف السلوك الملاحظ بحيث يكون أكثر اتساعاً ودلالة من مجرد إعادة صياغة العلاقات الإمبريقية في التعريف الإجرائي. ويمكن كذلك أن تكون التكوينات الفرضية عيانية Concrete مثل زمن الرجع، أو تكون مجردة Abstract مثل الشخصية أو الدافعية، أو الطلاقة اللغوية، أو الذكاء، أو التوافق ، أو تقدير الذات، أو القيادة التربوية .

فكلاهما كان التكوين الفرضي أو المفهوم عيانياً أو مفهوماً بسهولة التحقق من صدق المقياس المتعلق به، والعكس إذا كان المفهوم مجرداً. درجة صعوبة التتحقق من صدق التكوين الفرضي تتناسب ودرجة عيانية أو تجريد المفهوم المراد قياسه. فمثلاً يمكن بسهولة قياس مفهوم زمن الرجع Reaction Time وذلك عن طريق حساب الزمن الذي يستغرقه الفرد في استجابته بالضغط على زر جرس عند سماعه أو رؤيته مثيراً معيناً.

فالمتغير موضع الاهتمام في هذه الحالة هو سرعة الاستجابة التي يمكن قياسها بسهولة. وتعتمد دقة القياس على التقنية المستخدمة في تقديم المثير وتسجيل الزمن. أي أنها هنا نهتم اهتماماً مباشراً بإجراءات القياس، ولا تكون في حاجة إلى تقدير صدق المقياس، ومن ثم يمكن إجراء دراسات لتحديد علاقات معينة بين المقياس ومتغيرات أخرى.

ولكن كثيراً من المفاهيم السيكولوجية ليست بهذا القدر من البساطة، حيث إنها مفاهيم مجردة تمتد أبعد من حدود إجراءات قياس معينة. فمفهوم الدافعية Motivation وقوة الآنا Ego Strength يمكن أن تستدل عليهما من أنماط متعددة من السلوك مما

يتطلب التتحقق من صدق المقاييس التي نفترض أنها تقيس المفهوم المجرد، وهذا بالطبع يتطلب تصوراً فكرياً معيناً فيما يتعلق بالمفهوم وأبعاده المتكاملة. أى أن المفهوم أو التكوين الفرضي يُمثل فرضياً يشير إلى ارتباط أنماط متعددة من السلوك فيما بينها في دراسة الفروق الفردية.

لذلك يرى نانلى (Nunnally, 1978) ، وزيلر (Zeller, 1978) أن جميع النظريات العلمية تهتم أساساً بدراسة التكوينات الفرضية ولا تهتم كثيراً بالحالات الخاصة، أى تهتم بالعموميات دون الخصوصيات. فالعلماء لا يستطيعون إجراء بحوثهم بدون هذه التكوينات الفرضية والتغيير عنها والتواصل بها، ويسعون دوماً إلى تطوير مقاييس لهذه التكوينات والتوصل إلى علاقات وظيفية بين مقاييس التكوينات الفرضية المختلفة. فصدق التكوين الفرضي يعد من المشكلات الأساسية في التعميمات العلمية، فدراسة تكوين فرضي معين مثل الذكاء أو معدل التعلم يتطلب استخدام مقاييس أو أكثر بحيث تميز نتائجها بقابلية التعميم على فئة من المقاييس التي تتعلق بالتكوين الفرضي نفسه.

ولذلك فإن الأسس المنطقية التي تربط صدق التكوين الفرضي بالتعميمات العلمية تعد أكثر تعقيداً من الأسس التي يعتمد عليها الصدق المرتبط بمحل.

وتباين التكوينات الفرضية في درجة اتساع وتحديد نطاق السلوك الذي يستدل منه عليها. فأحياناً يكون هذا النطاق محدوداً حيث يشتمل على عدد قليل من التغيرات التي يمكن ملاحظتها وقياسها مثل زمن الرجع الذي أشرنا إليه، وأحياناً كثيرة يكون النطاق متسعًا بل مفتوحاً بحيث يشمل العديد من التغيرات المتداخلة مما يلقى بظلال من العموض على التكوين الفرضي. ومع هذا يحاول علماء النفس تحديد بعض التغيرات الأساسية المتعلقة بالتقوين الفرضي تحديداً وأوضحاً، ويصوغون افتراضات تتعلق بالتغيرات الأخرى التي ربما يرون أن التقوين الفرضي يمكن أن يشملها. ولتوسيع ذلك نجد أن علماء النفس قد انفقوا على أن قياس مفهوم الذكاء ينعكس في سلوك حل المشكلات التي تتطلب قدرات استدلالية، ولكنهم اختلفوا فيما إذا كانت القدرات الحسية الإدراكية والذاكرة تعد جزءاً من هذا المفهوم.

ويتطبق ذلك على كثير من المفاهيم السيكولوجية والتربيوية، حيث تكون حدود نطاق السلوك الملاحظ الذي يستدل منه على المفهوم غير واضحة وضوحاً تماماً، لذلك يفضل قياس المفاهيم أو التكوينات الفرضية استناداً إلى ضم نتائج مقاييس متعددة لأنماط السلوك المتعلقة بها بدلاً من الاعتماد على كل مقياس على حدة من أجل زيادة صدق التعميمات العلمية.

ولكن نظراً لما يتطلبه ذلك من جهد كبير، فإن الاعتماد على مقياس واحد لمفهوم معين يجعل هناك درجات من صدق التكوين الفرضي بحسب اقتراب نتائج هذا المقياس من نتائج مقاييس أخرى للمفهوم ذاته. وهذا يدل على أن تقييم صدق التكوين الفرضي يتطلب تجميع أدلة متعددة من مصادر متباعدة وفي مواقف مختلفة حيث إنه لا يوجد مؤشر كمي لهذا النوع من الصدق.

منهجية التحقق من صدق التكوين الفرضي:

إن عملية التوصل إلى أدلة عن صدق التكوين الفرضي يمكن اعتبارها محاولة للإجابة عن التساؤلين السابقين اللذين اقترحهما كرونباك Cronbach أو لوفنجر Loevinger ، وهذا يتطلب كما ذكرنا جمع أدلة من مصادر متعددة ويطرق مختلفة وتقييم هذه الأدلة تقييماً متكاملاً. إذ يمكن فحص مفردات الاختبار لمعرفة الشيرات التي تشتمل عليها ومحتها ونوعها، ومدى استقرار درجات الاختبار في ظروف متباعدة، ومدى تجانس مفرداته، وارتباط درجاته بدرجات اختبارات أخرى ويمتغيرات ربما تختلف عما يقيسه الاختبار، ومدى تمييزه بين مجموعات مختلفة من الأفراد فيما يقيسه الاختبار، وتأثير الظروف التجريبية على أداء المختبرين، وغير ذلك من البيانات التي تلقى الضوء على ما تدل عليه درجات الاختبار.

وقد اقترح كرونباك Cronbach وميهل Meehl في مقالتهما الرائدة في هذا المجال عام ١٩٥٥ توجهاً مختلفاً في هذا الشأن، حيث أشارا إلى منظور آخر لعملية جمع الأدلة عن صدق التكوين الفرضي يتركز حول بناء نظرية مصغرة Miniature Theory تتعلق بالاختبار في ضوء الخطوات الثلاث التالية:

(١) تبني نظرية معينة يستند إليها الاختبار، وتُتحقق منها فروضاً معينة، تتعلق بالسلوك المتوقع من الأفراد الذين يحصلون على درجات مختلفة في الاختبار.

(٢) جمع بيانات يستند إليها في التتحقق من صحة الفروض.

(٣) تقييم ما إذا كانت النظرية تفسر بالفعل البيانات المتجمعة تفسيراً كافياً، فإذا لم يكن الأمر كذلك ينبغي مراجعة النظرية، وتكرر الخطوات السابقة للتوصيل إلى تفسير أكثر كفاية .

وعلي الرغم من أن إجراء هذه الخطوات يبدو أمراً يسيرًا إلا أنها في واقع الأمر تتطلب جهداً كبيراً ودراسات متعددة وأحكاماً تقييمية تستند إلى نتائج بحوث تجريبية .

لذلك ربما يكون من المناسب أن نقدم مزيداً من التوضيح لهذه الخطوات في إطار مقترنات كرونباك Cronbach وميهل Meehl وبراون Brown. فالنظريات التي يستند إليها الاختبار تشمل على مجموعة من المفاهيم والقضايا والمبادئ المترابطة، لذلك يطلق عليها «شبكة القوانين المترابطة Monological Network».

وتفيد هذه الشبكة في الربط بين الخاصة أو السلوك الملاحظ وبين التكوين الفرضي النظري، أو بغيره من التكوينات الفرضية، أو بخاصة ملاحظة أخرى، أي أن هذه الشبكة بما تتضمنه من قوانين تعبر عن علاقات محددة أو احتمالية.

وهنا تأتي عملية الاستدلال على معنى التكوين الفرضي من البيانات التي تمثل السلوك الملاحظ. وهذه العملية يجب أن تكون محددة تحديداً دقيقاً بحيث يمكن لباحثين مختلفين التحقق منها والتواصل بها. وهذا القوانيين ينبغي أن تكون ذات صلة بسلوك ملاحظ في وقت ما. أما إذا كانت القوانين مشتقة من قوانين أخرى، فإن هذه القوانين الأخرى هي التي ينبغي أن تفي بهذا الشرط. ويمكن التعبير عن هذا السلوك الملاحظ كيفياً مثل متفائل أو مرح، أو كميًّا مثل درجات الاختبارات، فكلاهما يفيد في قياس التكوين الفرضي.

وهذا لا يتأنى إلا إذا استخدم هؤلاء الباحثين شبكة القوانين المترابطة المقترنة والمتعلقة بتكوين فرضي معين، وتتضمن عملية الاستدلال التتحقق من فرض معينة في ضوء هذه الشبكة واستناداً إلى البيانات التي يتم جمعها.

وينبغي توضيح تعريف التكوين الفرضي المراد التتحقق من صدقه، والكيفية التي اشتقت بها الفرض من النظرية المتبناة، والإجراءات التجريبية، والأسلوب الذي اتبع في الاستدلال على معنى التكوين الفرضي المعين من هذه الإجراءات، وتقسيم الأدلة التي أكدت صحة الفرض، والأدلة الواهنة.

ولعل المشكلة التي تواجه القائمين بذلك تتعلق بهذه الأدلة الواهنة، حيث إنه ربما يصعب تحديدها بدقة وتبريرها؛ إذ ربما يكون الاختبار مقياساً غير مناسب للتقوين الفرضي المعين، أو أن النظرية التي استند إليها يشوبها بعض أوجه القصور، أو أن التصميم التجاري أو إجراءاته غير مناسبة للتحقق من صحة الفرض.

لذلك يرى كرونباك Cronbach وثورنديك Thorndike وأستاري Anastasi أن هذه الخطوات مثالية وربما لا تتحقق بدرجة كافية في كثير من التكوينات الفرضية أو المفاهيم السيكولوجية والتربية، ويقترحون محاولة الاقتراب من هذه المثالية بأن

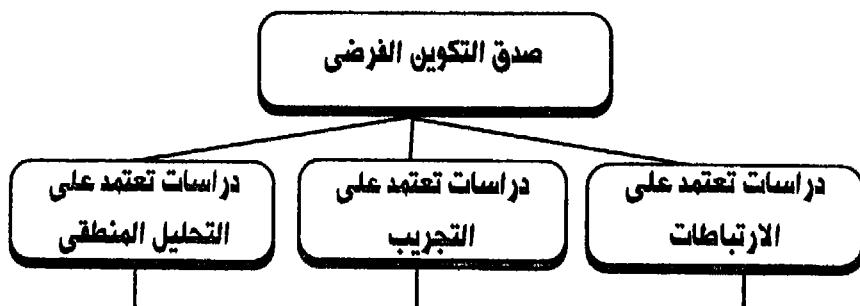
يستخدم الاختبار كمؤشر **Sign**, بحيث يقدم معلومات توضح طبيعة النطاق السلوكي المتعلق بمفهوم أو تكوين فرضي معين، وبالتالي يعد مؤشرًا لطبيعة هذا المفهوم أو هذا التكوين الفرضي.

أساليب جمع أدلة تتعلق بـ *نطاق التكوين الفرضي*:

اتضح مما سبق أن صدق التكوين الفرضي يتطلب جمع معلومات وأدلة من مصادر متعددة، وفحص نتائج الدراسات المتعلقة بمفهوم أو تكوين فرضي، معين والربط فيما بينهما.

فالنظرية المتكاملة لهذه الأدلة تلقي الضوء على طبيعة التكوين الفرضي، والعوامل التي تسهم في الكشف عنه وتطويره؛ لذلك ربما لا نعجب إذا لاحظنا تعدد أساليب جمع هذه الأدلة ، نظرًا لأن البحث في صدق التكوين الفرضي يعد عملية مستمرة وفقًا لما يطرأ على تكوين فرضي معين من متغيرات ، أو ما يتوصل إليه الباحثون من نتائج ربما تخالف النتائج التي اعتقاد البعض أنها أكدت صدق هذا التكوين الفرضي.

غير أنها سوف نركز مناقشتنا على بعض الأساليب الأساسية المستخدمة في دراسات الصدق ، وسوف نوضحها بالشكل التخطيطي (٥ - ٦) التالي :



شكل (٥ - ٦) يوضح بعض الأساليب المستخدمة في

دراسات صدق التكوين الفرضي

ويتضح من شكل (٥ - ٦) أنه يمكن تصنيف هذه الأساليب في ثلاثة مجموعات متكاملة هي :

- (١) دراسات تعتمد على الارتباطات .
- (٢) دراسات تعتمد على التجريب .
- (٣) دراسات تعتمد على التحليل المنطقي .

ومما لا شك فيه أن هذه الأساليب يغلب عليها الطابع التكاملى ، فكل منها يثرى الآخر ويترشد بتائجه . فالدراسات الارتباطية تهتم بتحديد طبيعة الفروق بين الأفراد الذين يحصلون على درجات مرتفعة في الاختبار والأفراد الذين يحصلون على درجات منخفضة ، وكذلك تحليل العلاقات القائمة بين درجات مجموعة من الاختبارات للكشف عن السمات التي تشتراك هذه الاختبارات في قياسها واحتلافها عن غيرها من السمات .

أما الدراسات التجريبية فتهدف لمعرفة التغيرات التي تحدث في أداء الفرد في اختبار معين نتيجة التأثير التجربى فى أحد المتغيرات المتعلقة بالفرد للتحقق من صحة فرض معين يتعلق بما يقيسه الاختبار ، بينما تركز الدراسات المنطقية على محتوى الاختبار وأساليب تصحيحه أو تقدير درجاته والعوامل التي تؤثر في هذه الدرجات . وسوف نوضح فيما يلى كل من هذه الأساليب بمزيد من التفصيل .

أساليب تختتم على الارتباطات :

تتعدد الأساليب التي تستخدم فى دراسات صدق التكوين الفرضى استناداً إلى الارتباطات ، فأحد هذه الأساليب يتعلق بدراسة الفروق بين مجموعات مختلفة من الأفراد Group Differences Method . وتهدف كثير من هذه الدراسات للمقارنة بين مجموعتين مختلفتين في ضوء الدرجات التي يحصل عليها الأفراد في اختبار أو مقياس معين ، حيث يتوقع أن المجموعتين سوف تختلف استجاباتها أو الدرجات التي يحصل عليها أفراد كل منهما في الاختبار أو المقياس ، ويمكن التتحقق من ذلك بطريقة مباشرة .

إذا أردنا التتحقق مثلاً من صدق التكوين الفرضى لاختبار يقىس العصبية Neuroticism ، فإنه يمكن اختيار مجموعتين من الأفراد استناداً إلى تشخيص الأطباء النفسيين بحيث تكون هذه السمة مرتفعة لدى إحدى المجموعتين ومتناقصة لدى المجموعة الأخرى ، ويطبق الاختبار بعد ذلك على كل من المجموعتين . فإذا ثبت أن هناك اختلافاً واضحًا بين درجات كل من المجموعتين ، فإنه يمكن اعتبار ذلك أحد أدلة صدق التكوين الفرضى للاختبار ، ويستباقى عندئذ على المفردات التي ميزت بدرجة أكبر بين المجموعتين .

ويشير براون (Brown, 1972) في هذا الشأن إلى أن النتائج التي نحصل عليها من مقياس للشخصية مثلاً ربما تعزى إلى الخاصة أو السمة المقاومة ، أو إلى طريقة القياس (استبيان ، قائمة ملاحظة ، بطاقة تقرير ذاتي ، مقابلة أو غير ذلك) ، أو إلى كليهما معاً .

فإذا درسنا السمة نفسها أو سمات متباعدة بطرق مختلفة، فإنه ربما نحصل على نتائج مختلفة، لذلك إذا أردناا عزل الإسهام النسبي لكل من السمة أو الطريقة يجب دراسة أكثر من سمة وأكثر من طريقة معاً، أي ندرس الصدق التقاري والصدق التمايزى. ففي الصدق التقاري ننظر إلى الارتباطات بين السمات نفسها إذا أجري قياسها بطرق مختلفة، وفي الصدق التمايزى ننظر إلى الارتباطات بين سمات متباعدة إذا أجرى قياسها بطريقة واحدة.

والأسلوب الثاني الذي يستند إلى الارتباطات يعتمد على إيجاد معامل الارتباط بين درجات اختبار ويفترض أنه يقيس تكويناً فرضياً معيناً، ودرجات اختبار آخر يثبت الأدلة المتعددة أنه يقيس التكوين الفرضي ذاته.

فمثلاً كثير من اختبارات الذكاء الجديدة التي يقوم الباحثون أو المؤسسات المتخصصة ببنائها تعتمد على إيجاد ارتباط درجات الاختبار الجديد بدرجات اختبار ستانفورد بينيه للذكاء Stanford - Binet، أو اختبار ويكسلر للذكاء Wechsler وهو اختباران شائعان الاستخدام وتالا كثيراً من دراسات الصدق على مدى زمني طويل.

فإذا وجد أن معامل الارتباط قيمته مرتفعة فإن هذا يعد أحد أدلة صدق التكوين الفرضي. ويمكن كذلك أن نستدل على أن درجات الاختبار الجديد يمكن أن ترتبط بمتغيرات أخرى بنفس قدر ارتباط كل من الاختبارين المذكورين بهذه التغيرات.

وعادة يمتد هذا الأسلوب بحيث يتراول العلاقات القائمة بين الاختبار ومجموعة أخرى من الاختبارات بدلاً من اختبار واحد. ويمكن إجراء ذلك باستخدام أحد الأساليب الإحصائية متعددة المتغيرات التي سبق أن أشرنا إليها، يسمى التحليل العائلي Factor Analysis . ويهدف هذا الأسلوب لتحليل هذه العلاقات وذلك بتكونين مصفوفة لقيم معاملات الارتباط بين درجات الاختبار الجديد، ودرجات عدة اختبارات أخرى مناسبة، وإجراء تحليل إحصائي لهذه المصفوفة لتحديد أقل عدد من العوامل (التكوينات الفرضية) التي تسهم في تفسير التباين في قيم معاملات الارتباط داخل المصفوفة، وكذلك العوامل التي تسهم في الأداء في كل من هذه الاختبارات.

ويذلك نستطيع التوصل إلى تجمع من الاختبارات التي تقيس تكويناً فرضياً معيناً، ودراسة محتوى هذه الاختبارات لتعرف طبيعة هذا التكوين الفرضي. ويمكن إجراء هذا التحليل باستخدام البرنامج الفرعى Factor ، وهو أحد برامج جزءة الحاسوب الإحصائية للعلوم الاجتماعية (SPSS) .



ولكن نظراً لأنه من الضروري عدم الاقتصار على تعرف العلاقات الموجبة بين الاختبار الجديد الذي يفترض أنه يقياس تكويناً فرضياً معيناً وغيره من الاختبارات المشابهة التي تقيس التكوين الفرضي نفسه، وإنما أيضاً تعرف علاقته ببعض الاختبارات أو المقاييس التي لا تقيس هذا التكوين الفرضي، فقد اقترح كامبل Campbell وفيسك Fiske عام ١٩٥٩ أسلوباً للتحقق من ذلك اعتماداً فيه على مصفوفة متصوفة أطلق عليها «مصفوفة السمات المتعددة والطرق المتعددة Multitrait - Multimethod Matrix» يمكن باستخدامها الكشف عن علاقات الاختبار باختبارات مشابهة للتوصيل إلى الصدق التقاري لل اختبار Convergent Validity، وعلاقته باختبارات مختلفة عنه للتوصيل إلى الصدق التمييزي لل اختبار Discriminant Validity.

وتعد مصفوفة السمات المتعددة والطرق المتعددة بمثابة تصميم تجربى منظم يقدم أدلة عن كل من هذين التوقيعين من الصدق، ويطلب إعداد هذه المصفوفة قياس عدد من التكوينات الفرضية بطرق متعددة ، ولتوضيح هذا الأسلوب نفترض أننا قمنا بقياس ثلاثة سمات مختلفة (الفهم اللغوى ، والاستدلال العددى ، والمعلومات العامة) بطرق ثلاثة (تقديرات المعلمين ، واختبارات ورقية ، واختبارات فردية) وبذلك تشتمل المصفوفة على تسعه متغيرات ، وندون فى خلايا هذه المصفوفة قيم معاملات الارتباط بين كل متغيرين منها ، ويوضح جدول (٣-٥) هذه القيم الافتراضية:

الطريقة الثالثة	الطريقة الثانية	الطريقة الأولى
٣ ب ٣ ج	٢ ب ٢ ج	١ ج ١ ب ١ ج
	القطر الرئيسي للمصفوفة	١١, ٩٣, ٤٨, ٠, ٢٢, ٠, ٤١, ٠, ٩٢, ٠, ١٣, ٠, ٢٠, ٠, ٥٧, ٠, ٢٠, ٠, ٦٢, ٠, ٣٣, ٠, ٥٤, ٠, ١٢, ٠, ١٨, ٠, ٢١, ٠, ٢٢, ٠, ٥٦, ٠, ١٥, ٠, ٦٢, ٠, ٢٠, ٠, ٥٣, ٠, ١٣, ٠, ١٢, ٠, ٢١
		الطريقة الأولى ج الثانية ج الثالثة ج
		الطريقة الأولى ج الثانية ج الثالثة ج

جدول (٣-٥) يوضح مصفوفة السمات المتعددة والطرق المتعددة لمجموعة من البيانات الافتراضية

ويتضح من جدول (٣ - ٥) أن هناك فيما لمعاملات الارتباط تقع على القطر الرئيسي للمصفوفة، وهذه تعد بمثابة قيم معامل الثبات؛ لأنها تشير إلى قياس السمة نفسها بالطريقة نفسها، أي (أ، أ، ب، ب، ج، ج، ...). ويلاحظ أنها فيما مرتفعة.

وهناك قيم تقع داخل مثلاثات تحدها خطوط متقطعة، وهذه تشير إلى قيم معاملات الارتباط بين السمات الثلاثة التي أجريت قياسها بكل من الطرق الثلاث، ومعظمها قيمًا منخفضة. كما أن هناك قيمًا تقع داخل مثلاثات تحدها خطوط متصلة (المثلاثات الثلاثة التي تقع على الخط الرئيسي للمصفوفة)، وهذه تشير إلى قيم معامل الارتباط عبر السمات المقاومة بطريقة واحدة، وهي قيم مرتفعة نسبيًا. أما القيم التي تقع على القطرين اللذين يفصلان بين المثلاثات التي تحدها خطوطًا متقطعة فإنها تشير إلى قيم معامل الارتباط بين السمة نفسها مقاسة بالطرق الثلاث، وهي قيم مرتفعة نسبيًا أيضًا. وهذه القيم تدل على معاملات الصدق ويقصد به الصدق التقاري، وينبغي أن تكون هذه القيم موجبة ومرتفعة.

كما ينبغي أن تكون قيم معامل الارتباط بين الطرق المختلفة التي تقاس السمة نفسها (معاملات الصدق) أكبر من قيم معامل الارتباط بين السمات المختلفة التي تقاس بالطريقة نفسها، وهي القيم المدونة داخل المثلاثات التي تحدها خطوطًا متصلة.

فالتمييز بين السمات المتباعدة أهم من التمييز بين طرق القياس المختلفة، كما أن قيم معاملات الصدق ينبغي أن تكون أكبر من قيم معامل الارتباط بين كل من السمات المتباعدة والطرق المختلفة. وهي القيم المدونة داخل المثلاثات التي تحدها خطوطًا متقطعة.

وهاتان الملاحظتان الأخيرتان تفيدان في تقديم أدلة عن الصدق التمايزى، فما يهمنا هو معرفة ما إذا كان الارتباط بين السمات كان نتيجة لاختلاف طريقة قياس هذه السمات مما قد يشير إلى تحيز طريقة القياس، أم أنه ارتباط فعلى. وهذا يعد من الأدلة المهمة التي تزودنا بها هذه المصفوفة في تقدير الصدق التقاري والتمايزى التي ربما لا تزودنا بها الأساليب الأخرى، نظراً لاعتماد هذه الأساليب على طريقة قياس واحدة.

أساليب تختتم على التجريب:

تعتمد هذه الأساليب على التدخل التجريبي لإحداث تغيرات في درجات الأفراد في اختبار ما كوسيلة لتعرف مدى تأثر الأداء بمعالجات أو متغيرات معينة مما يساعد في تأكيد بعض التفسيرات المتعلقة بنتائج الاختبار أو رفضها.

فإذا تأثرت درجات الاختبار بمؤثرات انتقالية، فإن هذا ربما يحد من تفسير الدرجات مما يضفي بعض الشك على صدق التكوين الفرضي للاختبار. فمثلاً إذا أردنا معرفة مدى تأثير اختبار يقيس فهم المقرؤه باستراتيجية تعامل الطالب مع المقرؤء، فإننا نطبق الاختبار على مجموعة من الطلاب بالطريقة المعتادة، ثم نطبق عليهم اختباراً مكافئاً بعد تدريبهم تدرباً سريعاً على استخدام استراتيجيات مختلفة لفهم المقرؤء، فإذا وجدنا تحسيناً في درجاتهم، فإن هذا يكون دليلاً على أن الاختبار يتأثر بالاستراتيجية المستخدمة.

ويمكن أيضاً الاستعانة بمجموعة ضابطة يتم تدريبيها بطريقة محدودة الفاعلية لمقارنة درجاتها بدرجات المجموعة التجريبية المشار إليها. كذلك يمكن التتحقق مما إذا كان الاختبار تعتمد الإجابة عن مفرداته على الدافعية. فعندئذ يمكن تصميم تجربة يقوم فيها الفاصل بتعزيز مجهودات الطلاب وتبيّن أثر ذلك على درجاتهم. فالتدخل التجاري في هذين المثالين يعد تدخلاً قصيراً للأمد؛ لأنّه يحدث أثناء العملية الاختبارية أو قبلها بوقت قصير، ولكن يمكن أيضاً أن يستغرق التدخل التجاري وقتاً أطول مثل أساليب العلاج النفسي أو الأساليب العلاجية التربوية في دراسات صدق التكوين الفرضي.

وكذلك للتحقق من صدق التكوين الفرضي لمقاييس «القلق»، من المهم تحديد ما إذا كان المقياس يزورنا بنتائج تتفق مع نظرية معينة تتعلق بالقلق. فإذا أوضحت النظرية أن القلق يزداد في مواقف معينة تثير الإحباط فإنه يمكن إجراء تجربة لتحديد ما إذا كان الأفراد الذين يواجهون أحد هذه المواقف يحصلون على درجات مرتفعة في المقياس أكثر من الذين لم يواجهوا هذا الموقف.

كما يمكن أن تشمل الأساليب التي تعتمد على التجربة أدلة من مواقف طبيعية. فمثلاً إذا كان تعريف تكوين فرضي معين يتضمن أن هذا التكوين الفرضي لا يتأثر بالنضج أو المتغيرات البيئية فإنه يمكن التتحقق من ذلك تجريبياً.

ولتوسيع ذلك نفترض أنها نود دراسة صدق التكوين الفرضي لاختبار جديد في الذكاء العام، فربما نفترض أن الذكاء العام يزداد بسرعة لدى الأطفال الصغار، بينما يزداد بدرجة أقل لدى المراهقين، ويقل ازيداده تدريجياً لحين بلوغ سن النضج.

لذلك نطبق الاختبار على عينات من الأفراد ممثلة لمجموعات عمرية مختلفة من سن ٣ أعوام إلى ٣٠ عاماً، ونرسم منحنيناً يمثل متسلقات درجات كل مجموعة عمرية، فإذا أكَدَ المنحنى الفرض السابق، فإن هذا يقدم أحد الأدلة عن صدق التكوين

الفرضي للاختبار، آخذين في الاعتبار أن الاختبار الذي يفترض أنه يقيس الذكاء ينبغي أن تعكس درجاته هذه التزعة النمائية.

أساليب تتحمّل على التحليل المنطقي:

تعتمد هذه الأساليب على الفحص الدقيق للاختبار والأداء الذي يتطلبه، واحداث تكامل بين نتائج هذا الفحص وبين النظرية التي يستند إليها الاختبار وأراء المُختبرين الذين سبق أن اختبروا باختبارات مشابهة، ويترتب على ذلك افتراض تفسيرات ربما تختلف عن التفسيرات التي اقترحها مُعد الاختبار.

ويرى كرونباك Cronbach أن هذا التحليل المنطقي من أهم مصادر التوصل إلى فرض بديلة فيما يتعلق بالأداء في الاختبارات والمقاييس، فالمحكم الذي لديه خبرة سابقة كبيرة بالخطاء التي شابت الاختبارات السابقة يمكنه أن يكتشف جوانب الضعف في أداة القياس الجديدة.

ومع هذا فإننا لا نستطيع باستخدام هذا التحليل المنطقي للمحتوى أن ندحض صدق أداة القياس، وإنما نستطيع أن نقدم فرض بديلة يمكن التحقق من صحتها بالأساليب الإمبريقية أو التجريبية.

كذلك من المفيد تحليل العمليات Process Analysis التي يستخدمها الأفراد في أدائهم أو في التوصل إلى إجاباتهم عن مفردات الاختبارات؛ لأن هذا التحليل يضيف أدلة جديدة تزيد من فهمتنا للتكتوبينات الفرضية التي تقيسها هذه الاختبارات. إذ يمكن باستخدام نتائج ملاحظة هذه العمليات تقييم ما تفهم به من تباين في درجات الاختبارات تقييماً أولياً.

إذا افترضنا أن اختباراً معيناً يقيس كفاية الطلاب في الرياضيات، ولكن ينت ملاحظة أداء مجموعة من الطلاب في الاختبار أن غالبيتهم يخطئون في قراءة الأسئلة، فعندئذ يصبح تفسير الدرجات المنخفضة موضوع تساؤل. وعلى الرغم من صعوبة ملاحظة العمليات التي يستخدمها الأفراد في إجابتهم عن مفردات الاختيار من متعدد، إلا أنه يمكن أن نطلب من كل منهم الإجابة بصوت مسموع، حيث يمكن أن يوضح الفرد جوانب الصعوبة ويفظه في كشف أخطائه، كما يمكن أن يطلب من الفرد بعد انتهاءه من الإجابة أن يير اختياره للإجابة الصحيحة عن كل مفردة واستبعاده البدائل الأخرى للإجابات التي تشتمل عليها المفردة. وهذا ربما يلقى الضوء على المؤشرات التي أدت بالأفراد إلى الإجابة الصحيحة بعمليات غير مناسبة.

ونظراً لصعوبة إجراء هذا النوع من التحليل لما يتطلبه من وقت وجهد وتنظيم معين، فقد اتجه بعض الباحثين إلى بناء برامج للحاسوب تتضمن شبكة العمليات النموذجية المتعلقة بحل مشكلة أو إجراء مهمة معينة، ويتم إدخال استجابات الأفراد لمفردات الاختبار في الحاسوب، ويجري البرنامج تصنيفاً لأنماط الأخطاء وترتيبها بحسب درجة أهميتها في حل المشكلة أو أداء المهمة.

والخلاصة أن صدق التكوين الفرضي يساعد في تحديد السمات التي يهدف الاختبار لقياسها، كما يفيد في بناء وتحقيق النظريات التربوية والسيكولوجية. وهذا يتطلب جمع أدلة ومعلومات متعددة لتلقي ضوءاً كافياً على التكوينات الفرضية باستخدام الأساليب السابقة.

وقد بين ساكس Sax وزيلر Zeller أنه يمكن إيجاز عملية التوصل إلى أدلة تتعلق بصدق التكوين الفرضي في الخطوات التالية على الأقل:

- (١) تبرير أهمية التكوين الفرضي من الناحية التربوية والنفسية على أن يكون التكوين الفرضي معرفاً تعريفاً إجرائياً ويدل على سمة قابلة لقياس.
- (٢) الاستناد إلى نظرية تربوية أو سيكولوجية أو اقتراح نموذج منطقى يوضح المفاهيم والعلاقات القائمة بينها.
- (٣) التمييز بين التكوين الفرضي والتكوينات الفرضية الأخرى المماثلة لها.
- (٤) التوصل إلى أدلة من مصادر متعددة باستخدام الأساليب الارتباطية والتجريبية والمنطقية لتأكيد التكوين الافتراضي. وعندما تستخدم اختبارات متعددة في هذا الشأن فإنه يمكننا الحصول على الصدق التقاري للتكوين الفرضي.
- (٥) التوصل إلى أدلة تتأكد منها أن التكوين الفرضي لا يرتبط بعوامل وقتية أو دخيلة لكي نحصل على الصدق التمايزى للتكوين الفرضي.
- (٦) إجراء تعديلات مستمرة في التكوين الفرضي بما يتفق والأدلة والمعلومات الجديدة المتجمعة.

منظور تكمالي لطرق تقيير التصريح والثبات :

اتضح لنا مما سبق أن صدق المحتوى والصدق المرتبط بمحك وصدق التكوين الفرضي ليست أنواعاً متباعدة، وإنما جوانب متكاملة لعملية متعددة تتعلق بالتحقق من صدق تفسيرات واستخدامات الدرجات المستمدة من أدوات القياس التربوى والنفسى.

لذلك فإن الفصل بين هذه الجوانب ، أى دراسة أحدها بمعزل عن الجوانب الأخرى يؤدى إلى أدلة غير كافية حول صدق التفسيرات التي توصل إليها. لذلك ينبغي أن نضع في اعتبارنا أن صدق أدوات القياس ليس صدقًا عاماً وإنما مرتبط بمواصفات أو استخدامات معينة لمجموعات معينة من الأفراد.

وقد سبق أن لاحظنا من الشكل التخطيطي (٥ - ١) أن الأدلة التي تجمع للتحقق من صدق التكوين الفرضي تتضمن أدلة عن صدق محتوى أداة القياس وصدقها المرتبط بمحك ، فصدق التكوين الفرضي إذن يعد بمثابة الإطار المتكامل الذي يجمع جوانب الصدق الثلاثة ، ويمكن أن يجرب عن التساؤلات الثلاثة التي ذكرناها في مستهل حديثنا عن مفهوم الصدق .

تفسير صدق المحتوى يتعلق بالاستدلال على أداء الفرد في النطاق الشامل للمفردات من أدائه في اختبار يشتمل على عينة مماثلة من المفردات ، ولكن إذا أردنا الاستدلال من ذلك على «قدرة الفرد» أو «معارفه» فإننا نكون بصدد الاستدلال على التكوين الفرضي الذي تتطوّر عليه درجات الفرد . كذلك إذا أردنا التنبؤ بحصوله على تقدير مرتفع في نهاية مدة دراسية معينة باستخدام درجاته في عينة المفردات فإن تفسيرنا يتعلق بمحك الأداء المستقبلي .

فالاختبار عندئذ تعدد جوانب الصدق فيه ، ولا يقتصر صدقه على جانب واحد . كما أن الاختبار في هذه الحالة يمكن استخدامه كأدلة تقييد في اتخاذ القرارات ، وفي توضيح طبيعة السمة أو المتغير الذي يقيسه .

وهذا المنظور التكاملى للصدق يلعب دوراً رئيساً في عمليات بناء وتطوير أدوات القياس التربوى والنفسى . فكثير من الباحثين والممارسين يظنون أن الصدق يرتبط بالصيغة النهائية للاختبار . ولكن إذا أمعنا الفكر قليلاً نجد أن صدق المحتوى يتعلق بالمرحلة الأولى في بناء الاختبار ، وهى تجميع عينة مماثلة من المفردات للنطاق السلوكي الشامل المراد قياسه ، ويمكن أيضاً انتقاء هذه العينة بأساليب الصدق المرتبط بمحك من حيث تمييزها بين الناجحين والراسيين ، أو بين المتمكنين وغير المتمكنين في ضوء مسحك أداء معين . ويمكن حذف المفردات التي لا تسحق متطلبات تكوين فرضي معين تقيسه المفردات ، وربما نجري ذلك بالتحقق من الاتساق الداخلى أو تجانس عينة المفردات . وينطبق ذلك بوضوح على اختبارات الذكاء والاستعدادات والاختبارات التحصيلية ، وربما ينطبق أيضاً على بعض مقاييس الجوانب الوجودانية والشخصية .

ولذلك فإن مراحل بناء وتطوير أدوات القياس تتفق بل وتكامل مع أساليب وإجراءات تقدير الصدق بجوانبه الثلاثة.

كذلك يرتبط مفهوم الصدق بمفهوم الثبات ارتباطاً وثيقاً، فقد سبق أن أوضحنا أن الاختبار لا يمكن أن يتسم بالصدق في استخدامات معينة دون أن تكون درجاته متسقة، كما أن قيمة مؤشر الثبات تعد بمثابة الحد الأعلى لقيمة معامل صدق الاختبار. وتؤثر قيم معامل الثبات في درجة العلاقة بين الاختبار التنبؤي والاختبار المحاك. وكذلك تسهم قيمة معامل الاستقرار لدرجات اختبار ما في زيادة ثقتنا بصدق التكوين الفرضي للاختبار؛ نظراً لأننا نهتم بالسمات التي تميز بالثبات النسبي.

كما أن معامل الاتساق الداخلي أو معامل التجانس للاختبار يقدم دليلاً آخر على صدق التكوين الفرضي.

لذلك يرى كرونباك Cronbach وجليسز Glaser أن مفهوم الصدق والثبات مفهومان متربطان ويمكن أن يندرج تحت اسم «مقاييس قابلية التعميم Measures of Generalizability»، فالفرق الرئيسي بينهما يكمن في الأبعاد التي نود التعميم عليها. لذلك تعد نظرية كرونباك المتعلقة بإمكانية التعميم Generalizability Theory هي التي أشرنا إليها في مناقشتنا لمفهوم الثبات من النظائر المعاصرة المهمة التي أسهمت في إبراز التكامل بين مفهومي الثبات والصدق.

ولذلك سوف نوضح هذه النظرية بمزيد من التفصيل كما سبق أن أشرنا في الفصل الرابع عشر.

خلاصة

بعد الصدق من الخصائص السيكومترية الأساسية للإختبارات والمقاييس، ونعني بالصدق مدى فائدة القياسات في اتخاذ قرارات موافمة لغرض معين. والإختبار لا يتسم بالصدق دون أن تكون درجاته متسقة، وهذا يعني أن مفهومي الصدق والثبات متراطمان. ونظراً لأن الإختبار يمكن أن يستخدم في أغراض متعددة، فإنه يمكن أن يكون للإختبار عدة مؤشرات للصدق وفقاً لأغراض استخداماته، ونوع القرارات التي يمكن اتخاذها استناداً إلى درجاته.

لذلك فإن هناك جوانب متعددة للصدق تناول ثلاثة أقسام رئيسة من القرارات، أحدها: صدق المحتوى، ويتعلق بعدي تمثيل محتوى الإختبار للنطاق السلوكي الشامل للسمة المراد الاستدلال عليها، والثاني الصدق المرتبط بمحك، ويتعلق بعدي صلاحية الإختبار أو المقاييس في أغراض التنبؤ من أجل اتخاذ قرارات عملية مثل قرارات الانتقاء أو التصنيف. ويمكن تمييز نوعين من هذا الصدق أحدهما: الصدق التنبؤى الذي يدل على مدى ارتباط أداء الأفراد في الإختبار بأدائهم في محك مستقبلي، والآخر: الصدق التلازمي، ويدل على اقتران تباين درجات اختبار بدرجات اختبار أو مقاييس آخر يطبق في الوقت نفسه تقريباً على مجموعة من الأفراد. وتتعدد طرق تقدير الصدق التنبؤى، من أهمها إيجاد الارتباط بين الإختبار التنبؤى واختبار المحك، أو استخدام معادلة الانحدار في التنبؤ بدرجات الإختبار المحك أو استخدام جداول الترافق ، أو الاستناد إلى نظرية المفعنة. غير أن هناك بعض المشكلات المتعلقة بالمحك، وما ينبغي أن تتوافر فيه من شروط، مثل ثبات درجاته، وموضوعيته، ومواءمته لنوع القرار المراد اتخاذه . كما أن الصدق المرتبط بمحك يتأثر بجموعة من العوامل من أهمها: مدى تجانس عينة الأفراد المختبرين وعددها، وثبات درجات المحك وعدد مفرداته، ومدى تأثيره بمتغيرات دخلية، والمدة الزمنية التي تفصل بين تطبيق الإختبار التنبؤى واختبار المحك .

أما الجانب الثالث من الصدق فيتعلق بصدق التكوين الفرضي الذي يتناول العلاقة بين نتائج الاختبار أو المقياس وبين المفهوم النظري الذي يهدف الاختبار لقياسه. ونظرًا لأن التكوينات الفرضية هي التي يتركز عليها الاهتمام في القياس، فإن هذا الجانب من الصدق قد نال قدرًا كبيراً من الجهود البحثية، وتعدّت منهجيات وأساليب التتحقق منه. فهناك أساليب تعتمد على الارتباطات، وأساليب تعتمد على التجريب، وأساليب تعتمد على التحليل المنطقي والفحص الدقيق لل اختبار والأداء الذي يتطلبه في ضوء النظرية التي يستند إليها الاختبار.

غير أنه لا يجب النظر إلى الجوانب الثلاثة للصدق على أنها جوانب متباعدة، وإنما تعد جوانب متكاملة لعملية متسعة تتعلق بالتحقق من صدق تفسيرات واستخدامات الدرجات المستمدّة من أدوات القياس. كما أن مفهومي الصدق والثبات تُعدان من المفاهيم المتراكبة، ويندرجَا تحت مقاييس «قابلية التعميم» التي سنوضحها في الفصل الرابع عشر.

الفصل السادس

معايير الاختبارات ومحکات الأداء

* مقدمة

- * معايير الاختبارات مرجعية الجماعة أو المعيار
- * أنواع الجماعات المرجعية
- * خصائص عامة للجماعة المرجعية
- * أنواع معايير الاختبارات
- * أولاً : الدرجات المحوّلة
- * الدرجات التائية الخطية والمعيارية
- * التساعيات المعيارية
- * ثانياً : معايير المئنيات والرتب المئنية
- * ثالثاً : معايير الارتفاع
- * معايير العمر الرمزي
- * معايير الفرق الدراسية
- * رابعاً : الصفحة النفسية أو التربية (البروفيل النفسي أو التربوي)
- * محکات الأداء في الاختبارات مرجعية المحک

مقدمة :

يعد مفهوم معايير الاختبارات Tests Norms من المفاهيم الأساسية المتعلقة بتفسير درجات الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests ، فالدرجة التي يحصل عليها فرد في اختبار ما والتي تسمى الدرجة الخام Raw score لا يكون لها معنى ويصعب تفسيرها ما لم يتم إسنادها إلى نظام مرجعي Reference system ، فهذا النظام هو الذي يسمح باستخلاص معلومات مفيدة من درجات الاختبار.

والنظام المرجعي المناسب لهذا النوع من الاختبارات يعتمد على استخدام معلومات يتم الحصول عليها من جماعة معينة تسمى الجماعة المرجعية Norm Group ، وهذه الجماعة تكون خصائصها محددة ومعلومة لمستخدمي الاختبار، لذلك تسمى الاختبارات التي تستخدم هذا النظام الاختبارات مرجعية الجماعة أو المعيار.

وهناك نظام آخر لا يعتمد على جماعة مرجعية، وإنما يعتمد على محك أداء متوقع Criterion ، ويتحدد هذا المحك في ضوء الكفايات التي تشتمل على معارف ومهارات متوقعة ينبغي أن يتلکها الفرد.

ونظراً لأن تفسير درجات الاختبارات في هذه الحالة يستند إلى هذا المحك فإن الاختبارات عندئذ تسمى الاختبارات مرجعية المحك Criterion- Referenced Tests . وقد أشرنا إلى التمييز بين هذين النوعين من الاختبارات في الفصل الأول، وسوف نتناولهما بالتفصيل في الفصل الثامن.

وينصب الاهتمام في النظام الأول على مقارنة الدرجة التي يحصل عليها فرد في اختبار مرجعى الجماعة بأداء أفراده بهدف ترتيب درجات الأفراد في الاختبار بالنسبة لبعضهم البعض، وبذلك يمكن تحديد المركز النسبي للفرد بين أفراده في ضوء معيار جماعته المرجعية .

إذ لا تستطيع إجراء هذه المقارنة باستخدام الدرجات الخام أو النسبة المئوية من العدد الكلى لأسئلة الاختبار التي يجب عنها كل فرد إجابة صحيحة. فلذلك نحدد على وجه الدقة مركز الفرد بالنسبة للجماعة المرجعية ينبغي تحويل هذه الدرجات الخام إلى نوع آخر من الدرجات يسمح بهذه المقارنة، مثل الدرجات المعيارية ، والدرجات التائية ، والمئويات ، والرتب المئوية ، وغيرها من الدرجات المحولة التي ناقشناها بإسهاب في الفصل الثالث.

ولتوضيح ذلك نفرض أن درجة فرد في اختبار للذكاء ١٣٢ ، فهنا ربما ينحدر إلى الأذهان أن هذا الفرد أجاب عن ١٣٢ مفردة إجابة صحيحة، أو أن الفرد حصل على ١٣٢ نقطة من المجموع الكلى للنقاط وهو ٢٠٠ مثلاً.

ولكن هذه الدرجة الخام لا تنقل لنا صورة واضحة عن مركزه بالنسبة لأقرانه في الذكاء، لذلك لابد من إجراء نوع من التحويل على هذه الدرجة لكي يتسعى لنا مقارنتها بغيرها من الدرجات.

ومن هنا تتضح أهمية معايير الاختبارات، فدرجات اختبارات الذكاء والشخصية والاختبارات التحصيلية المقننة وغيرها من الاختبارات والمقاييس تفسر عادة باستخدام جداول المعايير.

• معايير الاختبارات مرجعية الجماعة أو المعيار:

إن المعايير عبارة عن مجموعة من الدرجات المشتقة بطرق إحصائية معينة من الدرجات الخام كما ذكرنا بحيث تأخذ بعين الاعتبار توزيع الدرجات الخام المستمدة من تطبيق الاختبار على عينة عشوائية ممثلة للمجتمع المستهدف.

و قبل أن نوضح أنواع المعايير التي تستخدم في تفسير الدرجات ينبغي أن نلقي مزيداً من الضوء على أنواع الجماعات المرجعية التي يستند إليها في التوصل إلى هذه المعايير، وخصائصها وكيفية انتقاءها.

أنواع الجماعات المرجعية :

تعتمد معايير الاختبارات مرجعية الجماعة أو المعيار على الجماعة المرجعية التي تستمد منها هذه المعايير Norm Group . لذلك ينبغي أن تكون خصائص هذه الجماعة محددة تحديداً جيداً ليتسنى تفسير درجات هذه الاختبارات تفسيراً دقيقاً.

ولتوضيح ذلك نفترض أننا قارنا درجة طالب في اختبار تحصيلي في الفيزياء بدرجات مجموعة من الطلاب المبتدئين في دراسة هذه المادة في المرحلة الثانوية ، فإن ترتيب هذا الطالب سيكون مرتفعاً بالنسبة لهذه المجموعة ، أما إذا قارنا درجته بدرجات مجموعة أخرى من الطلاب المتخصصين في الفيزياء ، فإن ترتيبه سيكون منخفضاً. أي أنها إذا لم نعرف شيئاً عن طبيعة الجماعة التي نقارن الطالب بها ، فإنه لا يمكن التوصل إلى استنتاج سليم فيما يتعلق بالمركز النسبي للطالب في الفيزياء . ومن هنا تبدو أهمية تحديد خصائص الجماعة المرجعية التي سوف يقارن بها الطالب .

فخصائص الجماعة المرجعية ينبغي أن تكون مشابهة لخصائص أفراد الجماعة الذين سيطبق عليهم الاختبار فيما بعد من أجل تقويم السمة التي يقيسها لديهم في ضوء مركزهم النسبي .

لذلك يبرر التساؤل التالي : ما الخصائص التي ينبغي أن تتوافر في الجماعة المرجعية؟ والحقيقة أن هذه الخصائص تتباين بحسب طبيعة المعايير المستخدمة ، فالمعايير الوطنية National Norms تختلف عن المعايير المحلية Local Norms وكلاهما يختلف عن المعايير المتعلقة بموقع واحد مثل مؤسسة أو مدرسة معينة Special Norm Institution Average Norms Groups . لذلك ربما يكون من المناسب توضيح كل من هذه المعايير فيما يلى :

المعايير الوطنية : تقييد هذه المعايير في تقويم البرامج التعليمية باستخدام الاختبارات التحصيلية المقننة؛ وذلك للتحقق مما إذا كان أداء طلاب نهاية مرحلة تعليمية معينة أو فرقة دراسية أعلى أو أدنى من المتوسط على المستوى الوطني .

لذلك تقوم مؤسسات متخصصة بإعداد هذه الاختبارات نظراً لما تتطلبه عمليات بنائها وإعداد معاييرها الوطنية من وقت وجهد كبير وكلفة مرتفعة . فالجماعة المرجعية الوطنية National Norms Group يفترض أنها بمثابة عينة ممثلة لطلاب الوطن في مرحلة تعليمية معينة، غير أن تحديد هذه العينة وتمثيلها للوطن يختلف باختلاف وجهة نظر المؤسسة المعينة .

لذلك يصعب مقارنة المعايير الوطنية للختبارات التي تقوم مؤسسات مختلفة بإعدادها . غير أن انتقاء هذه العينة ينبغي أن يأخذ بعين الاعتبار الخصائص الديموغرافية المختلفة ، مثل المناطق الجغرافية ، والفرق الدراسية ، وحجم المدرسة وموقعها ، والحالة الاجتماعية والاقتصادية ، والنوع والتركيبة الطلابية ، وغير ذلك من الخصائص ذات الأهمية .

ويتم تحديد ذلك بالتفصيل في دليل الاختبار لكي يتحقق مستخدمو الاختبار من مدى تمثيل العينة الوطنية للمجتمع المستهدف . وأحياناً تعد المؤسسة المعنية معايير مستقلة للمجموعات الفرعية في كل من هذه الخصائص على حدة إذا تبين أن كلا من هذه المجموعات تختلف عن الأخرى في أي من هذه الخصائص .

ويتم اختيار هذه العينة باستخدام أساليب المعاينات العشوائية Random أو المعاينات العشوائية الطبقية التي تراعي الخصائص أو المتغيرات السابقة Startified Random Sampling . ويجب أن تكون العينة المتنقلة كبيرة بدرجة كافية لضمان ملاءمة المعايير للمجتمع المستهدف ، وتقليل أخطاء المعاينات Sampling Errors .

وستستخدم المعايير الوطنية أيضاً في اختبارات الاستعدادات، مثل اختبار الاستعداد الأكاديمي Scholastic Aptitude Test (SAT) الذي يستخدم في الولايات المتحدة الأمريكية للتنبؤ بالنجاح الأكاديمي، حيث اعتمدت معاييره على عينات من طلاب نهاية المرحلة الثانوية اشتملت على عدةآلاف من الطلاب من مختلف الولايات لكل اختبار فرعى.

غير أننا نفتقر في الوطن العربي إلى مثل هذه الاختبارات التي تعتمد على معايير وطنية أو إقليمية يمكن باستخدامها مقارنة أداء الطالب في مجال دراسي معين بمتوسط أداء أقرانه على هذين المستويين.

المعايير المحلية: تعد هذه المعايير في بعض الأحيان أكثر فائدة من المعايير الوطنية. فنادرًا ما نجد تطابقاً بين خصائص الجماعة المرجعية التي تستمد منها المعايير الوطنية وخصائص الجماعة التي يطبق عليها الاختبار فيما بعد، إذ إن هذه المعايير يختلف معناها بمرور الزمن وتغير خصائص الأفراد مما يؤدي إلى صعوبات في تفسير الدرجات تفسيراً ذي معنى.

فمثلاً التركيبة الطلابية لمجتمع محلى معين تختلف عادة عن نظيرتها فى الجماعة المرجعية، لذلك ربما تفضل المؤسسات التعليمية جمع بيانات من طلاب هذا المجتمع وإعداد معايير محلية استناداً إلى هذه البيانات وذلك لتفسير درجاتهم فى الاختبارات المرجعية الجماعة أو المعيار تفسيراً أكثر مواءمة لظروف المجتمع المحلى الذى تتوارد فيه المؤسسة التعليمية. أى أن المعايير المحلية تسمح بمقارنة أداء الطالب بأقرانه فى المنطقة التى يعيش فيها، ولا تسمح بالمقارنة خارج هذه المنطقة.

وكثير من اختبارات الذكاء مثل اختبار ويكسيلر Wechsler Test يتضمن معايير تستند إلى الراشدين بعامة مما يؤدي إلى صعوبات فى تفسير درجات بعض الأفراد. فإذا كانت درجة طالب فى الاختبار تدل على أن ذكاءه أعلى من متوسط أقرانه الذين فى مستوى العمرى اعتماداً على معايير الكبار بعامة، فإنه يمكن أن يكون ذكاؤه أقل من متوسط الطلاب الملتحقين بالكليات. فالتنبؤ بنجاحه عنده فى إحدى الكليات يتطلب مقارنة درجته بمتوسط أداء هؤلاء الطلاب بدلاً من مقارنتها بمتوسط أداء الراشدين بعامة، وهذا يتطلب معايير مشتقة من درجات هؤلاء الطلاب فى الاختبار.

وإعداد المعايير المحلية لا يحتاج إلى جهود مؤسسات متخصصة وإنما ربما يحتاج إلى استشارتها فقط، حيث إن هذه المعايير تعتمد على جماعة مرتجعية محلية مما ييسر عملية انتقاء العينة واختبارها وإعداد المعايير باستخدام الحاسوب .

كما أنها تيسر تفسير درجات الاختبار للمعلم والطالب وولي أمره والمجتمع المحلي، وتفيد المعلم في تبيّن التقدّم الدراسي لتلاميذه ومقارنة أدائه في المجالات الدراسية المختلفة بالنسبة لأقرانه.

ومن الجدير بالذكر أنه يمكن استخدام المعايير الوطنية والمعايير المحلية معًا حيث إنها متكاملان. فإذا أردنا مقارنة تقدّم مجموعات من الطلاب في برامج تعليمية متماثلة، فإن المعايير المحلية تكون مناسبة، ولكن إذا أردنا أيضًا مقارنة الأداء المحلي بالأداء على المستوى الوطني، فإن المعايير الوطنية تصبح ضرورية ومكملاً للمعايير المحلية.

معايير مؤسسة أو مدرسة معينة: تفيد هذه المعايير في مقارنة متوسط أداء طلاب فرق دراسية معينة في إحدى المدارس بمتوسط أداء الفرق المنشورة في مدارس أخرى. ونلاحظ أن الاهتمام هنا لا يتركز على مقارنة طالب معين بأقرانه وإنما مقارنة متوسط أداء الفرق بعامة.

لذلك ينبغي تحديد متوسطات أداء هذه الفرق لعينة مماثلة من المدارس على المستوى الوطني، وترتيب هذه المتوسطات ترتيباً نسبياً. وعادة يكون تباين توزيع هذه المتوسطات أقل من تباين توزيع متوسط درجات الطلاب.

كذلك يمكن أن تعد مؤسسة معينة معايير خاصة للاختبارات التي تستخدم في انتقاء المتقدمين لشغل وظائف بها، أو يقوم مكتب القبول بالجامعات بإعداد معايير لاختبارات الاستعدادات التي يطبقها على الطلاب المتقدمين للالتحاق بكلية معينة.

معايير فئات خاصة: تحتاج بعض الاختبارات إلى معايير لفئات خاصة من ذوي الإعاقات العقلية أو الجسمية مثل اختبارات الذكاء، أو لفئة معينة من الأفراد مثل طلاب الفرق الأولى بكلية الطب، أو الطلاب المتقدمين للالتحاق بكليات الهندسة، أو أفراد بعض المهن، وكذلك لمؤسسات معينة مثل المدارس الخاصة، أو المعاهد الدينية، أو الشركات الاستثمارية، وهكذا. ففي مثل هذه الحالات ربما تبدو ضرورة إعداد معايير خاصة لهذه الفئات للاسترشاد بها في اتخاذ قرارات مناسبة بشأنهم.

كما يمكن إعداد معايير خاصة لاختبارات الاستعدادات الموسيقية، والجمالية، والكتابية، والميكانيكية، وبعض الاختبارات التحصيلية المقننة، مثل اختبارات اللغات الأجنبية، والحسابات الآلية، وما شابه ذلك.

خصائص عامة للجامعة المرجعية :

على الرغم من اختلاف الجماعات المرجعية تبعاً لطبيعة هذه الجماعات ومدى اتساع عيناتها، إلا أنه ينبغي أن تتوافر فيها بعض الخصائص العامة لكي تكون هذه

الجماعات مناسبة للمعايير التي تستمد منها. ومن أهم هذه الخصائص التي حدتها الجمعية الأمريكية لعلم النفس (APA) ما يلى:

تمثيل عينة الأفراد للمجتمع المستهدف: عند انتقاء عينة من الأفراد من أجل بناء المعايير ينبغي مراعاة أن تكون هذه العينة ممثلة تمثيلاً كافياً للمجتمع المستهدف.

إذا اقتصرت العينة على إحدى المدارس أو المناطق الجغرافية، فإن العينة في هذه الحالة تكون متحيزة. إذ ربما تكون هذه المدرسة متميزة أو أداء طلابها أقل من متوسط أداء طلاب مجتمع هذه المدرسة أو هذه المنطقة، فنوع المعايير سواء أكانت وطنية أم محلية أم خاصة يفرض طبيعة العينة المستخدمة. فإذا كانت المعايير وطنية، فإنه ينبغي التأكد من تمثيلها تمثيلاً متوازناً للأفراد من مختلف المناطق، ومتناسباً مع المقيمين في الريف والحضر ، والذكور والإإناث ، وغير ذلك من المستويات الديموغرافية .

وعلى الرغم من أن العينة العشوائية ربما تناسب معايير الاختبارات التي تقتصر على مدينة أو منطقة واحدة، حيث يمكن انتقاء عدة مئات من الأفراد عشوائياً، إلا أنها تصبح غير ممكنة إذا كانت المعايير وطنية. فانتقاء عينة عشوائية كبيرة الحجم من مختلف محافظات الوطن يتطلب جهوداً كبيرة ونفقات باهظة، بل وربما تؤدي إلى زيادة خطاء المعاينات.

لذلك يفضل استخدام العينات العشوائية الطبقية حيث يتم تصنيف المجتمع المستهدف وفقاً لعدد من المتغيرات التي أشرنا إليها وانتقاء نسبة معينة من كل مجموعة تصنيفية بحسب عددها في المجتمع، وبذلك تقل أخطاء المعاينات ويقتصر الجهد والكلفة. غير أنه ينبغي مراعاة أن تمثل العينة للمجتمع لا يعني زيادة عددها فحسب، وإنما يعني أيضاً مراعاة عدم تحيزها استناداً إلى تصميم معاينات Sampling Design حيث العينة غير المتتحيز إذا كان مناسباً تكون المعايير المستعملة منها متميزة بالاستقرار النسبي بحيث إذا استخدمت عينة أخرى مشابهة نحصل على نتائج مماثلة ، وهذا ما يجب أن يتواوف في معايير الاختبارات التربوية والنفسية. وربما كان هذا هو أحد الأسباب التي تدعو إلى إعداد معايير محلية إلى جانب المعايير الوطنية.

حداثة المعايير: إن معايير الاختبارات تتغير بتغير خصائص الأفراد بمرور الزمن وبخاصة إذا كانت الاختبارات تقيس خصائص تتأثر تأثراً ملحوظاً بما يطرأ على البيئة من تطورات . فمثلاً المفردات اللغوية تنمو وتتغير دلالتها، والمناهج الدراسية تتتطور ، والبيئة التعليمية تحدث فيها تعديلات مستمرة. لذلك فإن معايير الاختبارات التحصيلية مثلًا تتأثر بجميع هذه العوامل التي تتعكس على الطلاب، بل وربما تتأثر اختبارات الاستعدادات ومقاييس الشخصية أيضاً ببعض هذه العوامل وغيرها.

وأحياناً يظل محتوى الاختبارات كما هو على الرغم من التغيرات التي ربما تحدث للأفراد نتيجة التطورات المستمرة التي تحدث في المجتمع مما يجعل معايير هذه الاختبارات غير صالحة . فالجامعة المرجعية التي استمدت منها معايير اختبار معين منذ عدة سنوات لا تصبح مناسبة للاختبار في الوقت الحاضر .

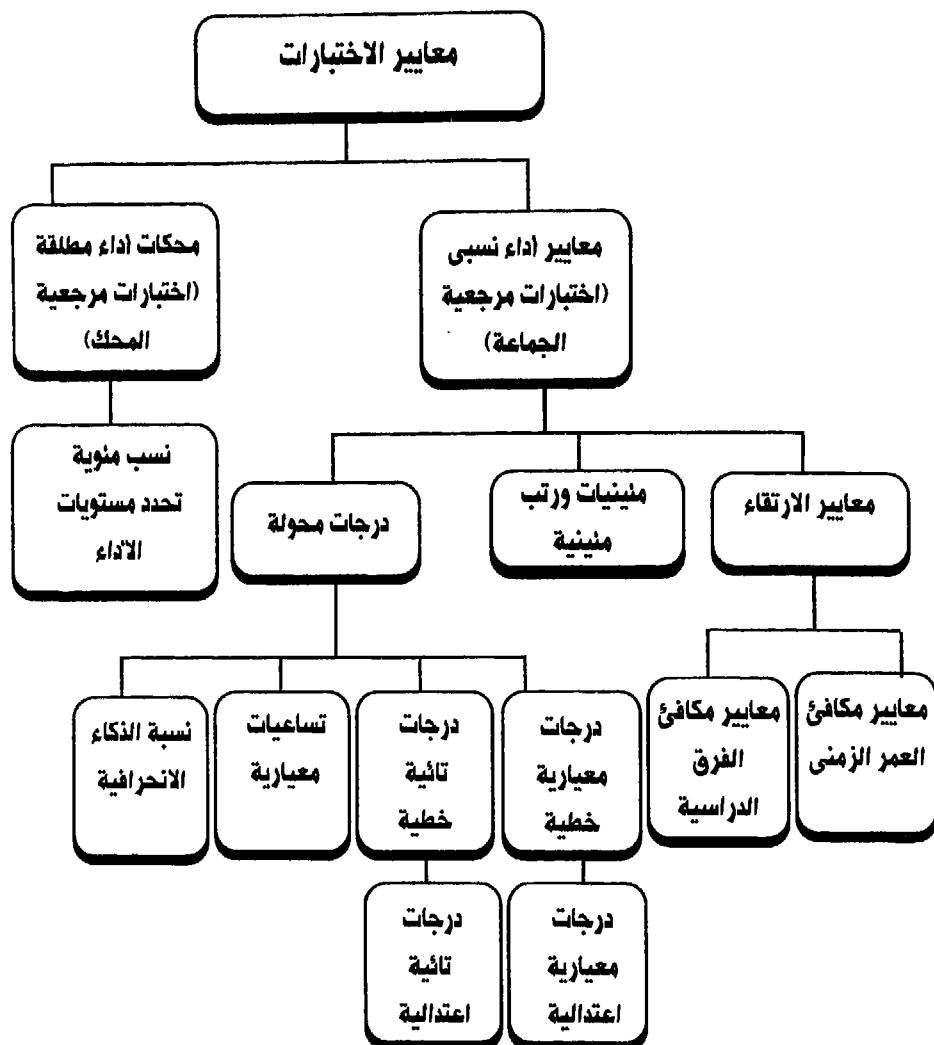
لذلك ينبغي تحديث معايير الاختبارات بصفة مستمرة وفقاً لما يطرأ على الخصائص المحددة للجامعة المرجعية من تغير ، أو ما يطرأ على محتوى الاختبار من تطوير . ولا ينبغي التتحقق من حداة المعايير بالنظر إلى تاريخ نشر الاختبار ، إذ ربما تصدر طبعات جديدة من الاختبار نفسه دون إعادة النظر في الجامعة المرجعية التي استمدت منها معايير الاختبار .

موائمة المعايير لمن سيطبق عليهم الاختبار؛ فالجامعة المرجعية تكون موائمة Relevant إذا اتفقت خصائصها مع خصائص من سيطبق عليهم الاختبار . فإذا كانت هذه الجماعة تمثل أفراد مهنة معينة ، فإنه ينبغي أن تكون خصائصها مماثلة لخصائص أفراد المهنة نفسها الذين نود اختبارهم . ولكن إذا استُخدم الاختبار في أغراض متعددة ، فإنه ينبغي أن تُستمد معاييره من أكثر من جماعة مرجعية واحدة مناسبة .

أنواع معايير الاختبارات :

تتعدد أنواع معايير الاختبارات التي تُستخدم في مقارنة درجة الفرد في اختبار ما بدرجات جماعة مرجعية من أقرانه . وتُتحقق جميع هذه المعايير من الدرجات الخام للاختبار . غير أننا سوف نركز في عرضنا على المعايير شائعة الاستخدام في الاختبارات والمقياس المنشورة لكي يتمكن الباحث ومستخدم الاختبار من تفسير الدرجات المستمدة منها في ضوء المعيار المناسب .

ويمكن تصنيف هذه الأنواع من المعايير بحسب طريقة تفسير كل منها للمرجع السببي للفرد فيما يقيسه الاختبار أو المقياس ، ويوضح الشكل التخطيطي (٦ - ١) التالي هذا التصنيف المقترن :



شكل (٦ - ١) يوضح أنواع معايير الاختبارات مرجعية الجماعة

ومحکات الأداء في الاختبارات مرجعية المحك

(من تصميم المؤلف)

أولاً: الدرجات المحولة :

تستخدم كثير من الاختبارات المنشورة في الآونة الأخيرة معايير تعتمد على الدرجات المعيارية المحولة مثل الدرجات المعيارية الاعتدالية Normalized Z Scores والدرجات التائية الخطية Linear T Scores ، والدرجات التائية المعيارية Normalized

T Scores، والتساعيات المعيارية Stanines، ودرجات معيارية محولة خاصة ببعض اختبارات القبول بالكلليات في الولايات المتحدة الأمريكية ، مثل اختبار الاستعداد الأكاديمي (SAT) Scholastic Aptitude Test ، واختبار سجل الدراسات العليا-Grad- uate Record Examintion (GRE) . وقد أعد هذان الاختباران مكتب امتحانات التساعيات (CEEB) College Entrance Examination Board . كما أن معايير القبول بالكلليات Wechsler Intelligence Test تعتمد على بعض اختبارات الذكاء مثل اختبار ويكسنر Deviation IQ (DIQ) ، واختبار التصنيف العام لأفراد الخدمة العسكرية Army General Classification Test يعتمد على معايير يشار إليها بالحروف الأولى لاسم الاختبار (AGCT) . كما تستخدم بطاريات الاختبارات الصفيحة التربوية أو النفسية Profiles . وبعض هذه الاختبارات تم تعريفه وتقنيته على البيئة المصرية وبعض البيانات العربية الأخرى مما سيتضح عند مناقشة الاختبارات والمقياسات التربوية والنفسية في فصول لاحقة . ويمكن أن يرجع القارئ إلى الفصل الثالث للاطلاع على الدرجات المحولة بأنواعها المختلفة وكيفية حسابها . وسوف نقتصر هنا على توضيع أهم هذه الدرجات المحولة التي تُستخدم كمعايير للاختبارات والمقياسات .

الدرجات التائية الخطية والمعيارية :

سبق أن ذكرنا في الفصل الثالث أن الدرجات التائية الخطية تحصل عليها بإجراء تحويل معين على الدرجات المعيارية الخطية ، وذلك بضرب الدرجة المعيارية في ١٠ وجمع ٥٠ على حاصل الضرب ، وبذلك يصبح متوسط الدرجة التائية الخطية ٥٠ ، وانحرافها المعياري ١٠ .

إذا جعلنا شكل توزيع الدرجات الخام اعتدالياً يصبح توزيع الدرجات المعيارية الخطية والدرجات التائية الخطية اعتدالياً ، بغض النظر عن شكل توزيع الدرجات الخام قبل جعله اعتدالياً ، وعندئذ نحصل على الدرجات المعيارية الاعتدالية ، والدرجات التائية الاعتدالية . وتدون في أدلة الاختبارات الدرجات الخام وما يقابلها من درجات تائية خطية أو اعتدالية ، فإذا كان متوسط درجات اختبار ما ١٠٠ وانحرافه المعياري ١٦ وحصل فرد على الدرجة ١١٦ في الاختبار ، فإن الدرجة المعيارية الخطية المقابلة لدرجته الخام = $\frac{116 - 100}{16} + 1$ ، وهذه بدورها تقابل الدرجة التائية الخطية $1 + (1 + 1)(1 + 1) = 60$. وهذا يدل على أن درجته في الاختبار تفوق المتوسط بقدر وحدة انحراف معياري واحد (١٠ وحدات تائية) بغض النظر عن شكل توزيع الدرجات الخام في الاختبار .

أما إذا جعلنا شكل توزيع الدرجات الخام اعتدالياً، فإنه يمكن تفسير درجة الفرد في ضوء معيار المئينيات الذي ناقشناه في الفصل الثالث، وسوف نوضح استخداماته في تفسير درجات الاختبارات بعد قليل.

محايد اختبار الاستعداد الأكاديمي :

أعد هذا الاختبار مكتب امتحانات القبول بالكليات في الولايات المتحدة الأمريكية (CEEB) عام ١٩٤١ ، ويطبق عادة لانتقاء الطلاب المتقدمين للالتحاق بالكليات. ويشمل الاختبار على اختبارين فرعيين أحدهما لغطي والآخر رياضي. وطبق على مجموعة كبيرة من الطلاب بلغت ١٠٦٥٤ طالباً. وتعتمد معايير كل من هذين الاختبارين الفرعيين على تحويل الدرجات الخام إلى درجات معيارية معدلة (CEEB) متوسطها ٥٠٠ وانحرافها المعياري ١٠٠ . ويجري ذلك بضرب الدرجات المعيارية الخطية (Z) المقابلة للدرجات الخام في ١٠٠ وإضافة ٥٠٠ إلى حاصل الضرب، وبذلك يتحول توزيع الدرجات المعيارية الخطية إلى توزيع جديد قيمة كل من متوسطه وانحرافه المعياري أكبر من قيمتهما في معيار الدرجة الثانية T Score .

وقد أجريت بعض التعديلات الإحصائية على درجات الطلاب الذين اخترعوا بعد عام ١٩٤١ لضممان توحيد تفسير الدرجات في ضوء معايير الاختبار من عام إلى آخر.

محايد اختبار التصنيف الخام لأفراد الخدمة العسكرية ، واختبار ويكسنر للذكاء :

فالاختبار الأول كان يستخدم لتصنيف الأفراد الملتحقين بالجيش في الولايات المتحدة الأمريكية وتسكينهم في الوحدات المختلفة، وبعد من الاختبارات شائعة الاستخدام في هذا الغرض. واعتمدت معاييره (AGCT) على تحويل الدرجات الخام إلى درجات معيارية معدلة متوسطها ١٠٠ وانحرافها المعياري ٢٠ ، وهما قيمتان اختياريتان.

أما اختبار ويكسنر للذكاء فيعد من الاختبارات المعروفة، واعتمدت معايير درجات اختباراته الفرعية على تحويل الدرجات الخام إلى درجات معيارية معدلة متوسطها ١٠ وانحرافها المعياري ٣ ، بينما اعتمدت معايير درجاته الكلية على تحويل الدرجات المعيارية المقابلة لها إلى درجات معيارية معدلة متوسطها ١٠٠ وانحرافها

المعيارى ١٥ ، وهمًا قيمتان اختياريتان أيضًا. وتسمى هذه المعايير نسبة الذكاء الانحرافية (IQ) (DIQ) .

فإذا حصل فرد في مستوى عمرى معين على الدرجة الخام الكلية ١٣٠ في هذا الاختبار، فإن معنى ذلك أن درجته تفوق متوسط مجموعته العمرية بقدر انحرافين معياريين.

وتفسر جميع هذه المعايير بنفس الطريقة التي تفسر بها معايير الدرجات التائية. وعادة تزودنا أدلة الاختبارات بجدواول تحول فيها الدرجات الخام مباشرة إلى هذه المعايير، وتسمى جدواول المعايير Tests Norms .

التساعيات المعيارية :

تعد من المعايير التي تعتمد أيضًا على تحويل الدرجات المعيارية إلى نظام آخر يختلف في المتوسط والانحراف المعياري عن المعايير السابقة، فالتساعيات المعيارية Stanines هي درجات معيارية اعتدالية متوسطها ٥ وانحرافها المعياري ٢ تقريبًا، واشتقت اسمها Stanine من اختصار الكلمة Standaradized أي مقمن أو معياري ، وكلمة nine أي تسعه ، وبذلك سميت التساعيات المعيارية .

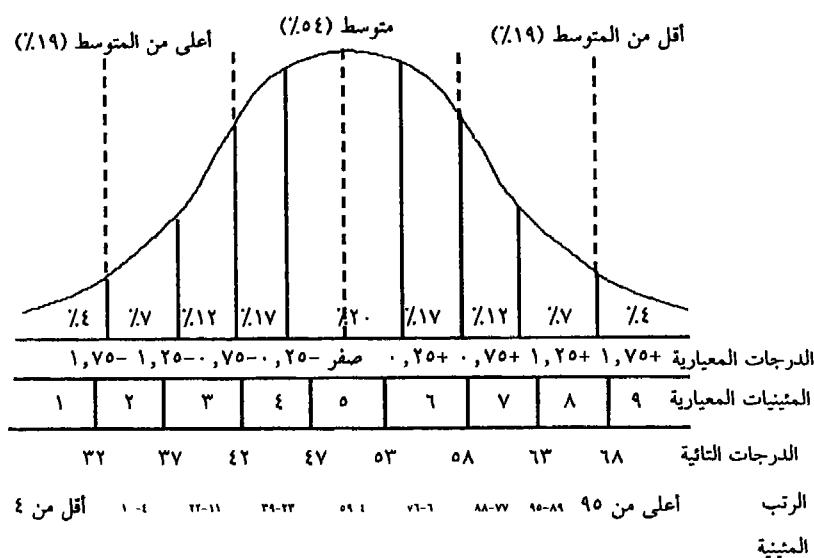
ولم يستخدم هذا المعيار بكثرة إلا قرب الحرب العالمية الثانية، حيث كانت القوات الجوية الأمريكية تتنفيذ برنامجاً نفسياً يتطلب تطبيق العديد من الاختبارات على آلاف من الجنود، وكان لابد من البحث عن معيار بسيط يمكن استخدامه في تحويل الدرجات الخام لهذه الاختبارات إلى نوع آخر من الدرجات يسهل فهمها وتفسيرها وتقرير نتائجها .

وقد وجد علماء النفس أن التساعيات المعيارية تفي بذلك ، وانتشر استخدام هذا المعيار بعد ذلك في كثير من الاختبارات التربوية والنفسية مرجعية الجماعة أو المعيار، بل أصبح يفضل استخدامه عن غيره من معايير الدرجات المحولة الأخرى . ولا يرجع ذلك فقط إلى سهولة اشتقاقه واستخدامه ، وإنما إلى مزايا أخرى تتعلق بتفسير الدرجات سوف نوضحها بعد قليل .

ويتبين من اسم هذا المعيار أنه يشتمل على تسع مجموعات من نقاط الدرجات بحيث تقع نسبة مئوية معينة من الأفراد المختبرين في كل مجموعة من هذه المجموعات . وبعبارة أخرى تتحصر نسبة مئوية معينة من الأفراد بين الانحرافات المعيارية المختلفة تحت المنحني الاعتدالي ، وبذلك ينقسم المنحني الاعتدالي إلى

تسعة أقسام بواسطة التساعيات المعيارية. ويحد كل قسم منها فترات تمثل نصف انحراف معياري، ويعين لكل قسم الأعداد الصحيحة المرتبة من 1 إلى 9. أى أن التساعيات المعيارية ليست نقاطا وإنما فترات، ولذلك فهي تختلف عن الدرجات المحوّلة السابقة.

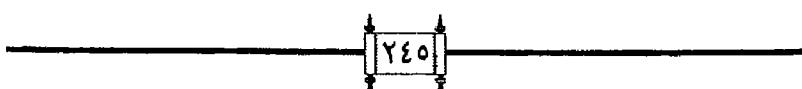
ويوضح الشكل (٦ - ٢) التالي الأقسام التسعة، والنسب المئوية التقريرية تحت المنحنى الاعتدالى، والدرجات المعيارية والتائية والرتب المئنية المتناظرة لها:



شكل (٦ - ٢) يوضح التساعيات المعيارية والنسب المئوية من الأفراد

والدرجات المعيارية والتائية المتناظرة لكل منها

ويتضح من شكل (٦ - ٢) أن سعة التساعيات المعيارية من الثنائي إلى الثامن متساوية مما يدل على وحدات درجات معيارية متساوية، بينما كل من التساعي الأول والتاسع أكثر اتساعاً من التساعيات الأخرى، ونهايتهما مفتوحة. كما يتضح أن التساعي المعياري الخامس يمتد من ربع انحراف معياري يسار المتوسط إلى ربع انحراف معياري يمين المتوسط، وتقع حوالي ٢٥٪ من الحالات بين هاتين القيمتين. وكذلك نلاحظ أنه كلما اتجهنا إلى أقصى اليمين أو أقصى اليسار نقل النسب المئوية لعدد الحالات، ويمكن استدلال هذه النسب المئوية بالطرق الرياضية.



ولتحويل الدرجات الخام إلى تساعيات معيارية ينبغي ترتيب الدرجات ترتيباً تنازلياً، ونعين للنسبة ٤٪ العليا من الدرجات التساعي المعياري ٩ ، وللنسبة المئوية ٧٪ التالية في الترتيب التساعي المعياري ٨ ، وللنسبة المئوية ١٢٪ التالية التساعي المعياري ٧ ، وهكذا.

وبهذا الشكل تكون قد فرضنا التوزيع الاعتدالى على الدرجات الخام أو جعلنا توزيع هذه الدرجات يقترب من التوزيع الاعتدالى ، بينما الدرجات المعيارية الخطية أو الدرجات التائية الخطية لا تغير من شكل التوزيع الأصلى للدرجات.

ولكن ينبغي مراعاة أنه إذا كان توزيع درجات سمة معينة لمجموعة من الأفراد ليس اعتدالياً بطبيعته ، فإنه لا يجوز إجبار هذا التوزيع ليكون اعتدالياً . ولكن إذا كان ابتعاد توزيع الدرجات عن الاعتدالية ناجم عن أخطاء القياس ، فعندئذ يجوز فرض التوزيع الاعتدالى عليه .

في جميع الأفراد الذين تقع درجاتهم داخل فترة معينة يعين لهم التساعي المعياري المناظر لهذه الفترة . فمثلاً يعين لجميع الأفراد الذين درجاتهم الخام تنازلي الدرجات المعيارية المحصورة بين +٢٥ ، +٧٥ ، أو الدرجات التائية المحصورة بين ٥٣ ، ٥٨ التساعي المعياري ٧ حيث تقع ١٢٪ من الحالات ، وهكذا في بقية الدرجات .

ومن ميزات التساعيات المعيارية أن تساوى فتراتها من التساعي الثاني حتى التساعي الثامن يعني أن التساعي ٨ أفضل من التساعي ٦ بنفس قدر أفضلية التساعي ٥ عن التساعي ٣ ؛ لذلك فإن أداء الأفراد في الاختبارات الفرعية المختلفة الذي يعبر عنه بالتساعيات المعيارية يوضح بدقة الجوانب النسبية لقوتهم وضعفهم في سمات مختلفة . ومما يدل على بساطة هذا المعيار أنه يبين مباشرة المركز النسبي لطالب في ضوء أداء أقرانه في الفرق الدراسية نفسها .

فعندما يعلم الطالب أن درجته في اختبار ما تقابل التساعي المعياري ٧ أو ٨ أو ٩ فإنه يفهم من ذلك مباشرة أن أداءه فيما يقيسه الاختبار يتميز على أداء الطالب المتوسط في فرقته .

أما إذا كانت درجته تقابل التساعي المعياري ٢ أو ٣ فإن أداءه يكون ضعيفاً نسبياً . ولذلك يفضل تفسير التساعيات المعيارية تفسيراً وصفياً كما هو موضح فيما بين الخطوط المتقطعة في شكل (٦ - ٢) فمثلاً يمكن تفسير درجات اختبار يقيس المهارات الأساسية في القراءة كالتالي :

وصف الأداء	التساعي المعياري
مرتفع (٪٤)	٩
أعلى من المتوسط (٪١٩)	٧ ، ٨
متوسط (٪٥٤)	٤ ، ٥ ، ٦
أقل من المتوسط (٪١٩)	٢ ، ٣
ضعيف (٪٤)	١

كما يمكن تفسير كل تساعي معياري على حدة من حيث وصفه للأداء، ولكن الوصف في هذه الحالة يكون أكثر تميزاً، مثل أداء مرتفع جداً، أداء مرتفع، أداء فوق المتوسط، أداء فوق المتوسط بقليل، متوسط، أداء أقل من المتوسط بقليل، أداء أقل كثيراً من المتوسط، أداء ضعيف، أداء ضعيف جداً، ويعتمد كل من التفسيرين على طبيعة الاختبار ودرجة التمييز المرجوة.

وبذلك تفيد التساعيات المعيارية في ترتيب أداء أفراد مجموعة معينة في اختبار واحد أو عدة اختبارات، ومقارنة أدائهم في هذه الاختبارات باستخدام هذا المعيار لمتابعة تقدمهم الدراسي أثناء العام أو في عدة أعوام متالية. فالتساعي المعياري ٣ مثلاً لا يختلف معناه من اختبار إلى آخر، حيث يدل على نسبة مئوية من الأفراد تتحصّر درجاتهم بين درجتين معياريتين معينتين.

كما يمكن تعين أوزان معينة للتساعيات المعيارية للحصول على درجة مركبة يمكن استخدامها في التقويم الشامل لأداء الطالب. وتفيد التساعيات المعيارية أيضاً في الإرشاد والتوجيه التربوي والمهني للطلاب، وبخاصة في نهاية المراحل التعليمية حيث تتخذ قرارات تتعلق بالشخصيات أو المناهج الدراسية المناسبة للطلاب وكذلك الاختبارات المهنية. إذ يمكن استخدامها في التمثيل البياني لمستوى أداء الطالب في مجموعة من الاختبارات المتنوعة المناسبة لهذه القرارات، بحيث تتضح جوانب القوة والضعف للطالب مما يساعد في اتخاذ قراره استناداً إلى بيانات موضوعية بدلاً من اعتماده على أحکامه الذاتية.

وعلى الرغم من مزايا التساعيات المعيارية سواء في إعدادها أو تقرير نتائجها باستخدام رقم صحيح واحد، إلا أن هناك بعض أوجه قصور تترجم عن تفسير درجات الاختبارات في ضوء هذا المعيار.

ويتلخص ذلك في أن جميع الأرقام المناظرة للتساعيات المعيارية تمثل مدى متسع نسبياً من مقدار السمة المقاسة كما رأينا. ولذلك يصعب تحديد مقدار النمو الذي حدث في سمة من السمات كالتحصيل أو الذكاء مثلاً باستخدام هذا المعيار، فالدرجة الخام التي يحصل عليها طالب في اختبار تحصيلي معين يمكن أن تزداد بقدر ١٥ نقطة إذا طبق عليه الاختبار قبل عملية التعليم وعقب الانتهاء منها، ومع ذلك تظل درجته في كل من مراتي التطبيق مقابلة لنفس التساعي المعياري نظراً لأن مدى التساعيات متسع كما ذكرنا.

وإذا افترضنا أن الطالب (أ) حصل على الدرجة الخام ٨٤ في التطبيق الأول لأحد الاختبارات ، وأن هذه الدرجة تعد أعلى درجة مقابلة للتساعي المعياري ٦ ، وأن الطالب (ب) حصل على الدرجة ٨٥ في التطبيق الأول للاختبار نفسه ، وأن درجته تعد أقل درجة مقابلة للتساعي المعياري ٧ ، فإذا أعيد تطبيق الاختبار على كل من الطالبين وحصلوا على الدرجتين الخام ٨٥ ، ٨٤ على الترتيب واستخدمنا جدول المعايير نفسه لتفسير هذه الدرجات ، فإننا نلاحظ أن الطالب (أ) قد ارتفعت درجته من ٨٤ إلى ٨٥ أي من التساعي المعياري ٦ إلى التساعي المعياري ٧ ، بينما انخفضت درجة الطالب (ب) من ٨٥ إلى ٨٤ أي من التساعي المعياري ٧ إلى التساعي المعياري ٦ .

وبذلك يبدو أن الطالب (أ) قد حدث له نمو على عكس الطالب (ب) ، مع أن هذا التغير في التساعيات المقابلة للدرجات الخام حدث نتيجة تغير الدرجات الخام نقطة واحدة ارتفاعاً أو انخفاضاً.

كذلك لا نستطيع باستخدامها التمييز بين الأفراد الذين تفوق درجاتهم في أحد الاختبارات التساعي المعياري ٩ أو تقل درجاتهم عن التساعي المعياري ١ ، نظراً لأن نهاية كل منها مفتوحة.

لذلك اقترح علماء القياس جعل عدد التساعيات ١١ بدلاً من ٩ لمزيد من دقة التمييز ، وأطلقوا على المعيار الجديد اسم المعيار ذو الأحد عشر قسمًا Sta - eleven . ويُقسم هذا المعيار كل من التساعي ١ إلى قسمين صفر ، ١ ، والتساعي ٩ إلى قسمين ٩ ، ١ بحيث تقع نسبة ١٪ من الدرجات في كل من القسمين صفر ، ١٠ تحت المنحنى الاعتدالي .

ونلاحظ مما سبق أن معايير الاختبارات والمقاييس التي تعتمد على الدرجات المحولة تختلف فيما بينها في القيم الاختيارية لكل من متوسط درجات الجماعة المرجعية والانحراف المعياري . ويمكن تلخيصها في الجدول (٦ - ١) التالي :

خصائص الجماعة المرجعية	الدرجات المحولة
سن = صفر ، ع = ١	الدرجة المعيارية
سن = ٥٠ ، ع = ١٠	الدرجة الثانية
سن = ٥٠٠ ، ع = ١٠٠	معيار اختبار الاستعداد الأكاديمي
سن = ١٠٠ ، ع = ٢٠	معيار اختبار التصنيف العام
سن = ١٠٠ ، ع = ١٥ أو ١٦	نسبة الذكاء الانحرافية
سن = ٥ ، ع = ٢	التساعيات المعيارية

جدول (٦ - ١) ملخص معايير الاختبارات التي تعتمد على

الدرجات المحولة

ومما هو جدير بالذكر أن الدرجات المحولة المستخدمة في الاختبارات والمقاييس التربوية والنفسية لا تقتصر فقط على ما هو موضح بجدول (٦ - ١)، وإنما يمكن التوصل إلى درجات محولة أخرى باختيار قيم مناسبة لكل من المتوسط والانحراف المعياري للدرجات الخام للجماعة المرجعية، وإجراء التحويل باستخدام الصيغة (٣ - ٧) المذكورة في الفصل الثالث.

ثانياً : معايير المئينيات والرتب المئينية :

تعد المئينيات والرتب المئينية نوعاً آخر من معايير الاختبارات والمقاييس التي تستخدم في أغراض انتقاء وتسكين الأفراد، ولا تعتمد على التحويلات الخطية التي وضحتها فيما سبق. فالمئينيات Percentiles هي نقط على توزيع الدرجات تقع دونها نسبة مئوية معينة من هذه الدرجات، أما الرتب المئينية Percentile Ranks فإنها تعين الموقع النسبي للفرد أو النسبة المئوية من درجات أقرانه التي تقل عن درجته. فالمائيني ٥٥ مثلاً يعني النقطة التي تقع دونها ٥٥٪ من الدرجات في التوزيع، وهذا لا يعني أن الفرد الذي درجته الخام في الاختبار تقابل المائيني ٥٥ أنه حصل على ٥٥ درجة، إذ ربما يكون قد حصل على ٧٠ درجة، ولكنها تعنى أن ٥٥٪ من الأفراد

المختبرين حصلوا على درجات تقل عن ٧٠ درجة، لذلك فإن الدرجة الخام ٧٠ تقابل المئينة الذي رتبته ٥٥ ، أي أن ٥٥ هي الرتبة المئينة.

وقد سبق أن أوضحنا بالعديد من الأمثلة في الفصل الثالث كيفية حساب المئينات والرتب المئينة ومزاياها وعيوبها.

ونظراً لسهولة حساب وفهم هذا النوع من المعايير فإنه شائع الاستخدام في كثير من الاختبارات التربوية والنفسية المنشورة. وتزودنا أدلة هذه الاختبارات بجدول معايير مئينة يمكن الرجوع إليها لتفسير الدرجات الخام التي يحصل عليها الأفراد في الاختبار. وأحياناً تحدد معايير مئينة مستقلة للمهن، والفرق الدراسية، والأعمار المختلفة، وكذلك معايير منفصلة لكل من الذكور والإإناث.

ويمكن استخدام هذه الجداول في إجراء مقارنات داخل الفرد وبين الأفراد، فمثلاً هناك جدول معايير المئينات لاختبار الاستعدادات الفارقة Differential Aptitude Test (DAT) الذي يشتمل على ثمانية اختبارات تقيس الاستدلال اللغطي، والقدرة العددية، والاستدلال المجرد، والاستعداد الكتابي، والاستدلال الميكانيكي، والعلاقات المكانية، والهجاء، والنحو.

ويوضح جدول (٦ - ٢) التالى الرتب المئينة المتاظرة لدرجات أحد الطلاب فى كل من هذه الاختبارات الفرعية:

الرتبة المئينة	الدرجة الخام	
٦٥	٣٣	الاستدلال اللغطي
٧٠	٢٧	القدرة العددية
٤٥	٣٤	الاستدلال المجرد
٤٠	٤١	الاستعداد الكتابي
٢	٤٠	الاستدلال الميكانيكي
٥٥	٣٥	العلاقات المكانية
٥٥	٦٢	الهجاء
٦٥	٢٩	النحو

جدول (٦ - ٢) يوضح الرتب المئينة الماظرة لدى أحد الطلاب في احسانات الاستعدادات .. المعرفة

وبالنظر إلى جدول (٦ - ٢) يمكن مقارنة الرتب المئينية للطالب في الاختبارات الفرعية، كما يمكن مقارنة درجاته في كل اختبار فرعى بدرجات الجماعة المرجعية. ونلاحظ أن رتبته المئينية المناظرة لدرجة الخام فى القدرة العددية ٧٠، أى أنه يتفوق على ٧٠٪ من أقرانه في هذه القدرة، كما أنها أعلى الرتب المئينية المناظرة لدرجاته في الاختبارات الفرعية المختلفة.

وهنا ربما يتساءل مستخدم الاختبار عن كيفية الحكم على رتبة مئينية معينة ولتكن ٨٥ مثلاً، أى هل هذه رتبة مئينية مرتفعة؟ والإجابة عن ذلك تعتمد على الغرض من استخدام الاختبار وطبيعة الجماعة المرجعية، فإذا كان الغرض انتقاء أفضل الأفراد للقيام بمهام معينة أو للالتحاق بدراسات تتطلب كفايات مميزة، فإن الرتبة المئينية ٨٥ ربما تبدو غير مقبولة.

أما إذا كان الغرض تحديد مكانة الفرد بين أقرانه في صف دراسي معين ، وكانت خلفية الجماعة المرجعية وخبراتها مشابهة لخبرات الفرد المختبر، فإن هذه الرتبة المئينية ربما تبدو جيدة. فالرتبة المئينية المقابلة لدرجة خام معينة لا تشير إلى ما يمتلكه الفرد من معارف أو مهارات أو خبرات تتعلق بمحض الاختبار، وإنما تحدد فقط مكانته النسبية بين أقرانه. فالفرد الذي رتبته المئينية ٥٠ في اختبار يقيس المهارات الحسابية لا ينبغي تفسيرها على أنه ضعيف التحصيل، وإنما تعنى أن درجته في الاختبار تقل عن ٥٠٪ من درجات أقرانه الذين طبق عليهم الاختبار نفسه.

وكما في معظم معايير الاختبارات، لا يجب عند تفسير الرتب المئينية اعتبارها قيمًا مطلقة أو غایة في الدقة؛ نظرًا لأن الأفراد الذين يحصلون على درجات تقابل رتبتين مئينيتين معينتين مثل ٤٢ ، ٤٤ تكون درجاتهم متقاربة.

لذلك يفضل ناشرو الاختبارات تدوين المئينيات في أدلة الاختبارات بحيث تتضمن مدى معين Percentile Bands بدلاً من اعتبارها نقطاً محددة وذلك مراعاة لأخطاء القياس وجعل تفسيرها أكثر تحفظاً. فبدلاً من القول بأن الدرجة الخام ٦٢ في اختبار ما تناظر المئيني ٥٧ يمكن اعتبارها مناظرة للمدى المئيني ٥٦ - ٥٨ مثلاً.

ولعل من ميزات المئينيات والرتب المئينية أنه يمكن الاستناد إليها في تفسير جميع الدرجات المحولة التي أوضحتها فيما سبق والمبينة في شكل (٢ - ٦) . فالدرجة الثانية ٣٧ مثلاً تقابل الرتبة المئينية ١١ ، والتاسعى المعيارى ٤ يقابل الرتب المئينية المحسوبة بين ٢٣ ، ٣٩ وهكذا.

ثالثاً : معايير الارتقاء :

يطُرَد نمو الفرد في كثير من السمات الإنسانية بمرور الزمن بطريقة منتظمة، فتحصيل الطالب يرتقي بانتقاله من فرقة دراسية إلى أخرى، وقدراته العقلية تنمو، ومهارات الأفراد تحسن من عام إلى آخر، وهكذا. فالترندة الارتقائية تميز مختلف جوانب سلوك الإنسان، لذلك اهتم علماء القياس ببناء معايير ارتقائية يمكن تفسير درجات الاختبارات استناداً إليها بحيث يمكن مقارنة أداء الفرد بمتوسط أداء المستويات النمائية المختلفة. ولعل أكثر هذه المعايير استخداماً وبخاصة في الاختبارات التحصيلية بعض اختبارات القدرات معايير العمر الزمني Age Norms، ومعايير الفرق الدراسية Grade Norms؛ وذلك لأنَّه يسهل على مستخدمي الاختبارات فهمها وتفسيرها.

معايير العمر الزمني:

تسمى هذه المعايير أحياناً العمر التعليمي Educational Age، أو مكافئ العمر الزمني Age Equivalent، وتفيد هذه المعايير إذا توقعنا أن نمو الفرد يحدث نتيجة لعوامل النضج.

إذا كان محمد تلميذًا في الصف الرابع الابتدائي وطوله ١٣٨ سم فربما نعتبره أكثر طولاً بالنسبة لتلاميذ هذا الصف، غير أنه ربما يكون من الأفضل مقارنته بالأطفال الذين في نفس مستوى العمر، بدلاً من مقارنته بتلاميذ فرقته. فإذا كان عمره الزمني ٩ سنوات ، ١١ شهراً ، ووسيط الطول لأطفال هذا العمر (الجامعة المرجعية) ١٤٠ سم فإنه يعد أقصر من وسيط طول أقرانه. لذلك إذا كان لدينا معايير العمر الزمني أو معايير مكافئ العمر للأطوال ، فإنه يمكننا تحديد المجموعة العمرية المناظرة لطول محمد، وبالتالي يمكن تقييم طوله.

فلكي نحدد مكافئ العمر لطول محمد، ينبغي أن نحدد جماعة مرجعية وسيط طولها ١٣٨ سم . فإذا افترضنا أن عمر أفراد هذه المجموعة ٩ سنوات ، ٥ شهور ، فإن مكافئ عمر محمد حينئذ يكون ٩ سنوات ، ٥ شهور ، وتكتب (٩ - ٥).

وقد استخدمنا الشرطة الأفقية للالفصل بين الأعوام والشهور بدلاً من وضع فاصلة، وذلك لأنَّ معايير مكافئ العمر تعتمد على نظام ١٢ شهراً، أما معايير العمر الزمني التي تعتمد على نظام ١٠ شهور فتستخدم فاصلة بين الأعوام والشهور ، أي (٩ ، ٥) مثلاً.

وللحصول على معيار مكافئ العمر في خاصة معينة من خصائص الفرد، نطوي الاختبار المتعلق بهذه الخاصة على مجموعات من الأفراد ممثلة لمستويات عمرية مختلفة، ونوجد قيمة وسيط توزيع الدرجات في كل مستوى عمرى على حدة. ويمكن عندها تفسير الدرجات الخام على أنها تمثل مستوى أداء مناظر لوسيط أداء عمر زمني معين.

وتفسر الدرجات الخام في بعض اختبارات القدرات العقلية على أنها تمثل العمر العقلى Mental Age ، وفي اختبارات القراءة مثلا تمثل العمر القرائي Reading Age وهكذا.

إذا كان متوسط درجات جميع التلاميذ الذين يبلغون من العمر عشرة أعوام في أحد الاختبارات ٤٢ ، فإن أي تلميذ يحصل على الدرجة ٤٢ في هذا الاختبار يكون مكافئه العمر ١٠ أعوام ، بغض النظر عن عمره الفعلى. ويفترض بعض مؤلفى الاختبارات أن درجات الاختبار تزيد زيادة متتظمة من مستوى عمرى إلى آخر؛ لذلك يُعدون معايير للمستويات العمرية التي تتحصر بين الأعوام المختلفة ، أي ١٠ أعوام ، ٤ أشهر مثلاً. وهنا يمكن إضافة ثلث الفرق بين متوسط درجات التلاميذ الذين تبلغ أعمارهم ١٠ أعوام ، ١١ عاماً إلى الدرجة التي حصل عليها التلميذ في الاختبار، ويكون مكافئها العمر ١٠ أعوام ، ٤ شهور.

ومعايير العمر الزمني يمكن تفسيرها بسهولة إذا ارتبط ما يقيسه الاختبار بالعمر الزمني مثل القدرة العقلية، والقدرة على القراءة ، وكثير من المهارات الأساسية التي تهدف المدرسة الابتدائية لتنميتها لدى التلاميذ.

غير أن معايير مكافئ العمر يشوبها بعض العيوب التي ينبغي مراعاتها عند تفسير درجات الاختبارات التي تستند إليها.

ومن بين هذه العيوب صعوبة تفسير الدرجات المرتفعة أو المنخفضة بدرجة كبيرة ؛ نظراً لأن هذه المعايير تقتصر على المستويات المتوسطة للدرجات الخام. فإذا ما أردنا تفسير الدرجات التي تخرج عن هذا النطاق ينبغي إجراء عملية استكمال Exter- polation للخط النمائي البياني اعتماداً على فرضيات ربما يصعب تبريرها. فإذا كانت المهارات التي تقيسها الاختبارات تنمو نمواً متظهماً بزيادة العمر ، فإن عملية الاستكمال البياني ربما تكون ممكناً، أما إذا تباين معدل النمو من عام إلى آخر ، فعندها يصعب إجراء ذلك. فمعدل نمو فهم المقرؤه مثلاً لدى الطلاب الذين يبلغون من العمر ١٧ أو ١٨ عاماً أقل منه لدى التلاميذ الذين يبلغون من العمر ٧ أو ٨ أعوام.

لذلك فإن تفسير فرق درجات عمر عام واحد يختلف باختلاف المستويات العمرية . ويفضل استخدام معايير مكافئ العمر في المرحلة الابتدائية ، حيث إن فرضية تساوى وحدات النمو يمكن أن تتحقق في هذه المرحلة بدرجة أفضل من غيرها من المراحل.

كما أن هذا النوع من المعايير يعتمد على السياسات التعليمية المتعلقة بالحد الأدنى لأعمار التلاميذ الذين يلتحقون بالمدارس ، وأساليب انتقالهم إلى صفوف أعلى أو تسرعهم .

إذا كانت الجماعة المرجعية المستخدمة لبناء معايير عمر ١٠ أعوام اشتملت على تلاميذ انتقلوا بالأساليب المعتادة إلى الصنوف الأعلى ، فإن هذه المعايير سوف تختلف عن المعايير التي تعتمد على تلاميذ انتقلوا بأساليب التسرع؛ لذلك ينبغي تحديد نسبة كل مجموعة من هؤلاء التلاميذ في مستوى عمرى معين تحديداً دقيقاً لكي تفسر معايير مكافئ العمر الزمني تفسيراً معقولاً.

معايير الفرق الدراسية :

يطلق على هذا النوع من المعايير «معايير مكافئ الفرق الدراسية Grade Equivalent» ويشبه إلى حد كبير معايير مكافئ العمر . وهذه المعايير تناسب الاختبارات التحصيلية التي يستخدمها المعلمون في الصنوف الدراسية المختلفة وبخاصة في مرحلة التعليم الابتدائي . فالدرجة التي يحصل عليها الطالب في الاختبار يتم مقارنتها بوسط درجات الجماعة المرجعية في فرقه دراسية معينة . ونحصل على هذه المعايير بنفس طريقة معايير مكافئ العمر الزمني ، غير أنها هنا نطبق الاختبار على جماعة مرجعية ممثلة لمختلف الفرق الدراسية وإيجاد وسط درجات كل فرقه على حدة . ونستخدم قيمة وسط كل من هذه الفروق في تفسير الدرجات الخام ، أي نحدد الفرقه الدراسية المكافئة لدرجة خام معينة في الاختبار .

ويوضح جدول (٦ - ٣) معايير مكافئ العمر الزمني ومعايير الفرق الدراسية لاختبار افتراضي في فهم معانى الكلمات لتلاميذ الصنوف الأولى للمرحلة الابتدائية .

معانى الكلمات

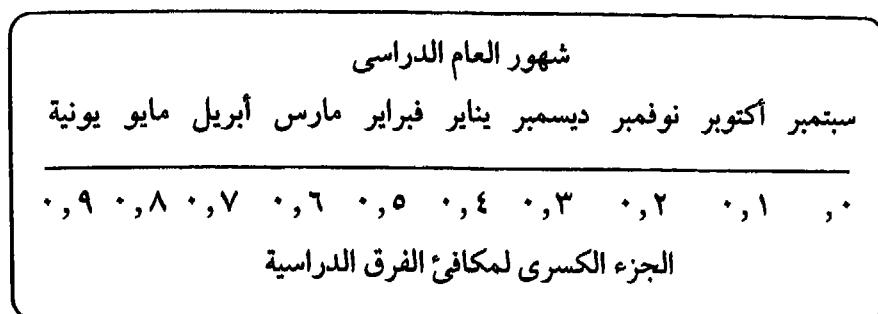
مكافيء الفرقة الدراسية	مكافيء العمر الزمني	الدرجة
١,٩	١ - ٧	١
٢,٠	٢ - ٧	٢
٢,١	٣ - ٧	٣
٢,١	٣ - ٧	٤
٢,٢	٥ - ٧	٥
٢,٢	٥ - ٧	٦
٢,٣	٦ - ٧	٧
٢,٤	٧ - ٧	٨
٢,٥	٨ - ٧	٩
٢,٦	١٠ - ٧	١٠
٢,٧	١١ - ٧	١١
٢,٨	٨ - صفر	١٢
٢,٨	٨ - صفر	١٣
٢,٩	٢ - ٨	١٤
٣,٠	٣ - ٨	١٥

جدول (٦ - ٣) يوضح معايير مكافيء العمر الزمني، ومكافيء

الفرق الدراسية لاختبار افتراضي في فهم معانى الكلمات

ويتضح من جدول (٣-٦) أن الدرجة الخام (١) تتناظر مكافيء العمر الزمني (١-٧) أي سبعة أعوام وشهر واحد، كما تتناظر مكافيء الفرقة الدراسية (١,٩) أي الصف الأول والشهر التاسع من العام الدراسي، وهكذا في بقية الدرجات.

وتعتمد هذه المعايير اعتماداً أساسياً على أساليب انتقاء المدارس والتلاميذ والسياسات التعليمية التي سبق أن أشرنا إليها، ويتم تدوين هذه المعايير على أساس عشرة أشهر وليس ١٢ شهراً، كما في معايير مكافئ العمر، أى تراوح مثلاً بين (٥ صفر) ، (٩ - ٥) لفترات شهرية للعام الدراسي . ويفضل وضع فاصلة بين الفترات ، فمثلاً الدرجة ٦،٧ تعنى الصف السادس ، والشهر السابع من العام الدراسي . والشكل (٦ - ٣) التالي يوضح الأجزاء الكسرية لمكافئ الفرق الدراسية :



شكل (٦ - ٣) يوضح الأجزاء الكسرية لمكافئ الفرق الدراسية

وكثيراً ما يخطئ البعض في تفسير مكافئ الفرق الدراسية، حيث يظنون أن هذا المعيار يشير إلى الفرق الدراسية التي ينبغي أن يلتحق بها التلميذ. فإذا كان مكافئ الفرق الدراسية للتلميذ في الصف الخامس الابتدائي ٣،٧ فهذا لا يعني أنه يجب نقل هذا التلميذ إلى الفرقة السابعة، أو أنه ينبغي إعطاؤه واجبات مدرسية مناسبة للفرقة السابعة، ولكنها تعنى فقط أن أداء هذا التلميذ يفوق متوسط أداء تلاميذ الصف الخامس.

ومن الأخطاء الأخرى الشائعة في تفسير درجات هذا المعيار مقارنته مكانة التلميذ في أحد الاختبارات بمكانته في اختبار آخر، فالانحراف المعياري يختلف اختلافاً ملحوظاً من اختبار إلى آخر، وهذا يجعل المقارنة غير ممكنة كما أوضحتنا في الفصل الثالث.

وتتميز معايير مكافئ الفرق بنفس ميزات معايير مكافئ العمر ويشوبها نفس عيوبها، فاستخدام معايير مكافئ الفرق الدراسية يكون مناسباً فقط إذا كانت الخاصة التي يقيسها الاختبار تباعيناً متزامناً من فرقه دراسية إلى أخرى، وافتراض تكافؤ الخبرات المدرسية، لذلك يفضل استخدامها في المرحلة الابتدائية.

والخلاصة أن معايير الارتفاع تقييد في تفسير درجات الاختبارات بوحدات يمكن فهمها ببساطة سواء كانت مكافئ العمر أو مكافئ الفرق الدراسية، ويمكن باستخدامها مقارنة أداء الفرد مباشرة بأداء أقرانه، ومقارنة أداؤه في اختبارات مختلفة لتعرف مقدار تقدمه أو نموه بمرور الزمن. ولكنها لا تناسب الشخصيات التي لا تتغير بانتظام بمرور الزمن أو من فرقة دراسية إلى أخرى، وتتأثر بتكوين الجماعة المرجعية وخليقياتها وخبراتها. كما أنه ينبغي إجراء عملية استكمال للخط البياني من عمر إلى آخر أو من فرقة دراسية إلى أخرى فيما يتعلق بالشهور، مما يتطلب افتراضات معينة ويقلل من دقة الدرجات وبخاصة أن وحدات هذه المعايير تختلف من عمر إلى آخر أو من فرقة دراسية إلى أخرى.

لذلك يفضل استخدام الرتب المئوية داخل كل فرقة دراسية بدلاً من وسيط درجات الفرق، فنقول مثلاً أن التلميذ (١) ربما تكون الرتبة المئوية المقابلة لدرجته في الاختبار ٩٢ بالنسبة لتلاميذ الصف الخامس الابتدائي.

رابعاً : الصفة النفسية أو التربوية (البروفيل النفسي أو التربوي) :

تشتمل بعض الاختبارات على عدة اختبارات فرعية تقيس سمات أو قدرات أو مهارات متعددة تشكل جميعها بطاريات من الاختبارات Tests Batteries وتسخدم في أغراض محددة مثل الانتقاء أو التسكين أو التشخيص. ومن أمثلة ذلك بطارية اختبارات الاستعداد الأكاديمي (SAT) ، واختبار ويكسنر لذكاء الراشدين (WAIS)، واختبار ويكسنر لذكاء الأطفال (WISC) ، واختبار الشخصية متعدد الأوجه (MMPI) ، واختبار الميول المهنية (SVIB) ، واختبار الاستعدادات الفارقة (DAT) ، وبطارية اختبارات الاستعداد العام (GATB) ، وبطارية اختبارات المهارات الشامل (CTBS) وبطارية اختبارات السجل الأكاديمي (GRE) وغيرها من بطاريات الاختبارات التي سنوضحها في الفصول التالية، فإذا طبقت أي من هذه البطاريات على أحد الأفراد، فإنه يحصل على عدة درجات وفقاً لعدد الاختبارات التي تشتمل عليها البطارية.

ويكون الغرض من استخدام هذه البطاريات ليس فقط معرفة أدائه النسبي في كل اختبار على حدة، وإنما أيضاً مقارنة أدائه في الاختبارات المختلفة.

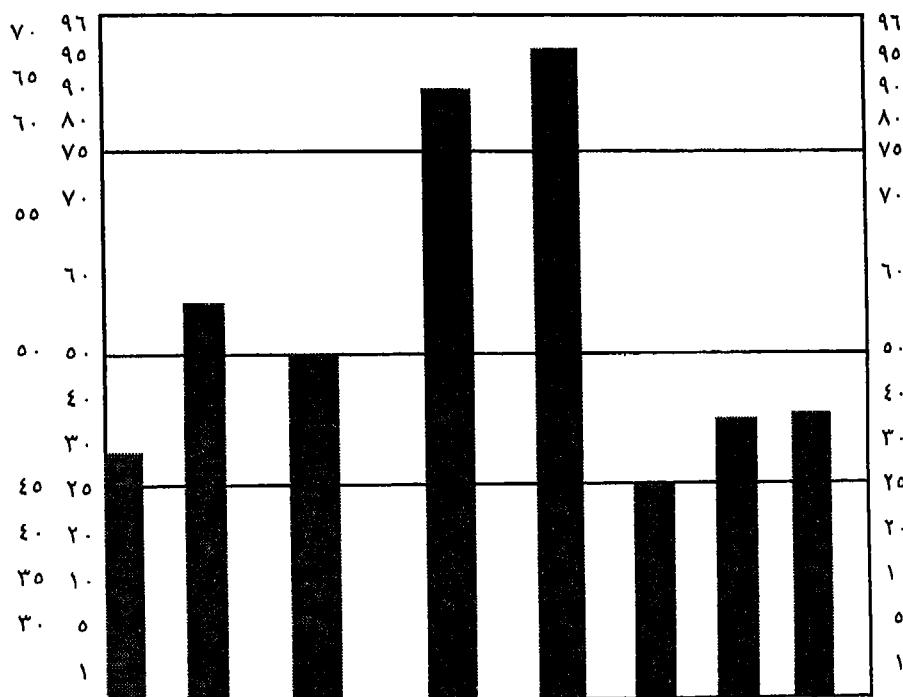
ولتوسيح الصورة العامة للأداء وتيسير عملية المقارنة يزود ناشرو الاختبارات مستخدمو هذه البطاريات ببطاقة أو ورقة رسم بياني معدة بطريقة خاصة تقييد في التمثيل البياني لدرجات الاختبارات التي تشتمل عليها بطارية معينة؛ وذلك لتوضيح موقع الفرد

أو مستوى أدائه في هذه الاختبارات التي تمثل أبعاداً نفسية أو معارف ومهارات تربوية معينة ، وتحديد جوانب القوة والضعف . ويسمى هذا التمثيل البياني النفسي أو التربوي الصفحة النفسية (البروفيل النفسي) Psychological Profile ، أو الرسم البياني التربوي (البروفيل التربوي) Educational Profile .

ويمكن إعداد الصفحة النفسية لمجموعة من درجات الاختبارات لفرد معين، حيث تمثل الاختبارات على المحور الأفقي، ويمثل مقدار كل درجة نقطة على ارتفاع مناسب مواز للمحور الرأسى . ولکى نلاحظ النقط بسهولة يمكن توصيل هذه النقط بخطوط مستقيمة مما يعطى شكل البروفيل، غير أن هذه الخطوط لا تعنى شيئاً حيث إن ما يهمنا هو النقط المناظرة لدرجات الاختبارات؛ لذلك ربما يفضل عدم توصيل النقط فيما بينها.

ويوضح شكل (٦ - ٤) التالي صفحة نفسية لدرجات أحد الطلاب في بطارية اختبارات الاستعدادات الفارقة مستمدة من دليل هذه البطارية الذى أصدرته المؤسسة السينكولوجية الأمريكية عام ١٩٦٣ .

العنوان: دليل استخدام اللغة العربية الاستعداد المكانى الاستعداد الميكانيكى الاستعداد المكانى الاستدلال المجرد القدرة العددية للنقطى المثبتات السداسية



شكل (٦ - ٤) يوضح صفحة نفسية لدرجات طالب فى اختبار الاستعدادات الفارقة

ويتضح من شكل (٦ - ٣) أن الصفحة النفسية تحدد المئينيات والدرجات المعيارية المقابلة للدرجات الخام للطالب في الاختبارات الفرعية الشمانية التي تشتمل عليها بطارية اختبارات الاستعدادات الفارقة، وتمثل الأعمدة الرأسية هذه الدرجات. والمئينيات والدرجات المعيارية المقابلة لدرجات جميع اختبارات البطارия مستمدة من جماعة مرجعية واحدة محددة الخصائص.

وهناك ثلاثة خطوط أفقية تمتد من النقط المقابلة لكل من المئيني ٢٥ ، والمئيني ٥ ، والمئيني ٧٥ ، أي الإرباعيات الأول والثاني والثالث. وذلك ليتسنى مقارنة درجات الطالب فيما بينها بمجرد النظر ، وكذلك مقارنتها بدرجات الطلاب الآخرين. وتقييد هذه الصفحة النفسية في تحديد جوانب القوة والضعف في استعدادات الطالب ، والاستناد إليها في التوجيه والإرشاد التربوي .

فإذا نظرنا إلى الصفحة النفسية للطالب الموضحة بشكل (٦ - ٤) نلاحظ أنه حصل على أعلى الدرجات في الاستعداد الميكانيكي حيث تقابل الرتبة المئينية ٩٥ تليها درجته في الاستعداد المكاني المقابلة للرتبة المئينية ٨٧ تقريباً ، وحصل على أقل الدرجات في الاستعداد الكتابي ، حيث تقابلها الرتبة المئينية ٢٥ أي الإرباعي الأول ، وبذلك تبين جوانب قوته وضعفه .

وتتميز الصفحات النفسية أو الرسوم البيانية التربوية بخصائص ثلاث هي: المستوى Level ، والانتشار Scatter ، والشكل Shape . «فالمستوى» يعني الارتفاع العام للأعمدة في الرسم البياني وهو يدل على مستوى الأداء العام للفرد أو المتوسط العام لدرجاته المعيارية في الاختبارات الفرعية ، و«الانتشار» يعني درجة تشتت مختلف النقط التي تمثل الدرجات حول الموقع المتوسط (المستوى) لدرجات جميع الاختبارات الفرعية ، ويدل ذلك على مدى اختلاف درجات الفرد فيما بينها . أما «الشكل» فيعني نمط الاستعدادات المتمثل في ترتيب درجات الاختبارات التي حصل عليها الفرد .

وهذه الخصائص يمكن أن تختلف إلى حد ما فيما بينها كل على حدة ، إذ يمكن أن تتفق صفحاتان نفسيتان لفردين في «المستوى» و«الانتشار» ، ومع ذلك تختلف كل منهما عن الأخرى في «الشكل» ، أي في ارتفاع وانخفاض النقط التي تمثل درجاتهم في اختبارات البطاربة .

ولا تقتصر الصفحات النفسية أو الرسم البياني التربوي Profiles على نوع واحد ، وإنما تتعدد أنواعها وطرق تمثيلها . فالنوع الذي قدمناه يمثل درجات بطاربة اختبارات تطبق على فرد واحد في وقت واحد لإعطاء صورة إستاتيكية مستعرضة

لدرجاته . بيد أنه يوجد نوع آخر يسمى الصفحات النفسية الارتقائية أو الطولية Longitudinal Profiles حيث تمثل درجات الاختبار نفسه أو عدة اختبارات في أزمنة مختلفة لتوضيح نمو الفرد في سمة أو سمات متعددة . فمثلاً يمكن تمثيل درجات تلميذ في اختبارات القراءة والحساب والمعلومات العامة من الصف الثالث إلى الصف السادس في رسم بياني تربوي Educational Profile للإفاده منها في تعرف التقدم الدراسي للتلميذ وتوجيهه توجيهًا تربويًا سليمًا .

وعادة يوضح مؤلف بطارية الاختبارات أمثلة لصفحات نفسية لبعض الحالات وكيفية استخدامها في الإرشاد والتوجيه التربوي أو النفسي لهذه الحالات للاسترشاد بها في استخلاص المعلومات من صفحة نفسية أخرى ، ويستخدم الحاسوب في طباعة وتنسيق هذه الصفحات .

غير أنه يجب مراعاة أن الفرق بين درجتين في اختبارين مختلفين يتأثر بأخطاء القياس المتعلقة بكل منهما ، ولذلك فإن ثبات هذا الفرق يكون أقل من ثبات درجات كل اختبار على حدة ، كما سبق أن أوضحنا في الفصل الرابع . ونظراً لأن الصفحات النفسية أو الرسوم البيانية التربوية تعكس الفروق بين درجات اختبارات فرعية مختلفة ، فإنه ينبغي الحبيطة قبل أن نعزى هذه الفروق إلى الاختلافات بين درجات الاختبارات الفرعية التي تمثلها الصفحة النفسية أو الرسم البياني التربوي وبخاصة إذا كان عدد مفردات هذه الاختبارات قليل مما يؤدى إلى انخفاض ثبات درجاتها .

كما يجب مراعاة بعض العوامل الأخرى لكي لا تقدم الصفحة النفسية صورة غير صادقة عن سمات الفرد . إذ ينبغي أن تُستخدم الجماعة المرجعية نفسها في إعداد معايير الاختبارات الفرعية ، أما إذا استخدمت جماعات مرجعية مختلفة ، فإننا لا نستطيع المقارنة المباشرة بين هذه الدرجات .

كما ينبغي أن تمثل درجات مجموعة الاختبارات في الرسم البياني بميزان أو تدريج واحد ، وأن يعبر عن الدرجات بالرتب المئينة أو أي نوع من أنواع الدرجات المحولة المقابلة لها . كما يفضل استخدام معيار لا يضم الفرق الضئيلة أو يصغر الفروق الكبيرة . ويمكن التغلب على ذلك إلى حد ما بأن تتضمن الصفحات النفسية مؤشرًا يبين مقدار أخطاء القياس لكي لا يساء تفسير الدرجات .

وقد سبق أن أوضحنا أنه يمكن التغلب على ذلك أيضًا بأن تمثل الدرجات بفترات ، أي تمحض في مدى معين Bands بدلاً من تمثيلها بنقط أو قيم مضبوطة ، وتكون الفروق بين الدرجات دالة إذا لم يحدث تداخل بين الفترات .

ومن هنا يتضح أن المعايير Norms تُعد أساساً لتفسير درجات الأفراد في الاختبارات والمقاييس مرجعية الجماعة أو المعيار، فتحويل الدرجة الخام التي يحصل عليها الفرد إلى نوع من أنواع المعايير التي أوضحتها تمكنا من تفسير هذه الدرجة في ضوء المركز النسبي للفرد بين أقرانه المماثلين له فيما يقيسه الاختبار.

وكذلك تمكنا المعايير من المقارنة بين أداء الفرد في مجالات متعددة يتم قياسها بواسطة بطارية من الاختبارات حيث يمكن تمثيل درجاتها برسم بياني تربوي أو بصفحة نفسية تداريجهها موحدة، وتنفيذ الأنواع المختلفة من المعايير التي أوضحتها في أغراض متعددة، ويمكن أن يتضمن الاختبار أكثر من نوع منها. فقد لاحظنا أن المئينيات يمكن أن تستخدم مع الدرجات المحولة، أو مع مكافئ العمر، أو مكافئ الفرقه الدراسية في اختبار واحد.

غير أنه في جميع الحالات ينبغي مراعاة خصائص الجماعة المرجعية التي تُعتمد منها المعايير، ومدى تشابهها مع خصائص الأفراد الذين ستُخذل قرارات بشأنهم في ضوء هذه المعايير، وأساليب انتقاء أفراد هذه الجماعة، وتاريخ تطبيق الاختبار عليهم، ومقدار خطأ القياس المتعلقة بأساليب المعايير، وغير ذلك من الاحتياطات الواجب مراعاتها عند تفسير معايير الاختبارات والمقاييس.

فهذه المعايير ليست مطلقة أو مستقرة بل هي معايير نسبية تعتمد اعتماداً أساسياً على جماعات مرجعية معينة، وتأثر هذه المعايير تأثيراً كبيراً بتغير خصائص هذه الجماعات التي يقارن بها الفرد في سمة معينة أو مجموعة من السمات التي يقيسها الاختبار.

ونظراً للخصائص النسبية للمعايير وعدم استقرارها مما يحد من عمومية استخدامها في تفسير درجات الاختبارات والمقاييس، فقد حاول علماء القياس التوصل إلى أسلوب جديد يهدف لجعل خصائص أدوات القياس مستقلة عن خصائص جماعات مرجعية معينة في إطار نظرية معاصرة يطلق عليها نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT) ، وسوف نوضح أهم المفاهيم الأساسية لهذه النظرية وبعض تطبيقاتها التربوية والنفسية في الباب الرابع.

مذكّرات الأداء في الاختبارات مرجعية المحكّم :

ذكرنا في مستهل هذا الفصل أنه يمكن تفسير درجات الاختبارات تفسيراً يمكننا من استخلاص معلومات مفيدة تتعلق بالأفراد بطرق متعددة. فإذا حولنا الدرجات الخام إلى نوع آخر من الدرجات، مثل المئينيات أو التساعيات المعيارية أو الدرجات التالية، فإنه يمكننا مقارنة أداء الفرد بأداء جماعة مرجعية معيارية Normative Group . وهذا يرتبط بالقياس مرجعى الجماعة أو المعيار.

ولكن عندما نستخدم الاختبارات مرجعية المحك Criterion - Referenced Tests ، فإننا لا نهتم بالمركز النسبي للفرد بين أقرانه، وإنما نهتم بمقارنته درجته في الاختبار بمستوى أداء **Performance Standard** يكون بمثابة محك Criterion يدل على المستوى المقبول لسلوك الفرد أو أدائه . فإذا كانت درجة الفرد تعادل أو تفوق مستوى الأداء المحدد فإن الفرد يكون قد حقق المستوى المطلوب ، أما إذا قلت درجته عن المستوى المحدد فإنه أداءه يعد ضعيفاً .

إذا لم تكن هناك قيود على عدد الأفراد الذين يمكن أن يحققوا مستوى الأداء المطلوب ، فإن الاختبارات مرجعية المحك تكون مناسبة في الكشف عن هؤلاء الأفراد . ولكن إذا اقتصرنا على انتقاء عدد محدود من الأفراد ، وليكن مثلاً نسبة ١٠٪ العلية منهم لتعيينهم في إحدى الوظائف أو قبولهم في إحدى الكليات فإن الاختبارات مرجعية الجماعة أو المعيار تكون أكثر ملاءمة في هذه الحالة حيث يتم انتقاءهم وفقاً لمعايير هذه الاختبارات التي أوضحتها فيما سبق .

فالإطار التفسيري لل اختبارات مرجعية المحك لا يعتمد إذن على جماعة مرجعية من الأفراد ، وإنما يعتمد على مستوى أو محك أداء في ضوء نطاق سلوكي Content Domain معين معرف تعرضاً جيداً ، يقيسه الاختبار . فمثلاً تحديد أداء الفرد فيما يتعلق بدرجة تمكّنه من حل تمارين هندسية معينة ، أو قراءة قطع نثرية محددة وهكذا .

فالتركيز في هذه الحالة ينصب على ما يستطيع الفرد أداؤه ، وما لا يستطيع أداؤه وليس على مقارنته بأقرانه دون تحديد لمعارفه أو مهاراته أو ما يجب أن يعرفه أو يؤديه . ويمكن تحديد محكمات أو مستويات أداء أو درجات قطع للمهن والوظائف المختلفة للتحقق من كفايات العاملين بعض النظر عن مقارنة أداء أحدهم بأداء الآخر . كما يمكن تحديد محكمات لمستويات كفاية الطلاب في أوقات مختلفة أثناء عملية التعليم للحصول على معلومات تتعلق بأدائهم ومدى تقدمهم الدراسي .

ففي مثل هذه الحالات يكون هناك تناقض بين أداء الفرد ومحك الإنجاز المطلوب ، وبذلك يمكن ترجمة درجة الفرد في الاختبار مباشرة إلى سلوك محدد يمكن ملاحظته وقياسه . أي أن الدرجة الخام في الاختبار يكون لها معنى في ذاتها دون إسنادها إلى جماعة مرجعية من الأفراد . فإذا قلنا مثلاً أن الدرجة الخام ٢٣ التي حصل عليها طالب في اختبار يشمل على ٢٥ مفردة اختيار من متعدد تقيس معرفة الرموز الكيميائية ، فإن درجته تدل مباشرة على مدى معرفته بهذه الرموز . وإذا اخترنا الحد

الأدنى لمستوى الأداء المقبول ٩٠٪ ، فإنه يمكن تقسيم أداء الطالب في ضوء هذا المستوى المحدد. ونظرًا لأن الطالب أجاب عن نسبة ٩٢٪ من المفردات إيجابة صحيحة ، فإن أدائه يكون جيداً ، دون أن نحدد ما إذا كان أداؤه مرتفعاً أو منخفضاً بالنسبة لأقرانه الذين طبق عليهم الاختبار نفسه.

فمحركات الأداء في الاختبارات مرجعية المحك تعتمد على النسب المئوية Percentages من المفردات التي تقيس سلوكاً معروفاً تعربياً دقيقاً والتي يجبر عنها الفرد إجابة صحيحة . بينما المعايير المئوية Percentiles للختارات مرجعية الجماعة أو المعيار تعتمد على النسب المئوية من الأفراد الذين تقل درجاتهم عن درجة معينة.

فالأولى تحدد ما يمتلكه الفرد من معارف ومهارات تتعلق بالأداء الجيد في مهنة أو دراسة معينة ، بينما تحدد الثانية ترتيب الفرد بين أقرانه فيما يقيسه الاختبار . فعلى الرغم من أن الطالب في المثال السابق أجاب عن ٩٢٪ من مفردات الاختبار إجابة صحيحة مما يدل على أداء جيد ، إلا أن درجاته ربما تقابل المئوي ٧٥ مثلاً ، مما يعني أنه يتضيق على ٧٥٪ فقط من أقرانه ، وهذا يدل على ارتفاع مستوى الجماعة المرجعية التي استمدت منها المعايير المئوية للختبار. أي أنه يمكننا تفسير درجات الاختبار والمقاييس باستخدام النظامين المرجعيين المطلق Absolute والنسبة Relative معًا ، أي النظام مرجعى المحك والنظام مرجعى الجماعة، للحصول على معلومات متکاملة تتعلق بمستوى كفاية الفرد في مجال معين ، وكذلك مكانته النسبية بين أقرانه في هذا المجال .

لذلك فإن كلا من هذين النظامين يقدم معلومات مناسبة تفييد في تقويم الأفراد والبرامج . و اختيار أيٌ من النظامين ينبغي أن يعتمد على الغرض من عملية التقويم واستخدامات الاختبار . وسوف نوضح بمزيد من التفصيل طرق تحديد مستويات الأداء في الاختبارات مرجعية المحك في الفصل الثامن .

خلاصة

لكى يتمكن مستخدم الاختبار أو المقياس التربوى والنفسى من تفسير درجات الأفراد ينبغى أن يستند هذا التفسير إلى نظام مرجعى مناسب . فالاختبارات مرجعية الجماعة أو المعيار يستند تفسير درجاتها إلى معايير جماعة مرجعية ، بينما يستند تفسير درجات الاختبارات مرجعية المحك إلى مسحكات أداء متوقع . والمعايير عبارة عن مجموعة من الدرجات المحولة أو المشتقة من الدرجات الخام بطرق إحصائية معينة . وتوجد أنواع متعددة من معايير الاختبارات ، مثل الدرجات المعيارية ، والدرجات التائية ، والمتينيات ، والرتب المئنية ، والتساعيات المعيارية ، ونسبة الذكاء الانحرافية ، ومعايير مكافئ العمر الزمنى ، ومعايير مكافئ الفرق الدراسية ، والصفحات النفسية والتربوية .

وتختلف هذه المعايير فى طريقة تفسيرها للمركز النسبى للفرد فيما يقىسه الاختبار ، وعلى خصائص وتكوين الجماعة المرجعية التى تستمد منها المعايير . أما مسحكات الأداء المتوقع فتدل على مستوى أداء مقبول لسلوك الفرد أو أدائه يمكن الحكم فى ضوئه على تحقق كفايات معينة لديه بغض النظر عن مقارنة أدائه بأداء أقرانه ، أى أن تفسير الدرجات فى هذه الحالة يستند إلى نظام مرجعى مطلق وليس إلى نظام مرجعى نسبى كما فى الحالة الأولى . ومع هذا يمكن إحداث تكامل بين النظائرتين فى تفسير درجات بعض الاختبارات .

الفصل السابع

تحليل مفردات الاختبارات

* مقدمة

- * أهمية تحليل مفردات الاختبار
- * صعوبة مفردات الاختبار
- * تصحيح قيمة معامل الصعوبة من أثر التخمين
- * أوجه قصور معامل الصعوبة وكيفية تلافيها
- * تمييز مفردات الاختبار
- * أساليب تقييم «تمييز» المفردات استناداً إلى محك داخلي
- * أسلوب يعتمد على معامل الارتباط
- * معامل تمييز المفردات
- * استخدام معامل الارتباط الثنائي المتسلسل الحقيقي
- * استخدام درجات المجموعتين الطرفيتين
- * تقييم «تمييز» المفردات استناداً إلى محك خارجي
- * تفسير معامل تمييز المفردات وعلاقته بمعامل الصعوبة
- * تحليل فاعلية بدائل الإجابات
- * استخدام تحليل المفردات في انتقاء مفردات الاختبار
- * استخدام الحاسوب في تحليل مفردات الاختبارات

مقدمة :

ناقشنا في الفصل الخامس الخصائص السيكومترية للاختبارات والمقاييس التربوية والنفسية، ومن أهمها الثبات والصدق، كما ناقشنا في الفصل السادس طرق تفسير درجات هذه الاختبارات والمقاييس باستخدام أنواع مختلفة من المعايير والمحكمات. ولكن نظراً لأن الاختبار أو المقياس يشتمل على مجموعة من المفردات أو الفقرات أو الأسئلة، فإن خصائص الاختبار تعتمد أيضاً على خصائص مفرداته. ولذلك نبني اختباراً يتميز بالفاعلية ينبغي فحص المفردات التي يشتمل عليها فحصاً دقيقاً باستخدام الأساليب الوصفية الكيفية، والأساليب الإحصائية الكمية. وقد سبق أن قدمنا بعض الأساليب الكيفية في تقدير صدق محتوى الاختبار، وسوف نتناول في هذا الفصل الأساليب الكمية في تحليل مفردات الاختبار Item Analysis ، وبخاصة الخصائص الإحصائية لكل مفردة من مفرداته.

وتتعلق هذه الخصائص بمستوى صعوبة المفردة Difficulty Level ، ودرجة تميزها Discrimination بين المستويات العليا والمستويات الدنيا في السمة أو القدرة التي يقيسها الاختبار في ضوء محك معين داخلي أو خارجي. فالمحك الداخلي ربما يتعلق بالأداء في مجموعة المفردات بعامة أو الاتساق الداخلي للمفردات ، والمحك الخارجي يتعلق بالأداء خارج نطاق الاختبار نفسه مثل النجاح في دراسة معينة ، أو في أداء عمل معين . والاهتمام في كلتا الحالتين ينصب على تقدير صدق كل مفردة من مفردات الاختبار .

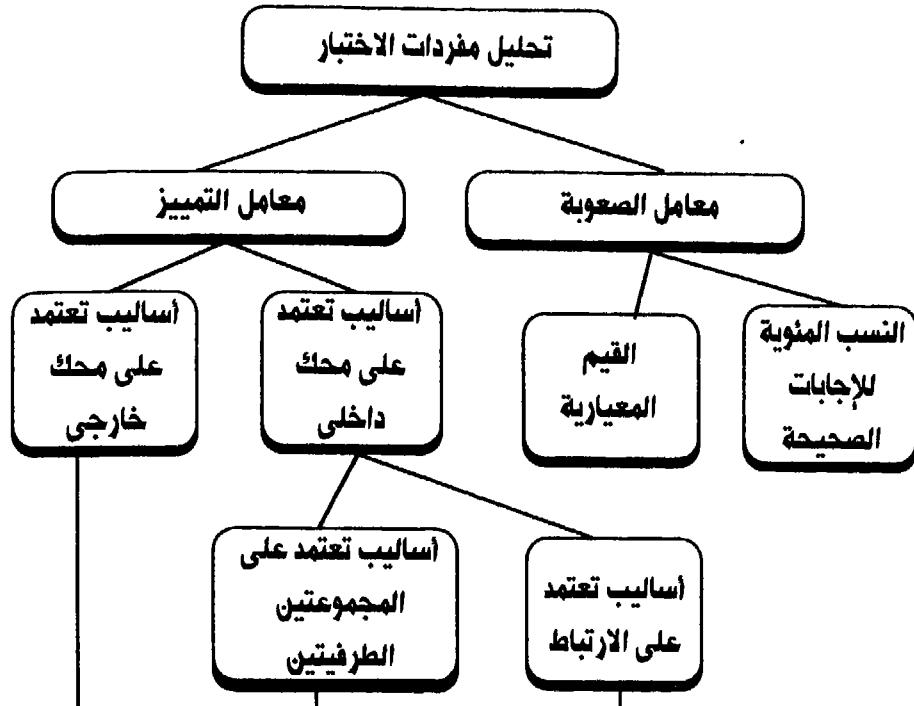
وسوف نناقش هاتين الخاصيتين استناداً إلى أسس ومبادئ النظرية الكلاسيكية للاختبارات مرجعية الجماعة أو المعيار ، ونرجئ مناقشتهما في ضوء نظرية الاستجابة للمفردات الاختبارية (IRT) إلى الفصل الرابع عشر . أما خصائص مفردات الاختبارات مرجعية المحك فسوف نوضحها في الفصل الثامن .

الأهمية تحليل مفردات الاختبار:

يهم القائمون بناء الاختبارات والمقاييس بكتابه أو انتقاء مفردات عالية الجودة لقياس السمات الإنسانية قياساً دقيقاً؛ لذلك يراعون كثيراً من الشروط في تكوين هذه المفردات وصياغتها والتحقق بالأساليب المنطقية وأحكام الخبراء من صدق محتوى كل مفردة على حدة وكذلك صدق محتوى الاختبار ككل. غير أنه مهمما بلغت دقة هذه الأساليب والأحكام فإنها لا تغني عن التجريب الميداني للاختبار وتحليل درجات مفرداته باستخدام الأساليب الإحصائية وتحديد العلاقة بين ما تقيسه المفردات وبين استجابات الأفراد لها، وذلك للتعرف على المفردات الغامضة أو المربكة أو التي تشجع على التخمين، والمفردات باللغة السهلة أو الصعوبة، حيث إنها لا تكشف عن الفروق الفردية في السمة التي يقيسها الاختبار مرجعى الجماعة أو المعيار وبذلك لا تميز بين مستويات هذه السمة، كما أنها لا تسهم بأى قدر في صدق الاختبار أو ثبات درجاته.

ويفيد ذلك في المراجعة الفنية للمفردات وتحسينها بحيث تسهم كل مفردة إسهاماً إيجابياً فيما يقيسها الاختبار. ويساعد تحليل المفردات أيضاً في تنمية مهارات القائمين بإعداد الاختبارات والمقاييس نتيجة الفحص المستنير للمفردات، وتعرف جوانب الضعف التي ربما تجعل بعض هذه المفردات غير صالحة، واستبقاء المفردات التي تفي بخصائص تححق الغرض من استخدام الاختبار أو المقاييس في مجالات التقويم المختلفة، وإعادة استخدامها إذا تطلب الأمر ذلك.

ولعل من المناسب قبل مناقشة أساليب تحليل مفردات الاختبار أن نقدم الشكل التخطيطي (١-٧) التالي الذي يوضح كل من خصائص الصعوبة والتميز وأساليب تقدير قيمة كل منها.



شكل (٧ - ١) يوضح أساليب تحليل مفردات الاختبار

(من تصميم المؤلف)

صعوبة مفردات الاختبار:

تعد صعوبة مفردات الاختبار من الخصائص التي تلعب دوراً مهماً في الاختبارات المرجعية الجماعية أو المعيار Norm - Referenced Tests وتأثر في إجابات الأفراد عن مفرداتها . فالفردات التي تشتمل عليها هذه الاختبارات ينبغي أن تميز تسييماً دقيقاً بين مستويات السمة المراد قياسها ، فالمرة التي يجب عنها جميع الأفراد ، أو التي لا يستطيع أحدهم الإجابة عنها لا تفيد في الكشف عن الفروق بينهم فيما يقيسه الاختبار . وقد بينت الدراسات السيكومترية والإمبريقية أن الاختبار يمكن أن يميز إلى أقصى حد ممكناً بين الأفراد المختبرين إذا كان متوسط مستوى صعوبة المفردات التي يشتمل عليها . ٥٠ . تقريراً ، أي يستطيع أن يجيب ٥٠٪ منهم عن كل مفردة من مفرداته . غير أن التوزيع الأمثل لمستويات صعوبة المفردات في هذا النوع من الاختبارات يعتمد أيضاً على العلاقات القائمة بين مفردات الاختبار ، فكلما زاد الارتباط بين المفردات تطلب الاختبار مدى متسع نسبياً لصعوبة مفرداته حول مستوى الصعوبة المتوسط . ٥٠ ، أما إذا قل الارتباط بين كل مفردة والأخرى فإننا متوقع أن يكون مدى صعوبة المفردات أقل اتساعاً ، لذلك يفضل انتقاء مفردات الاختبار بحيث يكون مدى صعوبتها معتدلاً ، ولكن متوسط صعوبة جميع المفردات يكون . ٥٠ ، تقريراً .

ويمكن الحصول على قيم تشير إلى مستوى صعوبة مفردات الاختبار بإيجاد نسبة عدد الأفراد في جماعة مرجعية محددة الذين يجيبون عن كل مفردة إجابة صحيحة أو يستطيعون حل مشكلة معينة. وكلما زادت هذه النسبة دل ذلك على سهولة المفردة، وكلما قلت دل ذلك على صعوبة المفردة.

ويطلق على هذا المؤشر معامل صعوبة المفردة (Item Difficulty Index).

$$\text{معامل صعوبة المفردة} = \frac{\text{عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة}}{\text{العدد الكلى لأفراد الجماعة المرجعية}}$$

$$\text{أى أن } M = \frac{C}{N} \quad \dots \dots \dots (1 - 7)$$

فإذا طبق اختبار يشتمل على ٥٠ مفردة اختيار من متعدد على مجموعة معينة من الأفراد ، وأجاب ٣٠ فرد إجابة صحيحة عن أحد مفردات الاختبار ، فإن معامل

$$\text{صعوبة هذه المفردة} = \frac{3}{5} = 0,60$$

وينبغي ملاحظة أن أقل قيمة لهذا المعامل (صفر) عندما تكون المفردة غاية في الصعوبة ولم يجب عنها أحد إجابة صحيحة، وأكبر قيمة (١,٠) عندما تكون المفردة غاية في السهولة، حيث يجب عنها جميع الأفراد. ولذلك فإن القيمة ٠,٦٠ تعني أن المفردة متوسطة الصعوبة إلى حد ما بالنسبة لجماعة مرجعية معينة.

وينبغي التأكيد هنا على أن خاصية «صعوبة المفردة» في إطار النظرية الكلاسيكية للاختبارات ليست مقصورة على المفردة؛ لأنها تعكس أيضاً قدرة أفراد الجماعة المرجعية الذين أجابوا عن تلك المفردة.

لذلك يفضل عند تفسير قيمة معامل الصعوبة تحديد الجماعة المرجعية التي استندنا إليها في تحديد هذه القيمة، فنقول مثلاً أنه إذا اخترنا مجموعة ممثلة للتلاميذ نهاية مرحلة التعليم الأساسي بمحافظة معينة في وقت ما بهذه المفردة ، فإن قيمة معامل صعوبتها ٠,٦٠ .

تصحيح قيمة معامل المعرفة من أثر التخمين :

عند استخدام الصيغة (١ - ٧) في إيجاد معامل معرفة مفردة الاختيار من متعدد، أو الصواب أو الخطأ التي تكون إجابتها إما صحيحة أو خطأ ينبغي مراعاة أثر التخمين Guessing .

فمن المعلوم أن بعض الأفراد يتوصلون إلى الإجابة الصحيحة عن طريق التخمين وحده أى بمحض الصدفة . فاحتمالية التخمين الصحيح للإجابة عن مفردة الاختيار من متعدد التي تشتمل على خمسة بدائل ٢٥ ، ، ، ، وأربعة بدائل ٥٠ ، ، بينما احتمالية التخمين الصحيح إذا كانت المفردة من نوع الصواب أو الخطأ ٢٠ ، ، ، نظراً لأنها تتضمن بديلين فقط مما يؤثر في قيمة معامل معرفة المفردة .

لذلك يبرز التساؤل التالي : هل يمكن اعتبار الدرجة الكلية التي يحصل عليها فرد في الاختبار مساوية لعدد إجاباته الصحيحة عن مفردات الاختبار؟ أم أنه ينبغي تصحيحها من أثر التخمين العشوائي لاستبعاد ذلك الجزء من الدرجة الكلية الذي يفترض حصول الفرد عليه بالصدفة، أى دون أن يعرف بالفعل الإجابة الصحيحة عن بعض المفردات؟ والحقيقة أنه لا يوجد اتفاق بين علماء القياس حول إجابة هذا التساؤل .

فالبعض يرى أنه ليس من الضروري إجراء هذا التصحيح، وحجتهم في ذلك ما يلى :

(١) لا يتغير ترتيب الأفراد فيما بينهم إذا أجرينا تصحيحاً لدرجاتهم الكلية في الاختبار من أثر التخمين .

(٢) يؤدي تصحيح درجات الأفراد الذين لا يلجأون إلى التخمين إلى خفض درجاتهم الكلية في الاختبار، بينما لا يحدث ذلك لمن اعتمد على التخمين في إجابته .

(٣) يعتمد تصحيح الدرجات من أثر التخمين على فرضية أن الإجابة الصحيحة إما تكون معلومة أو غير معلومة، فـى حين أن بعض الأفراد يلجأون إلى التخمين استناداً إلى معلومات جزئية لديهم .

(٤) تصحيح الدرجات من أثر التخمين ربما يؤدي إلى الافتراض الخاطئ بأن الدرجات الناتجة هي الدرجات الحقيقة للأفراد في الاختبار .

لذلك يقترح أصحاب هذا التوجه بعض الإرشادات التي يمكن أن تسهم في تقليل أثر التخمين بقدر كبير كالتالى :

(١) تحذير المختبرين من التخمين العشوائى فى إجاباتهم عن مفردات الاختبار.

(٢) السماح بوقت كاف لكل فرد لكي يتمكن من محاولة إجابة كل مفردة من مفردات الاختبار.

(٣) تضمين البادئات التى تشتمل عليها مفردات الاختيار من متعدد معلومات صحيحة جزئياً أو خطأ شائع تتعلق بالمجال الذى تقىسه المفردات.

أما البعض الآخر فيرى ضرورة تصحيح قيمة صعوبة مفردات الاختبار من أثر التخمين العشوائى؛ وذلك للحصول على تقدير نسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن كل مفردة استناداً إلى معلوماتهم الفعلية واستبعاد من يجيبون إجابة صحيحة عن طريق التخمين العشوائى.

ويستند هذا التقدير إلى افتراض أن اختيار الفرد إجابة غير صحيحة لمفردة اختبارية يرجع إلى افتقاره للمعلومات التى تتطلبها الإجابة الصحيحة ، وليس إلى خطأ معلوماته أو سوء فهمه، وكذلك افتراض أن الفرد الذى لا يعرف الإجابة الصحيحة يرى أن جميع البادئات التى تشتمل عليها المفردة يمكن أن تمثل الإجابة الصحيحة.

فإذا قبلنا هذين الافتراضين ، فإننا نتوقع توصل نسبة معينة من الأفراد الذين لا يعرفون الإجابة الصحيحة إلى هذه الإجابة بمحض الصدفة ، وهذه النسبة هي متوسط نسب عدد الأفراد الذين يختارون أحد البادئات الأخرى أو الإجابات الخطأ.

ويمكن الحصول على تقدير نسبة عدد الذين يجيبون إجابة صحيحة عن المفردة استناداً إلى معلوماتهم الفعلية بطرح هذه النسبة من النسبة الكلية لعدد الأفراد الذين يجيبون عن المفردة إجابة صحيحة.

فلتصحيح درجة الفرد من أثر التخمين (سـ)، نستخدم الصيغة (٢-٧) التالية:

$$سـ = ح - \frac{خ}{ك - ١} \quad (٢-٧).....$$

حيث (ح) ترمز إلى عدد الإجابات الصحيحة التي توصل إليها الفرد.

، (خ) ترمز إلى عدد إجاباته الخطأ.

، (ك) ترمز إلى عدد بادئات مفردة الاختيار من متعدد، أو مفردة الصواب أو الخطأ.

فإذا اشتمل اختبار على ٢٠ مفردة اختيار من متعدد لكل منها ٥ بدائل ، وأجاب فرد على ١٦ مفردة إجابة صحيحة، فإنه يمكن تصحيح هذه الدرجة من أثر التخمين باستخدام الصيغة (٧ - ٢) كالتالي:

$$\text{س} = \frac{16 - 4}{1 - 5} = 16 - 1 = 15$$

وإذا اشتمل اختبار على ٥٠ مفردة صواب أو خطأ، وأجاب فرد عن ٣٨ مفردة إجابة صحيحة، فإن درجته في الاختبار بعد تصحيحها من أثر التخمين تصبح كالتالي:

$$\text{س} = \frac{38 - 12}{1 - 2} = 12 - 38 = -26$$

ويذلك يمكن تقدير نسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن مفردة الاختبار استناداً إلى معلوماتهم الفعلية (م) باستخدام الصيغة (٧ - ٣) التالية:

$$م = \frac{\text{ح} - \text{خ}}{\text{ك} - 1} \quad \dots \dots \dots (7 - 3)$$

حيث (ح) ترمز إلى عدد أفراد مجموعة المختبرين الذين أجابوا عن المفردة إجابة صحيحة.

، (خ) ترمز إلى عدد أفراد مجموعة المختبرين الذين أجابوا عن المفردة إجابة خطأ.

، (ك) ترمز إلى عدد بدائل المفردة.

، (ن) ترمز إلى العدد الكلى لأفراد مجموعة المختبرين.

، (ن') ترمز إلى عدد أفراد مجموعة المختبرين الذين تركوا المفردة دون إجابة.

ويلاحظ أن هذه الصيغة أخذت بعين الاعتبار عدد الأفراد الذين لم يحاولوا الإجابة عن المفردة ، أي تركوها دون إجابة.

وإذا أردنا الحصول على قيمة معامل صعوبة مفردة دون إجراء تصحيح للدرجات من أثر التخمين ، فإننا نحذف النسبة $\frac{\text{ح}}{\text{ن}}$ من البسط، وتصبح (م) كالتالي:

$$م = \frac{\text{ح}}{\text{ن} - \text{n}} \quad \dots \dots \dots (7 - 4)$$

فإذا لم يترك جميع الأفراد المختبرين أى مفردة دون إجابة ، فإن $\text{n} =$ صفر، وتصبح (م) كالتالي:

$$م = \frac{\text{ح}}{\text{n}} \quad \dots \dots \dots (7 - 5)$$

وهي نفس الصيغة (٧ - ١) التي ذكرناها من قبل.

ولتوضيح كيفية تطبيق الصيغة (٣ - ٧) نفترض أن الاختبار الذي يشتمل على ٢٠ مفردة اختيار من متعدد ولكن منها ٥ بدائل طبق على ١٠٠ فرد حيث أجاب ٧٥ فرد منهم عن المفردة إجابة صحيحة، بينما ترك ١٠ أفراد المفردة دون إجابة، فعندئذ تكون صعوبة المفردة (م) كالتالي:

$$\frac{٣,٧٥ - ٧٥}{٩٠} = \frac{\frac{١٥}{١ - ٥} - ٧٥}{١٠ - ١٠٠} = \frac{٧٥}{٧١,٢٥} = ٠,٧٩$$

أما إذا تغاضينا عن تصحيح الدرجات من أثر التخمين، واستخدمنا الصيغة (٧-٥)، فإن صعوبة المفردة (م) تكون كالتالي:

$$M = \frac{٧٥}{١٠ - ٧٥} = ٠,٧٥$$

ويلاحظ أن المفردة تبدو أقل صعوبة عندما استخدمنا صيغة التصحيح من أثر التخمين واستبعدنا الأفراد الذين لم يحاولوا الإجابة عن المفردة، وذلك لأننا اعتمدنا على نسبة عدد الأفراد الذين أجروا عن المفردة استناداً إلى معلوماتهم الفعلية. أما إذا تغاضينا عن الأفراد الذين تركوا المفردة دون إجابة ، فإن قيمة معامل الصعوبة بعد تصحيحه من أثر التخمين تصبح:

$$0,71 = \frac{٧١,٢٥}{١٠٠} = \frac{٣,٧٥ - ٧٥}{١٠٠}$$

وهذا يعني أن حوالي ٧١٪ من المختبرين أجروا إجابة صحيحة عن المفردة استناداً إلى معلوماتهم الفعلية ، في حين أن ٧٥٪ من المختبرين أجروا إجابة صحيحة عن المفردة سواء باستخدام معلوماتهم الفعلية أو عن طريق التخمين العشوائي .

غير أنه ينبغي مراعاة أنه في الاختبارات الموقعة، أي التي تعتمد على السرعة في الإجابة ربما يترك بعض الأفراد إجابة بعض المفردات نظراً لضيق الوقت، ففي هذه الحالة لا تستطيع تحديد ما إذا كانوا قد تركوا الإجابة عن هذه المفردات بسبب ضيق الوقت أو لافتقارهم إلى المعلومات المطلوبة. فلذلك تميز بينهما يمكن افتراض أن جميع المفردات بما في ذلك آخر مفردة أجاب عنها الفرد قد حاول إجابتها استناداً إلى معلوماته ، أما بقية المفردات التي تركها دون إجابة فإننا نفترض أنه لم يحاول الإجابة عنها.

كما ينبغي عند استخدام الصيغة (٣ - ٧) مراعاة الافتراضين اللذين ذكرناهما من قبل، على الرغم من أننا قد لا نستطيع عملياً معرفة أيهما هو الذي أثر بالفعل في إجابات الأفراد بحيث تعبر القيمة التقديرية لمعامل الصعوبة بعد تصحيحها من أثر التخمين عن نسبة عدد أفراد مجموعة المختبرين الذين توصلوا إلى الإجابة الصحيحة استناداً إلى معلوماتهم الفعلية.

وعموماً، فإن هذه القيمة التقديرية تعد قيمة تقريرية محكومة بالافتراضين اللذين استندت إليهما، ومع هذا فإنه من المتوقع أن تكون أكثر دقة من القيمة التي نحصل عليها دون إجراء هذا التصحيح.

أوجه قصور معامل الصعوبة وكيفية تلافيها :

إن استخدام إحدى الصيغ السابقة في تقدير قيمة معامل صعوبة المفردات يتطلب مراعاة بعض الاعتبارات الأخرى نوجزها فيما يلى:

(١) ينبغي مراعاة أن قيم معامل صعوبة المفردات الاختبارية سواء أجرينا تصحيحها من أثر التخمين أو لم نجر ذلك تكون من المستوى الرتبى، فهذه القيم تدل على رتب الصعوبة النسبية للمفردات. فمثلاً إذا كانت قيم معامل صعوبة أربع مفردات ٤٥ ، ٤٠ ، ٣٥ ، ٣٠ على الترتيب ، فإن هذا يعني أن المفردة الأولى أسهل المفردات الأربع، يليها المفردة الثانية، ثم الثالثة، وأصعبها المفردة الرابعة.

ولكننا لا نستطيع استنتاج أن الفرق بين القيمتين ٤٥ ، ٤٠ ، ٣٥ يساوى الفرق بين القيمتين ٣٥ ، ٣٠ ، أي أن هذه الفروق المتساوية بين قيم معامل صعوبة المفردات لا تناهض فروقاً متساوية في مستوى صعوبة هذه المفردات، وهذا يعني أن وحدات قيم معامل الصعوبة غير متساوية.

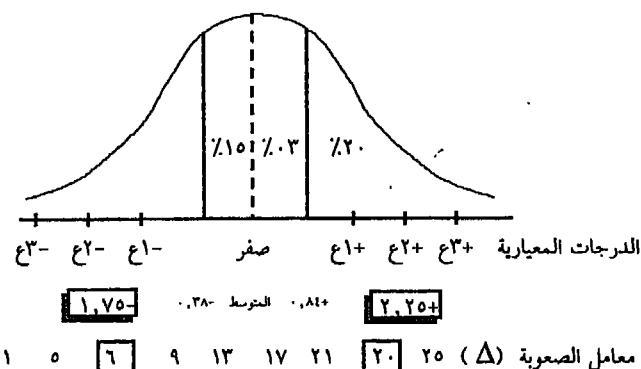
وهذا يعد أحد أوجه قصور معامل الصعوبة في إطار النظرية الكلاسيكية للإختبارات . أما وجه القصور الآخر فهو اعتماد قيم معامل الصعوبة على خصائص الجماعة المرجعية التي استند إليها في التوصل إلى هذه القيم مما يحد من إمكانية تعميم تفسيرها .

(٢) يرى ثورنديك (Thorndike, 1985) أنه يمكن التغلب على وجه القصور الأول بافتراض أن توزيع درجات السمة التي يقيسها الاختبار اعتدالى ، وبذلك يمكن تحويل النسب المئوية من الأفراد الذين يجيبون عن مفردة اختبارية إجابة صحيحة

(أى القيم الرتبية لمعامل الصعوبة) إلى الميزان المتدرج ذى الفترات المتساوية الذى يمثل المحور الأفقي للمنحنى الاعتدالى المعيارى (أى ميزان الدرجات المعيارية).

فقد سبق أن أوضحنا فى الفصلين الثالث والرابع أنه يمكن تحديد نسب مئوية معينة من المساحة تحت المنحنى الاعتدالى تنحصر بين المتوسط ووحدات الانحراف المعيارى التى يمثلها المحور الأفقي.

فمثلاً المساحة تحت المنحنى الاعتدالى التى تنحصر بين المتوسط وانحراف معيارى واحد موجب أو سالب تساوى ١٤,٣٤٪ وهكذا . فإذا كانت قيمة معامل صعوبة مفردة ٥٠، فإن هذه القيمة تناظر الدرجة المعيارية صفر على المحور الأفقي، وقيمة المعامل ٦٥، مثلاً تناظر الدرجة المعيارية -٣٨، وقيمة المعامل ٢٠، تناظر الدرجة المعيارية +٨٤، . ويمكن توضيح ذلك بشكل (٧ - ٢) التالى:



شكل (٧ - ٢) يوضح الميزان القرى لقيم معامل الصعوبة

فقيمة معامل الصعوبة ٦٥، . يعني أن ٦٥٪ من أفراد الجماعة المعيارية أجابوا إجابة صحيحة عن المفردة، وتتضمن هذه النسبة ٥٪ العليا مضافةً إليها ١٥٪ من الحالات التى تقل عن المتوسط كما هو موضح بالشكل.

لذلك فإن هذه القيمة تناظر -٣٨، . وحدة انحراف معيارى على التدرج الأفقي. وبالمثل إذا كانت قيمة معامل الصعوبة ١٦، . فإن هذا يعني أن ٢٠٪ أجابوا إجابة صحيحة عن المفردة، أي ٥٪ مطروحاً منها ٣٪ من الحالات التى تزيد عن المتوسط فنجد أنها تناظر +٨٤، . وحدة انحراف معيارى على التدرج الأفقي.

ويلاحظ أن القيم المرتفعة لمعامل الصعوبة تناظر درجات معيارية موجبة، أما القيم المنخفضة التى تدل على سهولة المفردات فإنها تناظر درجات معيارية سالبة.

كما يمكن إجراء بعض التحويلات الخطية المماثلة لتحويل الدرجات المعيارية إلى درجات تائية للتخلص من الإشارات السالبة. غير أن مركز الخدمات الاختبارية بالولايات المتحدة الأمريكية (ETS) قد توصل إلى نوع من التحويل الخطى يتحقق ذلك، وأطلق على معامل الصعوبة الناتج الرمز (Δ) ، وقرأ (دلتا) كالتالى:

$$\Delta = \frac{12}{4 + e^{-0.7 \times D}}$$

فالمفردة التى قيمة معامل صعوبتها صفر وحدة انحراف معياري أى تقع عند المتوسط تكون قيمة (Δ) المناظرة لها ١٣ . والمفردة التى قيمة معامل صعوبتها . ٣٨ تكون قيمة (Δ) المناظرة لها $13 + \frac{4}{e^{0.7 \times 38}} = 11.5$ ، أى ١٢ تقريرياً. وتتضمن الصيغة (٧ - ٦) ثابتين هما ١٣ ، ٤ ، أى أن قيمة متوسط الصعوبة (Δ) يساوى ١٣ ، وانحرافها المعياري = ٤ . وهاتان القيمتان مناسبتان للتخلص من الإشارات السالبة، فمن الناحية العملية تكون أعلى قيمة لمعامل الصعوبة (Δ) مساوية ٢ وتناظر الدرجة المعيارية ٢,٢٥+ ، وهذه تدل على مفردة صعبة؛ لأن نسبة من يجب عنها إجابة صحيحة حوالي ١٢٪ ، وأقل قيمة لها تساوى ٦ حيث تناظر الدرجة المعيارية - ١,٧٥ ، وهذه تدل على مفردة سهلة؛ لأن نسبة من يجب عنها إجابة صحيحة حوالي ٩٦٪.

(٣) للتغلب على وجه القصور الآخر المتعلق بارتباط قيم معامل الصعوبة بخصائص الجماعة المرجعية حاولت النظرية السيكومترية المعاصرة التى أصبح يطلق عليها نظرية الاستجابة للمفردة الاختبارية (IRT) معالجة هذا القصور فى ضوء أسس ومبادئ جديدة .

والخلاصة أنه يمكن إيجاد قيم معامل صعوبة مفردات الاختيار من متعدد، أو الصواب أو الخطأ باستخدام إحدى الصيغ التى تعتمد على نسبة عدد الأفراد الذين يجيبون عن المفردة إجابة صحيحة، وميزان هذه النسب يكون رتبياً مثل المئينيات، أو يمكن إيجاد هذه القيم على ميزان فترى باستخدام الدرجات المعيارية أو الدرجة المعيارية المعدلة (Δ).

وفي كلتا الحالتين ينبغي مراعاة الافتراضات التى تستند إليها كل منها والتي ناقشناها فيما سبق.

تمييز مفردات الاختبار:

من الخصائص الأخرى المهمة التي ينبغي أن تتوافر في مفردات الاختبارات مرجعية الجماعة أو المعيار خاصة «التمييز Discrimination» ، وتعنى بذلك مدى إمكانية قياس الفروق الفردية بواسطة مفردات هذه الاختبارات.

فإذا اختبرنا مجموعة من الطلاب باختبار تحصيلي معنٍ في مجال دراسي معين، ووجدنا أن عدد الطلاب الضعفاء في المجال الدراسي الذي يقيسه الاختبار مساوياً عدد الطلاب الأقواء الذين أجابوا إجابة صحيحة عن إحدى مفردات الاختبار، فإن هذه المفردة لا تصلح لقياس الفروق الفردية في التحصيل في هذا المجال الدراسي، حيث إنها لم تميز مطلقاً بين مجموعتي الطلاب.

غير أن هذا الاستنتاج يتطلب توفير محك نسند إليه في تحديد الطلاب الضعفاء والطلاب الأقواء في التحصيل. فعلى الرغم من أن هناك محكمات متعددة يمكن استخدامها في هذا الشأن، مثل آراء المعلمين وتقديرات الاختصاصيين في المجالات الدراسية المختلفة، إلا أن الدرجات الكلية التي يحصل عليها الطلاب في الاختبار نفسه تستخدم عادة كمحك مناسب سواء في الاختبارات التحصيلية المقسّنة أو الاختبارات الصافية التي يكتبها المعلم بنفسه لطلابه.

أي أن المحك المستخدم في هذه الاختبارات يكون محكًا داخلياً Internal Criterion . فإذا كان الطالب الذين حصلوا على درجات مرتفعة في الاختبار قد أجابوا إجابة صحيحة عن إحدى مفردات الاختبار، والطلاب الذين حصلوا على درجات منخفضة في الاختبار قد أجابوا إجابة خطأ عن هذه المفردة، فإن المفردة تكون قد ميزت بين المجموعتين.

وريما نتخذ ذلك دليلاً على صدق المفردة Item Validity مع الأخذ في الاعتبار التتحقق من صدق محتوى المفردة قبل تضمينها في الاختبار.

غير أنه في بناء الاختبارات التي تقيس الميول والصحة النفسية وبعض اختبارات الاستعدادات العقلية والجوانب المهارية يفضل أن يكون المحك خارجي External Criterion ، فمثلاً في اختبار الميول يمكن استخدام أفراد المهن المختلفة كمجموعات محكمة لتقدير صدق مفردات اختباراته الفرعية . وكل من هذه الاختبارات المتعلقة بالميول نحو مهنة معينة يشتمل على المفردات التي تميز بين الأفراد الذين يتمون إلى هذه المهنة والناس بعامة. وقد سبق أن تناولنا في الفصل الخامس بمزيد من التفصيل أنواع المحكمات وبعض المشكلات المتعلقة بها.

وسوف نوضح فيما يلى بعض الأساليب البسيطة التى يمكن استخدامها فى تقييم مدى «تمييز» مفردات الاختبارات مرجعية الجماعة أو المعيار، وسوف نقدم بعض الأساليب الأخرى التى تناسب الاختبارات مرجعية المحك فى الفصل الثامن.

أساليب تقييم «تمييز» المفردات استناداً إلى محك داخله:

تعدد أساليب تقييم مدى تمييز مفردات الاختبارات استناداً إلى محك داخلى ، إذ يوجد ما يقرب من مائة أسلوب لتقدير علاقة درجات مفردة اختبار بالدرجة الكلية فى الاختبار. غير أن معظم هذه الأساليب تعتمد على أنواع مختلفة من معاملات الارتباط لتحديد درجة هذه العلاقة، ويرجع اختلاف أنواع معاملات الارتباط المستخدمة إلى أسلوب معالجة الدرجات الكلية فى الاختبار. بعض الأساليب يعتبر هذه الدرجات متغيراً متصلةً، وبعضها يعتبرها متغيراً ثنائياً وذلك بتقسيم الدرجة الكلية إلى قسمين متساوين، أو يعتمد على المقارنات الطرفية حيث يقسم الدرجة الكلية إلى ثلاثة أقسام ويقارن بين المجموعتين الطرفتين . Extreme Groups

وتستند هذه الأساليب إلى افتراضين أساسيين متضمنين فيها كالتالى :

- (١) تعبير الدرجة الكلية عما يقيسه الاختبار بالفعل ، وتزداد جودة الاختبار إذا اشتمل على مفردات ترتبط درجاتها ارتباطاً مرتفعاً بالدرجة الكلية .
- (٢) تعبير الدرجة الكلية عن سمة أحادية البعد ، فإذا كان الاختبار يقيس سمات متعددة ، فإنه يصعب تفسير معنى الدرجة الكلية ، ويؤدى تحليل المفردات إلى نتائج غامضة .

ولتوبيح ذلك نفترض أن الاختبار يشتمل على عدد قليل من المفردات محتواها لفظى ، وعدد كبير آخر من المفردات محتواها عددي ، فإذا أجرينا تحليلاً لمفردات هذا الاختبار لتحديد علاقة درجات كل من هذه المفردات بالدرجات الكلية فى الاختبار ، فسوف تؤدى نتائج هذا التحليل إلى استبعاد المفردات التى محتواها لفظى ؛ نظراً لضعف علاقتها بالاختبار ككل ، مما يقلل من صدق الاختبار. كذلك ينبغي فى مثل هذه الحالة إجراء تحليل المفردات لكل محتوى على حدة ، أى نوجد علاقة درجات كل مفردة محتواها عددي بالدرجات الكلية للمفردات التى محتواها عددي ، وبالمثل فى المفردات التى محتواها لفظى .

وسوف نقتصر هنا على أسلوبين فقط من أساليب تقييم مدى تمييز المفردات الاختبارية أحدهما يعتمد على الارتباط ، والآخر يعتمد على المقارنة بين مجموعتين طفيفتين نظراً لشيوخ استخدامهما وسهولة تطبيقهما وتفسير نتائجهما.

أسلوب يعتمد على معامل الارتباط :

يمكن تقييم تمييز مفردات الاختبار بإيجاد معامل الارتباط بين درجات كل مفردة والدرجات الكلية في الاختبار ، غير أن نوع معامل الارتباط المستخدم يعتمد على رؤيتنا لتوزيع الدرجات الكلية في الاختبار أي المحك.

إذا اعتبرنا توزيع هذه الدرجات متصلاً، فإنه يكون لدينا متغيران : أحدهما ثانية الدرجة (درجات كل مفردة إما واحد صحيح أو صفر) ، والآخر فترى متصل (الدرجة الكلية في الاختبار) ، وعندئذ يمكن استخدام معامل الارتباط الثنائي المتسلسل Biserial Correlation. غير أن إيجاد قيمة هذا المعامل تتطلب عمليات إحصائية تستغرق كثيراً من الوقت والجهد.

لذلك ربما نلجأ إلى تقسيم المتغير المتصل. أي الدرجات الكلية في الاختبار إلى قسمين متساوين يشتمل أحدهما على الدرجات العليا والآخر على الدرجات الدنيا ، وعندئذ يكون لدينا متغيران أحدهما ثانية الدرجة (درجات كل مفردة) ، والآخر فترى متصل (الدرجات الكلية في الاختبار) ، ولكننا قسمناه اعتبارياً إلى قسمين ، أي اعتبرناه أيضاً متغيراً ثانياً الدرجة. وفي هذه الحالة يمكننا استخدام نوع آخر من معاملات الارتباط يسمى معامل الارتباط الثنائي المتسلسل الحقيقي Point Biserial Correlation .

ولكن هذا يعتمد أيضاً على ما إذا كنا نعتبر التقسيم الثنائي للمتغير المتصل تقسيماً حقيقةً أم مصطنعاً. فإذا اعتبرنا النجاح أو الفشل في الإجابة عن مفردات الاختبار متغيراً متصلاً، يمثل السمة أو القدرة التي تنطوي عليها درجات المفردات ، فإن نقطة تقسيم الدرجات الكلية تعتمد على مستوى صعوبة المفردات ، وعندئذ يفضل استخدام معامل الارتباط الثنائي المتسلسل؛ نظراً لأن قيم هذا المعامل لا تعتمد على نسبة عدد الدرجات العليا أو الدنيا ، مما يجعل تقييم «تمييز» المفردات لا يتأثر بمستوى صعوبتها ، وبذلك نحصل على تقديرتين مستقلتين لكل من التمييز والصعوبة.

أما إذا اعتبرنا أن النجاح أو الفشل في الإجابة عن مفردات الاختبار يمثل كل منها مجموعة متمايزة، وأن جميع أفراد المجموعة الأولى متكافئون في النجاح، وجميع أفراد المجموعة الثانية متكافئون في الرسوب، ففي هذه الحالة يتأثر مدى تمييز المفردات بمستوى صعوبتها.

لذلك فإنه في حالة تساوي جميع الظروف الأخرى تكون قيمة معامل الارتباط الثنائي المتسلسل الحقيقي أقل من القيمة المناظرة إذا استخدمنا معامل الارتباط الثنائي المتسلسل.

ويرى بعض علماء القياس أن تقسيم الدرجات الكلية في الاختبار إلى قسمين يعد تقسيماً مصطنعاً وليس حقيقياً، إذ إن هذه الدرجات تمثل مدى متصلاً من السمة أو القدرة التي تقييسها مفردات الاختبار.

كما أن هذا التقسيم يؤدى إلى تأثير «تمييز» المفردات بمستوى «صعوبة» كل منها، وهو ما ينبغي تلافيه. لذلك ربما يكون معامل الارتباط الثنائي المتسلسل أكثر ملاءمة في تقييم «تمييز» مفردات الاختبار، ومع هذا فإننا سوف نوضح كيفية استخدام معامل الارتباط المتسلسل الحقيقي في تقييم «تمييز» مفردات الاختبارات نظراً لسهولة حساب قيمة وشيوع استخدامه.

معامل تمييز المفردات:

أولاً: استخدام معامل الارتباط الثنائي المتسلسل الحقيقي :

يمكن استخدام معامل الارتباط الثنائي المتسلسل الحقيقي Point Biserial كما ذكرنا في إيجاد درجة العلاقة بين درجات مفردات الاختبار، والدرجات الكلية، وذلك بتقسيم توزيع الدرجات الكلية في الاختبار إلى مجموعتين، إحداهما تمثل المجموعة العليا والأخرى تمثل المجموعة الدنيا في السمة التي يقيسها الاختبار، وبذلك تكون نقطة التقسيم في منتصف التوزيع.

والقيمة الناتجة تسمى: «معامل تمييز المفردات Discrimination Index»، وتدل هذه القيمة على مدى اتساق درجات كل مفردة مع درجات الاختبار ككل. ويمكن اشتقاء الصيغة الرياضية لمعامل الارتباط الثنائي المتسلسل الحقيقي (رج) من صيغة معامل ارتباط بيرسون (٣ - ١٠) التي قدمناها في الفصل الثالث.

وهذه الصيغة كالتالى :

$$\text{د} \text{ح} = \frac{\sqrt{\text{ص} \text{ن} ١ - \text{ص} \text{ن}}}{\text{ع} \text{ص}}$$

$$(v - v) \dots \dots \dots$$

حيث صن ترمز إلى متوسط توزيع الدرجات الكلية للمجموعة التي أجبت إجابة صحيحة عن مفردة معينة (أى س = 1).

، صن ترمز إلى متوسط توزيع الدرجات الكلية في الاختبار.

، ع ترمز إلى الانحراف المعياري للدرجات الكلية في الاختبار.

، صن ترمز إلى نسبة عدد أفراد المجموعة الأولى التي حصلت على (واحد صحيح) في المفردة.

، ص. ترمز إلى نسبة عدد أفراد المجموعة الثانية التي حصلت على (صفر) في المفردة.

والمثال التالي يوضح كيفية إيجاد قيمة معامل تميز مفردة اختبارية باستخدام هذه الصيغة :

نفترض أننا طبقنا اختباراً تحصيلياً مقتناً يشتمل على مفردات اختيار من متعدد على خمسة طلاب، وأردنا إيجاد قيمة معامل تميز إحدى هذه المفردات .

ويوضح جدول (٧ - ١) التالي درجاتهم في هذه المفردة (س)، ودرجاتهم الكلية في الاختبار (ص)، وكذلك قيم (صن)، وقيم (صن) عندما (س=1).

ص (عندما س=1)	٢ ص	ص	س	الطلاب
-	٦٤	٨	صفر	١
٦	٣٦	٦	١	٢
٩	٨١	٩	١	٣
-	٤٩	٧	صفر	٤
٥	٢٥	٥	١	٥
٢٠	٢٠٠	٣٥	٣	المجموع

جدول (٧ - ١) يوضح درجات خمسة طلاب في مفردة اختبار من متعدد

ودرجاتهم الكلية في الاختبار

الخطوة الأولى: نوجد (ص)، أي متوسط الدرجات الكلية في الاختبار للمجموعة التي أجبت إجابة صحيحة عن المفردة ($S = 1$) من العمود الخامس كالتالي:

$$\bar{S} = \frac{20}{3} = 6,7$$

والخطوة الثانية: نوجد (ص) أي متوسط الدرجات الكلية في الاختبار من العمود الثاني كالتالي:

$$\bar{S} = \frac{35}{5} = 7$$

والخطوة الثالثة: نوجد (σ_S) أي الانحراف المعياري للدرجات من العمودين الثالث والرابع باستخدام الصيغة ($\sigma_S^2 = \frac{\sum (S - \bar{S})^2}{N}$) التي سبق أن ذكرناها في الفصل الثالث وهي:

$$\sigma_S^2 = \frac{\sum (S - \bar{S})^2}{N}$$

$$\frac{245 - 200}{5} = \frac{200 - 175}{5} =$$

$$\frac{1}{2} = \frac{1}{2} =$$

والخطوة الرابعة: نوجد (ص). أي نسبة عدد الطلاب الذين حصلوا على الدرجة (صفر) في المفردة ، وذلك من العمود الثاني كالتالي:

$$\bar{S} = \frac{2}{5} = 0,4$$

والخطوة الخامسة: نوجد (ص) أي نسبة عدد الطلاب الذين حصلوا على الدرجة (1) في المفردة، وذلك من العمود الثاني أيضاً كالتالي:

$$\bar{S} = \frac{3}{5} = 0,6$$



والخطوة السادسة : نطبق الصيغة (٧ - ٧) لإيجاد قيمة معامل الارتباط الثنائي المتسلسل كالتالى:

$$\frac{\frac{\text{ص}٢ - \text{ص}١}{\text{ص}٢}}{\sqrt{\frac{\text{ص}١}{\text{ص}٢}}} = \rho_{xy}$$

$$\frac{\frac{٠,٦}{٠,٤}}{\sqrt{\frac{٠,٦}{٠,٤}}} = \frac{٧ - ٦,٧}{١,٤}$$

$$\cdot ٣ - ١,٣ \times ١,٣ = ٠,٢٨$$

ويلاحظ أن قيمة معامل الارتباط الثنائي المتسلسل لهذه المفردة سالبة، أى أن معامل تميزها سالبًا.

وهذا يعني أن نسبة عدد طلاب المجموعة الدنيا أى (الطلاب الضعاف) ولكنهم أجابوا إجابة صحيحة عن المفردة أكبر من نسبة عدد طلاب المجموعة العليا (الطلاب الأقوياء)، أى الذين حصلوا على درجات كلية مرتفعة في الاختبار ، وأجابوا عنها إجابة صحيحة، وهذا ربما يدل على أن المفردة غامضة أو مضللة وينبغي حذفها من الاختبار أو مراجعتها من الناحية الفنية مراجعة دقيقة لتعرف وجه القصور فيها. وأحياناً يحدث ذلك عن طريق الصدفة لقلة العدد الكلى للطلاب المختبرين كما في هذا المثال.

ومن الجدير بالذكر أنه يمكن الحصول على هذه القيمة نفسها باستخدام صيغة معامل ارتباط بيرسون (٣ - ١٠) التي سبق أن ذكرناها في الفصل الثالث، وذلك لأن معامل الارتباط الثنائي المتسلسل بعد حالة خاصة من معامل ارتباط بيرسون حيث يكون أحد المتغيران من النوع الثنائي .

وينبغي التنوية إلى أن هناك أساليب أخرى لإيجاد معامل التمييز تعتمد على أنواع أخرى من معاملات الارتباط مشتقة من معامل ارتباط بيرسون مثل معامل الارتباط الرباعي الحقيقي الذي يطلق عليه معامل فاي (Φ) Phi Coefficient أو تعطى قيماً تقديرية لمعامل ارتباط بيرسون مثل معامل الارتباط الرباعي Tetrachoric Correlation وتعتمد جميعها على طريقة معالجة الدرجات الكلية في الاختبار كمتغير متصل يتم

تقسيمه اعتبارياً إلى قسمين، أو يكون متغيراً ثانياً حقيقياً، وقد أوضحنا هذه الأنواع من معاملات الارتباط في مرجع آخر.

كما ينبغي التنوية أيضاً إلى أن عامل التخمين العشوائي يؤثر في معامل التمييز مثلما يؤثر في معامل الصعوبة. ويرى بعض علماء القياس أنه ينبغي تصحيح الدرجات من أثر التخمين قبل حساب قيمة معامل الارتباط الثنائي المتسلسل بين درجات المفردات والدرجات الكلية في الاختبار أو غيره من العوامل. غير أن الدراسات الإمبريقية بينت أن تأثير استخدام الدرجات بعد تصحيحها على قيمة معامل الارتباط ضئيلاً عند حساب قيمة معامل تميز المفردات.

ثانياً : استخدام درجات المجموعتين الطرفيتين :

تعتمد هذه الطريقة على تقسيم الدرجات الكلية في الاختبار إلى قسمين متباينين، وبذلك يجعل التغيير المتصل Continuous Variable متغيراً ثانياً Dichotomus Variable من أجل تبسيط العمليات الحسابية التي يتطلبها إيجاد معامل تميز المفردات.

وتتميز هذه الطريقة بأنها لا تتطلب جهداً كبيراً، وتعطى قيمة قريبة من القيم التي نحصل عليها باستخدام معامل الارتباط الثنائي المتسلسل الحقيقي.

وعادة يمثل أحد القسمين المجموعة التي نالت أعلى الدرجات في الاختبار، ويتمثل القسم الآخر المجموعة التي نالت أقل الدرجات في الاختبار نفسه. وقد توصل كيلي Kelley إلى أفضل نسبة مئوية من الأفراد ينبغي أن تشتمل عليهما كل من المجموعتين لكي يكون معامل التمييز أكثر دقة.

واستخدم في ذلك أسلوب تحليل الانحدار للدرجات الكلية في الاختبار في تقديم الدرجات الحقيقة للأفراد، وأجرى بعض التحليلات العددية للدالة الجبرية الناتجة لمعرفة أقصى قيمة تصل إليها هذه الدالة، ووجد أن النسبة المئوية التي تقابل أقصى قيمة للدالة . ٢٧٠ ٢٦٧٨

لذلك أوصى كيلي Kelley عند تحليل مفردات الاختبار الاعتماد على النسبة ٢٧٪ من الأفراد في كل من المجموعتين الطرفيتين ، واستبعاد نسبة ٤٦٪ الوسطي. ولكل تكون قيمة معامل التمييز متسقة من عينة إلى أخرى ينبغي عند إجراء هذا التقسيم إلا يقل عدد أفراد عينة تحليل مفردات الاختبارات المقسمة عن ١٠٠ في كل من المجموعتين الطرفيتين، ومن الأفضل زيادة عدد الأفراد عن ذلك.

أما في حالة الاختبارات التحصيلية الصافية التي يكتبها المعلم لطلابه، فإن عدد طلاب الفصل يكون أقل من ذلك بكثير؛ لذلك يمكن ضم عدد من الفصول وتحليل

مفردات هذه الاختبارات استناداً إلى العينة الكلية لكي تكون أكثر تمثيلاً وتؤدي إلى نتائج أكثر دقة.

وتحميّز هذه الطريقة بأنها لا تتطلّب جهداً كبيراً، وتعطى قيمًا قريبة من القيم التي نحصل عليها باستخدام معامل الارتباط الثنائي المتسلسل الحقيقي.

ويُمكن إيجاد معامل تميّز مفردة اختبارية معينة بطرح نسبة عدد أفراد المجموعة الدنيا الذين أجابوا إجابة صحيحة عن المفردة (ن) من نسبة عدد أفراد المجموعة العليا الذين أجابوا إجابة صحيحة عن هذه المفردة (ن)، أي أن :

$$\text{معامل تميّز المفردة} = \frac{n}{N} - \frac{n}{n} \dots \dots \dots \quad (1 - 8)$$

ويتطلّب تطبيق الصيغة (1 - 8) اتباع ما يلى :

(١) تطبيق الاختبار على العينة المستهدفة، وتصحيحه، وإيجاد الدرجة الكلية التي حصل عليها كل فرد في الاختبار.

(٢) ترتيب الدرجات الكلية من الأعلى إلى الأدنى للمجموعة ككل.

(٣) تقسيم الدرجات الكلية إلى قسمين بحيث يشتمل كل منهما على ٢٧٪ من عدد هذه الدرجات، ويتمثل أحد القسمين بالمجموعة العليا، ويتمثل الآخر بالمجموعة الدنيا. وإذا وقعت بعض الدرجات على الحد الفاصل بين المجموعتين، فإنه يمكن تعين هذه الدرجات بطريقة عشوائية في كل من المجموعتين بحيث يشتمل كل منهما على عدد متساوٍ من الأفراد.

(٤) إيجاد عدد الأفراد الذين أجابوا عن كل بدائل من بدائل مفردة الاختبار من متعدد في كل من المجموعتين العليا والدنيا.

وهذا يتطلّب عمل جدول مستطيل الشكل لكل مفردة، ويقسم إلى ثلاثة صفوف، يخصص الصف الأول للحروف المناظرة لبدائل الإجابات التي تشتمل عليها المفردة (أ، ب، ج، د، هـ مثلاً)، ويخصص الصفان الثاني والثالث لعدد أفراد كل من المجموعتين العليا والدنيا الذين أجابوا عن كل من هذه البدائل على الترتيب.

وإذا كانت المفردة من نوع الصواب أو الخطأ، أو التكملة توجّد عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة في كل من المجموعتين.

(٥) لإيجاد معامل تميّز المفردة: نطرح نسبة عدد أفراد المجموعة الدنيا الذين أجابوا عن المفردة إجابة صحيحة من نسبة عدد أفراد المجموعة العليا الذين أجابوا عن المفردة إجابة صحيحة.

ويكن توضيح هذه الخطوات بالمثال التالي حيث طبق اختبار يتكون من مفردات اختيار من متعدد لكل منها خمسة بدائل على ٤٠ طالب.
وفيما يلى درجاتهم الكلية في الاختبار:

١٨	٢٢	٢٣	٣٣	٤٣	٢٦	٢٠	١٥
(١٣)	٢١	(٤٥)	١٧	٣٥	١٦	٢٧	٢٩
٣٢	٤١	٣٧	٣٦	٤٢	٤٤	١٩	٢٧
٤٠	٣٢	١٥	٤٠	١٧	٣٥	١٤	١٥
٣٧	٣٨	٣٤	٤١	٣٣	٣٠	٤١	١٨

فالخطوة الأولى: نُرتب الدرجات ترتيباً تناظرياً ، أي نبدأ بالدرجة ٤٥ وننتهي بالدرجة ١٣ .

والخطوة الثانية: نختار أعلى ٢٧٪ من الدرجات ، وأدنى ٢٧٪ من الدرجات لتمثل كل من المجموعتين العليا والدنيا على الترتيب، وذلك بضرب العدد الكلى للطلاب في ٢٧ ، أي: $40 \times 27 = 1080$ ونقرب هذا العدد بحيث يصبح ١١٠ . وبذلك تكون درجات المجموعة العليا هي : ٤٥ ، ٤٤ ، ٤٣ ، ٤٢ ، ٤١ ، ٤٠ ، ٣٧ ، ٣٨ .

ودرجات المجموعة الدنيا هي: ١٩ ، ١٨ ، ١٧ ، ١٦ ، ١٥ ، ١٤ ، ١٣ .

والخطوة الثالثة: نستخدم هاتان المجموعتان في إيجاد قيمة معامل التمييز للإحدى مفردات الاختبار. وهذا يتطلب تحديد عدد طلاب كل من المجموعتين الذين أجابوا عن كل بديل من البذائل الخمسة التي تشتمل عليها المفردة، علماً بأن البديل (ج) يمثل الإجابة الصحيحة، لذلك أشرنا عليه بعلامة (*) ، وتلخيص ذلك في جدول (٧ - ٢) التالي :

هـ	د	* جـ	بـ	أـ	بدائل الإجابات
صفر	١	٩	صفر	١	عدد طلاب المجموعة الدنيا
١	١	٢	٣	٤	عدد طلاب المجموعة العليا

جدول (٧ - ٢) يوضح عدد الطلاب الذين أجابوا عن

كل بديل من بدائل مفردة اختيار من متعدد

والخطوة الرابعة : نوجد قيمة معامل تميز المفردة بطرح نسبة عدد طلاب المجموعة الدنيا الذين اختاروا البديل (جـ) من نسبة عدد طلاب المجموعة العليا الذين اختاروا هذا البديل حيث إنه يمثل الإجابة الصحيحة كالتالي :

$$\text{معامل تميز المفردة} = \frac{٢ - ٩}{١١} = \frac{٧}{٦٤} .$$

وهذه القيمة موجبة مما يدل على أن المفردة تميز بين المجموعتين.

ومن الجدير بالذكر أن تفسير قيم معامل التمييز تفسيراً أكثر دقة يتطلب مراعاة بعض الاعتبارات سوف نوضحها بعد قليل.

تقييم «تميز» المفردات استناداً إلى محك خارجي:

ذكرنا فيما سبق أن المحك الذي يستخدم في كثير من الاختبارات وبخاصة الاختبارات التحصيلية يكون محكاً داخلياً يتعلق بالدرجات الكلية في الاختبار.

لذلك يرى ساكس Sax ونيتوكو Nitko أنه إذا كان الغرض من الاختبار استقاء الأفراد أو تصنيفهم، فإن كل مفردة من المفردات التي يشتمل عليها ينبغي أن تسهم في الدرجة الكلية للاختبار، أي يكون معامل تميزها موجباً، كما ينبغي أن تسهم أيضاً في درجات محك خارجي مستقل عن الاختبار.

ففى مثل هذه الحالة لا يجب أن نعرف فقط أن الأفراد الذين أجابوا إجابة صحيحة عن المفردة يتميزون عن الذين أجابوا إجابة خطأ بالنسبة لدرجاتهم الكلية في الاختبار، ولكن يجب أيضاً أن نعرف أن الأفراد الذين أجابوا إجابة صحيحة يتحملون درجة أكبر أن يحققوا نجاحاً في محك خارجي، لأن يحصلوا على تقديرات مرتفعة من مشرفيهم على العمل أو من معلميهم. فمعامل التمييز الذي يعتمد على محك

الدرجات الكلية في الاختبار، أي المحك الداخلي يدل على إسهام المفردات في الاختبار ككل، ولكنه لا يدل على ما إذا كان الاختبار منبئاً جيداً لمحك خارجي.

ويقترح Davis أنه في التطبيقات الميدانية ينبغي انتقاء مفردات الاختبارات التي تُستخدم في الانتقاء والتصنيف استناداً إلى قيمتين لمعامل التمييز إحداهما تعتمد على ارتباط درجات المفردات بالدرجات الكلية في الاختبار كمحك داخلي، والأخرى تعتمد على ارتباط الدرجات بالمتغير المراد التنبؤ به كمحك خارجي.

غير أنه في الواقع ربما ترتبط بعض المفردات بالمحك الخارجي ولكنها لا ترتبط بالمحك الداخلي، والعكس ربما يكون أيضاً صحيحاً، هذا بالإضافة إلى المشكلات المتعلقة بالمحاكمات التي أوضحتها في الفصل الخامس.

لذلك يرى نانلى (Nunnally, 1978) أنه بدلاً من بناء اختبار واحد يشتمل على مفردات تقيس جوانب مختلفة للمحك الخارجي، يُفضل بناء اختبارات متفصلة لكل جانب بحيث تكون مفردات كل منها مميزة تمييزاً مرتفعاً.

وعموماً، فإنه يندر استخدام ممحاكمات خارجية في تقسيم مدى تمييز مفردات الاختبارات، أي تقسيم صدق المفردات، وإذا أردنا استخدامها فإنه يمكن الاسترشاد بنفس الطرق السابقة في إيجاد قيم معامل التمييز.

تفسير معامل تمييز المفردات، وعلاقته بمعامل التهوية:

يعتمد تفسير معامل تمييز مفردات الاختبارات ومعامل صعوبتها على الغرض من استخدام الاختبار، وخصائص الجماعة المرجعية التي تستمد منها هذه القيم.

وتتراوح قيم معامل التمييز بين -1 ، +1 ، فإذا كانت القيمة موجبة لمفردة معينة، فإن هذا يعني أن تميزها موجباً، أي أن نسبة عدد الأفراد الذين حصلوا على درجات مرتفعة في الاختبار ككل وأجابوا إجابة صحيحة عن هذه المفردة أكبر من نسبة عدد الأفراد الذين حصلوا على درجات منخفضة في الاختبار وأجابوا عنها إجابة صحيحة أيضاً، والعكس إذا كانت القيمة سالبة. أما إذا كانت القيمة صفراء، فإن المفردة لا تميز بين المجموعتين، ويمكن أن يحدث ذلك إذا كانت المفردة سهلة جداً، أي أجاب عنها جميع الأفراد إجابة صحيحة، أو كانت صعبة جداً بحيث لم يجب عنها أحد إجابة صحيحة، أو أن المفردة كانت غامضة.

وإذا كانت قيمة معامل تمييز المفردة سالبة، فإن هذا ربما يحدث عن طريق الصدفة لقلة عدد الأفراد المختبرين، أو ربما يرجع إلى غموض المفردة، أو عدم دقة الإجابة الصحيحة، أو قلة فاعلية بدائل الإجابات التي تشمل عليها المفردة.

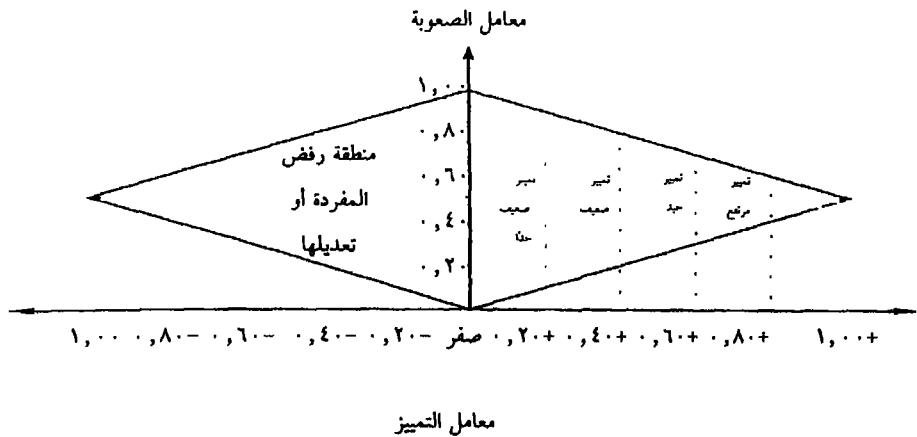
وبالطبع كلما زادت قيمة معامل تمييز المفردة كان ذلك أفضل، ولكن هذا يعتمد على الغرض من استخدام الاختبار كما ذكرنا. ففي الاختبارات مرجعية الجماعة أو المعيار التي تهدف للكشف عن الفروق الفردية ينبغي أن تميز مفرداتها تميزاً مرتفعاً بين مستويات السمة التي يقيسها الاختبار. أما في الاختبارات التشخيصية مرجعية المحك لا يجب أن نهتم كثيراً بمعامل تمييز مفرداتها، حيث يتركز الاهتمام هنا على جودة المفردات من حيث صياغتها، وقياسها لمكونات مهارات أو مهام معينة بحيث يمكن باستخدامها تحديد نقاط الضعف تحديداً دقيقاً.

وتعتمد قيم معامل تمييز المفردات على مستوى صعوبة المفردة، فكلما ابتعدت قيمة معامل صعوبة مفردة معينة عن .٥٠ ، بالزيادة أو النقصان قلت القيمة القصوى لمعامل التمييز. لذلك لا يجب أن تتوقع قيمة مرتفعة لمعامل التمييز إذا كانت مفردات الاختبار غاية في السهولة أو الصعوبة.

وعموماً ، إذا كانت قيمة معامل تمييز المفردة .٤٠ ، أو أكثر، فإن هذا يكون دليلاً على أن المفردة تميز بدرجة جيدة بين المجموعتين الطرفيتين، وإذا تراوحت بين .٢٠ ، .٤٠ ، فإن تمييز المفردة بين المجموعتين يكون لا بأس به، وإذا قلت هذه القيمة عن .٢٠ ، فإن تمييزها يكون ضعيفاً، أما إذا كان تمييزها صفرًا أو سالباً، فإنه ينبغي حذف هذه المفردة من الاختبار، أو مراجعتها مراجعة دقيقة وتعديلها.

ولكن ينبغي تأكيد أنه إذا كان الهدف من الاختبار قياس التمكّن من مهارة معينة مثلاً، فإن قيمة معامل الصعوبة ربما تساوى .٨٠ ، مثلاً، بينما قيمة معامل التمييز = صفر، ولا مانع من ذلك نظراً لأننا لا نهتم في هذه الحالة بإبراز الفروق الفردية وإنما قياس درجة التمكّن من المهارة المعينة.

ويوضح إيبيل (Ebel, 1965) وكريسبن وفيلدحسن (Kryspin, Feldhusen, 1979) العلاقة بين قيم كل من معامل الصعوبة ومعامل التمييز للمفردات الاختبارية بالشكل التخطيطي (٣ - ٧) التالي:



شكل (٧-٣) يوضح العلاقة بين معامل الصعوبة

ومعامل التمييز للمفردات الاختبارية

ويتبين من شكل (٧ - ٣) أن هناك علاقة بين قيم معامل صعوبة المفردات ومعامل تمييزها، فكلما ابتعدت قيمة معامل الصعوبة عن ٥٠ ، . إلى أعلى أو إلى أسفل على المحور الرأسى تقل القيمة القصوى لمعامل التمييز على المحور الأفقي.

لذلك فإن المفردات السهلة والمفردات الصعبة يكون تميزها ضئيلاً، والمفردات التي تتراوح قيمة معامل صعوبتها بين ٤٠ ، ٦٠ ، . يكون تميزها مرتفعاً كما يوضحه الخطان الأفقيان، وإذا بلغت قيمة معامل الصعوبة صفرًا أو واحداً صحيحاً، فإن معامل التمييز تكون قيمته صفرًا في الحالتين كما تتمثل نقطة تقاطع المحورين الأفقي والرأسى .

أما إذا كانت قيمة معامل تميز المفردة سالبة ، أي وقعت في منطقة الرفض التي إلى يسار المحور الرأسى، فإن هذه المفردة يمكن استبعادها أو مراجعتها وتدقيقها بالاسترشاد بنوع آخر من التحليل يسمى «تحليل فاعلية بدائل الإجابات» كما سيتبين بعد قليل .

إذا عدنا إلى المثال الذى قدمناه والمتعلق بمفردة الاختبار من متعدد الذى اشتغلت على خمسة بدائل (جدول ٧ - ٢) حيث وجدنا أن قيمة معامل التمييز ٦٤+ ، . ومعامل صعوبتها دون تصحيح أثر التخمين ٥٠ ، . مما يدل على أنها مفردة متوسطة الصعوبة، وإذا عدنا إلى شكل (٧ - ٣) نجد أن قيمة معامل التمييز تدل على أن المفردة تميز تميزاً مرتفعاً بين الطلاب الأقوية والطلاب الضعاف فى التحصيل، وبذلك تعد مفردة جيدة .

تحليل فاعلية بدائل الإجابات :

يهدف هذا التحليل لفحص البدائل الخطأ في مفردة اختيار من متعدد التي يطلق عليها المشتات Distractors ، ويمكن إجراء هذا التحليل إذا كانت قيمة معامل تميز المفردة سالبًا أو صفرًا أو قريباً من الصفر كما يوضحه الشكل (٧ - ٣) . كما يمكن إجراء ذلك إذا لاحظنا أن المفردة تتطلب مزيداً من التدقيق . فإذا تبين مثلاً أن أحد البدائل الخطأ (المشتات) في مفردة معينة لم يختاره أحد من الأفراد المختبرين ، فإن هذا ربما يكون دليلاً على عدم فاعلية هذا المشتت أو وضوح خطأه ، فمثل هذا المشتت لافائدة منه لأنه يضيع وقت الفرد في قراءته دون أن يقيس ما يهدف الاختبار لقياسه .

إذا وجدنا أن واحداً من المشتات اختاره معظم الأفراد وبخاصة الذين حصلوا على درجات مرتفعة في الاختبار ككل ، ولم يختاروا الإجابة الصحيحة ، فإن هذا ربما يكون دليلاً على أن هذا المشتت أكثر جاذبية من الإجابة الصحيحة ، أو ربما يكون هو الإجابة الصحيحة بالفعل ، وليس الإجابة التي حددها من أعد الاختبار . وهنا ينبغي مراجعة الإجابة الصحيحة والتأكد من صحتها وعدم صحة هذا المشتت . فإذا تأكدنا أن هذا المشتت يمثل إجابة خطأ ، فإنه ينبغي الإبقاء عليه؛ لأنه يساعد في كشف سوء الفهم وأنماط الأخطاء الشائعة لدى المختبرين مما يفيد في اقتراح أساليب العلاج .

أما إذا كانت إجابات الأفراد موزعة توزيعاً متعدلاً بين المشتات الثلاثة أو الأربع التي تشتمل عليها المفردة ، ولكن المفردة غير مميزة ، فإن هذا ربما يدل على أن هذه المفردة غامضة أو غاية في الصعوبة مما جعل المختبرين يلجأون إلى التخمين ، وهكذا .

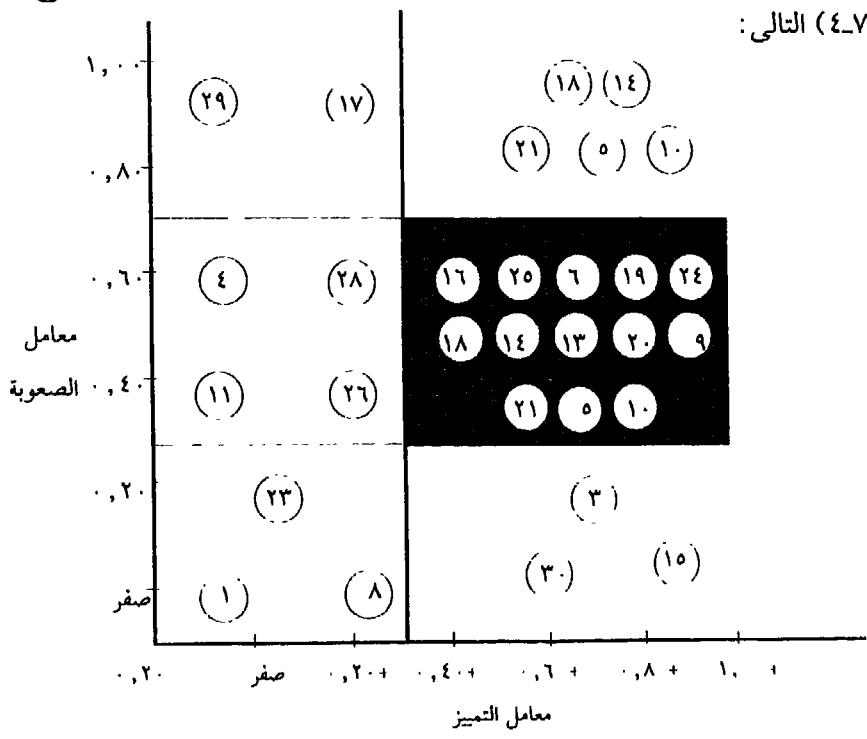
إذا عدنا مرة أخرى إلى مثال مفردة الاختيار من متعدد (جدول ٧ - ٢) نجد أن المفردة تميز تميزاً موجباً مرتفعاً بين المجموعتين العليا والدنيا حيث اختار الإجابة الصحيحة ٩ طلاب من المجموعة العليا ، وطالبان فقط من المجموعة الدنيا . ولكن إذا نظرنا إلى المشتت (أ) نجد أن عدد طلاب المجموعة العليا الذين اختاروه (٤) أكبر من عدد طلاب المجموعة الدنيا (١) مما يدل على أن هذا المشتت يحتاج إلى إعادة نظر ، وكذلك المشتت (ب) ، والمشتت (ه) حيث إنهما كانا أكثر جاذبية لطلاب المجموعة العليا . أما المشتت (د) فقد اختاره عدد متساو من كل من المجموعتين مما يدل أيضاً على عدم فاعليته .

من هذا يتضح كيف أن تحليل فاعلية المشتات لمفردات الاختيار من متعدد وغيرها من المفردات التي تتطلب الاختيار من بين بدائل معطاة يلقى الضوء على كل

وغيرها من المفردات التي تتطلب الاختيار من بين بدائل معطاة يلقى الضوء على كل مشتت تتضمنه المفردة، ويمكن القائم بناء الاختبار أو المعلم من فحص الإجابات فحصاً مدققاً يساعد في تحسين المفردة وزيادة فاعليتها في قياس السمة التي يقيسها الاختبار، وتشخيص الأخطاء وسوء الفهم لدى المختبرين في مجال معين.

استخدام تحليل المفردات في انتقاء مفردات الاختبار:

إن انتقاء المفردات في الاختبارات مرجعية الجماعة أو المعيار يعتمد اعتماداً كبيراً على خصائص هذه المفردات، أي مستوى صعوبتها ودرجة تميزها بين مستويات السمة المقاسة. ولذلك ينبغي عند انتقاء المفردات إجراء مقارنات فيما بينها بعد تجربتها ميدانياً للتحقق من ملاءمة مستوى صعوبتها للأفراد المختبرين وتميزها تميزاً جيداً بين الضعاف والأقوباء. ويمكن إجراء ذلك بتمثيل قيم كل من معامل التمييز ومعامل الصعوبة على المحورين الأفقي والرأسى على الترتيب. ويمكن أن تكون قيم معامل التمييز ناتجة عن استخدام معامل الارتباط الثنائي المتسلسل أو عن استخدام المجموعتين الطرفيتين، وتحديد نقطة تمثل قيمة كل من المعاملين لكل مفردة من مفردات اختبار يشتمل على ٣٠ مفردة اختيار من متعدد مثلاً كما هو موضح بشكل (٤-٧) التالي:



شكل (٤-٧) يوضح قيم معامل تميز ومعامل صعوبة ٣٠ مفردة اختيارية

ويتضح من شكل (٧ - ٤) أن المنطقة المظللة تمثل المفردات التي تزيد قيم معامل تميزها عن .٣٠ ، وتحصر قيم معامل صعوبتها بين .٣٠ ، .٧٠ ، فهذه المفردات يجب أن تستبق في الاختبار، أما المفردات التي تقع خارج هذه المنطقة فينبغي فحصها جيداً، نظراً لأن التحليل الإحصائي للمفردات لا يعني عن فحص محتوى المفردة وطريقة صياغتها وتكونيتها وإجراء التعديلات الضرورية عليها أو حذفها.

استخدام الحاسوب في تحليل مفردات الاختبارات:

يتطلب تحليل مفردات الاختبارات التربوية والنفسية وتحليل الاختبارات ككل كما رأينا كثيراً من التحليلات الإحصائية والعمليات الحسابية التي ربما يصعب إجراؤها يدوياً أو باستخدام الآلة الحاسبة، وبخاصة إذا كان عدد الأفراد المختبرين كبيراً كما في حالة الاختبارات المقنتة. لذلك ينبغي الإفادة من برامج الحاسوب المتوفّرة لإجراء كثير من التحليلات التي تتطلّبها عمليات بناء الاختبارات والمقييس.

فهناك برامج حديثة متعدّدة يمكن استخدامها في تحليل مفردات الاختبار وفحصها وتجميع مفردات الاختبار وتخزينها في ملف Item File يمكن استدعائه عند الحاجة وتعديل مفرداته وتحسينها وتنميّتها.

ولعل برنامج MicroCAT يعد من البرامج الجيدة الشاملة، حيث يمكن باستخدامه إجراء تحليل مفردات الاختبارات في إطار كل من النظرية الكلاسيكية للاختبارات (CTT) ، ونظرية الاستجابة للمفردة الاختبارية (IRT) ، وقد أعد هذا البرنامج مؤسسة نظم التقويم (ASC) Assessment Systems Corporation بالولايات المتحدة الأمريكية. وعلى الرغم من ارتفاع ثمن هذا البرنامج نسبياً، إلا أنه يعد بمثابة نظام اختباري متكامل حيث يطبق الاختبارات باستخدام الحاسوب، ويصحح نتائجها ويحلل درجات كل مفردة وكذلك درجات الاختبار ككل، ويعرض النتائج في جداول منتظمة ويمثلها برسوم وأشكال بيانية متنوعة موضحاً فيها أنماط استجابات كل طالب عن مفردات الاختبار.

كما أن هناك برنامجاً Iteman لتحليل المفردات أعدته المؤسسة السابقة نفسها، ويعد البرنامج ملفات درجات المفردات اعتماداً على الفاصل الضوئي للاختبار Optical Test Scanner أو اعتماداً على إدخال البيانات يدوياً. ويمكن باستخدام هذا

البرنامج تحليل درجات الاختيار من متعدد، والصواب أو الخطأ، وموازن التقدير، وما شابه ذلك. ويمكن إجراء تحليل لفردات الاختبارات التي تشتمل على اختبارات فرعية قد يصل إلى عشرة اختبارات.

ويقدم البرنامج معلومات إحصائية تتعلق بالمتوسط والتباين والانحراف المعياري، والتلواء وتفرط الدوائر الكلية في كل اختبار، والوسط وأعلى الدرجات وأقلها، والتوزيع التكراري للدرجات ، وقيم معامل كيودر وريتشاردسون لثبات درجات الاختبارات، والخطأ المعياري للقياس لكل اختبار فرعى ، وكذلك قيم كل من معامل صعوبة ومعامل تمييز المفردات للتعرف على المفردات التي تسهم في قياس سمة معينة. ويمكن كذلك إجراء هذه التحليلات لمجموعات فرعية من المختبرين كالذكور أو الإناث فقط. ويوضح شكل (٧ - ٥) التالى عينة من مخرجات هذا البرنامج :

MICROCAT (TM) Testing System Page 2 Copyright (C) 1987 - 1994 by Assessment System Corporation									
Item and Test Analysis Program ITEMAN Test Version 4.6									
Test analysis for data file: ITEMAN.DAT Date: 1-22-94 Time: 3:26 PM									
Item Statistics									
Bug					Alternative Statistics				
No.	Score	Prop.	Index	Point	Prop.	Performance	Prop.	Mean	Std. Dev.
1	20	.19	.35	.34	2	.19	.39	.35	.44
2	40	.68	.68	.68	3	.19	.39	.35	.44
4	64				5				

MICROCAT (TM) Testing System Page 8 Copyright (C) 1987 - 1994 by Assessment System Corporation									
Item and Test Analysis Program ITEMAN Test Version 4.6									
Test analysis for data file: ITEMAN.DAT Date: 1-22-94									
Item Statistics									
Bug					Alternative Statistics				
No.	Score	Prop.	Index	Point	Prop.	Performance	Prop.	Mean	Std. Dev.
1	1	.0	.0	.0	2	.0	.0	.0	.0
2	4	.0	.0	.0	3	.0	.0	.0	.0
4	9	.0	.0	.0	5	.0	.0	.0	.0
6	14	.0	.0	.0	7	.0	.0	.0	.0
8	19	.0	.0	.0	9	.0	.0	.0	.0
10	24	.0	.0	.0	11	.0	.0	.0	.0
12	29	.0	.0	.0	13	.0	.0	.0	.0
14	34	.0	.0	.0	15	.0	.0	.0	.0
16	39	.0	.0	.0	17	.0	.0	.0	.0
18	44	.0	.0	.0	19	.0	.0	.0	.0
20	49	.0	.0	.0	21	.0	.0	.0	.0
22	54	.0	.0	.0	23	.0	.0	.0	.0
24	59	.0	.0	.0	25	.0	.0	.0	.0
26	64	.0	.0	.0	27	.0	.0	.0	.0
28	69	.0	.0	.0	29	.0	.0	.0	.0
30	74	.0	.0	.0	31	.0	.0	.0	.0
32	79	.0	.0	.0	33	.0	.0	.0	.0
34	84	.0	.0	.0	35	.0	.0	.0	.0
36	89	.0	.0	.0	37	.0	.0	.0	.0
38	94	.0	.0	.0	39	.0	.0	.0	.0
40	99	.0	.0	.0	41	.0	.0	.0	.0
42	104	.0	.0	.0	43	.0	.0	.0	.0
44	109	.0	.0	.0	45	.0	.0	.0	.0
46	114	.0	.0	.0	47	.0	.0	.0	.0
48	119	.0	.0	.0	49	.0	.0	.0	.0
50	124	.0	.0	.0	51	.0	.0	.0	.0
52	129	.0	.0	.0	53	.0	.0	.0	.0
54	134	.0	.0	.0	55	.0	.0	.0	.0
56	139	.0	.0	.0	57	.0	.0	.0	.0
58	144	.0	.0	.0	59	.0	.0	.0	.0
60	149	.0	.0	.0	61	.0	.0	.0	.0
62	154	.0	.0	.0	63	.0	.0	.0	.0
64	159	.0	.0	.0	65	.0	.0	.0	.0
66	164	.0	.0	.0	67	.0	.0	.0	.0
68	169	.0	.0	.0	69	.0	.0	.0	.0
70	174	.0	.0	.0	71	.0	.0	.0	.0
72	179	.0	.0	.0	73	.0	.0	.0	.0
74	184	.0	.0	.0	75	.0	.0	.0	.0
76	189	.0	.0	.0	77	.0	.0	.0	.0
78	194	.0	.0	.0	79	.0	.0	.0	.0
80	199	.0	.0	.0	81	.0	.0	.0	.0
82	204	.0	.0	.0	83	.0	.0	.0	.0
84	209	.0	.0	.0	85	.0	.0	.0	.0
86	214	.0	.0	.0	87	.0	.0	.0	.0
88	219	.0	.0	.0	89	.0	.0	.0	.0
90	224	.0	.0	.0	91	.0	.0	.0	.0
92	229	.0	.0	.0	93	.0	.0	.0	.0
94	234	.0	.0	.0	95	.0	.0	.0	.0
96	239	.0	.0	.0	97	.0	.0	.0	.0
98	244	.0	.0	.0	99	.0	.0	.0	.0
100	249	.0	.0	.0	101	.0	.0	.0	.0
102	254	.0	.0	.0	103	.0	.0	.0	.0
104	259	.0	.0	.0	105	.0	.0	.0	.0
106	264	.0	.0	.0	107	.0	.0	.0	.0
108	269	.0	.0	.0	109	.0	.0	.0	.0
110	274	.0	.0	.0	111	.0	.0	.0	.0
112	279	.0	.0	.0	113	.0	.0	.0	.0
114	284	.0	.0	.0	115	.0	.0	.0	.0
116	289	.0	.0	.0	117	.0	.0	.0	.0
118	294	.0	.0	.0	119	.0	.0	.0	.0
120	299	.0	.0	.0	121	.0	.0	.0	.0
122	304	.0	.0	.0	123	.0	.0	.0	.0
124	309	.0	.0	.0	125	.0	.0	.0	.0
126	314	.0	.0	.0	127	.0	.0	.0	.0
128	319	.0	.0	.0	129	.0	.0	.0	.0
130	324	.0	.0	.0	131	.0	.0	.0	.0
132	329	.0	.0	.0	133	.0	.0	.0	.0
134	334	.0	.0	.0	135	.0	.0	.0	.0
136	339	.0	.0	.0	137	.0	.0	.0	.0
138	344	.0	.0	.0	139	.0	.0	.0	.0
140	349	.0	.0	.0	141	.0	.0	.0	.0
142	354	.0	.0	.0	143	.0	.0	.0	.0
144	359	.0	.0	.0	145	.0	.0	.0	.0
146	364	.0	.0	.0	147	.0	.0	.0	.0
148	369	.0	.0	.0	149	.0	.0	.0	.0
150	374	.0	.0	.0	151	.0	.0	.0	.0
152	379	.0	.0	.0	153	.0	.0	.0	.0
154	384	.0	.0	.0	155	.0	.0	.0	.0
156	389	.0	.0	.0	157	.0	.0	.0	.0
158	394	.0	.0	.0	159	.0	.0	.0	.0
160	399	.0	.0	.0	161	.0	.0	.0	.0
162	404	.0	.0	.0	163	.0	.0	.0	.0
164	409	.0	.0	.0	165	.0	.0	.0	.0
166	414	.0	.0	.0	167	.0	.0	.0	.0
168	419	.0	.0	.0	169	.0	.0	.0	.0
170	424	.0	.0	.0	171	.0	.0	.0	.0
172	429	.0	.0	.0	173	.0	.0	.0	.0
174	434	.0	.0	.0	175	.0	.0	.0	.0
176	439	.0	.0	.0	177	.0	.0	.0	.0
178	444	.0	.0	.0	179	.0	.0	.0	.0
180	449	.0	.0	.0	181	.0	.0	.0	.0
182	454	.0	.0	.0	183	.0	.0	.0	.0
184	459	.0	.0	.0	185	.0	.0	.0	.0
186	464	.0	.0	.0	187	.0	.0	.0	.0
188	469	.0	.0	.0	189	.0	.0	.0	.0
190	474	.0	.0	.0	191	.0	.0	.0	.0
192	479	.0	.0	.0	193	.0	.0	.0	.0
194	484	.0	.0	.0	195	.0	.0	.0	.0
196	489	.0	.0	.0	197	.0	.0	.0	.0
198	494	.0	.0	.0	199	.0	.0	.0	.0
200	499	.0	.0	.0	201	.0	.0	.0	.0
202	504	.0	.0	.0	203	.0	.0	.0	.0
204	509	.0	.0	.0	205	.0	.0	.0	.0
206	514	.0	.0	.0	207	.0	.0	.0	.0
208	519	.0	.0	.0	209	.0	.0	.0	.0
210	524	.0	.0	.0	211	.0	.0	.0	.0
212	529	.0	.0	.0	213	.0	.0	.0	.0
214	534	.0	.0	.0	215	.0	.0	.0	.0
216	539	.0	.0	.0	217	.0	.0	.0	.0
218	544	.0	.0	.0	219	.0	.0	.0	.0
220	549	.0	.0	.0	221	.0	.0	.0	.0
222	554	.0	.0	.0	223	.0	.0	.0	.0
224	559	.0	.0	.0	225	.0	.0	.0	.0
226	564	.0	.0	.0	227	.0	.0	.0	.0
228	569	.0	.0	.0	229	.0	.0	.0	.0
230	574	.0	.0	.0	231	.0	.0	.0	.0
232	579	.0	.0	.0	233	.0	.0	.0	.0
234	584	.0	.0	.0	235	.0	.0	.0	.0
236	589	.0	.0	.0	237	.0	.0	.0	.0
238	594	.0	.0	.0	239	.0	.0	.0	.0
240	599	.0	.0	.0	241	.0	.0	.0	.0
242	604	.0	.0	.0	243	.0	.0	.0	.0
244	609	.0	.0	.0	245	.0	.0	.0	.0
246	614	.0	.0	.0	247	.0	.0	.0	.0
248	619	.0	.0	.0	249	.0	.0	.0	.0
250	624	.0	.0	.0	251	.0	.0	.0	.0
252	629	.0	.0	.0	253	.0	.0	.0	.0
254	634	.0	.0	.0	255	.0	.0	.0	.0
256	639	.0	.0	.0	257	.0	.0	.0	.0
258	644	.0	.0	.0	259	.0	.0	.0	.0
260	649	.0	.0	.0	261	.0	.0	.0	.0
262	654	.0	.0	.0	263	.0	.0	.0	.0
264	659	.0	.0	.0	265	.0	.0	.0	.0
266	664	.0	.0	.0	267	.0	.0	.0	.0
268	669	.0	.0	.0	269	.0	.0	.0	.0
270	674	.0	.0	.0	271	.0	.0	.0	.0
272	679	.0	.0	.0	273	.0	.0	.0	.0
274	684	.0	.0	.0	275	.0	.0	.0	.0
276	689	.0	.0	.0	277	.0	.0	.0	.0
278	694	.0	.0	.0	279	.0	.0	.0	.0
280	699	.0	.0	.0	281	.0	.0	.0	.0
282	704	.0	.0	.0	283	.0	.0	.0	.0
284	709	.0	.0	.0					

ويمكن تحليل مفردات الاختبارات أيضاً اعتماداً على البرمجيات المتداولة بكثرة حيث يسهل استخدامها في تصميم الجداول الإلكترونية وتحليلها، ومن أهمها برنامج MICROSOFT EXCEL الذي يمكن إدارته في بيئة النوافذ Windows 97 أو windows 95. وهذا يتطلب تنظيم الجداول وفق ما يراه معد الاختبار، وكذلك برمجة العمليات الإحصائية المطلوبة لتحليل مفردات الاختبار أو تحليل درجاته الكلية، وتصميم الرسوم ، والأشكال البيانية التي توضح النتائج توضيحاً بصرياً مناسباً.

ويوضح شكل (٧ - ٦) التالي عينة من هذا التصميم أشار إليها وود ، وساfrit : (Safrit & Wood, 1995)

Student	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
1	1	1	1	1	1	5
2	1	1	1	0	1	4
3	1	1	1	0	1	4
4	1	1	1	1	0	4
5	1	1	1	1	0	4
6	1	0	1	0	1	3
7	1	0	1	0	1	3
8	0	1	1	0	1	3
9	1	0	0	0	1	2
10	0	1	0	1	0	2
11	0	1	0	0	1	2
12	0	1	1	0	0	2
Number Correct	8	9	9	4	8	
Difficulty	0.67	0.75	0.75	0.33	0.67	
Discrimination	0.66	0.29	0.68	0.42	0.12	
Decision	Retain	Revise	Revise	Revise	Discard	

شكل (٧ - ٦)

جدول إلكتروني لبرنامج EXCEL لتحليل المفردات

من هذا يتضح أن تحليل مفردات الاختبارات والمقاييس يُعد من العمليات الأساسية في جمع البيانات الإمبريقية المتعلقة بكل مفردة وتقييم خصائص كل منها

باستخدام قيم معامل الصعوبة ومعامل التميز، واستخدام نتائج هذا التقسيم في فحص المفردات التي يتبيّن قلة فاعليتها ومراجعتها وتدقيقها أو حذفها إذا تطلب الأمر ذلك. كما يمكن استخدام هذه النتائج في تقديم تغذية مرتبطة للمختبرين والقائم بإعداد الاختبار، وبخاصة إذا كان معلماً حيث يمكنه الإفادة من النتائج في إثراء تعلم طلابه وتنمية مهارته في بناء الاختبارات.

غير أنه ينبغي إعادة تأكيد أن تحليل مفردات الاختبارات لا يغنى عن العناية والحرص في بناء مفردات الاختبار وصياغتها صياغة دقيقة لا تحتمل الغموض أو التخيّل. فانتقاء مفردات الاختبارات يعتمد بدرجة أساسية على الأحكام الإنسانية - بالإضافة إلى ما نتوصل إليه من نتائج التحليل الإحصائي للبيانات الإمبريقية المتعلقة بالمفردات.

خلاصة

إن الغرض الأساسي من تحليل مفردات الاختبارات التتحقق من بعض الخصائص التي ينبغي توافرها في المفردات التي يتكون منها الاختبار. ومن أهم هذه الخصائص صعوبة المفردة، ودرجة تميزها بين المستويات العليا والدنيا في السمة أو القدرة التي يقيسها الاختبار. فمستوى صعوبة المفردة يتحدد بنسبة عدد الأفراد الذين أجابوا عن المفردة إجابة صحيحة، وكلما زادت قيمة هذه النسبة دل ذلك على سهولة المفردة، والعكس يكون صحيحاً . ولكن ينبغي مراعاة تصحيح هذه القيمة من أثر التخمين في المفردات التي تتطلب اختيار من متعدد، أو صواب أو خطأ. كما يمكن تلافي أوجه قصور معامل الصعوبة بأساليب متعددة كأن نجري على قيم هذا المعامل بعض التحويلات باستخدام خصائص المنحنى الاعتدالى، أو نستخدم أحد نماذج الاستجابة للمفردة الاختبارية التي سنوضحها في الفصل الرابع عشر. أما درجة تميز المفردة فيمكن تقييمها استناداً إلى محك داخلى باستخدام معامل الارتباط بين درجات كل مفردة والدرجات الكلية في الاختبار، أو باستخدام المجموعتين الطرفيتين، كما يمكن تقييمها استناداً إلى محك خارجى حيث يوجد ارتباط درجات المفردات بدرجات المتغير الخارجى المراد التنبؤ به.

وينبغي مراعاة العلاقة بين معامل صعوبة المفردات، ومعامل تميزها عند تفسير قيم كل منها. ويمكن الإفادة من نتائج تحليل المفردات في التتحقق من فاعلية بدائل الإجابات في مفردات الاختيار من متعدد، وتعرف جوانب القصور فيها من أجل مراجعتها وتعديلها أو استبعادها. ونظراً لما يتطلبه هذا التحليل من عمليات إحصائية وحسابية متعددة، وبخاصة إذا كان عدد المفردات أو الأفراد أو كلاهما كبيراً، فإنه يمكن إجراء ذلك باستخدام برمج الحاسوب مثل برنامج MicroCAT بطريقة أكثر مرونة وفاعلية .



تطبيقات القياس والتقويم التربوي والنفسى

(أولاً: قياس وتقدير الجوانب المعرفية



الفصل الثامن

قياس التحصيل والكفايات

* مقدمة

- * قياس التحصيل الدراسي أو الأكاديمي من منظور تاريخي
- * مفهوم التحصيل الدراسي وعلاقته بمفهوم الاستعداد
- * التمييز بين اختبارات التحصيل واختبارات الاستعداد
- * افتراضات أساسية يرتكن إليها قياس التحصيل
- * تصنيف اختبارات التحصيل واستخداماتها
- أولاً : اختبارات التحصيل مرجعية المحك
- * خطوات بناء الاختبارات مرجعية المحك
- * استخدامات الاختبارات مرجعية المحك
- ثانياً : اختبارات التحصيل مرجعية الجماعة أو المعيار
- * خطوات بناء اختبارات التحصيل المقنتة
- * بعض أوجه النقد التي توجه للاختبارات التحصيلية المقنتة
- * خلاصة

مقدمة :

عرضنا في الباب الأول أهم الأسس المنطقية والمبادئ الأخلاقية والمفاهيم والأساليب الإحصائية والمنهجية للقياس والتقويم التربوي والنفسى.

وسوف نتناول في هذا الباب بعض التطبيقات المتعلقة بقياس وتقدير الجوانب المعرفية للشخصية الإنسانية، وتوظيف البيانات المستمدة من الاختبارات التي تقيس هذه الجوانب في اتخاذ القرارات المتعلقة بالأفراد والبرامج. وتتضمن الجوانب المعرفية Academic Achievement Cognitive Attributes، والذكاء العام General Intelligence، والاستعدادات الخاصة Special Aptitudes، حيث يتباين الأفراد تبايناً ملحوظاً في كل منها. فالفرق الفردية في التحصيل والذكاء والاستعدادات تبدو أكثر وضوحاً للمعلم في الصنوف المدرسية وبخاصة في مراحل التعليم العام.

ففي الصف المدرسي تكون الفرصة متاحة للمعلم للاحظة طلابه في ظروف لا توافر مثلها في أي مكان آخر، فالمعلم يนาش طلابه في الصف ويتفاعل معهم ويضع تقييمات لأدائهم إما بواسطة اختبارات تحريرية أو شفوية، وبذلك يلمس الفروق الفردية في تحصيل طلابه بوضوح أثناء قيامه بعملية تعليمهم وتقديرهم. وهذا يجعل من الضروري على المعلم تفهم طبيعة هذه الفروق وكيفية قياسها لكي يتمكن من الكشف عنها ومعرفة مداها، وبذلك يستطيع توجيه العملية التعليمية توجيهها يناسب قدرات واستعدادات طلابه بما يمكنهم من اكتساب الكفايات المرجوة من العملية التعليمية.

وهذا يتطلب توضيح مفهوم التحصيل الدراسي أو الأكاديمي والعلقة بينه وبين مفهوم الاستعداد والتمييز بين الاختبارات التي تقيس كل منهما.

لذلك سوف نفرد هذا الفصل لمناقشة كل من المفهومين، ونوضح أساليب قياس التحصيل الدراسي أو الأكاديمي باستخدام الاختبارات مرجعية الجماعة أو المعيار، والاختبارات مرجعية المعلم، ونبين كيفية الإفاده من نتائج هذه الاختبارات في تقويم كفايات المتعلمين وفاعلية البرنامج التعليمي. وتناول أساليب قياس الذكاء والاستعدادات الخاصة في الفصلين التاليين.

ولعل من المناسب أن نستهل هذا الفصل بإلقاء الضوء على تطور قياس التحصيل الدراسي أو الأكاديمي من منظور تاريخي.

قياس التحصيل الدراسي أو الأكاديمي من منظور تاريخي :

إن استيعاب الأساليب المختلفة التي تستخدم في وقتنا الحاضر في قياس التحصيل الدراسي أو الأكاديمي استيعاباً مستثيراً ينبع أن يستند إلى رؤية واضحة للتطورات التاريخية التي مرت بها هذه الأساليب حتى أصبحت كما نراها اليوم.

فالمتحانات المدرسية كانت تعتمد اعتماداً أساسياً على الاختبارات الشفوية حتى نهاية النصف الأول من القرن التاسع عشر. وفي عام ١٨٤٥ تزعم هوراس مان Horace Mann الذي كان يعد من القادة التربويين البارزين في ولاية بوسطن الأمريكية حركة تطوير التعليم العام، وأكد ضرورة استخدام الاختبارات التحريرية التي تسم بال موضوعية والاتساق بدلاً من الاختبارات الشفوية في تقويم المتعلمين. ولكن استمر استخدام الاختبارات الشفوية في التقويم إلى أن بدأت في الاضمحلال وحلت محلها الاختبارات التحريرية. وفي عام ١٨٦٤ أعد المربى الإنجليزي جورج فيشر G. Fisher الذي كان يعمل مدير مدرسة أول اختبار تحصيلي تحريري يتكون من عدة مقاييس متدرجة وأمثلة ومواصفات متنوعة يمكن باستخدامها تقييم جودة الخط والنحو والتعبير والهجاء والرياضيات وغيرها من المواد الدراسية.

وأطلق على هذه البطارية من الاختبارات «كتاب الموارين Scale Book ». وقد وضع في هذا الكتاب كيفية تقدير درجات كل اختبار فرعى تشمل عليه البطارية، والدرجات الكلية والمعايير التي يمكن استخدامها في مقارنة الطالب بنفسه في هذه المواد المختلفة، ورسم صفحة تربوية تبين جوانب قوته وضعفه. وبذلك وضع هذا الكتاب الاختباري الأساس لبناء الاختبارات التحريرية الموضوعية. ولعل هذه البطارية من الاختبارات كانت الركيزة التي استندت إليها حركة بناء الاختبارات مرحلة المحك فى وقتنا الحاضر.

وبعد مضي ما يقرب من ربع قرن من الزمان أعد رئيس J. Rice عام ١٨٩٥ أول اختبار تحصيلي موضوعي في الولايات المتحدة الأمريكية يقيس قدرة تلاميذ المدارس الابتدائية على الهجاء وطبقه على ما يقرب من ١٦٠٠٠ تلميذ في الصفوف من الرابع حتى الثامن، كما أعد اختباراً تحصيلياً في الحساب واللغة الإنجليزية. وتعد هذه الاختبارات لبيات لبناء الاختبارات التحصيلية المقنتة حيث شهد أوائل القرن العشرين ظهور عدد من هذه الاختبارات المقنتة متمثلة في اختبار الحساب الذي أعده ستون Stone عام ١٩٠٨ ، واختبار جودة الخط الذي أعده ثورنديك Thorndike عام ١٩٠٩ ويزيد عدد هذه الاختبارات زيادة كبيرة قرب نهاية عام ١٩٢٠ بما في ذلك بطاريات الاختبارات، مثل بطارية اختبارات ستانفورد التحصيلية للمرحلة الابتدائية عام ١٩٢٣ ، وبطارية امتحانات أيوا للمحتوى الدراسي للمرحلة الثانوية عام ١٩٢٥ .

وقد أسهمت هذه البطاريات وما أكدته كثیر من الدراسات التي أجريت في ذلك الوقت حول عدم اتساق تقديرات المعلمين لطلابهم، في التوجه المتزايد نحو الاختبارات الموضوعية لتقويم التحصيل الدراسي. وتزايد عدد الاختبارات التحصيلية المقنتة المنشورة التي استخدمت المفردات الموضوعية حتى وصل إلى المئات، وتضمنت اختبارات مادة دراسية واحدة، وبطاريات تشتمل على عدد من اختبارات مواد دراسية متعددة، واختبارات تشخيصية وتنبؤية. كما ازداد الاهتمام ببناء مفردات موضوعية تقيس الفهم والاستبطاط وتطبيق المعلومات، وغيرها من العمليات المعرفية العليا، وكذلك الأهداف التعليمية الشاملة. وشهد عقد الثلاثينيات تصميم آلية تصحيح الاختبارات الموضوعية لتيسير تطبيق وتصحيح الاختبارات التحصيلية المقنتة على نطاق واسع.

وعلى الرغم من الخلاف القائم حتى الآن حول فاعلية كل من الاختبارات الموضوعية التي تشتمل على مفردات اختيار من متعدد، أو صواب أو خطأ، أو مزاجة، أو تكملة، أو اختبارات المقال، إلا أن تطوير هذه المفردات الموضوعية بحيث تشمل العمليات المعرفية العليا التي يفترض أنها تقاس ف aras فقط بأسئلة المقال ساعد في انتشار استخدامها في الاختبارات التحصيلية المقنتة لتقويم المتعلمين.

وقد أدى تصميم برامج اختبارية على المستوى القومي والإقليمي في الولايات المتحدة الأمريكية إلى دفعه قوية لحركة تطوير الاختبارات التحصيلية وبخاصة الاختبارات المقنتة، ولعل أهمها برنامج مكتب اختبارات القبول بالكلجات (CEEB) عام ١٩٥٢ الذي أشرنا إليه في الفصل السادس، وبرنامج تقويم التعليم العام General Education Evaluation Program الذي اكتمل تفريغه عام ١٩٥٤، وكذلك البرنامج الاختباري للكلجات الأمريكية American Colleges Testing Program عام ١٩٥٩ لفرز المتقدمين للالتحاق بالكلجات بغض انتقاء الطلاب المتميزين للحاقدتهم ببعثات دراسية.

ولم يقتصر استخدام الاختبارات التحصيلية المقنتة على الأغراض التربوية، وإنما استُخدمت أيضًا في انتقاء العاملين بالمصانع والمؤسسات الحكومية.

كذلك أدى التوسيع في استخدام هذه الاختبارات - بالإضافة إلى مشاركة خبراء علم النفس في القياس النفسي وبناء الاختبارات المقنتة - إلى التشابه الكبير في عمليات بناء هذه الاختبارات وتقييمها سواء كانت تحصيلية أو استعدادات أو قدرة عقلية عامة. فالاختبارات التحصيلية المقنتة التي تقيس أهدافًا تعليمية شاملة أو متسعة أصبحت تشبه اختبارات الذكاء العام، فالاختلاف بينهما يكمن في مدى اتساع المحتوى وعموميته وما يتطلبه من توافر معارف أو مهارات تعليمية معينة لدى الأفراد المختبرين.

وعلى الرغم من انتشار استخدام الاختبارات التحصيلية المقنية ما يزيد عن نصف قرن في مختلف مجالات التقويم التربوي والنفسى ، إلا أنه ظهر في أوائل السبعينيات في الولايات المتحدة الأمريكية كثير من الآراء المعارضه لهذا النوع من الاختبارات.

وحجتهم في ذلك أن هذه الاختبارات تستند في تفسير درجاتها على جماعة مرجعية معينة تعد بمثابة معيار يقارن في ضوئه أداء الأفراد، لذلك أطلق عليها الاختبارات مرجعية الجماعة أو المعيار كما أشرنا في الفصول السابقة، أى أنها تحدد المركز النسبي للفرد بين أقرانه، دون أن تهم بتحديد ما يمكن للفرد أن يؤديه أو يفعله.

وعلى الرغم من أن ثورنديك Thorndike عام ١٩١٣ ، وفلانagan Flanagan عام ١٩٥١ وغيرهما ميزاً بين المعيار النسبي ومحلk الأداء المطلوب في القياس التربوي والنفسى ، إلا أن هذا التمييز لم يسهم إسهاماً ملحوظاً في فلسفة بناء الاختبارات التحصيلية إلى أن نشر جيلزر Glaser مقالته المهمة حول الاختبارات مرجعية المحرك في دورية عالم النفس الأمريكي American Psychologist عام ١٩٦٣ ، وأعقبها مقالة بابام Babam وهاسك Husek في دورية القياس التربوي عام ١٩٦٩ التي حفظت المشغلين بالقياس على الاهتمام بتطوير أساليب التقويم باستخدام الاختبارات مرجعية المحرك. وتربت على ذلك حركة بحثية كبيرة في هذا المجال ، حتى أنه بحلول عام ١٩٧٨ تم نشر ما يزيد عن ٦٠٠ دراسة وورقة تتعلق بهذا النوع الجديد من الاختبارات . واستمرت هذه الحركة البحثية ولا تزال حتى وقتنا الحاضر.

وقد شجع على ذلك الحركات التربوية التي نادت بالمسؤولية التربوية Accountability ، والتقويم بالأهداف Evaluation by Objectives وعقود الأداء ، Mastery Learning ، والتعلم من أجل التمكن Performance Contracting ، والتعلم من أجل التميز Learning for Excellence ، وغيرها من التجديدات التربوية المعاصرة . غير أن كثيراً من علماء القياس والمريين أصبحوا الآن يؤكدون أهمية إحداث تكامل بين نوعي الاختبارات ، والإفادة منها في الأغراض المختلفة للتقويم التربوي والنفسى .

مفهوم التحصيل الدراسي وعلاقته بمفهوم الاستهداف :

يمكن تعريف التحصيل الدراسي Achievement بأنه درجة الاكتساب التي يحققها فرد أو مستوى النجاح الذي يحرزه أو يصل إليه في مادة دراسية أو مجال تعليمي أو تدريسي معين . فالاختبارات التي يطبقها المعلم على طلابه على مدار العام

الدراسي مثل اختبار اللغة العربية أو الكيمياء أو الحاسوب يفترض أنها تقيس التحصيل الدراسي أو الأكاديمي.

والهدف من تصميم هذه الاختبارات التحصيلية هو قياس مدى استيعاب الطلاب بعض المعرف والمفاهيم والمهارات المتعلقة بالمادة الدراسية في وقت معين أو في نهاية مدة تعليمية معينة.

أما الاستعداد Aptitude فيمكن تعريفه بأنه مدى قابلية الفرد Readiness للتعلم، أو مدى قدرته على اكتساب سلوك معين أو مهارة معينة إذا ما تهيأت له الظروف المناسبة. ومن هنا تبرز أهمية الاستعدادات لأنها تدل على إمكانية اكتساب أو تعلم الفرد سلوك معين. وربما يختلف هذا السلوك المتعلّم في درجة تعلقه، فقد يكون مهارة عقلية معقدة مثل تعلم اللغات الأجنبية والرياضيات المتقدمة، أو تعلم أنشطة حركية أو جسمية بسيطة. ولذلك فإن تعريف الاستعداد يتضمن القدرة على تعلم مهارات متعددة وسلوك متعدد. فالعبرة هنا بالقدرة على التعلم وليس نمط السلوك المتعلّم أو نوع المهارة المكتسبة.

فالتحصيل الدراسي إذن يدل على الوضع الراهن لأداء الفرد أو ما تعلمه أو اكتسبه بالفعل من معارف ومهارات في برنامج تعليمي معين، أما الاستعداد فيدل على الأداء المستقبلي المتوقع. ولكن إذا تأملنا اختبارات التحصيل واختبارات الاستعداد نلاحظ أنها جميعاً تقيس ما تعلمه الفرد. وبعض اختبارات الاستعداد تقيس مجالات أكاديمية، أي تقيس تذكرة المفاهيم والحقائق، وعمليات الاستباط وحل المشكلات وغيرها من العمليات المعرفية التي تقيسها اختبارات التحصيل. لذلك ربما لا يستطيع التمييز بين اختبارين أحدهما يقيس التحصيل والأخر يقيس الاستعداد الدراسي مثلاً بمجرد النظر إلى محتواهما. وأحياناً تستخدم اختبارات التحصيل في التنبؤ بالأداء المستقبلي نظراً لأن الأداء في وقت مضى أو الأداء الحالى يستخدم عادة في التنبؤ بدرجة جيدة بالأداء المستقبلي، وفي مثل هذه الحالة يمكن اعتبار اختبار التحصيل مقياساً للاستعداد.

ويبدو من ذلك أن هناك تشابهاً بين التحصيل والاستعداد والاختبارات التي تقيس كل منهما. إلا أنهما يختلفان في مدى اتساع وعمومية الخبرات المتعلمة، فالاستعداد يعتمد على الخبرة التعليمية العامة، أي يعكس التأثير التجمعي للخبرات المتعددة التي يكتسبها الفرد في سياق حياته اليومية، أما التحصيل فيعتمد على خبرات تعليمية محددة في أحد المجالات الدراسية أو التدريبية.

التمييز بين اختبارات التحصيل واختبارات الاستعدادات :

يمكن التمييز بين اختبارات التحصيل واختبارات الاستعدادات على أساس التعريفات السابقة. فيمكن اعتبار اختبار ما مقياساً للتحصيل إذا كان يقيس نواتج برنامج تعليمي أو تدريسي معين، واعتبار اختبار آخر مقياساً للاستعداد إذا كان يقيس خبرات تعليمية واسعة، أي يعكس تأثير خبرات الفرد في مجموعها التي اكتسبها في حياته اليومية بهدف التنبؤ بما يستطيع الفرد أن يتعلم في المستقبل.

لذلك يمكن القول بأن الاختبارات التحصيلية تقيس التعلم الذي يتم تحت شروط محددة بدرجة نسبية وفي ظروف يمكن التحكم فيها، مثل التعلم الذي يتم داخل الصد المدرسي أو في برنامج تدريسي معين، وأن التركيز يكون على الحاضر أو الماضي، أي ما تم تعلمه بالفعل.

أما اختبارات الاستعدادات مثل اختبارات الذكاء العام، وبطاريات الاستعدادات المتعددة، واختبارات الاستعدادات الخاصة، فهي تقيس تأثير التعلم تحت شروط يصعب التحكم فيها، وأن التركيز يكون على المستقبل، أو ما يمكن للفرد أداؤه فيما بعد إذا ما أتيحت له الفرصة المناسبة.

والتمييز الثاني بين اختبارات التحصيل واختبارات الاستعدادات يرتبط باستخدامات كل منها، فالاختبارات التحصيل تستعمل كما ذكرنا في تقويم الفرد عند انتهاءه من تعلم مادة دراسية معينة، أو موضوعاً دراسياً معيناً، أو التدرب على مهارة معينة، وهي تبين ما يستطيع الفرد أداؤه في وقت معين، أي أنها تركز على الحاضر أو الماضي وما تم تعلمه بالفعل. أما اختبارات الاستعدادات فتركز على المستقبل، أي تساعد في التنبؤ بالأداء اللاحق، أو ما يمكن للفرد أداؤه مستقبلاً إذا ما أتيحت له الظروف المناسبة.

غير أن هذا التمييز غير قاطع، فقد يعتمد اختبار الاستعداد على ما تعلمه الفرد تحت شروط منتظمة إلى حد ما، وقد يشتمل اختبار التحصيل على خبرات تعليمية أو تدريبية شاملة وغير منتظمة، كما قد يستعمل كوسيلة تنبؤية للتعلم اللاحق. وفي مثل هذه الحالات ربما يستخدم اختبار التحصيل لنفس الغرض الذي يستخدم فيه اختبار الاستعداد.

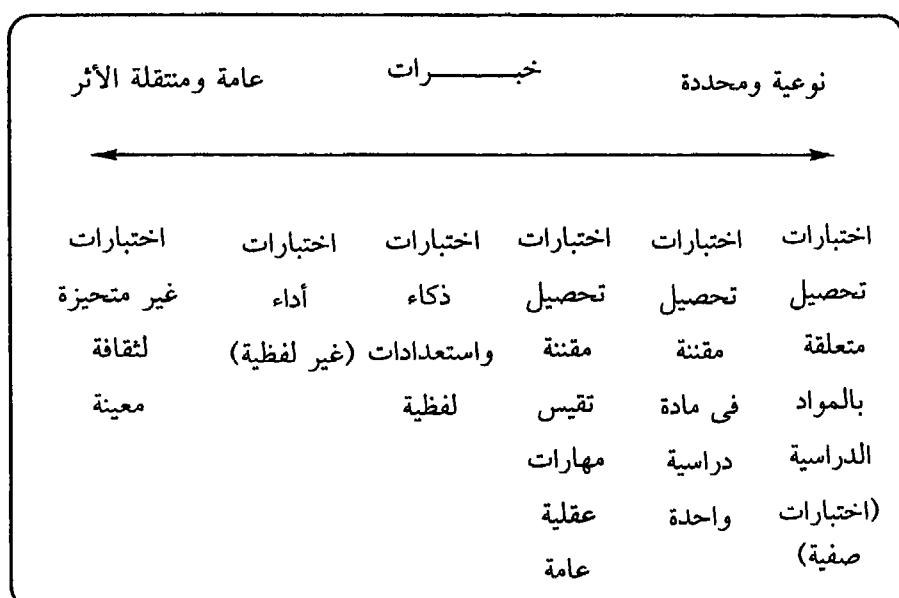
فمثلاً ربما نستخدم درجة التلميذ في اختبار تحصيلي في الحساب للتنبؤ بنجاحه اللاحق في الجبر، أو ربما نستخدم اختبارات التحصيل في بعض المواد العلمية مثل الكيمياء والفيزياء والتاريخ الطبيعي للتنبؤ بأدائهم في كليات الاقتصاد والعلوم السياسية، وهكذا.

وعند التمييز بين اختبارات التحصيل واختبارات الاستعدادات يجب أن يكون واضحاً أن جميع الاختبارات النفسية تقيس الواقع الراهن لسلوك الفرد، وتأثير درجة الفرد في الاختبار بالخبرات السابقة لا يعني بالطبع أنه لا يكون لها أثر لاحق أو مستقبلي، فـ، أدائه.

وقد بين Cronbach Anastasi أن علماء النفس اتجهوا إلى استخدام مفهوم جديد يحل محل مفهوم التحصيل والاستعداد هو مفهوم «القدرات النامية أو المستمرة» *Developing Abilities*.

فجميع اختبارات القدرات سواء كانت اختبارات ذكاء عام أو بطاريات استعدادات متعددة أو اختبارات استعدادات خاصة أو اختبارات تحصيل تقيس جميعها مستوى النمو الذي وصل إليه الفرد أو حققه في قدرة واحدة أو أكثر. أي أنه يمكن اعتبار أن الاستعداد والتحصيل يقعان على متصل واحد - Achievement Continuum Aptitude ، ويتحدد موقع كل منهما على هذا المتصل وفقاً لنوعية وشمول الخبرات التي يفتقر إليها.

ويوضح شكل (٨ - ١) التالى ترتيب القدرات النامية على هذا المتصل:



ويتبين من شكل (٨ - ١) أن اختبارات التحصيل، المتعلقة بالمواد الدراسية وهي

الاختبارات الصفيّة التي يكتبها المعلم لطلابه لقياس معارف ومهارات محددة تحديداً ضيقاً مثل اختبار قياس معانٍ الكلمات في اللغات أو اختبار قياس المهارات الحسابية تقع في أقصى الطرف الأيمن.

ويلي ذلك الاختبارات التحصيلية المقننة التي تستخدم في قياس مدى تحقق أهداف تعليمية عريضة ومحتوها يكون متسعًا. وبعض هذه الاختبارات يتعلق بمادة دراسية واحدة، والبعض الآخر يقيس مهارات عقلية عامة مثل التفكير المنطقى وحل المشكلات، وهذه الاختبارات الأخيرة تشبه اختبارات الذكاء والاستعدادات. أما اختبارات الذكاء والاستعدادات اللغوـية مثل اختبار ستانفورد بينيه Binet أو بطارية اختبارات الاستعدادات الفارقة (DAT) فإنها على الرغم من كونها مشابهة لاختبارات التحصيل ، إلا أنها أقل ارتباطاً بممواد دراسية معينة. ويلي ذلك الاختبارات غير اللغوـية التي تتطلب أداء عملياً، ثم الاختبارات غير المتحيزـة لثقافـة معينة Culture Fair Tests والتي تستخدم لمقارنة أداء الأفراد من ذوي أطر ثقافية متباعدة . وهذه الأنواع الأخيرة من الاختبارات لا ترتبط بممواد دراسية معينة ، وإنما تتميز بالعمومية وإمكانية انتقال أثرها لمواقف حياتية مختلفة، ولذلك تقع في أقصى الطرف الأيسر لمتصـل الجوانـب المعرفـية .

وقد بيـنت الـدراسـات الإـمـبرـيقـية اـرـتـباطـ الاـختـبارـاتـ المـتـجـاـورـةـ عـلـىـ المـتـصـلـ اـرـتـباطـاـ مرتفـعاـ، ويـقـلـ الـارـتـباطـ كـلـماـ اـبـتـعدـتـ عـنـ بـعـضـهاـ عـلـىـ المـتـصـلـ نـفـسـهـ . فـمـثـلاـ وـجـدـ أـنـ قـيمـ مـعـاـلـمـ الـارـتـباطـ بـيـنـ بـعـضـ اـخـتـبارـاتـ التـحـصـيلـ وـالـذـكـاءـ مـرـتفـعـةـ بـدرـجـةـ كـبـيرـةـ بـعـدـ أـخـذـ أـخـطـاءـ الـقـيـاسـ فـيـ الـاعـتـارـ .

افتراضات أساسية يرتکن إليها قياس التحصيل :

تصمـمـ اـخـتـبارـاتـ التـحـصـيلـ كـمـ سـبـقـ أنـ ذـكـرـناـ لـقـيـاسـ الـمعـارـفـ وـالـمـهـارـاتـ التـىـ اـكـتـسـبـهاـ الـطـلـابـ فـيـ الـمـجـالـاتـ الـدـرـاسـيـةـ الـمـخـتـلـفـةـ، وـكـذـلـكـ لـقـيـاسـ مـسـتـوـيـاتـ الـمـهـارـاتـ فـيـ الـمـجـالـاتـ الـتـدـرـيـسـيـةـ لـلـعـامـلـيـنـ فـيـ الـمـصـانـعـ وـالـمـؤـسـسـاتـ الـحـكـومـيـةـ وـالـمـجـالـاتـ الـعـسـكـرـيـةـ وـغـيرـهـ . وـيـتـرـكـزـ الـاـهـتـمـامـ فـيـ مـشـلـ هـذـهـ الـحـالـاتـ عـلـىـ قـيـاسـ مـكـتـسـبـاتـ الـأـفـرـادـ فـيـ نـهـاـيـةـ مـدـةـ دـرـاسـيـةـ مـعـيـنـةـ يـقـدـمـ خـلـالـهـ مـجـمـوعـةـ مـنـ الـخـبـرـاتـ الـمـنـظـمـةـ .

لـذـكـ يـرـتـکـنـ قـيـاسـ التـحـصـيلـ عـلـىـ بـعـضـ الـاـفـرـاضـ الـاـسـاسـيـةـ لـكـىـ تـكـونـ نـتـائـجـ الـقـيـاسـ مـتـسـقـةـ وـصـادـقـةـ وـغـيرـ مـتـحـيـزـ وـهـىـ كـالـتـالـىـ :

- (1) أن يقيـسـ الاـخـتـبارـ نـطـاقـاـ سـلـوكـيـاـ يـمـكـنـ تـحـديـدـهـ بـدـقـةـ . فالـاخـتـبارـ التـحـصـيلـيـ سـوـاءـ أـعـدـهـ الـمـعـلـمـ لـطـلـابـهـ أـوـ أـعـدـهـ جـهـاتـ أـخـرىـ يـنـبـغـىـ أـنـ يـسـتـنـدـ إـلـىـ أـهـدـافـ

تربيوية يمكن صياغتها صياغة سلوكية محددة بحيث يمكن قياسها. وعلى الرغم من أن الاختبارات التحصيلية المقننة يكون محتواها أكثر اتساعاً من الاختبارات التي يعدها المعلم لطلابه إلا أنه ينبغي أن يكون هناك اتفاق على ما تقيسه من أهداف تشتراك فيها مجموعة من المدارس أو المناطق التي سوف تستخدم الاختبار.

(٢) أن يقيس الاختبار الأهداف المتعلقة بالمادة الدراسية أو محتوى البرنامج التدريسي، ولا تقىس أهدافاً عارضة أو غير مهمة. وهذا يعني أن يكون اختبار التحصيل صادقاً في محتواه. فإذا صمم اختبار لقياس مهارة الطالب في رسم الأشكال الهندسية مثلاً يجب أن يتضمن مفردات اختبارية تقيس هذه المهارة، ولا تقىس المهارة اللغوية.

(٣) أن تكون المعالجة التعليمية مناسبة للأفراد الذين سيطبق عليهم الاختبار بحيث تراعي الفروق الفردية بينهم، وأن تناح لهم الفرصة لتعلم المجال الذي يقيسه الاختبار. وعلى الرغم من إمكانية توفير هذه الظروف في الصفوف المدرسية التي يخترعها المعلم باختبارات من إعداده، إلا أن عدم تكافؤ هذه الفرص والخبرات التعليمية تعد مشكلة في حالة اختبارات التحصيل المقننة نظراً لتباين المحتوى التعليمي الذي تقيسه هذه الاختبارات من مدرسة إلى أخرى، أو من مؤسسة عمل إلى أخرى، واختلاف خبرات الأفراد المختبرين عن خبرات الجماعة المرجعية التي أعدت معايير الاختبار استناداً إليها.

تصنيف اختبارات التحصيل واستخداماتها :

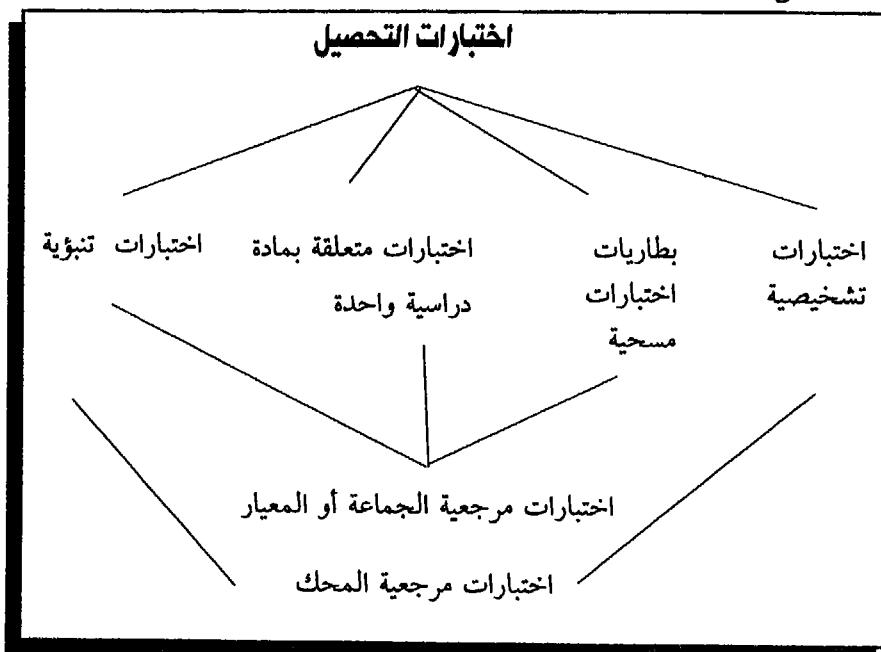
تعدد أنواع اختبارات التحصيل وتتنوع استخداماتها في المجالات التربوية والتدرستية. وعلى الرغم من إمكانية تصنيف هذه الاختبارات بطرق مختلفة، إلا أن التصنيف التالي يتضمن بعض الأبعاد الرئيسية التي يمكن في ضوئها التمييز بين الأنواع المختلفة من هذه الاختبارات:

(١) **محتوى الاختبار**: على الرغم من أن معظم اختبارات التحصيل تقيس عينة من محتوى تعليمي أو تدريسي معين، إلا أن هذه الاختبارات تختلف في مدى ضيق أو اتساع المحتوى، ومدى تمثيل عينة مفرداته لهذا المحتوى. فاختبارات التحصيل التي يعدها المعلم لطلابه تكون محدودة في محتواها حيث تقيس عادة وحدة تعليمية معينة في إطار منهج دراسي محدد. إذ يمكن أن يقيس مثل هذا الاختبار معرفة الطالب الصيغ الرياضية أو الرموز

الكيميائية ، أو يمكن أن يقيس مهارات الطلاب في القراءة أو النجبر ، وهكذا.

أما الاختبارات التحصيلية المقننة فإن محتواها يكون أكثر اتساعاً ويشتمل على معلومات أساسية تتعلق بمادة دراسية واحدة لإحدى الفرق بغض النظر عن المنطقة التي تنتهي إليها المدرسة . كما أن بعض هذه الاختبارات يكون على شكل بطاريات Battery of Tests تقيس محتوى مواد دراسية مختلفة وليس مادة واحدة . وفي مثل هذه الحالة يصنف الاختبار على أساس الفرقة أو المرحلة الدراسية ، أي بطارية اختبار تحصيلي للمرحلة الأولى أو المرحلة الإعدادية أو الثانوية .

(٢) الغرض من استخدام الاختبار، تختلف اختبارات التحصيل تبعاً للأغراض التي تستخدم فيها . ويوضح الشكل التخطيطي (٨-٢) التالي الأنواع الرئيسية من هذه الاختبارات :



شكل (٨ - ٢) يوضح الأنواع الرئيسية من اختبارات التحصيل
(من تصميم المؤلف)

ويتضح من شكل (٨ - ٢) أن هناك أربعة أنواع رئيسية من اختبارات التحصيل هى: الاختبارات التشخيصية Diagnostic Tests ، وبطاريات الاختبارات المسحية

Survey Test Batteries ، والاختبارات المسحية المتعلقة بمحتوى مادة دراسية واحدة . **Prognostic Tests** ، والاختبارات التنبؤية **Single Survey Tests**

الاختبارات التشخيصية: تضم هذه الاختبارات لتشخيص صعوبات تعلم الطالب في مادة دراسية معينة، أي تهدف لتحديد جوانب القوة والضعف في تحصيل الطالب للمعارف والمهارات المتعلقة بهذه المادة.

ويتطلب بناء اختبار تشخيصي في المهارات الأساسية في القراءة أو الهجاء أو الحساب تحليل هذه المهارات إلى مكوناتها الفرعية، وقياس كل من هذه المكونات. لذلك فإن هذا النوع من الاختبارات يقيس التحصيل بدرجة أكثر تفصيلاً وعمقاً، وينقسم الاختبار عادة إلى عدد من الاختبارات الفرعية، وتقدّر درجة لكل طالب في كل اختبار فرعي. فالاختبار التشخيصي في القراءة مثلاً يستغرق وقتاً طويلاً نسبياً في تطبيقه، ويشتمل على عدد كبير من المفردات المتنوعة. إذ ربما يتكون من عدد من الاختبارات الفرعية أحدها يقيس تعرف الكلمات، وأخر يقيس فهم معانى الكلمات، وثالث يقيس سرعة القراءة، ورابع يقيس فهم الموضوع المقرروء، وغيرها، بل ربما يشتمل على أجهزة تقيس حركة العين، ومدى الانتباه، ووسع الذاكرة بواسطة عرض صور لمدة ثانية واحدة واحتفائها، وتآثر اليد والعين، وهكذا. وتفيد هذه الاختبارات في التشخيص من أجل تقديم تعليم علاجي، لذلك تكون عادة من نوع الاختبارات . **Criterion - Referenced Tests** مرجعية المحك

كما تفيد هذه الاختبارات المرشد النفسي والأخصائى الكلينيكى فى تحديد صعوبات التعلم لدى بعض الأطفال، وتعرف أسبابها، واقتراح أساليب التدريب أو التعليم العلاجي المناسب. فالطفل البطئ في القراءة يحتاج إلى تدريب مناسب لزيادة سرعته، والطفل الذي يسلك في الصف المدرسي سلوكاً غير مرغوب فيه بسبب ضعف تحصيله أو تكرار رسوبه، والطفل المتفوق في التحصيل ربما يعاني أيضاً من الملل مما يؤدي به إلى سوء التوافق وربما الهروب من المدرسة، فكلاهما يحتاج إلى تعديل سلوكه بأساليب علاجية منظمة استناداً إلى نتائج الاختبارات التشخيصية.

بطاريات الاختبارات المسحية: تتكون كل منها من مجموعة من الاختبارات المسحية يتعلق كل منها بمادة دراسية مختلفة لإحدى فرق مرحلة تعليمية معينة. فمثلاً بطارية الاختبارات المسحية لإحدى فرق المرحلة الابتدائية ربما تشتمل على اختبارات في الحساب، والقراءة، والدراسات الاجتماعية، والعلوم، وربما الهجاء.

ويطارية الاختبارات المسحية للمرحلة الثانوية ربما تشتمل على اختبارات في اللغة العربية، والعلوم، والمواد الاجتماعية. وهذه البطاريات عبارة عن مجموعة متناسقة من اختبارات التحصيل المسحية التي يمكن تطبيق كل منها على حدة،

وُتُستخدم في تحديد المستوى العام لأداء الطالب في المواد الدراسية ومقارنته أدائه بأداء أقرانه ، لذلك تعدد من نوع الاختبارات مرجعية الجماعة أو المعايير *Norm-referenced Tests* . ويمكن أيضًا تمثيل درجات الطالب في صفة تربوية *Referenced Profile* للمقارنة بين أدائه في مختلف المواد الدراسية التي تقيسها بطارية الاختبارات بعد تحويل الدرجات الخام إلى درجات معيارية ذات تدريج موحدة على الرغم من اختلاف عدد مفردات كل اختبار تشمل عليه البطارئ كما سبق أن أوضحنا في الفصل السادس . ومثال ذلك بطارية كاليفورنيا للاختبارات المساعدة *California Achievement Tests* (CAT) ، وبطارية متابعة التقدم التربوي للطلاب من الصف الرابع حتى نهاية مرحلة التعليم الثانوى (STEP) ، وبطارية متابعة التعلم *Stanford Sequential Tests of Educational Progress* (STEP) ، وبطارية ستانفورد لاختبارات التحصيل لجميع صفوف التعليم العام *Stanford Achievement Tests Battery* ، وبطارية ستانفورد للمهارات الأكاديمية *Tests of Academic Skills (TASK)* ، وتشتمل على مستويين أحدهما للمرحلة المتوسطة والآخر للمرحلة الثانوية ، وتقيس مهارات القراءة ، وفهم المقصود ، واللغة الإنجليزية ، والرياضيات .

وتختلف بطاريات اختبار التحصيل المصححة عن الاختبارات التشخيصية في أن محتواها يكون متسعًا ، ويشتمل كل اختبار على عينة محددة من المحتوى والمهارات في المواد الدراسية المختلفة ، لذلك يصعب باستخدامها تشخيص جوانب القوة والضعف في تحصيل الطلاب .

ولعل تطبيق بطارية الاختبارات المصححة لتحديد المستوى العام لأداء الطالب بعد الخطوة الأولى المنطقية ، حيث يليها تطبيق الاختبارات التشخيصية المتعلقة بالمجال أو المهارة التي يفتقر إليها الطالب لكي يتضمن التحديد الدقيق لجوانب القصور في هذا المجال أو هذه المهارة واقتراح العلاج المناسب .

الاختبارات المصححة المتعلقة بمادة دراسية واحدة : وتحدف هذه الاختبارات المصححة إلى قياس التحصيل في مادة دراسية واحدة . فعلى الرغم من إمكانية تطبيق أحد اختبارات بطاريات الاختبارات المصححة التي أشرنا إليها بمفردها ، إلا أنه يفضل استخدام الاختبارات المتعلقة بمادة دراسية معينة ؛ وذلك لأن عينة محتواها تكون أكثر تفصيلاً لهذه المادة ، وعدد مفردات الاختبار يكون أكبر مما يسمح بتقويم التحصيل بدرجة أكثر تفصيلاً . غير أن هذه الاختبارات تزودنا بدرجة كلية واحدة فقط بدلاً من عدة درجات كما في بطاريات الاختبارات المصححة ، وبذلك لا تساعد المعلم في تعرف مجال ضعف الطالب في المادة الدراسية .

الاختبارات التنبؤية : على الرغم من أن اختبارات التحصيل تقيس الوضع الراهن لأداء الطالب في مادة دراسية معينة ، إلا أنه يمكن استخدام هذه الاختبارات في التنبؤ بالتحصيل اللاحق في هذه المادة، وبذلك تؤدي وظيفة اختبارات الاستعدادات . فإذا استخدمنا اختباراً تحصيلياً في فهم المقرئ للتنبؤ ب مدى إفادة الطالب من عملية تعليم اللغة العربية أو غيرها من المواد التي تتطلب ذلك الفهم ، فإن هذا الاختبار يعد بمثابة اختبار تنبؤي .

ويمكن أيضاً أن نستخدم اختبارات التحصيل في انتقاء الطلاب للالتحاق بالكلليات ، وكذلك انتقاء العاملين في المجالات المختلفة . غير أن هذا الاستخدام ينبغي أن يبني على أساس منطقية اعتماداً على ارتباط درجاتها بدرجات محك مستقبلى مثل النجاح في الكلية ، أو في مهنة معينة . ولذلك يكون التنبؤ باستخدام هذه الاختبارات جيداً ينبغي أن تقيس أنشطة مماثلة لمتطلبات دراسة أو مهنة معينة .

درجات اختبارات التحصيل لطلاب نهاية مرحلة التعليم الثانوى ربما نستطيع باستخدامها التنبؤ بدرجات الطلاب في السنة الأولى الجامعية بدرجة أفضل من درجات اختبارات الذكاء . وبالطبع ربما تزداد القيمة التنبؤية إذا استخدمنا نوعاً من الاختبارات معًا .

كما نستطيع باستخدام بطاريات الاختبارات التحصيلية انتقاء أفضل المتقدمين لشغل وظائف معينة تتطلب معارف ومهارات يمكن قياسها بهذه الاختبارات . وكذلك يمكن استخدامها في مجالات التوجيه والتدريب المهني لاتخاذ قرارات تتعلق بالالتحاق بدراسة معينة أو برنامج تدريسي معين ، ولكن هذا يتطلب التحقق من الصدق التنبؤى للاختبار التحصيلي المستخدم .

(٣) طريقة الاستجابة : يمكن أيضاً تصنيف الاختبارات التحصيلية تبعاً لطريقة الاستجابة المرجوة ، فكثيراً من هذه الاختبارات تعتمد على الورقة والقلم ، وتنقسم إلى اختبارات تعتمد على التعرف Recognition ، وأخرى تعتمد على الاسترجاع Recall . فاختبارات التعرف تشتمل عادة على مفردات اختيار من متعدد ، أو صواب أو خطأ ، أو مزاوجة حيث تتطلب الاختيار من بين بدائل معطاة ، أو تحديد ما إذا كانت عبارة معينة صحيحة أو خطأ . ويستخدم هذا النوع من المفردات بكثرة في اختبارات التحصيل المقننة ، كما يستخدمه المعلم في الاختبارات التي يُعدها بنفسه لطلابه . أما اختبارات الاسترجاع فتشتمل على أسئلة تتطلب تكملة أو إجابة قصيرة أو أسئلة مقال . وفي هذه الحالة يقدم السؤال ويكتب الطالب الإجابة التي يراها مناسبة .

كما أن بعض اختبارات التحصيل تقتصر على قياس معرفة الحقائق والمبادئ المتعلقة بمادة دراسية معينة دون الاهتمام الكافى بقياس قدرة الطالب على استخدام

وتوظيف هذه المعرفة في مواقف تتطلب التفكير المثير وحل المشكلات. وربما يرجع ذلك إلى صعوبة بناء اختبارات تشتمل على مفردات اختيار من متعدد لقياس هذه المستويات المعرفية العليا. غير أنه إذا علمنا أن المفردات الاختبارية التي تقيس هذه المستويات تسهم في انتقال أثر التعلم المدرسي إلى مواقف أخرى متنوعة، فإن ذلك يشجع على الاهتمام بقياس تحققها لدى الطلاب.

كما أن اختبارات التحصيل يمكن أيضًا أن تقيس المهارات العملية. فالاختبارات العملية أو اختبارات الأداء تقيس مهارة الطالب في تطبيق ما تعلمه بأن يقوم بتشغيل جهاز، أو يجري تجربة، أو يعزف على آلة موسيقية، أو يؤدي تمارين رياضية معينة، وهكذا.

(٤) طريقة تصميم وبناء الاختبار: سبق أن أوضحنا في الفصل الأول أن أدوات القياس يمكن تصنيفها إلى أدوات يُعدها المعلم وأخرى مقتنة، ونؤكد هنا أن هذا التصنيف ينطبق على اختبارات التحصيل. فالاختبارات التحصيلية الصافية يقوم المعلم أو مجموعة من المعلمين في المدرسة بإعدادها لطلابهم لقياس الأهداف التعليمية للمواد الدراسية، وبذلك يكون محتواها مرتبطة ارتباطاً مباشرًا بما تم تدريسه في المدرسة.

أما اختبارات التحصيل المقتنة فيقوم بإعدادها خبراء القياس ومؤسسات نشر الاختبارات، وتُستخدم في مدارس مختلفة نظرًا لأنها تقيس أهدافاً تعليمية عريضة، ومحتها يكون متسعاً. فهي تقيس معارف اتفق خبراء التربية على ضرورتها لمعظم الطلاب في مادة دراسية لفرقة معينة، ويتم تجريب الاختبار على عينة عشوائية كبيرة نسبيًا من الطلاب بحيث تكون خصائص هذه العينة محددة تحديدًا دقيقًا، وتحسب معايير الاختبار لاستخدامها في مقارنة أداء الطالب بأقرانه.

(٥) تفسير درجات الاختبار : أوضحنا في الفصل السادس أن درجات الاختبار يمكن تفسيرها في ضوء معيار نسبي أو محك أداء مطلق. فاختبارات التحصيل التي تهتم بمقارنة أداء الطالب بمعيار جماعة مرجعية لتحديد مركزه النسبي أو ترتيبه بين أقرانه تعد من نوع الاختبارات المرجعية الجماعية أو المعيار Norm - Referenced Tests أما اختبارات التحصيل التي تهتم بتحديد مدى تمكن الطالب من مهارة محددة في إطار محتوى دراسي معين، فإن درجته في الاختبار يتم مقارنتها بمحك أو مستوى أداء محدد مسبقاً دون مقارنة درجته بمتوسط أداء جماعة مرجعية معينة ، وتعد هذه الاختبارات من

نوع الاختبارات مرجعية المحك Criterion - Referenced Tests ، ويتميز هذا النوع من الاختبارات بأن له قيمة تشخيصية لتحديد جوانب القوة والضعف في تحصيل الطالب؛ لذلك فإنها تكون أقرب إلى الاختبارات التي يُعدها المعلم لطلابه.

وسوف نناقش فيما يلى كيفية بناء كل من هذين النوعين من الاختبارات نظراً لأهميتهما في قياس وتقدير التحصيل الدراسي أو الأكاديمي، والتحقق من فاعلية البرامج التعليمية والتدريبية.

أولاً: اختبارات التحصيل مرجعية المحك :

يُستخدم هذا النوع من الاختبارات في كثير من الأغراض التربوية التي تتعلق بتقدير كفايات الأفراد وما يمتلكون من معارف ومهارات وظيفية مكتسبة من برامج تعليمية أو تدريبية محددة الأهداف والنواتج.

وتحتاج هذه الاختبارات بأنها تصف السلوك وصفاً دقيقاً بما يمكن من تشخيص جوانب القوة والضعف واقتراح أساليب العلاج. لذلك انتشر استخدام الاختبارات مرجعية المحك في المؤسسات التربوية والتدريبية في الولايات المتحدة الأمريكية وغيرها من دول العالم.

فهناك اهتمام متزايد في هذه الدول بتطوير برامج التعليم والتقويم الذي يستند إلى الكفايات Competencies والمودولات التعليمية Instructional Modules ، وبرامج التعليم بالاستعانة بالحاسوب Computer - Assisted Instruction حيث يحصل الطالب على قدر كبير من التعليم من خلال التفاعل المتبادل مع الحاسوب. فكل طالب يتعلم وفقاً لسرعته الذاتية وإمكاناته في برنامج تعليم فردي. وعقب انتهاء من تعلم وحدة صغيرة أو مهارة معينة يجري البرنامج اختباراً مرجعياً المحك لتحديد مدى تقدم الطالب. فإذا حقق المستوى المطلوب يمكن أن ينتقل إلى الوحدة الصغيرة أو المهمة التالية، أما إذا لم يحقق هذا المستوى يمكن أن يعيد تعلم الوحدة مرة أخرى حتى يصل إلى المستوى المطلوب مما يجعله يُوظف أقصى إمكاناته في التعلم.

كما أن المعلم يمكنه الإفادة من طرق بناء هذا النوع من الاختبارات في تطوير الاختبارات الصفيحة التي يكتبها لطلابه مما يزيد من فاعليتها في تقدير كفاياتهم ومكتسباتهم وتشخيص الصعوبات التي يواجهها غالبيتهم. ويفيد تطبيق هذه الاختبارات

بصورة دورية في متابعة التقدم الدراسي للطلاب وحفظهم على التعلم المستمر، كما يفيد في تقويم فاعلية برنامج تعليمي معين استناداً إلى البيانات المجتمعة من التطبيق الدورى لهذه الاختبارات مما يقدم مؤشراً لنمو الطالب ومدى تحقيقهم للأهداف التعليمية المتعلقة بهذا المحتوى الدراسي.

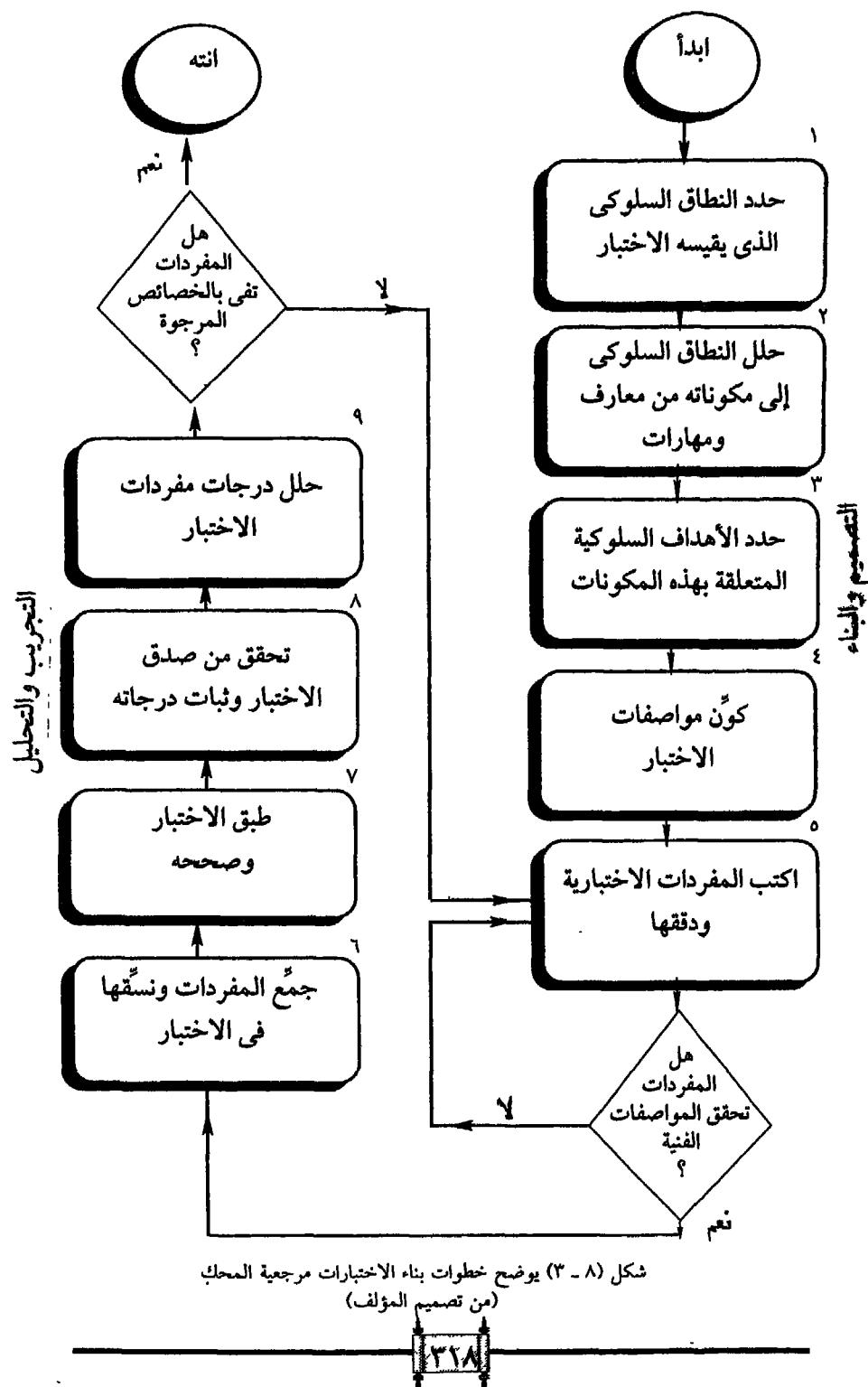
ونظراً لأهمية هذا النوع من الاختبارات فقد قام مؤلف هذا الكتاب منذ عام ١٩٨٤ ببناء مجموعات من هذه الاختبارات في مجالات متعددة، كما أعد مرجعاً متخصصاً للإختبارات التشخيصية مرجعية المحك (علام ، ١٩٩٥) وأدلة معلم لبناء هذه الاختبارات (علام ، ١٩٩٥ ، ١٩٨٣ ، ١٩٩٧)، كما قام تلاميذه بإعداد عدد آخر من هذه الاختبارات في مواد دراسية مختلفة في رسائلهم للماجستير والدكتوراه.

لذلك سنتناقش بإيجاز كيفية بناء هذه الاختبارات وتفسير نتائجها، وعلى القارئ المهتم الرجوع إلى المراجع السابقة وغيرها من المراجع المتخصصة للاستزادة في هذا المجال.

وكذلك يمكن الرجوع إلى الكتاب المصدر الذي أعده مركز دراسات التقويم التابع لجامعة كاليفورنيا (CSE، 1989) المتعلق بالإختبارات مرجعية المحك الذي يشير إلى مجموعة متنوعة من هذه الاختبارات في المجالات المختلفة التي تم نشرها، كما يشير إلى تعليق المركز على كل اختبار منها.

خطوات بناء الإختبارات مرجعية المحك :

يتطلب بناء الاختبارات مرجعية المحك إعداد خطة تفصيلية دقيقة قبل البدء الفعلى في كتابة المفردات أو الأسئلة التي يشتمل عليها الاختبار. ويوضع الشكل التخطيطي (٨ - ٣) التالي خطوات بناء هذه الاختبارات.



وينتضح من شكل (٨ - ٣) أن بناء الاختبارات مرجعية المحك يمر بخطوات متعددة سوف نناقشها فيما يلى :

الخطوة الأولى : تحديد النطاق السلوكي الذي يقيسه الاختبار وتحليله إلى مكوناته من معارف ومهارات :

فالتعريف الدقيق للنطاق السلوكي Behavioral Domain الذي يقيسه الاختبار مرجعي المحك يعد متطلباً أساسياً من متطلبات هذا النوع من الاختبارات، فالنطاق السلوكي يعد بمثابة الإطار المرجعي الذي تنساب إليه درجة الفرد في الاختبار. ونقصد بالنطاق السلوكي المعرف والمهارات المحددة تحديداً دقيقاً بما يمكننا من معرفة ما يستطيع الفرد أداءه وما لا يستطيع أداءه. ويدرك بابام Popham أن التحديد الدقيق للنطاق السلوكي يعد الركيزة الأساس لمفهوم المحك Criterion.

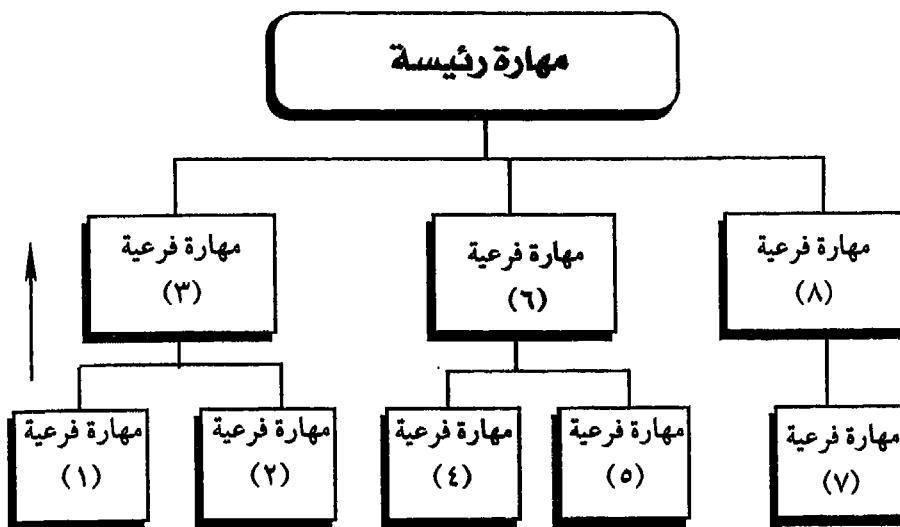
وبالطبع تختلف النطاقات السلوكية اختلافاً ملحوظاً في درجة تعقدتها ونوعية الأداء الذي تتضمنه ومتطلبات النمو أو التعلم .

وكلما زاد تعقد النطاق السلوكي تطلب تحليله إلى مكونات يمكن قياسها جهداً ملحوظاً. فنطاق مهارة فهم المفروع ربما يكون أكثر تعقداً من نطاق مهارة رسم الخرائط ، وهذا بدوره ربما يكون أصعب من نطاق معرفة جمع الأعداد البسيطة. ويفضل في مثل هذه الحالات الاستعانة بخبراء المادة الدراسية أو اختصاصي تصميم البرامج التدريبية في تحديد النطاق السلوكي المراد قياسه ، وتحليله إلى مكوناته وما تشمل عليه من معارف ومهارات وسلوكيات .

وهناك طرق متعددة لإجراء هذا التحليل لعل أهمها طريقة تحليل بنية المحتوى Structural Analysis التي اقترحها سكاندورا Scandura, 1977) وتحليل المهام Task Analysis الذي اقترحه جاجنى ويريجز (Gagne & Briggs, 1974).

ويهدف هذا التحليل لتحديد المتطلبات أو المكونات السلوكية Prerequisite Behavior اللارم توافرها لدى المتعلم لكي يتحقق الكفاية أو المهارة الرئيسة التي يقيسها الاختبار ، ونكرر عملية تحليل هذه المتطلبات حتى نصل إلى السلوك المدخلى Entery Behavior للتعلم الذي اكتسبه بالفعل . ثم تُرتب هذه المتطلبات أى المعرف والمهارات المساعدة ترتيباً هرمياً بنائياً بحسب إسهاماتها في تكوين المهارة الرئيسة . ويمثل السلوك المدخلى قاعدة البناء الهرمى ثم تدرج مستويات السلوك من الأبسط إلى الأكثر صعوبة حتى نصل إلى قمة البناء الذى يمثل الكفاية أو المهارة الرئيسة .

ويوضح شكل (٨ - ٤) التالى تخطيطاً لنواتج مثل هذا التحليل:



شكل (٨ - ٤) يوضح التخطيط الهرمي البنائى لمهارة رئيسة افتراضية

وإجراء هذا التحليل يتطلب معرفة أنماط التعلم الرئيسية ومستوياتها المتدرجة ونواتجها لكي يكون التحليل منطقياً ومتكاملاً (علام، ١٩٩٥).

الخطوة الثانية: صياغة الأهداف السلوكية المتعلقة بنواتج تحليل النطاق السلوكي؛

فالمهارات الفرعية التي نتجت عن تحليل النطاق السلوكي ينبغي صياغتها صياغة سلوكية بحيث يمكن قياسها بمفردات اختبارية. وينبغي أن تشتمل عبارات الأهداف السلوكية على وصف السلوك المتوقع بأفعال تشير إلى نشاط معين يمكن ملاحظته، مثل يصنف، يميز، يقترح، يستخرج، ينسق، يخطط، يبرر، يدقق، يتعرف، وهكذا. ويراعى أن يتاسب الفعل مع نمط التعلم المطلوب ومستوى المعرفى.

كما يجب أن تشير هذه العبارات إلى المحتوى المرجعى الذى سوف يتعامل معه الفرد أثناء اختباره، مثل محتوى الصيغة الجبرية، أو المعارك الحربية، أو برامج الحاسوب، أو الصيغ الكيميائية، أو خدمات الإنترنت، أو دوائر التلفاز، وما شابه

ذلك. ويمكن أيضاً أن تُحدد الشروط والإجراءات التي يتم تقييم الأداء في ضوئها، أي تحديد ما إذا كان تحقيق الهدف يتطلب سرعة أو دقة معينة، أو استجابة تحريرية أو شفهية أو حركية، وغير ذلك من الشروط التي تتطلبها طبيعة الهدف السلوكي.

الخطوة الثالثة: تكوين المواصفات التفصيلية للاختبار

يختلف الاختبار مرجعى المحك عن غيره من الاختبارات في تحديد مواصفاته تحديداً تفصيلاً دقيقاً.

ف نوع مفردات الاختبار وطبيعة التشيرات التي تشتمل عليها وما تتضمنه من محتوى، وخصائص الاستجابات وتعليماتها، وطريقة ومحكمات تقدير درجاتها ينبغي أن تحدد تحديداً دقيقاً. والهدف من هذه التفاصيل المطلوبة تحديد النطاق السلوكي للاختبار بحيث يمكن الاسترشاد به في بناء مفردات متتجانسة تقيس جميعها هدفاً سلوكيًّا محدداً وارتباطها به ارتباطاً مباشراً بما يجعل هناك تزامن بين الهدف السلوكي والمفردة الاختبارية التي تقيسه، ويضمن صدق محتواها وصدق محتوى الاختبار ككل. وربما يتطلب ذلك أيضاً التحقق من صدق محتوى المواصفات التفصيلية للاختبار بواسطة مجموعة من خبراء المجال الدراسي أو التدريسي المعين.

ويرجع الاهتمام بهذه المواصفات التفصيلية إلى أنه يُستند إليها في كتابة مفردات متتجانسة وبناء اختبارات متوازية أو متكافئة على الرغم من اختلاف الاختصاصيين الذين يقومون بإعداد هذه الاختبارات.

الخطوة الرابعة: كتابة أو تكوين مفردات الاختبار

فمفردات الاختبار مرجعى المحك تتبثق مباشرة من مواصفات الاختبار، فإذا كانت هذه المواصفات مبرمجة فإنه يمكن استخدام الحاسوب في تكوين مفردات الاختبار Computer Items Generation ، وهناك بعض الأساليب التي يمكن الاسترشاد بها في إجراء هذه البرمجة مثل صيغ المفردات Item Forms ، وقواعد التحويلات السيمانتية Semantic Transformations ، والحمل الراسمة Mapping Sentences ، وغيرها من الأساليب التي ناقشها المؤلف في مراجع أخرى (علام ١٩٨٦ ، ١٩٩٥) أما إذا استرشد بمواصفات تفصيلية دون محاولة برمجتها ، فإنه ينبغي كتابة المفردات Item Writing بواسطة القائمين ببناء الاختبار .



وعندئذ ينبغي تصميم خطة معاينات المفردات لانقاء نوع وعدد المفردات التي تمثل النطاق السلوكي تمثيلاً كافياً مع مراعاة محتوى النطاق ومكونات المهارة التي يتضمنها. فإذا كان النطاق السلوكي يقيس مهارة فهم المقرؤه مثلاً، فإنه ينبغي مراعاة طول القطع المقرؤه وعدها، وصعوبة كلماتها، ومجال الموضوعات التي تتناولها، والمثيرات أو المؤشرات التي تتضمنها، وغير ذلك من المتغيرات. ويجب الاسترشاد بهذه الخطة عند كتابة مفردات الاختبار نظراً لأن الإجراءات المنظمة لمعاينة النطاق السلوكي Domain Sampling تُعد من أهم خصائص الاختبارات مرجعية المحك.

فتمثل مفردات الاختبار للنطاق السلوكي الشامل في ضوء هذه الإجراءات يجعل من الممكن الاستدلال على درجة الفرد في هذا النطاق من درجته في عينة المفردات التي يشتمل عليها الاختبار.

ويتطلب كتابة مفردات الاختبار مهارة كبيرة وتمكن من محتوى النطاق السلوكي الذي يقيسه الاختبار.

وهناك نوعان من المفردات تستخدم في قياس الجانب المعرفي والمهارى، أحدهما يتطلب إجابة مقيدة والأخر يتطلب إجابة حرة. ومن أمثلة النوع الأول مفردات الاختيار من متعدد، والصواب أو الخطأ، والمزاوجة. ومن أمثلة النوع الثاني أسئلة المقال، والإجابة القصيرة، والتكميلة. ويتطلب كتابة كل من هذه الأنواع مراعاة بعض الشروط الفنية وبخاصة المفردات التي تتطلب إجابة مقيدة لقياس المهارات المعرفية العليا. وينبغي تدقيق المفردات عقب الانتهاء من كتابتها بواسطة مجموعة من خبراء المجال الذي يقيسه الاختبار وخبراء القياس والتقويم.

وقد أوضح المؤلف شروط وفنيات صياغة الأنواع المختلفة من المفردات بأمثلة من مختلف المجالات الدراسية في أكثر من مرجع آخر (عام ١٩٨٣ ، ١٩٩٥ ، ١٩٩٦ ، ١٩٩٧).

الخطوة الخامسة: تجميع مفردات الاختبار:

بعد الانتهاء من كتابة المفردات وتدقيقها ينبغي تنسيقها في ضوء الأهداف التي تقيسها لتكوين الاختبار. وهذا يتطلب مراعاة التسلسل المنطقي للمفردات وتجميع المفردات التي من نوع واحد معًا في ضوء مكونات المهارة التي تم تحليلها ومستوياتها المعرفية. وينبغي أن يسبق كل نوع منها تعليمات واضحة ودقيقة تخبر الفرد كيف

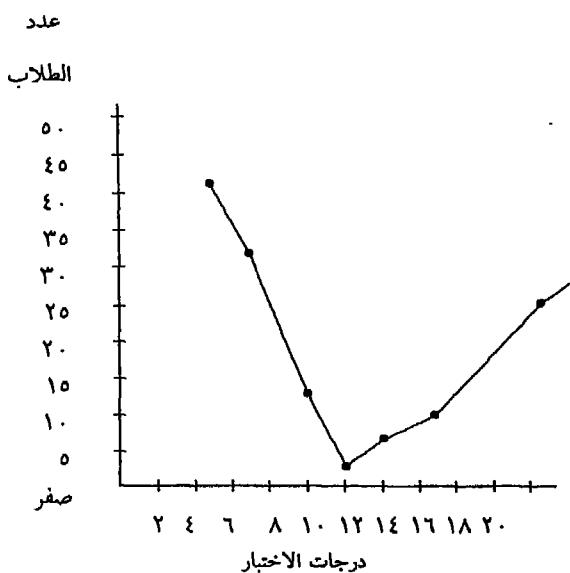
يستجيب وأين بدون استجابته، حيث إن هذه التعليمات تعد جزءاً من الاختبار. كما ينبغي تحديد زمن الاختبار إذا كان اختباراً مسوقاً في ضوء طبيعة مفرداته ونوعها ومحتراماً، ومن ثم يمكن طباعة الاختبار ومشتملاته.

الخطوة السادسة: تجريب الاختبار ميدانياً وتحليل مفرداته :

بعد الانتهاء من تكوين أو كتابة مفردات الاختبار وتجميعها ينبغي تجريب الاختبار ميدانياً على عينة مناسبة أو أكثر من الأفراد لتعرف مدى ملاءمة المفردات للمختبرين ودرجة تميزها بين الأفراد الذين حققوا هدفًا معيناً والأفراد الذين لم يتمكنوا من تحقيقه، وذلك لتقييم صدق المفردات في قياس النطاق السلوكي المتعلق بالأهداف المحددة. ويستطلب تحديد خصائص المفردات استخدام بعض الأساليب الإحصائية للبيانات المستمدلة من التجريب الميداني للاختبار. غير أن هذه الأساليب ربما تختلف عن تلك التي ناقشناها في الفصل السابع المتعلقة بتحليل مفردات الاختبارات مرجعية الجماعة أو المعيار.. ويرجع هذا الاختلاف إلى خصوصيات الاختبارات مرجعية المحك واستخداماتها، فهذا النوع من الاختبارات لا يهدف لإبراز الفروق الفردية في السمة المراد قياسها، وإنما يهتم بقياس تمكّن الأفراد من نطاق سلوكي معين يتعلّق بمهارة أو مهمة محددة نتيجة عملية التعليم أو التدريب. وهنا يفترض أن يحقق جميع الأفراد الأهداف التي يتضمنها النطاق السلوكي إذا كانت عملية التعليم أو التدريب فاعلة مما يجعل درجاتهم في الاختبار متقاربة. ونظرًا لأن أساليب تحليل مفردات الاختبارات مرجعية الجماعة أو المعيار - كما سبق أن ذكرنا - تعتمد على معاملات الارتباط التي تعتمد بدورها على تباين الدرجات، فإن هذه الأساليب لا تناسب استخدامات الاختبارات مرجعية المحك.

فإيجاد كل من معامل صعوبة المفردات ومعامل تميزها في هذه الحالة ينبغي أن يكون مناسباً للقرارات التي تتخذ في ضوئها. وهذه القرارات تتعلق بتصنيف الأفراد بحسب مستويات تمكّهم من النطاق السلوكي الذي تقيسه مفردات الاختبار. فإذا طبق الاختبار على مجموعة من الأفراد قبل بدء برنامج تعليمي أو تدريسي معين نتوقع انخفاض درجاتهم، بينما إذا طبق عليهم الاختبار نفسه أو صيغة مكافئة له عقب انتهاء البرنامج، فإننا نتوقع حصول غالبيتهم على درجات مرتفعة.

ويوضح شكل (٨ - ٥) التالي التوزيع المتوقع للدرجات قبل عملية التعليم وعقب الانتهاء منها:



شكل (٨ - ٥) يوضح توزيع افتراضي متوقع للدرجات ٥٠ طالبًا قبل عملية التعليم وعقب الانتهاء منها ويتبين من شكل (٨ - ٥) أن توزيع الدرجات يتبع شكلاً مفتوحًا إلى أعلى، حيث يشير الخط البياني الأيسر إلى توزيع درجات الطلاب قبل عملية التعليم نظرًا لأنها درجات منخفضة، بينما يشير الخط البياني الأيمن إلى توزيع درجاتهم عقب الانتهاء من عملية التعليم نظرًا لارتفاع درجاتهم.

ونقطة التقائه المخططين (الدرجة ١٢) تعد نقطة فاصلة للطلاب قبل عملية التعليم وبعدها. وتمثل هذه النقطة درجة القطع Cut-off Score للاختبار، فالدرجات الأقل منها تدل على عدم التمكن من النطاق السلوكي الذي يقيسه الاختبار، والدرجات الأعلى منها تدل على التمكن.

ويعتمد تحليل مفردات الاختبارات مرجعية المحك على تقسيم هذا التغير الذي حدث نتيجة عملية التعليم.

لذلك ينبغي إيجاد معامل صعوبة المفردات لمجموعة الأفراد قبل التعليم وعقب الانتهاء منه، أو للمجموعة التي تلقى التعليم والمجموعة التي لم تلق التعليم. فيما يهمنا في هذه الحالة هو درجة حساسية المفردة Item Sensitivity لعملية التعليم أو لفاعلية المعالجة التعليمية. إذ تتوقع أن يكون هناك فرق موجب بين قيمة معامل صعوبة كل مفردة بعد التعليم وقبله.

ولإيجاد معامل تمييز المفردات اقترح كوكس Cox وفارجاس Vargas أسلوباً بسيطاً يعتمد على طرح نسبة عدد من أجاب إجابة صحيحة عن كل مفردة من مفردات الاختبار قبل بدء عملية التعليم من نسبة عدد من أجاب إجابة صحيحة عن كل من هذه المفردات عقب الانتهاء من عملية التعليم. وتتراوح قيم هذا الفرق بين $-1, +1$. فكلما اقترب الفرق من $(+1)$ دل ذلك على أن المفردة تميزاً مرتفعاً بين الأفراد في مرتب تطبيق الاختبار أي قبل بدء عملية التعليم وعقب الانتهاء منها. وإذا اقترب الفرق من الصفر أو كان سالباً دل ذلك على عدم فاعلية المفردة في التمييز بين الأفراد مما يتطلب مراجعتها للتعرف على جوانب القصور فيها، والاسترشاد بأراء المختبرين حول فاعلية البرنامج التعليمي أو التدريبي بالنسبة لهم قبل إجراء التعديلات المناسبة على المفردة.

ومن الجدير بالذكر أن هناك العديد من الطرق الأخرى التي تستخدم في تحليل مفردات الاختبارات مرجعية المحك يمكن أن يرجع إليها القارئ في المصادر التي سبق أن أشرنا إليها.

الخطوة السابعة: التتحقق من صدق الاختبار وثبات درجاته :

تختلف أساليب التتحقق من صدق الاختبارات مرجعية المحك وثبات درجاتها اختلافاً ملحوظاً عن أساليب تقدير صدق وثبات الاختبارات مرجعية الجماعة أو المعيار التي ناقشناها بإسهاب في الفصلين الرابع والخامس، وذلك على الرغم من عدم اختلاف المعنى العام لمفهومي الصدق والثبات في كل من نوعي الاختبارات. فمفهوم الصدق يتعلق بمنطقية تفسير الدرجات وما يتربّع عليها من قرارات صائبة، أما مفهوم الثبات فيتعلق باتساق الدرجات. وكما أن هناك أنواعاً مختلفة من الصدق وثبات درجات الاختبارات مرجعية الجماعة أو المعيار، توجد أيضاً أنواع من صدق وثبات درجات الاختبارات مرجعية المحك. غير أنها سوف نقتصر هنا على الصدق المتعلق بالنطاق السلوكي Domain - Referenced Validity ، والثبات المتعلق باتساق تصنيف الأفراد بحسب درجة تمكّنهم من نطاق سلوكي معين.

فالاختبار مرجعي المحك يقيس الأداء أو السلوك المتعلق بنطاق سلوكي من المعارف أو المهارات يُعد بمثابة السلوك المحك Criterion Behavior ، وكما سبق أن ذكرنا ينبغي تحديد السلوك المحك تحديداً دقيقاً. ونقصد بصدق النطاق السلوكي للاختبار أن المحتوى والمهام التي تشتمل عليها عينة مفرداته تمثل السلوك المحك تمثيلاً كافياً، فالنطاق السلوكي يشير عندئذ إلى هذا السلوك المحك.

ولعل هذا النوع من الصدق يشبه إلى حد ما صدق محتوى الاختبارات مرجعية الجماعة أو المعيار. غير أن مفهوم صدق المحتوى أكثر اتساعاً من مفهوم صدق النطاق السلوكي نظراً لأن الاختبار مرجعى المحك يقيس عادة نطاقاً سلوكياً محدوداً، مثل رسم أشكال هندسية، أو تكوين جمل صحيحة، أو إجراء عمليات حسابية، أو استنباط نتائج من تجربة عملية، وهكذا.

والتحقق من صدق النطاق السلوكي يتطلب تقييم إجراءات تحليل النطاق السلوكي إلى مكوناته من معارف ومهارات، ونواتج هذا التحليل (الخطوة الأولى في بناء الاختبار التي سبق ذكرها)، وخطة معاينة مفردات الاختبار لهذه المكونات في ضوء المواصفات التفصيلية للاختبار (الخطوتان الثالثة والرابعة). ويعتمد هذا التقييم على أحكام الخبراء فيما يتعلق بنتائج هذه الخطوات للتحقق مما إذا كانت مواصفات النطاق السلوكي محددة تحديداً تفصيلاً واضحاً ودقيقاً ومكتملاً ومرتبطة ارتباطاً وثيقاً بمكونات هذا النطاق، وكذلك ما إذا كانت مفردات الاختبار تعد بمثابة عينة كافية وممثلة لهذة المكونات. لذلك فإن صدق النطاق السلوكي يتطلب وصفاً دقيقاً للنطاق المعين أو السلوك المحك، ومؤشرات عن مدى كفاية تمثيل مفردات الاختبار لهذا النطاق.

وللحذر من اتساق تصنيف الأفراد المختبرين إلى متمكينين أو غير متمكينين، فإنه ينبغي تطبيق الاختبار عليهم وإعادة تطبيقه أو تطبيق اختبار آخر مكافئ له. ويقتصر الثبات بمقارنة نسبة عدد الأفراد الذين يتم تصنيفهم على أنهم متمكينين في كل من مرتبى التطبيق. وكلما زاد اتساق قرار التصنيف في المرتبين دل ذلك على ارتفاع درجة الثبات.

ولكن تصنيف الأفراد ينبغي أن يستند إلى درجة فاصلة أو درجة قطع محددة Cut - off Score كما ناقشناها في الخطوة السادسة. كذلك ينبغي مراعاة أخطاء التصنيف الموجبة والسلبية التي أشرنا إليها في الفصل الخامس في سياق مناقشتنا للصدق التبؤى باستخدام جداول التوقع. والأسلوب الإحصائى الذى يأخذ هذه الأخطاء بعين الاعتبار هو معامل كابا Kappa Coefficient، حيث يحدد نسب اتساق التصنيف وفقاً لدرجة قطع معينة، وتصحيح هذه النسب من أخطاء التصنيف التى تعزى إلى الصدفة. إذ ربما يكون قد تم تصنيف بعض الأفراد إلى متمكينين أو غير متمكينين عن طريق الصدفة.

والصيغة الرياضية لمعامل كابا كالتالى:

$$\text{معامل كابا (K)} = \frac{\frac{L}{M} - 1}{1 - \frac{L}{M}}$$

حيث (L) ترمز إلى نسبة اتفاق التصنيف. وهى تساوى مجموع نسب أعداد الأفراد الذين يتم تصنيفهم فى المجموعة نفسها فى مرتب التطبيق ، (M) ترمز إلى نسبة الاتفاق المتوقع عن طريق الصيغة ، وهى تساوى مجموع النسب الهاشمية لأعداد الأفراد الذين يتم تصنيفهم فى المجموعة نفسها فى مرتب التطبيق.

ولتوضيح ذلك نفترض أننا طبقنا اختباراً مرجعى المحك على مجموعة تكون من 100 فرد، وأعدنا تطبيقه فى اليوم التالى ، واختربنا درجة قطع ٩٠٪ ، وأجرينا تصنيفاً لهؤلاء الأفراد فى مجموعتين إحداهما متمنكة والآخر غير متمنكة استناداً إلى درجة القطع المحددة. وأوجدنا نسب أعداد من أجابوا إجابة صحيحة عن مفردات الاختبار فى مرتب التطبيق، ونسب أعداد من أجابوا إجابة خطأ.

ويوضح جدول (٨ - ١) التالى هذه النتائج:

إعادة تطبيق الاختبار

		غير متمنك	متمنك	
		(ب)	(ج)	متمنك
		(د)	(هـ)	غير متمنك
٧٢ (٠, ٧٢)	١٧ (٠, ١٧)	٥٥ (٠, ٥٥)		الاختبار
٢٨ (٠, ٢٨)	٢٢ (٠, ٢٢)	٦ (٠, ٦)		
	٣٩ (٠, ٣٩)	٦١ (٠, ٦١)		

جدول (٨ - ١) يوضح كمية إيجاد قيمة معامل كابا (k)

فالخطوة الأولى: نوجد مجاميع النسب الهاامشية لكل من الصفين، أي مجموع النسبتين في الخليتين (أ) ، (ب) وكذلك (ج) ، (د) أي $72\% + 28\% = 100\%$ ، وكذلك مجاميع النسب الهاامشية لكل من العمودين أي مجموع النسبتين في كل من الخليتين (أ) ، (ج) ، والخليتين (ب) ، (د) أي $61\% + 39\% = 100\%$. ويلاحظ أن 55% من الأفراد تم تصنيفهم إلى متمنكين ، 22% تم تصنيفهم إلى غير متمنكين في مرتب التطبيق.

$$\text{وبذلك تكون } L = 55\% + 22\% = 77\%.$$

كما يلاحظ أن $72\% + 61\% = 133\%$. هما المجموعين الهاامشيين للتصنيف «متمن» ، $39\% + 28\% = 67\%$. هما المجموعين الهاامشيين للتصنيف «غير متمن». وبذلك تكون

$$L = (72\%, 61\%) + (28\%, 39\%) = 100\%.$$

والخطوة الثانية: نعرض في الصيغة (٨ - ١) لإيجاد قيمة معامل كابا (K) :

$$\text{معامل كابا (K)} = \frac{77\% - 54.84\%}{100\% - 54.84\%} = 49\%.$$

ولتفسير هذه القيمة ينبغي ملاحظة أن قيم معامل كابا لا تزيد عن (١+) ولا تقل عن (١-)، ونحصل على القيمة الأولى عندما تكون النسب الهاامشية متساوية في مرتب التطبيق، أما القيمة الثانية فنحصل عليها إذا اختلف تصنيف الفرد في المرتدين.

ونظراً لأن القيمة التي حصلنا عليها في هذا المثال 49% . فإن هذا يعني أن نسبة اتساق التصنيف منخفضة مما يدل على انخفاض قيمة الثبات.

وتأثر قيمة معامل كابا باختلاف درجة القطع التي نختارها، وكذلك تتأثر بعدد مفردات الاختبار ومستوى صعوبتها، وعدد الأفراد المختبرين. لذلك ينبغي تدوين المعلومات المتعلقة بهذه المتغيرات إلى جانب قيمة معامل كابا للاسترشاد بها عند تفسير هذه القيمة. وينبغي الإشارة إلى أننا سبق أن ناقشنا مستويات الأداء ودرجات القطع في نهاية الفصل السادس.

ويمكن الرجوع إلى المصادر التي أشرنا إليها في هذا الفصل للاطلاع على مزيد من طرق تحديد هذه المستويات أو الدرجات.

استخدامات الاختبارات مرجعية المحك

نشأت حركة القياس مرجعى المحك كما أوضحتنا في مستهل هذا الفصل مرتبطة بفلسفة تربوية للتعليم تختلف عن فلسفة القياس مرجعى الجماعة أو المعيار، فقد تبين قصور هذا النوع الأخير من القياس في التقويم التشخيصي، وكذلك قصوره في تقويم فاعلية البرامج التعليمية من حيث مدى تحقيقها للأهداف المرجوة، وبذلك لا يساعد في اتخاذ قرارات تفيد في تطوير هذه البرامج وزيادة فاعليتها.

فالأساليب والممارسات التربوية التقليدية أدت إلى جعل وظيفة المؤسسات التعليمية تقصر بدرجة كبيرة على تصنيف المتعلمين والحكم على ضعاف التحصيل بالفشل واستبعادهم تدريجياً خلال المراحل الدراسية، وتشجيع القليلين على متابعة التعليم دون الاهتمام بتقديم العون الكافي لجميع الأفراد للتعلم الفاعل واستكمال التعليم. فالنظام التعليمي السائد يوحد أساليب التدريس والمدة الزمنية للتعلم، وتُصمم الاختبارات بطريقة تجعل الطلاب الضعاف أو من يعانون من صعوبات التعلم يحصلون على درجات منخفضة حيث إن معيار الأداء في هذه الاختبارات يعتمد على الموازنة بالطالب المتوسط. وهذا يتنافى مع وظيفة المؤسسات التعليمية التي ينبغي أن يكون هدفها الجوهرى هو التعليم والتعلم الذى هو حق للجميع. ومع التطورات الكبيرة الحادثة في المجالات العلمية والتكنولوجية يقل الطلب على العمالة الماهرة ويلاقي أمثال هؤلاء الطلاب الذين لم يتع لهم قدرًا كافياً من التعليم صعوبات جمة في الحصول على أعمال، حيث إن الأعمال المتاحة التي تكون مناسبة لهم محدودة جداً. لذلك ينبغي أن ترتبط عمليات التعليم والتقويم بالأهداف التربوية المستقبلية المتسعة النطاق و تعمل على تحقيقها وتنمية إمكانات كل متعلم إلى أقصى حد ممكن بما يجعله قادرًا على مواجهة احتياجات ومتطلبات هذا القرن. ولعل حركة الاختبارات مرجعية المحك وما ارتبط بها من تقنيات وتطبيقات والتي أصبحت شائعة الاستخدام في العديد من الدول المتقدمة وبعض الدول النامية وما ترتب عليها من منهجية جديدة تؤكد التكامل بين عملية التعليم وعملية التقويم من خلال الأهداف التربوية المستقبلية تعد من التطورات المهمة التي وجَّه علماء القياس وخبراء التصميمات التعليمية جهودهم البحثية نحوها لتوسيع نطاق تطبيقاتها داخل الصف المدرسي.

تقييم الكفايات :Competence Assessment

إن منهجية القياس مرجعى المحك باعتبارها منهجة إديومترية مستحدثة أثرت تأثيراً واضحاً في النظر إلى التحصيل من منظور جديد، فأصبح ينظر إلى قياس التحصيل على أنه تقييم للكفايات Competence Assessment، أي تقويم القدرة على تنفيذ مهام محددة بنجاح بدلاً من اختزان بعض المعلومات والمعرف المترفرقة، وهذا يجعلنا نفسر التحصيل بأنه أداء Performance أو قدرة على إنجاز شيء ما. فالكفاية تتضمن توسيع نطاق المعرفة والفهم وتنمية مهارات متعددة ومتكاملة يصعب قياسها بواسطة الاختبارات التي تصمم وفق الأساليب السيكومترية الكلاسيكية. وقد أسهم هذا التوجه في تطوير تصميمات ونماذج تعليمية مبنية على أساس الكفايات Competency-Based Models من أجل التمكن والإتقان في ضوء فلسفة تربوية مؤداها أن كل طالب مهما كانت قدرته يستطيع أن يتعلم تعلمًا فاعلاً ويتقن الأهداف والكفايات المحددة إذا أتيح له الوقت الكافي والأسلوب التعليمي الذي يناسب مستوى قدرته (Carroll, 1970) (Guskey, 1985). وهذا يتطلب تنوعاً في أساليب التعليم وأدوات التقويم وتقنياتها. وقد استخدمت هذه التصميمات والنماذج التعليمية في كثير من المجالات التربوية والمهنية في الدول المتقدمة بهدف إكساب الكفايات وتقويمها تقويمًا بنائياً وتشخيصياً وختامياً وتحسين الممارسات في هذه المجالات.

وتعُد حركة تقويم المهارات الأساسية Basic Skills Assessment من التطورات المرتبطة بالقياس مرجعى المحك، وتهتم هذه الحركة بالكفايات الوظيفية المتعلقة بالقراءة والكتابة والرياضيات والحاسوب، وكذلك بمهارات العمل والوظائف، وتقويم التحصيل المرتبط بالحياة الفعلية والمهام التي يواجهها المتعلم في تعامله اليومي تقويمًا مرجعى المحك.

ذلك من أهم تطبيقات الاختبارات مرجعية المحك منع التراخيص المهنية. ففى الولايات المتحدة الأمريكية حددت كثير من المنظمات المهنية اختبارات مدخلية ينبغي أن يجتازها الفرد المتقدم للعمل فى مهنة معينة أو يحقق حدًا أدنى من مستوى محدد للسماح له بمزاولة المهنة. كما يمكن أن يعاد تطبيق الاختبار على الفرد إذا أخفق فى المرة الأولى. وعادة تقيس هذه الاختبارات الكفايات المتعلقة بالأدوار المهنية، وتفسر درجات الفرد في ضوء مستويات الأداء المحددة.

كما أن اختلاف أنواع المؤسسات التعليمية وأهدافها يتطلب الحصول على أدلة متعددة عن التحصيل وفق معايير محددة. ويمكن أن يتمثل ذلك في الموهبة الفنية، والإنتاج الأدبي، والقيادة، والتفكير العلمي، والمهارات البدنية، وخدمة المجتمع، وفنون صنع الأشياء، ومهارة تحليل المشكلات المعقدة، مما يمكن قياسه وتقييمه باستخدام الاختبارات مرجعية المحك.

وكذلك يمكن الإفادة من هذه الاختبارات في تشخيص أخطاء مهارات العمليات Process Skills Errors المتعلقة بإجراءات تنفيذ مهمة معينة، وأخطاء فهم المقرئ Reading Errors ، وأخطاء التحويل Transformation Errors التي تعرقل المتعلم عن انتقاء العملية أو سلسلة العمليات اللازمة لحل المشكلة، وأخطاء فك الترميز Encoding Errors التي تجعله لا يستطيع التعبير عما توصل إليه من حل للمشكلة بصورة مكتوبة مقبولة على الرغم من معرفته للحل.

عقود الأداء Performance Contracting

تسعى كثيرون من الدول إلى ضبط جودة الأداء التربوي، وتحديد مستويات هذا الأداء بانتظام، وتُعرف جوانب الضعف فيه وتشخيصه ومعالجته، والإفادة من ذلك في تطوير نوعية المناهج المدرسية والممارسات التربوية. لذلك ينادي كثير من المسؤولين في المجال التربوي في بعض دول العالم المتتطور وكذلك أولياء أمور الطلاب بتقويم نواتج العملية التعليمية في ضوء مدى تحقيقها للأهداف التربوية والمجتمعية. ويصبح المعلم مسؤولاً عن تقديم أدلة تبرهن على أن تلاميذه قد حفظوا هذه الأهداف بعد مرور مدة دراسية محددة. وتُعتمد هذه الأدلة من نتائج تطبيق اختبارات مرجعية المحك أثناء عملية التعليم وعقب الانتهاء منها. وهذه الاختبارات تتراffer فيها خصائص معينة تجعلها أكثر ملاءمة لقياس الأهداف التعليمية قياساً مباشراً في ضوء مستوى أداء محدد مسبقاً، كما أوضحنا في هذا الفصل.

ويرى مؤيدو حركة المسئولية التربوية Accountability في الولايات المتحدة الأمريكية وغيرها من الدول، أن عقود الأداء هي الوسيلة التي تضمن التزام جميع الأطراف المعنية بتحقيق الأهداف المرجوة للبرنامج التعليمي على أساس زمني محدد. غير أنه لا توجد مؤشرات واضحة تدل على مدى فاعلية تطبيق هذا النظام في المجالات التعليمية المختلفة، وبخاصة أنه يتطلب تنفيذ أنظمة إدارية متقدمة

تعتمد على التقنيات الحديثة، مثل تقويم ومراجعة البرامج Program Evaluation and Revision System.

مشاركة الجمهور في العمليات الاختبارية :

أدت التوجهات السابقة إلى زيادة مشاركة الجمهور في العمليات الاختبارية، وعمليات التقويم في بعض الدول المتقدمة، وما ترتب على ذلك من حوار ومناقشات، ونقد وتشريعات، ومراجعة التنظيمات في ظل هذا المناخ الديمقراطي الذي أدى إلى حركة المسئولة التربوية Accountability Movement التي أشرنا إليها.

وعلى الرغم من الآثار الإيجابية لهذا التطور الذي أفادت هذه الدول من نتائجه في السنوات الماضية مثل زيادة عدالة الاختبارات وغيرها من أساليب تقويم التحصيل، وتوسيع نطاق المشاركات في تحديد محتوى الاختبارات مرجعية المحك ، والأهداف التعليمية التي تقيسها هذه الاختبارات ، وعدم إساءة استخدامها ، وتقديم معلومات أفضل للمختبرين ومستخدمي الاختبارات في ضوء هذه الأهداف إلا أن هذا لم يخل من مشكلات تتعلق باتخاذ قرارات اعتبارية أو اعتماد مخططات ربما تتنافي مع ممارسات التقويم السليم .

تقويم البرامج التربوية Educational Program Evaluation

يهدف تقويم البرنامج التربوي لتعريف مدى تحقيقه لأهدافه المرجوة. والاختبارات مرجعية المحك تفيد ليس فقط في التقويم الخاتمي للبرنامج لتعريف مدى فاعليته في تحقيق تلك الأهداف، وإنما تفيد بدرجة أكبر في التقويم البنائي للبرنامج وذلك بالكشف عن جوانب القوة والضعف في أهدافه وتصميمه وإجراءاته وأساليب تنفيذه بحيث يمكن تعديل مسار البرنامج بناء على نتائج هذه الاختبارات.

كما يمكن إعادة النظر في أهداف البرنامج أو محتواه إذا ثبت عدم ملاءمتها لمستوى قدرات الطلاب والإمكانيات المادية والفنية المتوفّرة.

ثانياً : اختبارات التحصيل مرجعية الجماعة أو المعيار :

يرتبط هذا النوع من الاختبارات ارتباطاً وثيقاً باختبارات التحصيل المقrite Standaradized Achievement Tests . وتعُد هذه الاختبارات أدوات قياس تربوية مهمة نظراً لأنه يمكن أن تتخذ قرارات متعددة تتعلق بالأفراد والبرامج التربوية والتدرية استناداً إلى نتائجها .

وتحتفل اختبارات التحصيل المقتنة عن الاختبارات مرجعية المحك أو الاختبارات التي يكتتبها المعلم لطلابه في أنها تعتمد اعتماداً أساسياً على إجراءات تقييم منظمة وجداول معايير يُستند إليها في مقارنة أداء فرد أو مجموعة من الأفراد بأداء جماعة مرجعية Norm Group مواصفاتها محددة بدقة.

لذلك تقوم بإعداد هذه الاختبارات مؤسسات متخصصة تحت توجيه وإشراف خبراء القياس التربوي واحتياطي المناهج، وذلك لأنها تتطلب كثيراً من الإمكانيات المادية والفنية. وينتشر استخدام الاختبارات التحصيلية المقتنة في الولايات المتحدة الأمريكية بخاصة حيث يُتفق على بنائها واستخدامها ملايين الدولارات، وتقوم المؤسسات المتخصصة بنشر وتوزيع هذه الاختبارات في كافة أنحاء العالم. غير أنها نعاني في الوطن العربي من قلة هذه الاختبارات وندرة استخدامها في تقويم المستوى العام للطلاب والمدارس، وفي فحص وانتقاء العاملين بالمهن والوظائف المختلفة. وربما يرجع ذلك إلى عدم وجود مؤسسات عربية متخصصة يوكل إليها مهام إعداد هذه الاختبارات بأنواعها المختلفة التي سبق أن أشرنا إليها.

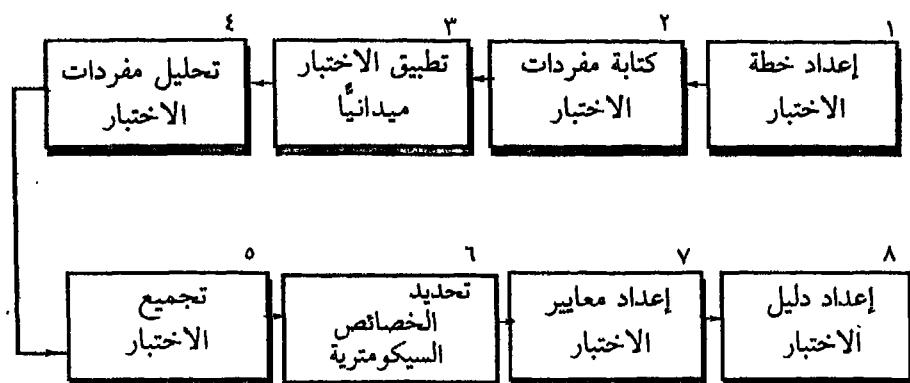
ففي الولايات المتحدة الأمريكية تقوم هذه المؤسسات بإعداد مئات من الاختبارات في المواد الدراسية المختلفة لقياس التحصيل العام لطلاب جميع المراحل التعليمية تكون عادة بطاريات من الاختبارات، وكذلك إعداد اختبارات مقتنة لمهارات أساسية في المراحلتين الابتدائية والمتوسطة، مثل القراءة ، وفهم المقروء ، والهجاء ، والعمليات الحسابية والهندسية ، وتوسيف الأرقام والأشكال البينية ، وحل المسائل ، وغيرها من مهارات التعبير والتواصل ، والتعامل مع المقادير الكمية . كما تقوم بإعداد اختبارات مقتنة في مجالات مهنية معينة تتطلب تقييماً لمعارف ومهارات العاملين فيها أو المتقدمين للالتحاق بها.

وتنتهي المدارس أو المؤسسات المهنية الاختبارات والمقياس المقتنة التي تناسب المرحلة الدراسية أو مجال العمل المطلوب ، وخصائص الأفراد، والغرض من التقويم . ولا يكون للمعلم أو المسؤول عن العمل دور في إعداد هذه الاختبارات، بل ولا يقوم بتصحيحها في كثير من الأحيان، وإنما تقوم المدرسة أو الجهة المعنية بإرسال أوراق الإجابة إلى المؤسسة الناشرة للاختبار حيث تُجرى عملية تصحيح الإجابات باستخدام أجهزة إلكترونية مخصصة لهذا الغرض ، وذلك لأن هذه الاختبارات تشتمل عادة على مفردات اختيار من متعدد . وتسجل هذه الأجهزة الدرجات الخام للأفراد في الاختبار ، وكذلك المئويات أو الدرجات المعيارية المعدلة أو الساعيات المعيارية المناظرة لهذه

الدرجات الخام لكي يتسمى مقارنة أداء كل فرد بأداء الجماعة المرجعية التي استمدت منها معايير الاختبار . ويعاد إرسال النتائج إلى الجهة المعنية بأحد وسائل الاتصال العادية أو الإلكترونية .

خطوات بناء اختبارات التحصيل المقننة :

يم ببناء اختبارات التحصيل المقننة بخطوات متعددة . وعلى الرغم من أن هذه الخطوات ربما تختلف بحسب طبيعة الاختبار ونوعه واستخداماته ، إلا أن هناك تسلسلاً منطقياً معيناً ينبغي مراعاته يوضحه الشكل التخطيطي (٨ - ٦) التالي :



شكل (٨ - ٦) يوضح التسلسل المنطقي لمراحل إعداد الاختبار التحصيلي المقنن

(من تصميم المؤلف)

وسوف نوضح فيما يلى هذه الخطوات بمزيد من التفصيل :

إعداد خطة الاختبار : فبناء الاختبار التحصيلي المقنن يتطلب العناية بإعداد خطة دقيقة ومنظمة يُستند إليها في تنفيذ الخطوات المتالية الموضحة بشكل (٨ - ٦) ، ويقوم بإعداد هذه الخطة مجموعة متكاملة من خبراء القياس والتقويم وختصاصي المواد الدراسية أو المجالات المهنية والتدريرية ومستشاري المناهج . ولإعداد هذه الخطة ينبغي تحديد الغرض من الاختبار وفيما تُستخدم نتائجه . فقد سبق أن ذكرنا أن هناك أنواعاً متعددة من هذه الاختبارات ، وتحديد النوع الذي تبين الحاجة إليه بعد أمراً ضرورياً . غير أنه في جميع الأحوال ينبغي تحديد محتوى الاختبار وذلك بفحص عينات من الكتب المدرسية في المجال المعين حيث إنها تعد مصدرًا أساسياً للمقرر أو المنهج المدرسي للفرقة المستهدفة في النظام التعليمي في مختلف أنحاء القطر . وكذلك

استشارة خبراء المادة الدراسية والمعلمين المتميزين إما فرادى أو جماعة. أما إذا كان المجال يتعلق بهم معينة فإنه ينبغي تعرف مواصفات هذه المهن ومتطلباتها الأساسية استرشاداً بآراء اختصاصى هذه المهن والمطبوعات المتعلقة بمواصفاتها. والهدف من ذلك تحديد الموضوعات والمشروعات والإجراءات التي تتضمنها، والوقت اللازم لتعلم كل منها. وهذا يتطلب تحليلًا دقيقاً للمحتوى ومتعلقاته، والاتفاق على نواتج هذا التحليل من جانب هؤلاء الخبراء، نظراً لأن الاختبارات التصصصية المقترنة تستخدم على نطاق واسع. ويتم في ضوء ذلك إعداد جدول مواصفات الاختبار Table of Specifications ، وهو جدول ثانٍ يبعُد ثُمَّ عنِّيهِ المجلالات أو الموضوعات المتعلقة بالمحظوي على الْبُعد الأفقي ، والمعارف والمهارات المرجو تحقيقها في مستويات معرفية مختلفة حيث تصاغ في عبارات عامة على الْبُعد الرأسى مثل معرفة المصطلحات والحقائق، فهم المقصود ، حل المسائل ، اقتراح تفسيرات ، تقدير مهام أو مطالب ، وهكذا. وتدون في خلايا الجدول أوراناً نسية لكل من المحتوى والمهمام وفقاً لأهميتها ، والزمن المناسب لتعلمها ، وإسهامها في الموضوعات اللاحقة ، مع مراعاة تحجيز التحzierات الفردية .

ويوضح جدول (٨ - ٢) التالى هيكل جدول مواصفات افتراضى لاختبار تحصيلي مقنن في القراءة:

جدول (٨-٢) يوضح هيكل جدول مواصفات اختبار تحصيلي مقتن في القراءة

ويُعد جدول (٨ - ٢) بمثابة خطة تنظيمية لمجالات المحتوى ومستويات المعرفة التي يهدف الاختبار لقياسها باستخدام عينة مماثلة من المفردات التي يقترح عددها في حلبنا الجدول في ضوء الاعتبارات السابقة.

كتابة مفردات الاختبار: بعد الانتهاء من إعداد جدول المواصفات يبدأ فريق الخبراء في كتابة مفردات الاختبار التي تقيس المعرفات والمهارات الموضحة على بعد

الرئيسي في ضوء المحتوى الموضع على **البعد الأفقي**. وتُعد هذه الخطوة من أصعب خطوات بناء الاختبارات المقننة نظراً لما تتطلبه من جهد كبير، وما تستغرقه من وقت طويلاً في كتابة مفردات ذات نوعية جيدة لقياس المعرفة أو المهارات المرجوة قياساً صادقاً.

ويتطلب ذلك تنظيم ورش عمل مكثفة لمجموعات الخبراء في أيام متالية لكتابة المفردات وصياغتها ومناقشتها ومراجعتها مراجعة دقيقة وإعادة صياغتها وانتقاء أفضلها. كما يتطلب أحياناً أن يكتب كل خبير على حدة مفردات مقترحة يجري مراجعتها وتقييمها في ورش العمل. وعادة يتم كتابة عدد من المفردات أكبر مما هو مدون في خلايا جدول المواصفات تنظرأ لأن بعض هذه المفردات ربما يستبعد عند فحصها وتدقيقها أو بعد تجربتها، كما أنه عادة تكون هناك صيغ متكافئة للاختبار مما يتطلب توافر عدداً كبيراً من المفردات.

وعلى الرغم من كثرة استخدام مفردات الاختيار من متعدد في الاختبارات المقننة، إلا أنه ليس هناك ما يمنع من استخدام أنواع أخرى من المفردات أو الأسئلة. فقد سبق أن ذكرنا أن المفردات الاختبارية يمكن تصنيفها إلى مفردات تتطلب إجابة مقيدة وتسمى مفردات موضوعية، ومفردات تتطلب إجابة حرية أو مفتوحة، ومفردات تتطلب أداءات عملية، ويعتمد انتقاء نوع المفردة على المهام أو المهارات المراد قياسها. غير أن مركز الخدمات الاختبارية بالولايات المتحدة الأمريكية Educational Testing Services (ETS) الذي يعد من المراكز المجموعة في بناء مختلف أنواع الاختبارات التي يتم نشرها في كل دول العالم أصبح يستخدم أنواعاً مختلفة من المفردات عند بناء الاختبارات المقننة. إذ يستخدم مفردات الاختيار من متعدد التي تتطلب من المستجيب وضع علامة معينة في ورقة الإجابة تشير إلى اختياره لأحد البذائل التي تشتمل عليها المفردة. وهذه المفردات تميز بامكانية تطبيقها بحيث تقيس المستويات المعرفية العليا، ويمكن أن يجيب الفرد عن عدد كبير منها في وقت أقل مما يجب فيه عن سؤالين أو ثلاثة أسئلة مقالية، كما أنها محددة الإجابة مما يجعل تصحيحها أيسر وأقل كلفة وأكثر اتساقاً، بل ويتم تصحيحها عادة باستخدام الحاسوب.

أما أسئلة المقال فإنها تناسب قياس الأصالة والطلاقة والتفكير المتدرج، وتوضح استيعاب الحقائق والتعبير عن ذلك في إطار تنظيمي مناسب، وعلى الرغم من عمقها في القياس إلا أن ذلك يكون على حساب دقة القياس. فتقدير الدرجات يشوهه عدم الاتساق، مما يتطلب الحرص والعناية عند صياغة أسئلة المقال بحيث يكون الإطار المرجعى للإجابة عنها محدداً وفي الوقت نفسه يسمح للفرد بأن يُعبر عما في حوزته من معارف تتعلق بمضمون السؤال مما يجعل عملية تقدير الدرجات أكثر موضوعية.

ويمكن تصميم أنواع من المفردات تضم ميزات المفردات الموضوعية وأسئلة المقال كما في الاختبارات التي تقيس التعبير اللغوي، وتسمى المفردات المتداخلة خطياً Interlinear Items ، حيث يطلب من المستجيب أن يقوم بتحرير قطعة نثرية تحتوى على بعض الأخطاء سواء في النحو أو الصياغة أو الفكرة التي تتناولها القطعة.

كما يمكن تصميم أنواع أخرى من المفردات غير التقليدية لتقييم نوعية مهارات معينة، مثل مهارة العلاقات الاجتماعية لدى الأطفال. وهنا يُقدم للأطفال موقف يتطلب حل مشكلة بطريقة جماعية كأن يطلب منهم تصميم مجسم معين باستخدام عدد محدد من المكعبات التي توزع عليهم. ويسجل القائم باللاحظة سلوك كل منهم أثناء العمل المشترك، ويتترجم هذه الملاحظات إلى درجات يمكن في ضوئها تصنيف مجموعة الأطفال في قسمين فيما يتعلق بمهارة العلاقات الاجتماعية، مثل المجموعة المتعاونة في التخطيط، والمجموعة المتعاونة في التنفيذ.

وكذلك يمكن تصميم مفردات من نوع آخر غير تقليدي تتعلق بالكفاءة الموقفية Situational Proficiency Items ، حيث يتركز الاهتمام على تقييم كفاءة الأفراد في أداء مهام محددة تحاكي مواقع العمل الفعلية، وتسمى هذه المفردات «مفردات في السلة In - Basket Items». وتُستخدم هذه المفردات عادة لتقدير العاملين في الوظائف التي تتطلب كفاءات ومهارات عليا معينة مثل الوظائف الإدارية أو القيادية.

فلتقدير المهارات الإدارية تصمم مفردات تتضمن محاكاة المواقف الإدارية بما في ذلك المراسلات والمذكرات والأدلة الإرشادية والتقارير التي تتطلب اتخاذ قرارات إدارية متعددة. كما يستخدم هذا النوع من المفردات الاختبارية في دراسة أساليب تدريب وانتقاء المتقدمين للالتحاق بهذه الوظائف.

أى أنه يمكن تحويل وتطوير أنواع المفردات المألوفة وابتكار أساليب جديدة لقياس الأهداف التعليمية والتدريبية التي تتناولها الاختبارات التحصيلية المقنة.

تطبيق الاختبار ميدانياً وتحليل مفرداته : بعد إعداد خطة الاختبار وكتابة مفرداته، فإن الخطوة التالية تطبيق الاختبار لتجريمه ميدانياً وتحليل مفرداته . فمهما بلغت عناية الخبراء في إعداد خطة الاختبار، ومهما كانت خبرتهم وحذفهم في بناء أنواع المفردات الاختبارية، فإن هذا لا يضمن أن المفردات سيكون وقوعها الميداني مماثل لما توقعه الخبراء.

لذلك ينبغي إعداد صيغة أولية للاختبار Pre-Test تشمل على عدد من المفردات أكبر مما هو مطلوب وذلك لتجريبيها ميدانياً، ويطبق الاختبار على عينة أو عينات من الأفراد تمثل المجموعة الفعلية التي سيطبق عليها الاختبار في صيغته النهائية فيما بعد.

لذلك يجب أن تمثل العينات المناطق الجغرافية، والمستويات الاجتماعية والاقتصادية والذكور والإإناث، وغير ذلك من الخصائص الديموغرافية.

وتعد هذه الخطوة من الخطوات المهمة في بناء هذا النوع من الاختبارات لأنها سيسجّر بناء على نتائج هذا التطبيق تحليل مفردات الاختبار باستخدام الأساليب الإحصائية التي سبق أن أوردناها تفصيلاً في الفصل السابع. ويؤدي هذا التحليل إلى إلقاء الضوء على صعوبة تمييز كل مفردة.

ويفيد ذلك في تحليل فاعلية مكونات المفردات من أجل تحديد جوانب القصور فيها وتعديلها وحذف المفردات غير الصالحة، وربما إضافة مفردات جديدة وتجربتها ميدانياً مرة أخرى، كما أنه يُفيد في الحكم على الأنواع الجديدة أو المبتكرة من المفردات مثل التي أشرنا إليها من حيث إمكانية الاستمرار في استخدامها، وتعزّز أفضل استخداماتها، وما إذا كانت تقيس شيئاً يصعب قياسه باستخدام الأنواع المألوفة من المفردات، ومدى فاعليتها في ذلك، وعدد المفردات الذي ينبغي استخدامه في الاختبار لجعل الدرجات أكثر ثباتاً. وبذلك يمكن انتقاء أفضل المفردات المتعلقة بالأهداف المراد قياسها لكي يضمّها الاختبار في صورته النهائية.

وبيني مراعاة أنه يفضل أن يتراوح مدى صعوبة المفردات بين السهولة والصعوبة لكي تميّز المفردات تميّزاً أفضل بين المختبرين.

كما ينبغي أثناء التجريب الميداني للمفردات تحديد الإجراءات المقترنة لتطبيق الاختبار لكي تضمنها تعليمات الاختبار عندما يوضع في صورته النهائية.

تجميع الاختبار: بعد إجراء تحليل المفردات والتحقق من فاعليتها ينبغي تجميع الاختبار في صورته النهائية. وهذا يتطلب التتحقق من أن المفردات المتقدمة تمثل جميع خلايا جدول الموصفات في ضوء الأوزان النسبية لكل خلية منها. كما ينبغي إعادة مراجعة الاختبار مرة أخرى بواسطة أحد اختصاصي المادة التي يقيسها الاختبار للتحقق من عدم تمييز مفردات الاختبار لمناطق معينة، مثل المدن والريف أو لنوع الفرد وغير ذلك من عوامل التحيز.

كما ينبغي التتحقق من أن عدد المفردات التي يشتمل عليها الاختبار مناسب للزمن الذي يتم تحديده عن طريق تطبيق الاختبار على مجموعة عائلة من الأفراد مع مراعاة ما إذا كان الاختبار من نوع اختبارات القوة Power Test أو اختبارات السرعة Speed Test. فإذا كان الاختبار من النوع الذي يعتمد على القوة ينبغي أن يسمح لنسبة تبلغ حوالي ٩٠٪ من الأفراد أن يحاولوا الإجابة عن جميع المفردات ، أما إذا كان من النوع الذي يتطلب السرعة فإنه يتوقع أن تحاول نسبة أقل الانتهاء من إجابة جميع

مفردات الاختبار. وفي أي الأحوال ينبغي حساب متوسط الزمن اللازم للإجابة عن الاختبار.

كما ينبغي إجراء عملية تكافؤ للصيغ المختلفة من الاختبار بحيث لا تختلف درجات الفرد من صيغة إلى أخرى اختلافاً دالاً.

ويعد هذا التوازن في عدد المفردات وصعوبتها وتميزها وملاءمتها للمختبرين من العمليات الدقيقة التي يجب مراعاتها.

تحديد الخصائص السيكومترية للاختبار: وتتضمن هذه الخطوة تقدير ثبات درجات الاختبار، أي اتساق درجاته إذا أعيد تطبيق الاختبار على مجموعة الأفراد نفسها بعد مرور مدة زمنية مناسبة، وكذلك تقديم أدلة تتعلق بتكافؤ صيغتي الاختبار. أي إيجاد قيمة كل من معامل الاتساق ومعامل التكافؤ، وكذلك التحقق من الاتساق الداخلي لمفردات الاختبار إذا تبيّنت الحاجة إلى أدلة عن هذا النوع من الثبات.

ذلك ينبغي التتحقق من صدق نتائج الاختبار في تحقيق الغرض الذي يُبْتَدِئُ الاختبار من أجله. بعض اختبارات التحصيل المقننة تستخدَم في أغراض الانتقاء والتسلكين والتشخيص والتنبؤ، وهذه الاستخدامات تتطلب التتحقق من الصدق المرتبط بمحك الذي يعتمد على بيانات إمبريقية.

ولكن كثيراً من هذه الاختبارات تستخدم في تقويم إنجازات الطلاب وفي عملية التعليم، لذلك فإن محتوى الاختبار يُعد العنصر الأساسي. وهذا يتطلب التتحقق من صدق المحتوى لتقدير مدى تعلق مفردات الاختبار بالأهداف التعليمية العامة التي يهدف الاختبار لقياسها، ومدى تمثيلها لحالياً جدول مواصفات الاختبار. وقد سبق أننا نقاشنا مفهومى الثبات والصدق بإسهاب في الفصلين الرابع والخامس.

إعداد معايير الاختبار: بعد الانتهاء من تجريب الاختبار وتحليل مفرداته وانتقاء أفضل المفردات وتجميعها في الاختبار، وتحديد خصائصه السيكومترية، فإن الخطوة التالية إعداد معايير الاختبار.

على الرغم من السجهد الكبير الذي يُبذَل في انتقاء المفردات وفي تكوين الاختبار، إلا أن الاختبار لا يستفاد منه ما لم يكن هناك إطار مرجعي يُسْتَند إليه في تفسير الدرجات نفسياً له معنى ودلالة.

لذلك لابد من إعداد معايير الاختبار التي تتحقق ذلك. وهذه المعايير تمكّناً من مقارنة الدرجة أو الدرجات التي يحصل عليها الفرد بمتوسط درجات الجماعة المرجعية.

أو عينة التقنين Norm Group ، كما تمكّنا من المقارنة بين درجات الأفراد في المجالات المختلفة. لذلك فإن العناية بانتقاء عينة أفراد الجماعة المرجعية التي تستمد منها معايير الاختبار تعد من الأمور الأساسية في بناء الاختبارات المقنتة.

وقد سبق أن ناقشنا بالتفصيل خصائص الجماعة المرجعية، ومعايير الاختبارات وأنواعها وكيفية حسابها في الفصل السادس. وأوضحنا في هذا الفصل المفاهيم المتعلقة بالمعايير المحلية، والمعايير الوطنية، ومعايير مؤسسة معينة وفئات خاصة. وعلى الرغم من أن الاختبارات التحصيلية المقنتة تُستخدم على مستوى وطني مما يتطلب أن تكون عينة التقنين ممثلة لجميع المناطق الجغرافية وتراعي المتغيرات الديموغرافية الأخرى، إلا أنه يمكن أيضًا أن تحدّد معايير محلية للاختبار أو معايير تتعلق بجماعات معينة من الأفراد.

وتتنوع معايير الاختبارات المقنتة وفقاً لطبيعة الاختبار واستخداماته، فكثير من هذه الاختبارات تستخدم معظم أنواع المعايير التي ناقشناها في الفصل السادس، حيث يجري تحويل الدرجات الخام إلى درجات ثانية، أو تساعيات معيارية، أو مئينيات، أو مكافئ العمر الزمني، أو مكافئ الفرق الدراسية، وتُعرض نتائج هذا التحويل في جداول معايير أو صفحات تربوية (بروفيل Profile)، أو أشكال بيانية وغير ذلك من الأساليب المبتكرة في عرض معايير الاختبارات؛

إعداد دليل الاختبار، إن بناء الاختبار وإعداد معاييره ليس نهاية المطاف، وإنما لابد من إعداد دليل أو أدلة الاختبار، وأوراق الإجابة، والملاحق المعينة لمستخدمي الاختبار بحيث يكون لديهم معلومات كافية يمكنهم في ضوئها الحكم على ما إذا كان الاختبار يناسب الغرض الذي يستعنون إليه، وما إذا كانت نتائجه يمكن استخدامها استخداماً مناسباً.

وعادة يلتزم ناشرو الاختبارات المقنتة في إعدادهم لأدلة هذه الاختبارات بالمعايير التي حدتها الجمعية الأمريكية لعلم النفس (APA) المتعلقة بالاختبارات التربوية والنفسية وأداتها. وقد أوضحنا ذلك بالتفصيل في الفصل الثاني.

فجميع الاختبارات المقنتة ينبغي أن يصاحبها أدلة Test Manuals تزود مستخدم الاختبار بمعلومات تفصيلية تتعلق بكيفية استخدام الاختبار، وتفسير درجاته، وخصائص مفرداته ، وصدقه ، وثباته، وتعليمات تطبيقه وتصحيحه. أى أن دليل الاختبار يعد المصدر الرئيسي لمستخدم الاختبار الذي يود أن يحصل على معلومات محددة عن اختبار معين. لذلك ينبغي أن يكون الدليل حديثاً ومكتملاً.

وبانتهاء جميع هذه الخطوات يصبح الاختبار صالحًا للنشر. غير أن عملية بناء الاختبارات المدققة لا تنتهي بمجرد نشر الاختبار، وإنما تتطلب المراجعة والتدقير المستمر في ضوء ما يستجد من مطالب. وما تكشف عنه التطورات في المجالات التربوية والنفسية، وما تتوصل إليه الدراسات والبحوث في مختلف هذه المجالات. فالاختبار الذي ييلو مرضياً اليوم ربما لا يكون كذلك غداً، إذ ربما يلزم مراجعته أو بناء اختبار آخر بديلاً عنه مع تكرار جميع الخطوات السابقة التي تتضمن إعداد خطة الاختبار، وكتابة مفراداته، وتجريمه ميدانياً، وتحليل مفراداته، وتجميعه، والتحقق من صدقه وثبات درجاته، وإعداد معاييره وأدلسته التي يُسترشد بها في تطبيقه وتصحيحه وتفسير درجاته.

وهذا يؤكد ما سبق أن ذكرناه في مستهل هذا الجزء من ضرورة وجود مؤسسات متخصصة في الوطن العربي يوكل إليها مهام بناء الاختبارات المقننة بأنواعها المختلفة، وكذلك الاختبارات مرجعية المحك ، واختبارات السلفايات في مختلف المجالات والمهام والأعمال.

بعض أوجه النقط التي توجه للإختبارات التجريبية المقترنة:

على الرغم من مزايا الاختبارات التحصيلية المقتنة واستخداماتها المتعددة في مختلف المجالات التربوية والمهنية فيما يتعلق بإمكانية مقارنة أداء الأفراد في مجال معين أو مجالات متعددة بأداء آخرين، ومقارنته التحصيلي والأداء بين الصنوف والمدارس المختلفة، وتتيح التقدم الدراسي للطلاب والنمو المهني للعاملين، وتقسم فاعلية المناهج الدراسية، والإرشاد والتوجيه التربوي والمهني، إلا أن علماءقياس وجهوا كثيراً من النقد لهذا النوع من الاختبارات لأسباب متعددة ذكرنا بعضها في الفصل الثاني تحت عنوان «الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربوية». وفيما يلي بعض أوجه النقد الأخرى:

(١) تشمل بطاريات الاختبارات التحصيلية المقننة عادة على مفردات اختيار من متعدد فقط حيث يصعب استخدامها في قياس قدرة الطالب على تنظيم وعرض الأفكار، ونقد الموضوعات، وتحليل وتركيب المعلومات وعرضها عرضاً متكاملاً، وغير ذلك من المهارات العقلية العليا التي ينبغي العناية بتقديتها لدى الطلاب. وبالطبع لا نستطيع إنكار عيوب الأسئلة التي تتطلب إجابات مفتوحة، مثل أسئلة المقال وغيرها من الأسئلة التي تعتمد على التقديرات الذاتية، والتي تقيس هذه المهارات، حيث إنها لا تمثل المحتوى المدقق بالذات كافياً سبب الزمـن الذي تتطلب الإجابة عنها، وصعوبة

تصحيحها وعدم اتساق درجاتها ، مما يجعلها لا تناسب البرامج الاختبارية التي تطبق على نطاق واسع وإن كانت تصلح للتطبيق في نطاق المدرسة. غير أن الاتجاه نحو تطوير هذا النوع من الأسئلة وابتکار أنواع جديدة من المفردات الاختبارية وتضمينها في اختبارات التحصيل المقنية ربما يعالج هذه المشكلة ولو معالجة جزئية .

(٢) تقيس هذه الاختبارات في كثير من الأحيان نواتج تعليمية تختلف عن النواتج التي يبغى المعلم تحقيقها في سلوك طلابه ، فهى تقيس ما يهم خبراء الجهة أو المؤسسة المتخصصة الناشرة للاختبار بغض النظر عن الأهداف التربوية التي يهتم المعلم في منطقة أو مدرسة معينة بتحقيقها . وبالطبع يصعب على هذه المؤسسة بناء اختبار تحصيلي مقنن يناسب جميع المناطق وجميع المدارس ويمكن تطبيقه على جميع الطلاب في ظروف موحدة ومحددة .

(٣) يرتبط بالمشكلة السابقة مشكلة أخرى تتعلق بما إذا كانت المناهج المدرسية تتحدد بناء على الاختبارات التحصيلية المقنية ، وما إذا كان المعلم يلجأ إلى التدريس من أجل الاختبار . فإذا كان تقويم المعلم يعتمد على نتائج طلابه في هذه الاختبارات ، فإن المعلم سوف يؤكد في عملية التدريس ما يحتويه الاختبار ويغفل الموضوعات الأخرى التي ربما تكون أكثر أهمية ، وذلك لكي يبدو تدرسيه فاعلاً .

(٤) تبني اختبارات التحصيل المقنية على فلسفة تربوية تؤكد إبراز الفروق الفردية بين الطلاب ، وجعل توزيع درجاتهم في الاختبار يتخذ شكل التوزيع الاعتدالي الذي تستند إليه معايير هذه الاختبارات . ولعل الاتجاه إلى استخدام الاختبارات مرجعية المحك أدى إلى تخفيف حدة هذه المشكلة ، حيث إن هذه الاختبارات تؤكد أن الغرض من العملية التعليمية إكساب كل طالب الكفايات أو المعارف والمهارات ، وتنمية قدراته واستعداداته ، وتقويم هذه المكتسبات تقويمًا مباشرًا دون الحاجة إلى التركيز على الفروق الفردية بين الطلاب وبغض النظر عما اكتسبه كل منهم بالفعل . فإذا حقق كل طالب الأهداف المرجوة من العملية التعليمية فإننا نستطيع عندئذ الحكم على فاعليتها ، وإبراز الفروق الفردية إذن في مثل هذه المواقف التعليمية لا يكون مرغوبًا ولا مطلوبًا .

لذلك نجد أن المؤسسات المتخصصة في بناء اختبارات التحصيل المقنية أصبحت تكرّس جانبًا كبيرًا من جهودها وإمكاناتها في بناء اختبارات مرجعية المحك في

مختلف المجالات الدراسية والتدرية والمهنية لقياس مدى تحقق الأهداف المتعلقة بها، وتشخيص جوانب الضعف التي ربما تعوق المتعلم عن اكتساب المعارف والمهارات المرجوة.

وأحياناً تُتلقى بعض مفردات اختبارات التحصيل المقتنة وتُستخدم كمفردات مرجعية المحك. فالمفردات التي تقيس مهارة معينة يمكن استخدامها في تقويم هذه المهارة بحيث إذا أجاب الطالب عنها إجابة صحيحة دل ذلك على اكتساب هذه المهارة دون الحاجة إلى تفسير درجته استناداً إلى معايير اختبار التحصيل المقتن الذي تعلمه الجهة الناشرة، وإنما يمكن تحديد مستوى الأداء أو محك النجاح المطلوب الذي يستند إليه في تقويم الطالب. فمفردات اختبارات التحصيل المقتنة تعد مفردات عالية الجودة نظراً لمرورها بفحص منظم ومراجعة دقيقة مما يجعلها صالحة لقياس الجوانب المعرفية.

وفحص إجابة كل طالب عن كل مفردة أو مجموعة من مفردات الاختبار مرجعي المحك على حدة بدلاً من الاهتمام بالدرجات الكلية فقط أو درجات الاختبارات الفرعية يقدم للمعلم معلومات تفصيلية عن كل طالب، ربما يصعب الحصول عليها باستخدام اختبار التحصيل المقتن. وبذلك يمكن الإفادة من كل من نوعي الاختبارات وإحداث تكامل فيما بينهما في عملية تقويم الطلاب وإثراء عملية التعليم.

خلاصة

تهدف اختبارات التحصيل الدراسي أو الأكاديمي لقياس وتقدير مستوى الكفاءات التي حققها المتعلم نتيجة للخبرات التعليمية، وعلى الرغم من أن هذه الاختبارات قد مرت بتغيرات متلاحقة، إلا أن استناد عملية التقويم إلى أساليب علمية وإجراءات منهجية أدى إلى تطوير أساليب وتقنيات بناء هذه الاختبارات.

فقد أصبح هناك أنواع متعددة من اختبارات التحصيل تُستخدم في اتخاذ قرارات تربوية تتعلق بالأفراد والبرامج. فاختبارات التحصيل المقيدة بأنواعها المختلفة تُستخدم في وضع تقييمات للطلاب ومقارنة الأداء بين صفوف فرق معينة في مدارس أو مناطق مختلفة، وتُقدم أدلة مجتمعة عن نمو الطالب تفيد في الإرشاد التربوي واتخاذ قرارات الانتقاء والتصنيف والتسكين وكذلك في تقويم فاعلية البرامج التعليمية. لذلك تقوم ببناء هذه الاختبارات مؤسسات متخصصة نظراً لكلفتها المرتفعة وما تتطلبه من خبرات فنية متخصصة. وتعتمد عملية البناء على خطة تنظيمية دقيقة تحدد الخطوات المتتالية لإعداد مفردات الاختبار وتجريبيها ميدانياً وتحليلها، وكذلك أساليب التحقق من صدق وثبات درجات الاختبار، وإعداد معاييره التي يستند إليها في تفسير الدرجات.

أما الاختبارات مرجعية المحك فإنها تعد من التطورات المعاصرة التي تهدف لتقدير الكفاءات المتعلقة بمختلف المجالات الدراسية والتدرية والمهنية في ضوء مستوى أو محك أداء يتم تحديده مسبقاً دون الاهتمام بتحديد المركز النسبي للفرد بين أقرانه في المجال الذي يقيسه الاختبار.

لذلك يُعنى في بناء هذه الاختبارات بتحديد النطاق السلوكي الذي يشتمل على المعارف والمهارات المرجوة تحديداً دقيقاً بحيث يمكن ملاحظة السلوك وقياسه. لذلك تختلف أساليب تحليل مفردات هذا النوع من الاختبارات وأساليب تقدير صدق الاختبار وثبات درجاته عن الأساليب المتبعة في الاختبارات مرجعية الجماعة أو المعيار. ويعتمد استخدام أي من النوعين على الغرض من استخدام الاختبار، وطبيعة القرار المراد اتخاذها نظراً لأن النوعين يختلفان في كيفية تفسير درجاتهم.

لذلك يمكن الإفادة من كل من النوعين بحيث يكون هناك تكامل فيما بينهما لإنارة عملية التقويم وترشيد القرارات التربوية استناداً إلى بيانات ومعلومات متنوعة.

غير أن هذا يتطلب تدريب المعلمين والمسؤولين عن اتخاذ القرارات المتعلقة بالأفراد على كيفية انتقاء الاختبارات المناسبة وإجراءات تطبيقها وتقدير نتائجها للإفاده المثلى من اختبارات التحصيل بأنواعها المختلفة.

الفصل التاسع

قياس الذكاء الإنساني

* مقدمة

* قياس الذكاء من منظور تاريخي

* مفهوم الذكاء والمشكلات المتعلقة بتعريفه

* الذكاء كمفهوم وصفى

* الذكاء كمجموعة من القدرات المترابطة

* الذكاء كعامل عام

* الذكاء كعوامل متعددة

* الذكاء كبنية ثلاثة الأبعاد

* الذكاء كنظام هرمى من العوامل

* الذكاء ومدخل تجهيز المعلومات ومعالجتها

* قياس الذكاء الإنسانى

* اختبارات الذكاء العام

* أولاً : اختبارات الذكاء الفردية

* مقاييس ستانفورد - بينيه للذكاء

* مقاييس ويكسنر للذكاء

* مقارنة بين مقاييس ويكسنر ومقاييس ستانفورد - بينيه

* بعض مقاييس الذكاء الفردية الأخرى

* ثانياً : اختبارات الذكاء الجماعية

* أنواع اختبارات الذكاء الجماعية

* أمثلة لبعض اختبارات الذكاء الجماعية

* اختبارات ذكاء جماعية لفظية متعددة المستويات

* اختبارات الذكاء الجماعية غير اللفظية

* تعقيب عام على مقاييس الذكاء

* بعض القضايا المتعلقة بقياس الذكاء

* بناء اختبارات الذكاء

* بعض محددات الذكاء والعوامل المؤثرة فيه

* ثبوت نسبة الذكاء

* خلاصة

مقدمة :

عرضنا في الفصل السابق طرق قياس التحصيل الدراسي والكفايات باستخدام الاختبارات مرجعية الجماعة أو المعيار، والاختبارات مرجعية المحك ، واتضح لنا أن هذه الاختبارات تهدف لقياس المعارف والمهارات المتعلقة بمجال دراسي أو تدريسي محدد. غير أن اقتصار هذه الاختبارات على قياس الأداء الحالى للأفراد فى مواقف مدرسية أو تدريبية معينة يجعلها محدودة الفائدة فى التنبؤ بالأداء المستقبلى ، وبخاصة إذا كانت المعارف والمهارات المرجو اكتسابها فى المستقبل تختلف عما اكتسبه الفرد فى الوقت الحالى ، أو إذا كان صدق اختبارات التحصيل المستخدمة وثبات درجاتها منخفضاً . فالأداء المستقبلى يصعب التنبؤ به باستخدام متغير منبئ واحد ، وهو درجات اختبار التحصيل ، وإنما يتطلب استخدام متغيرات متعددة لكي يتسمى اتخاذ قرارات صائبة تتعلق بالأفراد استناداً إلى الدرجات المتنبأ بها . فالذكاء أو القدرة العقلية العامة ، والاستعدادات الخاصة المتعلقة ب مجالات معينة تعد من المتغيرات المعرفية الأساسية التي يمكن استخدامها إلى جانب متغير التحصيل الدراسي فى التنبؤ بالأداء المستقبلى للأفراد .

ونظراً لأن متغير الذكاء أكثر عمومية من متغير الاستعدادات الخاصة فسوف نفرد هذا الفصل لتوضيح مفهوم الذكاء وما يتعلق بتعريفه من مشكلات منهجرية وتطبيقية ، وطبيعة الذكاء ونمادجه النظرية ، وأساليب وأدوات قياسه وطرق بنائها ، ومحددات الذكاء ، والعوامل المؤثرة فيه . وتناول مفهوم الاستعدادات الخاصة في الفصل العاشر . ولكي يتضح مفهوم الذكاء بدرجة أفضل سوف نستهل الفصل الحالى بتوضيح تطور هذا المفهوم وأساليب قياسه من منظور تاريخي في إطار تطور حركة القياس العقلى .

قياس الذكاء من منظور تاريخي :

على الرغم من أن ابتكار اختبارات الذكاء ربما يُعد أحد التطورات السيكولوجية المهمة في القرن العشرين ، إلا أن محاولات التفكير في طبيعة الذكاء لها تاريخ قديم . فالذكاء حتى بداية القرن التاسع عشر كان يعد من المفاهيم الفلسفية التي لا تخضع للدراسة العلمية . ومن أوائل المحاولات التي أدت إلى وضع المعالم الرئيسة لقياس الذكاء قياساً كمياً على أساس علمية منظمة جهود كل من عالم البيولوجي الإنجليزي فرانسيس غالتون Galton (١٨٢٢ - ١٩١١) وهو من أتباع داروين Darwin البارزين ، وعالم النفس الأمريكي جيمس ماكين كاتل Cattell (١٨٦٠ - ١٩٤٤) .

وقد درس كاتل علم النفس التجربى فى جامعة ليتزرج فى ألمانيا، وكان واحدا من تلاميذ ويلهلم فنرت Wundt مؤسس معمل علم النفس التجربى فى ألمانيا عام ١٨٧٩ ، وأجرى بحوثه بعد حصوله على درجة الدكتوراه بالاشتراك مع جالتون، ثم انتقل إلى جامعة كولومبيا فى الولايات المتحدة الأمريكية رئيساً لقسم علم النفس، وكان أول من استخدم مفهوم «الاختبار العقلى Mental Test» عام ١٨٩٠ . غير أنه اعتبر القدرة العقلية العامة تعكس فى التمييز الحسى و زمن رد الفعل وغير ذلك من الوظائف النفسية البسيطة.

وقد أفاد كاتل من جهود جالتون واهتمامه بالأساليب الإحصائية فى استخدام هذه الأساليب كوسيلة مهمة للبحث فى مجال الفروق الفردية.

وفي عام ١٩٠٤ طلب وزير التعليم فى فرنسا من عالم النفس ألفرد بينيه Binet وزميله الطبيب النفسي سيمون Simon اقتراح كيفية الستعرف على الأطفال الذين لا يستطيعون التعلم بكفاءة فى الفصول المدرسية العادلة. لذلك قام كل من بينيه وسيمون ببناء أول مقياس للذكاء يشتمل على ٣٠ مشكلة مرتبة تصاعدياً - بحسب درجة صعوبتها - وتطلب القدرة على الحكم، والفهم، والتعميل. ونشر هذا المقياس عام ١٩٠٥ ، وتمت مراجعته وتعديلاته عامي ١٩٠٨ ، ١٩١١ استناداً إلى البيانات الإمبريقية التى استمدت من تطبيق المقياس على طلاب المدارس الفرنسية. وأصبح المقياس يشتمل على عدد أكبر من الاختبارات الفرعية التى تم تصنيفها فى مستويات عمرية مختلفة.

وبذلك استند المقياس فى قياسه للذكاء إلى الوظائف العقلية العليا بدلاً من الوظائف الحسية البسيطة التى بينت دراسات كاتل عدم ارتباطها بالتحصيل الدراسي. واقتراح بينيه وسيمون مفهوم «العمر العقلى Mental Age» الذى يحدد أداء الفرد بالنسبة لأقرانه فى نفس مجموعة العمر استناداً إلى درجات المقياس. فإذا كان أداء طفل مساوياً متوسط أداء أطفال الثامنة من العمر مثلاً، فإن عمره العقلى يكون مساوياً ٨ سنوات، وقد يختلف عمر الطفل العقلى عن عمره الزمنى سواء بالزيادة أو النقصان. ولعل مفهوم العمر العقلى يُعد من الإسهامات المهمة فى مقياس الذكاء بعد تعديله.

غير أن ويلهلم شترن Stern وهو أحد علماء النفس الألمان يبدو أنه أول من اقترح مفهوم نسبة الذكاء IQ Intelligent Quotient الذى تدل على النسبة بين العمر العقلى إلى العمر الزمنى، وضرب القيمة الناتجة فى ١٠ ، واستخدمت هذه النسبة فى المقياس الذى تعديله عام ١٩٠٨ .

وقد أدى اختبار بينيه إلى انتشار فكرة قياس الذكاء فى العالم وبخاصة فى الولايات المتحدة الأمريكية. حيث قام لويس تيرمان Terman وهو أحد علماء النفس

بجامعة ستانفورد عام ١٩١٦ بمراجعة مقياس بيته وترجمته إلى اللغة الإنجليزية،
وأطلق عليه مقياس ستانفورد بيته للذكاء . Stanford - Binet Intelligence Scale

وقام تيرمان بمراجعة المقياس عام ١٩٣٧ بحيث اشتمل على صيغتين متكافئتين (L) ، (M) واستخدم مفهوم نسبة الذكاء التي اقتربها شترن، وأعاد مراجعته عام ١٩٦٠ ، وأصبح صيغة واحدة (L - M) ضمت أفضل مفردات كل من الصيغتين، وتوسيع المقياس بحيث أصبح يقيس مستوى الرائد المتفوق . واستبدل مفهوم نسبة الذكاء (IQ) بمفهوم نسبة الذكاء الانحرافية Deviation IQ (DIQ) التي أشرنا إليها في الفصل السادس ، والتي اقتربها ويكسنر Wechsler عالم النفس الأمريكي في مقياسه الذي أعده لقياس الذكاء، وهو من الاختبارات الفردية، ويسمى مقياس ويكسنر - بلفيو Wechsler - Bellevue Intelligence Scale للذكاء إلى مستشفى بلفيو للأمراض النفسية التي كان يعمل بها بمدينة نيويورك .

ونظرًا لأن مقياس بيته يُعد من المقاييس الفردية Individual Test التي تطبق على فرد واحد في وقت واحد، فقد اتجهت جهود مجموعة من علماء النفس وبخاصة آرثر أوتيس Otis وكان مسؤولاً عن دور نشر عالمية ، وميلر Miller بجامعة مينيسوتا، وثورنديك Thorndike بجامعة كولومبيا ، وتيرمان Terman، إلى بناء اختبارات جماعية Group Tests .

وقد شجع على ذلك قيام الحرب العالمية الأولى التي تطلب فرز وتصنيف مئات الآلاف من المجندين في الجيش في وقت قصير . لذلك تم إعداد أول اختبار جماعي لقياس الذكاء عام ١٩١٧ ويسمى اختبار الجيش ألفا Army Alpha Test وهو اختبار لفظي . ونظرًا لأن بعض المجندين في ذلك الوقت لم يكن مستواهم في القراءة جيداً فقد تم إعداد اختبار جماعي آخر مناسب للأميين ويسمى اختبار الجيش بيتا Army Beta Test وهو اختبار غير لفظي .

وقد اختبرت أعداد كبيرة بهذه الاختبارين عامي ١٩١٧ ، ١٩١٨ ، ويحلول عام ١٩٢٢ بلغ عدد اختبارات الذكاء الجماعية المنشورة ٤٤ اختباراً .

وعندما قامت الحرب العالمية الثانية انهز علماء النفس العاملين بالجيش التطور الذي حدث في بحوث القياس العقلاني وأعدوا اختباراً جماعياً للذكاء يعرف باسم اختبار التصنيف العام لأفراد القوات المسلحة Army General Classification Test (AGCT) طبق على ملايين الجنود عند التحاقهم بالجيش، وكان يشتمل على أربع صيغ متكافئة يتطلب تطبيق كل منها ساعة واحدة تقريباً . كما أعدَّ صيغ أطول للاختبار لكن يمكن تقسيم أداء المختبر إلى أربعة أقسام هي : القدرة اللفظية ، والفهم المكاني ، وإجراء العمليات الحسابية ، والاستدلال الحسابي . وقد حل اختبار التأهيل

للقوات المسلحة (AFQT) Army Forces Qualification Test محل هذا الاختبار بعد ذلك.

وقد أسهم في الانتشار السريع لاختبارات الذكاء الجماعية تعميم التعليم وتبني الأساليب الكمية في تطوير الإدارة المدرسية في الولايات المتحدة الأمريكية.

فقد أعدَّت كثير من اختبارات الذكاء الجماعية لجميع الأعمار من الحضانة إلى الجامعة ، وأدى ذلك إلى تعميم وتنفيذ برامج اختبارية على نطاق واسع متمثلة في اختبارات القبول بالكليات ، واختبارات انتقاء الأفراد وتصنيفهم وتسكينهم ، واختبارات الفئات الخاصة وغيرها.

ولعل أهم مصدر من مصادر المعلومات المتعلقة بالاختبارات العقلية هو الكتاب السنوي للقياس العقلي Mental Measurement Yearbook الذي حررَه أوскаر بيروس Buros عام ١٩٣٨ ، ولا يزال إصداره مستمراً حتى وقتنا الحالى . ويقدم هذا الكتاب السنوي قوائم مراجع شاملة تتعلق بالقياس النفسي والتربوى ، ومراجعات ناقدة للاختبارات المنشورة والاختبارات الجديدة ، والمؤلفات التي تتناول القياس النفسي والتربوى .

فمثلاً الكتاب السنوي السابع الذى صدر عام ١٩٧٢ وأشار إلى ١١٥٧ اختبار ، ٧٩٨ مراجعات اختبارات بواسطة ٤٣٩ ممحكم ، ١٨١ خلاصة مراجعات ، ٩٣ دورية سينكولوجية ، ١٢٣٧٢ مصدرًا من مصادر اختبارات سينكولوجية وتربوية .

غير أن هذا الانتشار المتزايد لاختبارات الذكاء لم يواكب تطورات موازية في أساليب ومنهجيات القياس النفسي بما يسهم في التغلب على العديد من المشكلات وأوجه النقد بل وأحياناً الهجوم على اختبارات الذكاء، وبخاصة مفهوم نسبة الذكاء كما سيتضح في سياق مناقشتنا لمفهوم الذكاء وكيفية قياسه.

مفهوم الذكاء والمشكلات المتعلقة به تعريفه :

يُعد مفهوم الذكاء من أكثر المفاهيم السينكولوجية التي يدور حولها الحوار والجدل ليس فقط بين علماء النفس والمشغلين بالقياس وإنما أيضاً بين مستخدمي اختبارات الذكاء والمختبرين بها، بل امتد الجدل خارج النطاق الأكاديمي إلى المسؤولين عن اتخاذ القرارات في مختلف المجالات التطبيقية التربوية والنفسية والاجتماعية. ويرجع ذلك إلى ما ترتب على هذا المفهوم وأساليب قياسه من مشكلات متعددة اجتماعية وتربوية، مثل إثارة الضيق بين أفراد المجتمع الواحد نتيجة تحيز مفردات اختبارات الذكاء ضد مجموعات أقلية تعيش في المجتمع نفسه، والتفرقة

العنصرية بين الأجناس ، والتشخيص غير الدقيق للأطفال المختلفين عقلياً ، والانتقام غير الصائب لأفضل الأفراد للالتحاق بدراسات متقدمة أو القيام بمهام معينة ، وغير ذلك من المشكلات . فمفهوم الذكاء يتسم بتنوعه وتجهاته وتتنوعها نظراً لعدم وضوح المقصود منه على وجه التحديد مما أدى إلى اختلاف وجهات نظر علماء النفس حوله . وعلى الرغم من أن الجيل الأول من اشغلاوا بالقياس انصب اهتمامهم على مشكلة تعريف الذكاء ، إلا أنهم لم يتوصلا إلى تعريف علمي لهذا المفهوم ، بل ولم يتفقوا على تعريف يمكن الإفاده منه عملياً .

غير أن هذا لم يمنع علماء القياس من الاستمرار في محاولاتهم لقياس الذكاء ، وبنائهم اختبارات تتميز درجاتها بالثبات وفاعليتها في التنبؤ ، مما يدل على أن هذه الاختبارات تقيس شيئاً معيناً . وقد أدى هذا الوضع إلى زيادة حدة الجدل حول هذا المفهوم أكثر مما ساهم في إلقاء الضوء عليه . فمفهوم الذكاء يصعب تعريفه تعريفاً واضحاً ومبشراً ليس بسبب طبيعة الذكاء ، وإنما بسبب طبيعة المفاهيم المجردة بعامة (Neisser 1979) .

وقد أسهم في هذا الوضع النظر إلى الذكاء على أنه شيء مادي أو كينونة لها وجود فعلى يمكن فحصها ، بدلاً من النظر إليه على أنه مصطلح يشير إلى سلوك له دلالته في حياة الإنسان . فالبعض يظن أن المصطلحات أو المفاهيم التي تقترح لوصف الظواهر السلوكية أو غيرها من الظواهر تحمل بين طياتها الشيء الذي يمثلها مما يؤدي إلى الخلط بين مصطلح معين وما يمثله من سلوك أو أفعال معينة ، فكلمة «طقس» مثلاً تصف فعل أو سلوك ظاهرة طبيعية معينة ، وكلمة «ذكاء» أيضاً تصف ظاهرة سلوكية معينة .

ففي كلتا الحالتين نقيس الفعل أو السلوك ولا نقيس المفهوم ذاته . غير أنه في حالة «الطقس» يمكن قياس ظواهر متعددة قياساً مباشراً ، وتشير في مجملها إلى ما اصطلاح على تسميته «طقس» بينما في حالة «الذكاء» ربما يكون الأمر أكثر صعوبة . وعادة نصدر في ضوء هذه القياسات أحکاماً تتعلق بالظاهرة أو ما تتضمنها .

فنقول مثلاً أن الطقس جاف أو بارد أو حار ، كذلك يمكن القول أن ذكاء الفرد مرتفع أو منخفض أو أن السلوك اللغوی للفرد أفضل من سلوكه غير اللغوی ، وهكذا . فقياس الشيء أو إمكانية قياسه إذن لا يعني أن هذا الشيء له وجود مادي .

الذكاء كمفهوم وصفى :

في ضوء ما تقدم عمل بعض العلماء الذين حاولوا تعريف الذكاء على تجنب التعريفات الصورية ، أو الجامدة للذكاء ، لذلك لجأوا إلى وصفه أو تحديد خصائصه أو

ما يترتب عليه من سلوك إنسانى ، وذلك على الرغم من أن بعضًا آخر منهم استخدم مفهوم الذكاء على أنه يعني الاستعداد الفطري للفرد مما لا يمكن قياسه قياساً مباشراً. ولعل ما أشار إليه كاتل (Cattell 1963) حول مفهوم الذكاء المائع Intelligence الذى اعتبره استعداداً أساسياً للتعلم وحل المشكلات مستقل عن التعليم والخبرة يعبر عن هذا التوجه .

أما الذكاء المتبلور Crystalized Intelligence الذى أشار إليه أيضاً فقد اعتبره ينمو نتيجة تفاعل الذكاء المائع للفرد مع بيئته أو ثقافته، ويشتمل على المعارف والمهارات المتعلمة . ومن أمثلة تعريفات الذكاء فى ضوء ما يترتب عليه من سلوك تعريف بيئي Binet وسيمون Simon (١٩٠٥) على أنه تجمع من القوى أو القدرات العقلية ، والحكم ، والحسن العملى ، والمبادرة ، والقدرة على التكيف مع الظروف المحيطة . وتعريف تيرمان Terman (١٩٢٠) على أنه القدرة على التفكير المجرد ، وتعريف ويكسنر Wechsler (١٩٩٨) على أنه قدرة الفرد الكلية على العمل الهداف ، والتفكير المنطقي ، والتعامل مع بيئته بفاعلية . غير أن بياجيه Piaget (١٩٨٩) يرى أن الذكاء امتداداً للتكييف البيولوجي ويتضمن عمليات التمثيل Assimilation وهى العمليات التى تعتمد على مخططات معرفية داخلية تتوافق لدى الفرد ، وعمليات المواجهة Accommodation وهى العمليات التى تهدف لتعديل المخططات المعرفية الداخلية لكي تتواءم مع المثيرات البيئية الجديدة التى يواجهها الفرد .

فعمليات التمثيل تسهم في الابتعاد عن السلبية في التعامل مع البيئة ، بينما تسهم عمليات المواجهة في ضبط توجُّه الذكاء نحو التمثيل المعرفي الذي يتفق مع العالم الواقعي .

وعلى جانب آخر نلاحظ أن Thorndike (١٩٢٨) حاول الابتعاد في أعماله الرئيسية المتعلقة بقياس الذكاء عن تعريف الذكاء تعريفاً صورياً Formal Definition ، حيث يبدو أنه لم يقترح تعريفاً محدداً في سياق هذه الأعمال ، فقد رأى أن الذكاء يتعلق بمقارنة الفرد بأقرانه من حيث قدرته على إنجاز تسلسل من المهام التي اعتبرها نواتج عقلية .

وفي ضوء هذا الإرباك في تعريف الذكاء اقترح بيرت Burt (١٩٤٧) أنه من المناسب تذكر التمييز بين التعريف الاسمي Nominal Definition ، والتعريف الواقعي Real Definition .

فالتعريف الاسمي يتعلق بكيفية استخدام المفهوم أو المصطلح، أما التعريف الواقعي فيتعلق بتفسير طبيعة المفهوم المراد تعريفه، ويرى أنه ربما يكون من الضروري تعريف الذكاء تعريفاً اسمياً قبل التفكير في كيفية قياسه حيث إن هذا يتطلب مزيداً من البحث.

ويبدو أن هذا الاقتراح كان له صدى واسعاً بعد ذلك في الأوساط السيكولوجية، والدليل على ذلك تعدد المؤتمرات المتخصصة التي انعقدت في الولايات المتحدة الأمريكية وكندا لمناقشة مجال الذكاء الإنساني. ومثال ذلك مؤتمر تورonto بكندا «حول الذكاء On Intelligence» عام ١٩٧٩، وسلسلة المؤتمرات التي عقدها مركز بحوث وتطوير التعلم بجامعة يتسيرج بالولايات المتحدة الأمريكية المتعلقة ب المجالات مهمة في سيكولوجية التعليم، وكان أحد هذه المجالات «مجال الذكاء وطبيعته The Nature of Intelligence» وذلك عام ١٩٧٤ وأسهم فيه نخبة من علماء القياس وختصاصي علم النفس المعرفي، والندوات المتعددة التي نظمتها الجمعية الأمريكية للبحث التربوي من خلال اجتماعها السنوي، مثل الندوة التي عقدت في تورonto بكندا عام ١٩٧٨، وندوة سان فرانسيسكو عام ١٩٧٩ حول الذكاء الإنساني Human Intelligence، وكذلك إصدار دورية الذكاء Intelligence وتُعد من الدوريات السيكولوجية المتخصصة التي تهتم بنشر البحوث والدراسات المبتكرة التي تسهم في إثراء الفكر في مجال الذكاء الإنساني من حيث طبيعته ووظائفه وقياسه.

وقد اهتمت هذه المؤتمرات اهتماماً خاصاً باقتراح التوجهات المستقبلية لمفهوم الذكاء الإنساني ، وأساليب قياسه من منظور علم النفس المعرفي Cognitive Psychology، وقد تبنت بعض الدراسات التي قدمت في هذه المؤتمرات مفهوم «الشخص الذكي Intelligent Person» بدلاً من مفهوم «الذكاء Intelligence» حيث اعتبرت أن هذا الشخص يعد مثلاً نموذجياً تصوريًّا Prototype يقارن به الأشخاص الآخرين لمعرفة مدى مماثلتهم له، وتعتمد ثقتنا في الحكم بأن أحد الأشخاص الآخرين «ذكي» على مماثلته بوجه عام لهذا المثال النموذجي، كما هو الحال عندما نطلق على أحد الأشياء اسمًا معيناً مثل «مقدع»، فثقتنا في هذه التسمية تعتمد على مماثلة هذا الشيء للمثال النموذجي للمقدع. وتفضل روش Rosch وزملاؤها استخدام لفظ «قسم Category» في تحليل بنية المفاهيم Concepts (Rosch, 1978 ; Rosch , Mervis , Gray , Johnson & Boyes , 1976).

وينطبق هذا أيضاً على مفهوم «الشخص الذكي». وعلى الرغم من أن هؤلاء الباحثين يرون أنه لا توجد مجموعة معينة من الخصائص تُعرف هذا «القسم» تعريفاً دقيقاً، إلا أنه يجب أن تكون هناك خصائص يشارك فيها أعضاء هذا «القسم» لكي يُطلق عليه «قسمًا Category».

ولعل «المثال النموذجي Prototype» للشخص الذكي يُعد أحد العناصر التي تنتهي إلى هذا القسم إذا توافرت فيه جميع الخصائص المشتركة، غير أنه لا توجد دراسات علمية منظمة لهذه الخصائص، ولكن يمكن تتبع بعض الدراسات الاستطلاعية في هذا الشأن مثل دراسة ديب (Dube, 1977) وغيرها ، حيث بينت عدم الاتفاق على هذه الخصائص . ولعل هذا يرجع إلى عدم وجود مいく محمد للذكاء نظراً لتنوعه، إذ يمكن أن يتميز شخصان بالذكاء على الرغم من اختلافهما في كثير من السمات، أى أنهما يمثلان «المثال النموذجي التصوري للشخص الذكي» في أبعاد مختلفة .

لذلك يرى نايسر (Neisser, 1979) أن ما نطلق عليه «مقعد Chair» لمماثلته لمودج مقعد مثالي تصوري Prototype لا يعني أن هناك خاصية تسمى «مقعديّة Chairness» ، كذلك ما نطلق عليه «شخص ذكي» لمماثلته «النموذج شخص ذكي مثالي تصوري» لا يعني أن هناك شيئاً ما يسمى «ذكاء Intelligence» . فالمماثلة تعد حقيقة خارجية وليس مضموناً أو جوهراً داخلياً، وبذلك لا نستطيع تعريف الذكاء تعريفاً عملياتياً Process - Based نظراً لأنه متعدد الخصائص، ولكنه يعد مماثلة بين شخصين أحدهما شخص فعلى والآخر نموذج مثالي تصوري، وهذه المماثلة هي التي تعمل على قياسها في أبعاد متعددة.

ومن الجدير بالذكر أن ثورنديك Thorndike (١٩٢٤) باعتباره أحد الرواد الأوائل المرموقين في القياس النفسي كان قد اقترح منظوراً مماثلاً لهذا المنظور الجديد لمفهوم الذكاء .

ولعل هذا المنظور يلفت النظر إلى أنها لا نستطيع تعريف الذكاء تعريفاً لفظياً، ومع هذا يمكن قياسه وذلك بضم قياسات جميع المكونات المناسبة بأوزان اختيارية لكل مكونة منها للحصول على مؤشر عام .

غير أن كثيراً من هذه المكونات لا نستطيع قياسها بأى أسلوب مقتن، فخصائص المثال النموذجي للشخص الذكي لا تقتصر على الطلاقة اللغوية ، والقدرة المنطقية، والمعلومات العامة ، وإنما تتضمن أيضاً الفطنة ، والذوق العام ، والتفكير الابتكاري ، وعدم التحيز ، والاستقلال الفكري ، والتفتح العقلى ، وغير ذلك .

وبعض هذه الخصائص تمثل في مواقف حياتية مميزة ، والبعض الآخر لا نستطيع تقويمه إلا في ضوء حياة الفرد ككل ، ومع هذا فإن علماء القياس يسعون إلى قياس الخصائص الأكثر تعلقاً بالمثال النموذجي للشخص الذكي (Neisser , 1979) .

الذكاء كمجموعة من القدرات المتوازنة :

لم يغفل تاريخ القياس النفسي والتربوي البحث عن مؤشرات كمية مقبولة للذكاء الإنساني، فمعظم علماء النفس الرواد في مجال الذكاء وقياسه اعتبروا الذكاء مكون من قدرة عامة وقدرات خاصة متعددة.

فنسبة الذكاء IQ كانت وسيلة لهم في تقدير كم هذه القدرات وتلخيصها بقيمة عددية واحدة للفرد تمثل ذكاءه، كما هو الحال عند تمثيل توزيع مجموعة من الدرجات بقيمة عددية واحدة تعبّر عن متوسطها الحسابي. غير أن هؤلاء العلماء لم يتلقوا على القدرات الأساسية التي تمثل الخصائص التكوينية للذكاء. ولتوسيع ذلك نعرض فيما يلي بإيجاز بعض التوجهات التي كانت لها أهمية تاريخية ويتناولها الباحثون في وقتنا الحاضر بالدراسة.

الذكاء كعامل عام :

إن أول محاولة علمية لوصف العناصر التي تشتراك فيها جميع اختبارات الذكاء استناداً إلى الأساليب الإحصائية للدرجات هذه الاختبارات قام بها شارلز سبيرمان Spearman عام ١٩٠٤ ، وأعاد المحاولة عام ١٩٢٧.

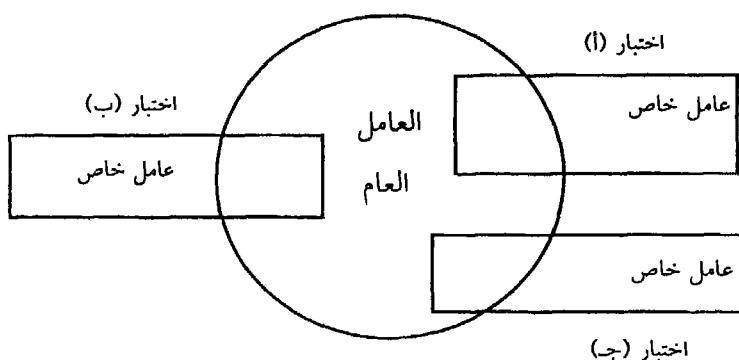
ولعله كان من أوائل الذين استخدمو أسلوب التحليل العاملى Factor Analysis الذي أشرنا إليه في الفصل الخامس في تحليل البيانات التي استمدتها من تطبيق اختبارات الذكاء على تلاميذ المدارس الإنجليزية.

وقد توصل إلى أن جميع مظاهر النشاط العقلي كما تُقاس باختبارات الذكاء تشتراك في عامل عام (g) General Factor عرّفه بأنه القدرة العامة على استنباط العلاقات المجردة. وعلى الرغم من أن بعض المهام المعرفية تتطلب هذه القدرة العامة أكثر من غيرها من المهام ، إلا أن جميع هذه المهام تتضمن هذا العامل العام إلى حد ما.

غير أن سبيرمان لاحظ أنه إذا كان العامل العام (g) بمفرده يفسر الذكاء ، فإنه من المتوقع أن يكون معامل الارتباط بين اختبارات الذكاء المختلفة تاماً ، أي أن جميع قيمه تساوى الواحد الصحيح. غير أنه وجد أن بعض هذه القيم أقل من ذلك بكثير ، مما أدى به إلى افتراض وجود عوامل أخرى تسهم في هذا التفسير أطلق عليها العوامل النوعية (S) . Specific Factors

فكل اختبار للذكاء يقيس نشاطاً عقلياً يمكن تحليله إلى عاملين أحدهما عام (g) مشترك بين هذا النشاط أو محتواه ، وهو ما أسماه الذكاء العام General Intelligence وعوامل خاصة (S) تتعلق بهذا النشاط دون غيره من الأنشطة.

فالعامل العام (g) إذن يسهم في الارتباط بين اختبارات الذكاء، بينما تعمل العوامل الخاصة على تحجيم الارتباط حتى لا تصل قيمته إلى الواحد الصحيح. لذلك فإن اختبارات الذكاء التي تبني في ضوء هذا التوجّه يتبع عنها درجة كلية واحدة للذكاء بحيث تعبّر بقدر الإمكان عن العامل العام دون الاهتمام بالعوامل الخاصة نظراً لأنها غير مشتركة بين الاختبارات المختلفة. ويوضح الشكل التخطيطي (٩ - ١) نظرية العامل العام لسييرمان:



شكل (٩ - ١) يوضح نظرية العامل العام لسييرمان

ويتضح من شكل (٩ - ١) أن كل مستطيل يُمثل اختباراً يقيس العامل العام (g) بقدر الإمكان. والمساحة الناتجة من تقاطع كل مستطيل مع الدائرة التي تمثل العامل العام تدل على مقدار ما يقيسه الاختبار من العامل العام. أما أجزاء المستطيلات التي تقع خارج الدائرة فإنها تمثل العامل الخاص بكل اختبار.

ومن الجدير بالذكر أن سييرمان أجرى تعديلاً على نظرية العامل العام بحيث تأخذ بعين الاعتبار العوامل التي لم تفسّر بواسطة العامل العام أو العامل الخاص، فقد وجد أن هناك بعض المهام العقلية الخاصة ولكنها تكون مرتبطة فيما بينها. لذلك أطلق عليها «العوامل الطائفية Group Factors»، وهذه العوامل أقل اتساعاً من العامل العام وأقل تخصصاً من العوامل الخاصة، أي أنها تتوسط بين كل من نوعي العوامل.

ولعل هذه النظرية قد أسهمت في الفكر التربوي في ذلك الوقت عندما كان يفترض أن التلميذ الذكي أو الموهوب يكون كذلك في جميع المجالات الدراسية دون تخصيص.

وقد قوبلت هذه النظرية بعد ذلك بهجوم شديد من جانب علماء النفس في الولايات المتحدة الأمريكية بحجة أن هذه النظرية قد اعتمدت في تحليلاتها الإحصائية

على عدد ضئيل من التلاميذ الذين طُبّقت عليهم الاختبارات ، وهذه العينة لا تصلح أساساً للتوصل إلى ما أطلق عليه سبيرمان «العامل العام» . كما أن الأسلوب الإحصائي الذي اتبّعه سبيرمان لا يوضح الأوزان النسبية للعوامل التي تتدخل في كل اختبار .

وقد اعترض جيلفورد (Guilford, 1976) على فكرة أن جميع اختبارات الذكاء ترتبط درجاتها ارتباطاً موجباً، حيث أكد أن البيانات الإمبريقية التي جمعها لم تؤيد ذلك . وأشار إلى مئات من الارتباطات الصفرية بين درجات هذه الاختبارات المستمدّة من عينات بلغ عدد أفراد كل منها ما يقرب من ٢٢٥ فرداً . كما أشار إلى أن قيم الارتباطات الموجبة التي حصل عليها سبيرمان ربما ترجع إلى استخدامه عدداً قليلاً من الاختبارات المحدودة في محتواها العاملى . وأهم من ذلك أن هذه الاختبارات يبدو أن تكوينها العاملى مرکب مما أدى إلى التداخل فيما بينها .

ويرى جيلفورد أن التوصل إلى ما يسمى «العامل العام» (g) يتطلب عينات غير متجانسة من الأفراد ، نظراً لأن هذا العامل ربما يمثل التباين في العمر الزمني أو النوع أو التعليم وما إلى ذلك . فإذا ارتبطت درجات اخبارين بالعمر الزمني فإن الاختبارين سوف يرتبطان بعضهما البعض نتيجة لذلك .

الذكاء كعوامل متعددة :

ينظر علماء النفس الأمريكيون إلى الذكاء نظرة مختلفة . فمثلاً ربما يكون فرد لديه قدرة لفظية مرتفعة ، وقدرة مكانية وأخرى عددية منخفضة . ويعتمد عدد القدرات الأساسية التي تفترض في هذه الحالة على الأساليب الإحصائية المستخدمة لاستخلاص هذه القدرات التي تسمى «عوامل Factors» ، كما يعتمد عددها على الإطار النظري الذي يتبنّاه الباحث في القدرات العقلية .

ولعل لويس ثيرستون Thurstone أحد علماء النفس البارزين بجامعة شيكاغو الأمريكية (١٩٣٨) هو الذي اهتمّاً ملحوظاً بمدخل العوامل المتعددة في الذكاء . حيث تركّزت كثير من دراساته وبحوثه في هذا المجال . فقد امتد نشاطه إلى تطوير أساليب ومناهج البحث في مشكلة التكوين العقلى وبخاصة منهج التحليل العاملى .

فقد سبق أن ذكرنا أن التحليل العاملى يُعد طريقة منتظمة لفحص مكونات الاختبارات وارتباطها بعضها البعض . وبعد تطبيق مجموعة من الاختبارات على عينة من الأفراد نحاول باستخدام منهج التحليل العاملى تحديد أقل عدد من العوامل أو القدرات المميزة التي يمكن قياسها بدرجات متسقة والتي تكمّن وراء أداء أفراد العينة في مجموعة الاختبارات .

وقد توصل ثيرستون إلى مجموعة من القدرات العقلية الأولية Primary Mental Abilities نتيجة تطبيق منهج التحليل العاملى على ٦٥ اختباراً طبقها على مجموعة من الطلاب في مرحلة التعليم الثانوى، واستطاع تحديد مجموعة من العوامل المستقلة نسبياً عن بعضها البعض الآخر، وإعداد اختبار لقياس هذه العوامل يسمى اختبار القدرات العقلية الأولية.

وهذه العوامل هي :

(١) القدرة على الطلاقة اللغوية : القدرة على استخدام الكلمات في مواقف مختلفة.

(٢) القدرة على الفهم اللغوى : فهم معانى الكلمات وال العلاقات القائمة بينها فى الجمل.

(٣) القدرة العددية: القدرة على إجراء العمليات الحسابية البسيطة بسرعة ودقة.

(٤) القدرة المكانية : القدرة على إدراك العلاقات الهندسية والمكانية الثابتة والتي يجري عليها تحويلات معينة.

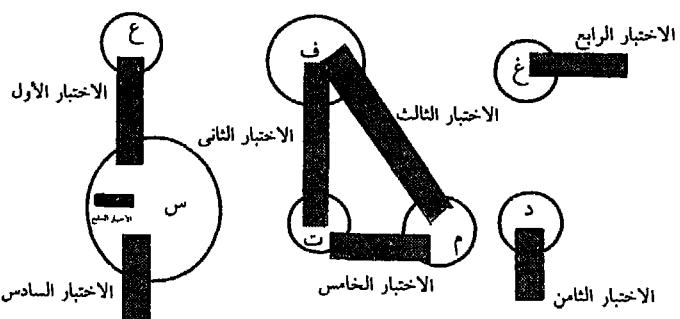
(٥) السرعة الإدراکية : الملاحظة السريعة والدقيقة لتفاصيل الأشياء المنظورة.

(٦) القدرة على التذكر : القدرة على الاسترجاع الفورى لمثيرات محددة.

(٧) القدرة على الاستدلال الاستقرائي : ويقصد به الاستباط العام، أي القدرة على استخلاص القواعد أو المبادئ.

ويمكن تمثيل نظرية ثيرستون أو نموذجه متعدد العوامل بالشكل التخطيطي

: (Sax, 1969) التالى (٢_٩) :



شكل (٩ - ٢) يوضح نظرية العوامل المتعددة لثيرستون

ويتبين من شكل (٩ - ٢) أن كل دائرة تمثل عاملًا مستقلاً تماماً عن غيره من العوامل. فالاختبار الأول يقيس بدرجة ضئيلة العامل العددي (ع)، وعامل

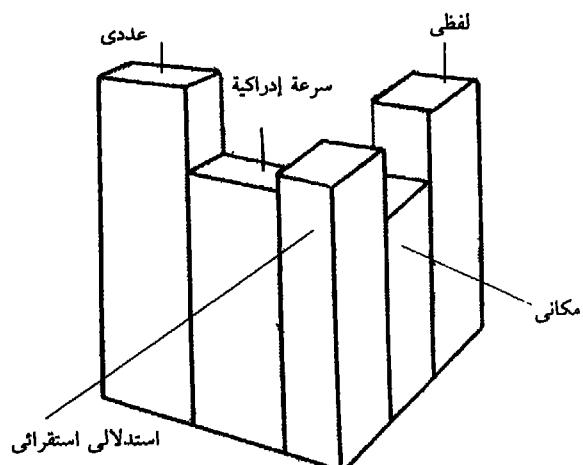
الاستدلال الاستقرائي (س)، والاختبار الثاني يقيس بدرجة كبيرة كل من العامل اللفظي (ف)، وعامل التذكر (ت)، ولكنه لا يصلح لقياس العامل المكانى (م)، والاختبار الثالث يقيس بدرجة أكبر العامل اللفظي وبدرجة أقل العامل المكانى، أما الاختبار الرابع فلا يصلح إلا لقياس العامل اللغوى (غ)، والاختبار الخامس يقيس كل من عامل التذكر (ت) والعامل المكانى (م)، أما الاختباران السادس والسابع يُعدان مقياسين جيدين لعامل الاستدلال الاستقرائي (س)، بينما يُعد الاختبار الثامن مقياساً تقنياً نسبياً لعامل السرعة الإدراكية (د).

وقد أعيد تقييم وتحليل القدرات العقلية الأولية عام ١٩٦٢ وتم التوصل إلى خمسة عوامل هي:

العامل اللفظي، والعامل العددى، وعامل الاستدلال الاستقرائي، وعامل السرعة الإدراكية، والعامل المكانى.

غير أن العوامل الثلاثة الأولى استُخلصت من نتائج تطبيق الاختبارات على تلاميذ الصفوف من الرابع الابتدائى إلى الثانوية العامة، أما العامل الرابع فقد اعتمد على نتائج تطبيق الاختبارات على الصنفوف من الحضانة إلى الصف السادس.

ويوضح الشكل التخطيطى (٩ - ٣) التالى هذه العوامل:



شكل (٩ - ٣) القدرات العقلية الأولية بعد إعادة تقييم وتحليل اختبار ثيرستون

وقد حاول ثيرستون فيما بعد تحليل العلاقات بين العوامل الأولية التي استخلصها لكي يتوصل إلى ما أسماه العوامل العامة من الدرجة الثانية - Second Order General Factors بغرض تفسير الارتباطات بين القدرات العقلية الأولية التي كانت قيمها متوسطة حيث وجد أن وسيط هذه القيم 35% . وعلى الرغم من أن نظرية ثيرستون استبعدت في البداية العامل العام (g) الذي افترضه سبيرمان كمكون له دلالتها في الوظائف العقلية، إلا أن الارتباطات بين القدرات العقلية الأولية التي وجدتها ثيرستون في تحليله الأخير جعله يفترض وجود عامل من الدرجة الثانية ربما يكون له علاقة بالعامل العام (g).

وبالطبع يمكن استخلاص «عوامل Factors» أخرى إذا أضفنا مهامًا عقلية مختلفة إلى ما تتضمنه الاختبارات التي استخدمناها ثيرستون وتطبيقاتها على عينات أخرى من الأفراد من ذوى خلفيات ثقافية متباعدة.

وعمومًا فإن استخدام نظرية العوامل المتعددة لثيرستون لتقييم الذكاء يتطلب تطبيق بطارية من الاختبارات البسيطة التي يصمم كل منها لتقييم عامل خاص أو أولي من العوامل السابقة، مثل اختبار القدرات العقلية الأولية لثيرستون ، وبطارية الاستعدادات الفارقة التي سوف نوضحها في الفصل العاشر.

الذكاء كبنية ثلاثية الأبعاد :

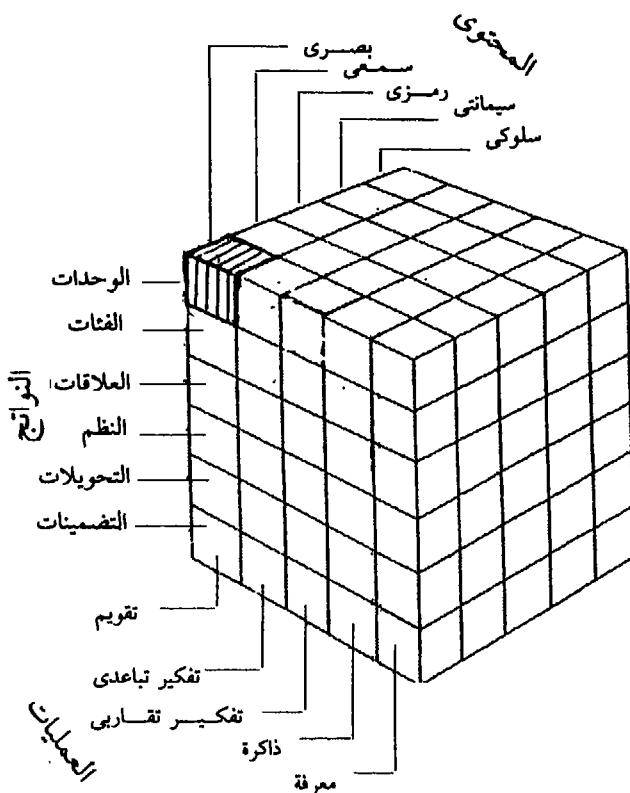
توجد محاولات أخرى من جانب علماء القياس النفسي لتصنيف العوامل التي استخلصتها البحوث المختلفة نتيجة استخدام منهج التحليل العاملى، ومن أهم الجهود فى هذا المجال بحوث جليفورد Guilford وزملاؤه بجامعة جنوب كاليفورنيا التي نمى توسيعها فى سلسلة من التقارير بعنوان عام «تقارير من معهد علم النفس».

وتتضمن هذه السلسلة أعمال جليفورد وزملاؤه، والأساليب والمنهجيات والبحوث التي أنجزها الفريق البحثي على مدى عقدين من الزمان. وفي كتابه الذى نشر عام 1967 بعنوان «طبيعة الذكاء الإنساني» اقترح جليفورد نموذجه المعروف «بنية العقل (SOI) Structure of Intellect» الذى يسهم فى الوظائف الكشفية التنبؤية Hueristic Model لتكوين الفروض المتعلقة بعوامل جديدة للذكاء. لذلك افترض بهذه البنية نظاماً ثلاثي الأبعاد أو الأوجه Facets يتضمن العديد من القدرات العقلية التي صنفت استناداً إلى هذا النظام كالتالى:

- (1) المحتوى Content: أي المادة التي يجري تناولها.
- (2) العمليات Processes أو الإجراءات Procedures التي تؤدي على المحتوى.

(٣) النواتج Products : أي ما ينتج عن أداء العمليات على المحتوى، أو صيغة الفكر الناتج.

وقد قسم كل من هذه الأبعاد أو الأوجه بعد تحليلها إلى أقسام فرعية، أربعة للمحتوى، وخمسة للعمليات، وستة للنواتج، وبذلك يصبح العدد الكلى للأقسام : $4 \times 5 \times 6 = 120$ قسمًا . غير أن جيلفورد عام ١٩٧٠ أعاد النظر في البُعد المتعلق بالمحتوى حيث قسم أحد مكوناته وهو المحتوى الشكلي إلى محتوى بصري، ومحتوى سمعي ، وبذلك أصبح عدد أقسام المحتوى خمسة ، والعدد الكلى للأقسام : $5 \times 5 \times 6 = 150$ قسمًا موضحة بالشكل التخطيطي (٩ - ٤) التالي :



شكل (٩ - ٤) يوضح نموذج جيلفورد ثلاثي الأبعاد أو الأوجه

وي يكن توضيح مكونات الأبعاد الثلاثة التي يمثلها شكل (٩ - ٤) كالتالي :

(١) المحتوى Content : ويعتمد أساساً على التمييز بين الأنواع المختلفة من

المعلومات التي يمكن أن تتضمنها مشكلة أو مهمة معينة يجدها الفرد وي العمل عليها العقل.

وينقسم المحتوى في ضوء ذلك إلى خمسة أقسام هي :

المحتوى السلوكي Behavioral : ويتضمن معلومات تتعلق بسلوك الأفراد والتفاعل فيما بينهم، ويكون السلوك أساساً غير لفظي حيث يتركز الاهتمام على الوعي أو الانتباه، والإدراكات ، والرغبات، والمشاعر، والأمزجة، والانفعالات، والتصرفات. وقد أضيف هذا القسم على أساس نظرى بحث ليمثل مجالاً عاماً يطلق عليه أحياناً «الذكاء الاجتماعي».

المحتوى السيمانتي Semantic : يتضمن معانى لفظية وأفكار.

المحتوى الرمزي Symbolic : يتضمن معلومات رمزية مثل الحروف والأرقام والإشارات المتعارف عليها.

المحتوى الشكلي Figural: يتضمن مواد عيانية يمكن إدراكتها من خلال الحواس، وينقسم إلى محتوى بصري ومحنتوى سمعى.

٢) العمليات processes : هي أساليب النشاط العقلى التي يفترض إجراؤها على محتوى أو معلومات معينة. وتتنقسم العمليات إلى خمسة أقسام هي :

الإدراك المعرفي Cognition : يتعلق بالأنشطة العقلية المعرفية، مثل الفهم والاستيعاب، واكتشاف المعلومات بمختلف أنواعها، وإعادة اكتشافها أو التعرف عليها.

الذاكرة Memory : تتعلق بالاحتفاظ بالمعلومات المتعلمة بحيث يمكن استدعائها بشكلها الأصلى عند الحاجة إليها.

الفكر التقاري Convergent Thinking : يتعلق بالبحث عن حل صحيح لمشكلة معينة، أو التوصل إلى معلومات محددة من معلومات أخرى معطاه، ويكون التأكيد هنا على المعلومات المتعارف عليها أو التي تجوز قبولاً، وربما تحدد المعلومات المعطاة الاستجابة تحديداً كاماً.

التفكير التباعدى Divergent Thinking: يتعلق بالبحث عن حلول ابتكارية متعددة تتميز بعنصر الجدة على الرغم من وحدة مصدر المعلومات.

التقويم Evaluation يتعلق بإصدار أحكام قيمة على المعرفة والفكير من حيث مدى صلاحية المعلومات وملاءمتها وكفايتها وال الحاجة إليها وما إلى ذلك.

(٣) النواتج Products : هي نتائج أداء العمليات على المحتوى، أي شكل الفكر الناتج. وتنقسم هذه النواتج إلى ستة أقسام هي :

الوحدات Units : تتعلق بإنتاج كلمة أو تعريف أو معلومات بسيطة متمايزه.

الفئات Classes : تتعلق بإنتاج وحدات تشتراك في بعض الخصائص مثل المفاهيم، وللحظة أوجه التشابه بين الوحدات.

العلاقات Relations : تتعلق بإنتاج أشكال مختلفة من العلاقات مثل أوجه الاتفاق أو الاختلاف أو التشابه.

النظم Systems : وتعمل بإنتاج مجموعة من التصنيفات متعددة الأشكال أو المحتوى بحيث تتميز بالاتساق الداخلى، مثل نظام بناء الجمل أو نظام العمليات الرياضية.

التحوليات Transformations : تتعلق بتغيير المعانى والتنظيمات والترتيبيات وتحويل المعلومات من صيغة إلى أخرى.

التضمينات Implications تتعلق بإنتاج معلومات تتخطى حدود البيانات المعطاة مثل التوقع أو التنبؤ المستقبلى.

ويتبين من شكل (٩ - ٤) أن إسهامات جيلفورد الرئيسة تتعلق بتطوير نظام تصنيفى Taxonomy لعوامل الذكاء الإنسانى فى ثلاثة أبعاد أو أوجه لمكعب يشتمل على ١٥ خلية تمثل كل منها عاملًا مركبًا تركيبيًا فريدًا من محتوى، وعملية، وناتج. فمثلاً الخلية الأولى المظللة تمثل عامل معرفة وحدات الأشكال، وهكذا فى بقية الخلايا، والوجه العلوى للمكعب يمثل طبيعة المحتوى أو المادة التى تتضمنها مفردات الاختبار.

فالفردات التى محتواها شكلى تتناول مثلاً تشابه الأشكال، وال العلاقات المكانية، والتصور البصرى المكانى، والغلق الإدراكي. والفردات التى محتواها رمزى تتناول الكلمات المختلفة وغير المختلفة ، وتقىس القدرة على إنتاج أكبر من عدد من الكلمات التى تبدأ بحرف معين، كما تتناول الأعداد والعمليات الحسابية. والفردات التى محتواها سيمانتى تتناول المعانى اللفظية مثل المشابهات اللفظية. والفردات التى محتواها سلوكي تتناول تعبيرات الوجه، وال العلاقات الاجتماعية، والتفاعل الاجتماعى، وتضمينات تغير صورة وجه شخص بصورة أخرى مثلاً، وما يتربى على ذلك من تغيير المعانى نتيجة تغير الشخصيات والمشاعر والمقاصد والأفكار، وكذلك التنبؤ من

موقف تمثيلي معين ب موقف آخر، استناداً إلى مشاعر ومقاصد شخصيات الموقف التمثيلي، وهكذا.

وقد أعد جيلفورد وزملاؤه اختبارات تقيس كثيراً من هذه العوامل التي عددها ١٥ عاملاً. فحتى عام ١٩٧١ كان عدد العوامل التي أمكنهم قياسها ٩٨ عاملاً (Guilford & Hoepfner, 1961).

وعلى الرغم من أن كثيراً من علماء النفس أشاروا إلى وجود ارتباطات مرتفعة بين الخلايا أي بين العوامل المختلفة، إلا أن النموذج قد استثار الفكر السيكولوجي والتربوي، والدليل على ذلك كثرة البحوث التي أجريت حول كل من النموذج وسلسلة الاختبارات التي صممت لقياس كثير من العوامل التي يتضمنها.

وقد أكد النموذج فكرة الإنسان المتفاعل ، كما أكد مفهوم «المعلومات Information» التي يعمل الإنسان على التمييز بينها في ضوء أنواع مختلفة من النواتج، مثل الوحدات والعلاقات والنظم. فالأقسام الخمسة للمحتوى عندما يتم ضمّها في جميع الطرق الممكنة لأقسام النواتج ستؤدي إلى ٣٠ قسماً فرعياً من المعلومات.

فهذا التصنيف المنظم يمكن النظر إليه على أنه نظرية سيكولوجية للمعرفة تتضمن ٣٠ قسماً فريداً تُعرف الأنواع الرئيسة للمعلومات. وهذا التصنيف للمعلومات لا يفيد فقط في إثراء النظرية السيكولوجية وإنما أيضاً له تطبيقاته في مجال التربية وتقنيات الحاسوب.

إذ يمكن اعتبار الإنسان كالحاسوب في عمليات تجهيز ومعالجة المعلومات Information Processing ، وهذا التجهيز يستند إلى الأنواع الخمسة من العمليات التي أشار إليها النموذج. غير أن فعل التجهيز Processing Action يختلف باختلاف نوع المعلومات التي يتم تجهيزها أو تناولها .

وهذا التفاعل بين نوع العملية ونوع المعلومات يُعد مسؤولاً عن التمايز بين القدرات التي تمثلها خلايا المكعب الذي يمثل النموذج (Guilford, 1967). والتشابه بين الإنسان والحاسوب، أو إمكانية استخدام نموذج بنية العقل (SOI) في تقنيات الحاسوب ليست أفكاراً خيالية، فقد اقترح كريستنسن (Christensen, 1963) الاسترشاد بأقسام هذا النموذج في توضيح وظائف الحاسوب، وأشار إلى التوازي بين بعض العوامل العقلية وبعض برامج الحاسوب.

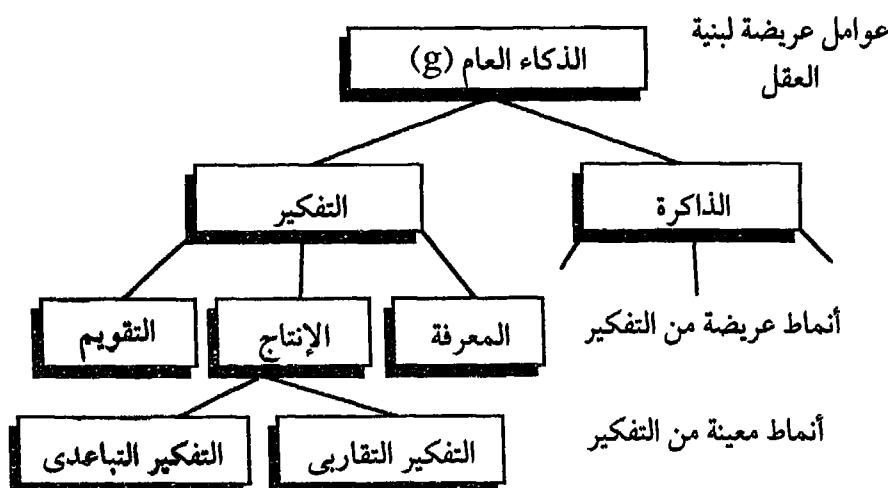
أما في المجال التربوي، فقد أشارت كثير من الدراسات إلى تطبيقات متعددة لهذا النموذج، حيث استخدمت ميكر (Meeker, 1964) النموذج في تقويم كل من

المنهج وقدرات الأطفال، وبينت أنه عند الحديث عن الذكاء يجب على المربين تحديد مجال معين. فعندما يكون الطفل متوسط القدرة ويُخفق في التعلم فإن نسبة ذكائه (IQ) لا تفسر هذا الإخفاق ولا تبرره. وترى أنتا نحتاج إلى اختبارات تستند إلى أساس نظرية تزود المعلم بمعلومات تشخيصية تساعد في اتخاذ قرارات تعليمية تتعلق بطفلك معين.

لذلك فإن نموذج جيلفورد يُعد أحد طرق تنظيم القدرات في المناهج المدرسية بدلاً من تركيز هذه المناهج على بعض القدرات وبخاصة قدرات الذاكرة على حساب قدرات التفكير وبخاصة التفكير التباعدي والتقويم، وكذلك تركيزه على المحتوى السيمانتي والمحتوى الرمزي، على حساب المحتوى الشكلي، والمحتوى السلوكى، وتركيزه على نوائح الوحدات، والفتات، على حساب نوائح العلاقات، والنظام، والتحوليات، والتضمينات.

وعلى الرغم من أن أبعاد أو أوجه النموذج الثلاثة لم يتتأكد انتظامها في نسق هرمي Hierarchical إلا أن جيلفورد قارن نموذجه بتصنيف بلووم Bloom's Taxonomy للاحظ أوجه شبه كثيرة بينهما مما جعله يقترح استخدام نموذجه في تصنيف الأهداف التربوية، وذلك لأنك أكثر دقة ويسعى بتحديد هذه الأهداف تحديداً إجرائياً.

ويوضح الشكل التخطيطي (٩ - ٥) التالي اقتراح جيلفورد فيما يتعلق بالنسق الهرمي لنموذجه (Guilford, 1959) :



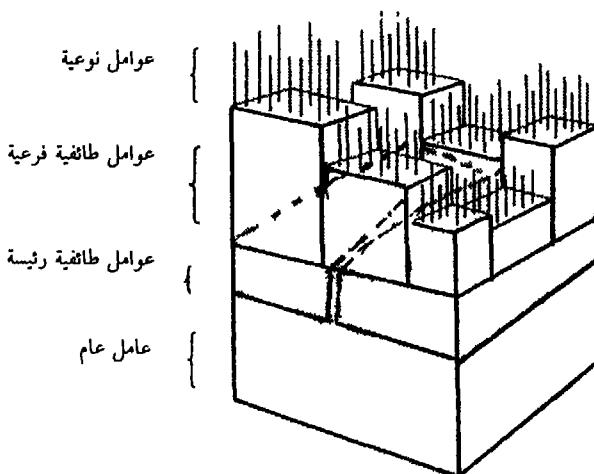
شكل (٩ - ٥) نموذج هرمي لعوامل بنية العقل لجيلفورد

الذكاء كنظام هرمي من العوامل :

اتضح لنا أن نموذج بنية العقل جيلفورد قدّم أساساً نظرياً لطبيعة العقل وأدى إلى تصنيف عدد كبير من العوامل المفصولة، ولكنه لم ي بين العلاقات القائمة بين هذه العوامل مما يتطلب التوصل إلى نموذج رياضي يُستند إليه في كشف هذه العلاقات أو توضيحها.

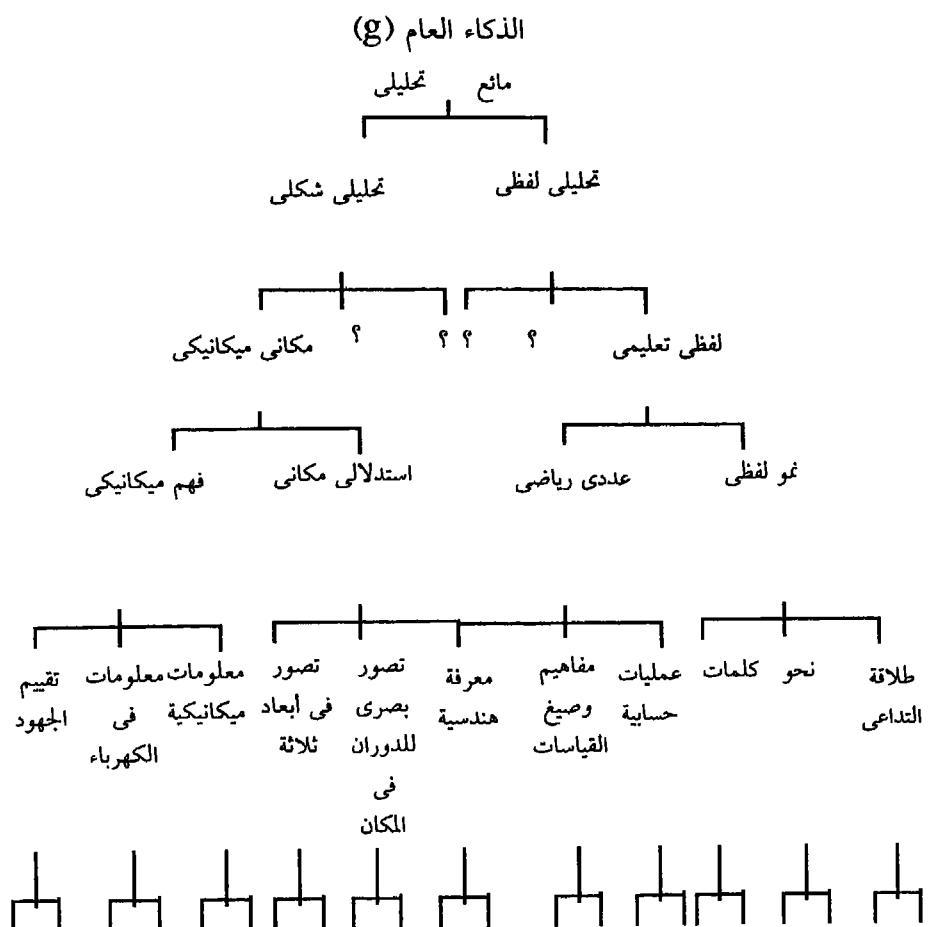
لذلك اقترح كثير من الباحثين استخدام النماذج الهرمية Hierarchical Models التي تمثل مستويات من العوامل متدرجة في عموميتها. ولعل جيلفورد كان أحد مؤلاء الباحثين حيث اقترح النموذج الهرمي لعوامل بنية العقل الذي أشرنا إليه فيما سبق، والموضح بالشكل التخطيطي (٩ - ٥) .

وقد بينَ فيرنون (Vernon, 1950) عالم النفس الإنجليزي أنه يمكن وصف القدرات العقلية وتصنيفها بوضوح أكثر وذلك بتنظيمها تنظيماً هرمياً في أربعة مستويات Major بالعامل العام (g) General Factor ، وتندرج إلى العوامل الطائفية الرئيسية Group Factors التي تنقسم إلى عاملين أحدهما لغوی - تعليمي ، والآخر عملی - ميكانيكي - مکانیکی ، وينقسم كل من هذين العاملين بدوره إلى عوامل طائفية فرعية Minor Group Factors ، وهذه تنقسم إلى عوامل نوعية Specific Factors تتعلق باختبارات معينة. ويُعد هذا التصنيف توسيعاً من أعمال سيرمان وثيرستون اللذين أشرنا إليهما من قبل. ويوضح الشكل التخطيطي (٩ - ٦) التالى هذا التصنيف الهرمي :



شكل (٩ - ٦) يوضح نموذج فيرنون للتصنيف الهرمي للقدرات العقلية

وقد اقترح كرونباك (Cronbach, 1970) بجامعة ستانفورد الأمريكية نموذجاً تصفيفياً يعد تلخيصاً للنماذج السابقة وبخاصة نموذج فيرنون يوضحه الشكل التخطيطي (٧-٩) التالي :



شكل (٩ - ٧) يوضح بنية هرمية مقترحة للقدرات العقلية اقترحها كرونباك

ويتضح من شكل (٩ - ٧) أن العامل العام (g) أو الذكاء المائع الذي أشار إليه كاتل Cattell أو الذكاء التحليلي يقع في قمة التصنيف، وهو مركز الشبكة العاملية بأسرها يليه العوامل العريضة المتعلقة به ، وتحتل العوامل النوعية قاعدة التصنيف.

الذكاء ومدخل تجهيز المعلومات ومعالجتها :

إن دراسة الذكاء لم تفصل عن مبحث الفروق الفردية أكثر من نصف قرن من

الزمان، لذلك نلاحظ أن جميع النماذج العاملية لبنية العقل التي نقاشناها فيما سبق استندت استناداً أساسياً في جمع البيانات وتحليلها إلى مدخل الفروق الفردية Differential Approach . ويتركز الاهتمام في هذا المدخل على قياس الفروق بين الأفراد في الذكاء والقدرات وليس على تفسير هذه الفروق.

وقد أدى ذلك إلى توجيهه بحوث الذكاء في الآونة الأخيرة وجهة جديدة تركز ترکیزاً أساسياً على الوظائف والعمليات المعرفية Cognitive Processes & Functions وهذا لا يعني أن قياس الفروق الفردية ليس ضرورياً ، وإنما يعني توجيه مزيد من الاهتمام إلى العمليات المعرفية التي تتحدد تجريبياً.

وستند الاختبارات العقلية إلى هذه العمليات بحيث يمكن أن تزودنا نتائجها بصفحات نفسية (بروفيل قدرات الأفراد) متعلقة بهذه العمليات تفيد في التقويم التشخيصي والتعليم العلاجي للمهارات الأساسية. وبذلك توظف الاختبارات التشخيصية بدرجة أكبر من توظيف الاختبارات المعيارية التنبؤية. ولعلنا قد أوضحنا ذلك عند تمييزنا بين الاختبارات مرجعية المحك والاختبارات مرجعية الجماعة أو المعيار في الفصل الثامن، عندما ذكرنا أن النوع الأول من الاختبارات يتميز بقيمة تشخيصية إذا حاولنا إحداث تكامل بين علم القياس وعلم النفس المعرفي في بناء هذه الاختبارات، وهو ما سوف نوضحه في الفصل الرابع عشر.

وقد أدى هذا التوجه الجديد إلى الاهتمام بمدخل تجهيز المعلومات ومعالجتها Information Processing Approach من أجل فهم أفضل لمجال الذكاء الإنساني. ويسعى الباحثون في هذا المجال إلى الكشف عن التمثيل Representations ، والعمليات Processes ، والاستراتيجيات Strategies التي يستخدمها الأفراد في حل مشكلات محددة تحوز الاعتراف بأنها تتطلب «ذكاء» في حلها.

كما يحاولون البحث عن كيفية إدماج كل من المدخل الفارق أو المدخل السيكومترى، ومدخل تجهيز المعلومات معًا في دراسة الذكاء الإنساني. وربما يرجع السبب في ذلك إلى أن كلا من المدخلين يشوبه بعض القصور، فالمدخل الأول يغفل العمليات والاستراتيجيات التي تُشكل جزءاً كبيراً بما نعنيه بالذكاء، أما المدخل الثاني فلا يزودنا بوسائل الدراسة المنظمة للمتغيرات المرتبطة بالفروق الفردية في الأداء، وبخاصة أن هذا المدخل ربما يؤدي إلى تعظيم قيمة مكونات التجهيز Processing على الرغم من أن هذه المكونات تتعلق بأداء مهمة محددة Task Components . Specific

فمن المعلوم أن الباحثين في علم النفس الفارق يختلفون عن الباحثين في علم النفس المعنى في استراتيجيات التنبؤ في مجال الذكاء الإنساني. فالباحثون في علم النفس الفارق يحاولون التوصل إلى نظريات متكاملة للذكاء، أما الباحثون في علم النفس المعنى فإنهم يبدعون بنظريات فرعية صغيرة أو نماذج تتعلق بأداء مهام معينة، أو أحياناً مجموعة من المهام.

ولعل مزايا أيٌّ من المدخلين تُعد عيباً للآخر مما يتطلب مزيداً من البحث حول كيفية إدماجهما والإفادة من ذلك في تطوير بحوث الذكاء الإنساني. ويرى ستربنبرج (Sternberg, 1979) أن البحوث المعاصرة التي تعمل على التكامل بين المدخلين تتجه إلى بناء نظريات فرعية محدودة للذكاء Subtheories. وهذه النظريات تتخذ شكل نظريات جوانب من الذكاء، مثل القراءة، والاستدلال، والتصور البصري المكانى، وما إلى ذلك ، أو مجموعات من هذه الجوانب .

ولعل نظرية ستربنبرج المتعلقة بتحليل مكونات الاستدلال القياسي Componential Theory of Analogical Reasoning وما تتضمنه من نماذج أساسية تتناول مظاهر القياس الاستدلالي Analogies والتصور الفكري لعملية فحص البائع Option Scanning تُعد مثالاً لهذا التوجه (Sternberg, 1977 , 1989) .

وهذه النظرية وأمثالها لا تفيد فقط في فهمنا للذكاء ، وإنما في تطوير اختبارات تستند إلى أفضل «المكونات Components» التي أطلق عليها ستربنبرج «العمليات المعرفية الدقيقة Cognitive Microprocesses» التي يستخدمها الفرد في حل مشكلات محددة يشتمل عليها اختبار أو ربما مفردة اختبارية .

وخلال ذلك أن تعدد نظريات ونماذج التكوين العقلى يُعد نتاجاً طبيعياً لتعقد وتشابك مفهوم الذكاء، ولا يجب أن ننظر إليها على أنها تناقض وجهات نظر متباعدة، حيث إنها تعد نوافذ متعددة على ظاهرة الذكاء الإنساني .

فهذه النماذج تزودنا بمنهجيات مختلفة لدراسة هذه الظاهرة التي تتطلب تكاتف جهود علماء النفس المعنى وعلماء القياس النفسي والتربوى من أجل التوصل إلى نموذج جديد لبنية العقل استناداً إلى نظرية معرفية .

ولعل هذا النموذج المأمول يسهم في تعريف أفضل لما تقيسه اختبارات الذكاء بالفعل ، ويقدم أساساً أكثر رسوحاً لتقدير المضمدين الاجتماعية والتربوية لهذه الاختبارات .

قياس الذكاء الإنساني :

بعد عرضنا السابق لأهم نظريات ونماذج التكوين العقلي وتوجهاتها المستقبلية ننتقل إلى كيفية قياس الذكاء الإنساني، فقد تبين لنا من هذا العرض مفهوم الذكاء العام وما يمكن أن يتضمنه من قدرات عقلية تستطيع قياسها بواسطة الاختبارات.

وسوف نتناول فيما يلى بعض اختبارات الذكاء العام الشائعة الاستخدام، ومعظم هذه الاختبارات تركز على بعض العوامل مثل الفهم اللغوي والاستدلال العام بدرجة أكبر من تركيزها على السرعة الإدراكية والتصور المكاني نظراً لأنه يعتقد أن العوامل الأولى تمثل السلوك الذكي تمثيلاً أفضل.

لذلك يلاحظ أن محتوى كثير من اختبارات الذكاء العام يتعلق بعامل الفهم اللغوي، وعامل إجراء العمليات الحسابية، وأنواع مختلفة من عوامل الاستدلال، بالإضافة إلى محتويات متاثرة تتعلق بعوامل الذاكرة، والمكان، والإدراك.

وعلى الرغم من وجود ارتباطات مرتفعة بين اختبارات الذكاء العام تصل أحياناً إلى .٨٠ .٠٠ بين بعض منها، إلا أن هذه الارتباطات ليست تامة مما يدل على أنها تقيس قدرات مختلفة إلى حد ما (Nunnally , 1978) . وهذا ربما يؤيد وجهات نظر بعض علماء النفس الذين يعارضون استخدام اختبارات الذكاء العام نظراً لأنه يمكن أن يتوصل الأفراد إلى حل مشكلة معينة باستخدام أساليب وعمليات مختلفة.

إذ ربما يعتمد أحد الأفراد اعتماداً أساسياً على قدرته المكانية في حل تمارين هندسية، ويعتمد فرد آخر على حل هذه التمارين حلاً مجرداً اعتماداً على قدرته الاستدلالية. لذلك ربما يقل استخدام اختبارات الذكاء العام مستقبلاً عندما تتوصل البحوث إلى طرق جديدة لبناء بطاريات من الاختبارات تقيس العوامل المتعددة للذكاء.

اختبارات الذكاء العام :

يمكن تصنيف اختبارات الذكاء العام إلى نوعين وفقاً لطريقة التطبيق. فالاختبارات التي تطبق على فرد واحد في وقت واحد تسمى الاختبارات الفردية Individual Intelligence Tests ، أما الاختبارات التي تطبق على عدد كبير من الأفراد في وقت واحد فتسمى الاختبارات الجماعية Group Intelligence Tests ، ومعظم الاختبارات الفردية مثل مقياس ستانفورد بينيه Stanford - Binet ، ومقاييس ويكسيلر Wechsler تحتاج إلى شخص مدرب على تطبيقها وتفسير نتائجها حيث يهتم بكيفية استجابة الفرد أو المفحوص. لذلك تطبق عادة الاختبارات الفردية بواسطة الأخصائيين

النفسين بالمدارس والجامعات والمؤسسات حيث يُستفاد من نتائجها في التشخيص الكlinيكي، واتخاذ قرارات مهمة مثل تشخيص حالات الضعف العقلي.

أما الاختبارات الجماعية ، فإنها تُعد أفضل وسائل تقييم القدرات العقلية لعدد كبير من المختبرين في هذه المؤسسات بفرض التوجيه التعليمي والمهني .

ونظراً لأنه يصعب بالطبع تناول جميع اختبارات الذكاء العام المستخدمة في وقتنا الحاضر ، فإننا سوف نوضح بإيجاز أمثلة لبعض الاختبارات الفردية والجماعية الشائعة الاستخدام .

أولاً : اختبارات الذكاء الفردية :

مقياس ستانفورد بينيه للذكاء(SBIS) :

سبق أن أشرنا إلى هذا المقياس في سياق عرضنا لقياس الذكاء من منظور تاريخي . وهو يُعد من أشهر الاختبارات الفردية التي تقيس الذكاء العام وأكثرها استخداماً.

ويُعد مقياس ستانفورد - بينيه كما ذكرنا امتداداً لمقياس بينيه Binet الأصلي . وقد نشرت الطبعة الأولى لهذا المقياس عام ١٩١٦ في الولايات المتحدة الأمريكية وأجريت عليه عدة تعديلات في الأعوام ١٩٣٧ ، ١٩٧٢ ، ١٩٦٠ ، ١٩٨٤ . وسوف نلقي الضوء فيما يلى على هذا المقياس :

مقياس ستانفورد - بينيه (طبعة عام ١٩١٦) :

قام تيرمان Terman بجامعة ستانفورد الأمريكية بترجمة مقياس بينيه - سيمون Binet - Semon وأجرى تطبيقه على عينة تشمل على ١٠٠ طفل ، ٤ فرد راشد . وكما في مقياس بينيه الأصلي اعتمد تيرمان على ميزان العمر الزمني Age Scale ، حيث قام بتجميع الاختبارات التي يشتمل عليها المقياس في مستويات عمرية .

وашتمل المقياس على مفردات متقدمة من مقياس بينيه وإضافة مفردات جديدة تمثل عينة واسعة من المهام التي تتطلب ذكاء ، ولكن هذه المهام لا تعتمد على خبرات مدرسية معينة ، وأحد محركات انتقاء مفردة في اختبار فرعى هو تزايد النسبة المئوية من الأطفال في المستويات العمرية المتالية الذين يستطيعون أن يحيبوا عنها إجابة صحيحة أى التمايز العمري Age Differentiation، وتوضع المفردة عندئذ في المستوى العمري للاختبار الفرعى الأكثر ملاءمة له .

ومن المحكّات الأخرى لانتقاء المفردات الارتباط بين درجات كل مفردة والدرجات الكلية في المقياس. وهذا يعني أن كل مفردة تقيس ما يتعلّق بالمقياس ككل، وهو ما سبق أن أشرنا إليه باسم «تجانس المفردات Item Homogeneity » من حيث محتواها، وذلك في الفصل الرابع.

ويعتمد العمر العقلي (MA) ، ونسبة الذكاء Intelligent Quotient (IQ) في القياس على عدد الاختبارات الفرعية التي يجتازها المفحوص في المستويات العمرية المختلفة . وتحسب نسبة الذكاء باستخدام الصيغة :

$$\text{نسبة الذكاء (IQ)} = \frac{\text{العمر العقلي}}{\text{العمر الزمني}} \times 100 \quad (1-9) \dots$$

وذلك بعد تحويل كل من العمر العقلي وال عمر الزمني إلى شهور.

مقياس ستانفورد - بينيه (طبعة عام ١٩٣٧) :

أعيد تقييم المقياس على أطفال من مستويات عمرية مختلفة تبدأ من عام ونصف حتى خمسة أعوام ونصف بفارق عمرى قدره نصف عام بين كل مستوى وأخر، ومن أعمار تبدأ من ٦ أعوام حتى ١٤ عاماً بفارق عمرى قدره عام واحد، وكذلك من أعمار تبدأ من ١٥ عاماً حتى ١٨ عاماً بفارق عمرى قدره عام واحد. واشتملت العينة على الذكور والإناث من مستويات اجتماعية واقتصادية تفوق المتوسط، وبذلك لم تمثل المجتمع الأمريكي، وتم تجميع الاختبارات لكل فئة من هذه الفئات العمرية ، وأضيف مستوى الرشد المتوسط، وثلاثة مستويات للراشد المتفوق.

ويوضح شكل (٩ - ٨) بعض المواد الاختبارية غير اللفظية التي يشتمل عليها مقياس ستانفورد - بينيه (Nunnally , 1970) .



شكل (٩ - ٨) يوضح بعض المواد الاختبارية غير اللفظية في مقياس ستانفورد - بينيه

وعند اختبار طفل ينبغي على الفاحص أن يحدد المستوى العمري القاعدي Basal Age (BA) وهو أعلى مستوى عمرى يستطيع عنده الطفل أن يجتاز جميع الاختبارات الفرعية المتعلقة به . وعندئذ يستمر اختبار الطفل حتى يصل إلى أعلى مستوى عمرى (CA) يفشل عنده الطفل في اجتياز أيٌ من الاختبارات الفرعية المتعلقة به . ويحسب العمر العقلى بإضافة العمر القاعدى (BA) إلى عدد الشهور التى يحصل عليها الطفل من الاختبارات الفرعية التى اجتازها من أعلى مستوى عمرى وصل إليه (CA) ، حيث إن لكل من هذه الاختبارات عدد محدد من الشهور موضح فى دليل المقياس .

وبذلك يمكن حساب نسبة الذكاء (IQ) باستخدام الصيغة (١-٩) التي ذكرناها . ومن الجدير بالذكر أن المقياس أصبح يتكون من صيغتين متكافتين (L) ، (M) بعد أن قام تيرمان وميريل بإجراء التعديلات المناسبة عليه ، وذلك إذا تطلب الأمر إعادة تطبيق المقياس على الفرد نفسه .

مقياس ستانفورد- بيئي (طبعة عام ١٩٦٠) :

أعاد تيرمان مراجعة المقياس حيث انتهى أفضل مفردات كل من الصيغتين (L) ، (M) واتسع المدى العمري من عامين حتى الرشد .

واشتمل المقياس على مواد اختبارية تناسب الطفل الصغير مثل اللعب والصور وأشياء عيانية أخرى . وأعد دليل المقياس وكراسة تسجيل الإجابات وغير ذلك من المطبوعات التى يسترشد بها مستخدم المقياس فى تطبيقه ، وفقاً للتعليمات المحددة بدقة .

ومن التعديلات الأساسية فى هذه الطبعة استخدام نسبة الذكاء الانحرافية IQ (DIQ) التى أشرنا إليها فى الفصل السادس ، نظراً لأوجه القصور المتعددة التى تتعلق بنسبة الذكاء (IQ) وبخاصة عدم تساوى الوحدات العمرية ، وصعوبة تفسيرها فى المستوى العمرى للراشد ، حيث إنه لا يوجد اتفاق على العمر الزمنى الذى يستخدم فى مقام صيغة نسبة الذكاء فى هذه الحالة ، أي العمر الذى يفترض أن يتوقف عنده النمو العقلى للفرد ، حيث إن هذه تعد من المشكلات التى تحوطها اعتبارات كثيرة .

مقياس ستانفورد- بيئي (طبعة عام ١٩٧٢) :

لم تتضمن عينات الأفراد التى استندت إليها المعايير فى الطبعات السابقة فئات الأقليات Minority Group فى المجتمع الأمريكى ، لذلك أعيد تقييم المقياس عام

١٩٧٢ على عينة ممثلة لجميع الفئات بما في ذلك الأقليات، واشتملت على ٢١٠٠ طفل بمعدل ١٠٠ طفل لكل مستوى عمرى (٢ - ١٨ عاماً) ، وظل متوسط نسبة الذكاء الانحرافية ١٠٠ وانحرافها المعياري ١٦ . ووُجد أن قيمة معامل ثبات درجات القياس باستخدام طريقة التجزئة النصفية تفوق ٩٠ ، ولكن قيم معامل الارتباط بين درجات المقياس ، ودرجات تحصيل التلاميذ أو تقديرهم العام في المواد الدراسية تراوحت بين ٤٠ ، ٧٥ ، ٠ ، وتعكس هذه القيمة الموجبة حقيقة أن مقياس ستانفورد - بيئيه مُشبع بمواد لفظية وبخاصة في المستويات العمرية العليا ، والدليل على ذلك أن قيمة معامل الارتباط تزداد بالنسبة لمواد القراءة ، واللغة الإنجليزية بينما نقل بالنسبة للرياضيات (Sax , 1964) .

مقياس ستانفورد - بيئيه (طبعة عام ١٩٨٤) :

هذه الطبعة الرابعة للمقياس لا تختلف كثيراً عن الطبعة السابقة فيما عدا بعض التغييرات المهمة حيث أُعدت للأغراض التالية:

(١) الإفاده في التمييز بين الطلاب المتخلفين عقلياً والطلاب الذين يعانون من إعاقات معينة في التعلم.

(٢) معاونة المربين والخصائص النفسيين فهم أسباب الصعوبات التي ربما تواجه بعض الطلاب في المدرسة.

(٣) المعاونة في التعرف على الطلاب المهوبيين.

(٤) دراسة نمو المهارات المعرفية لدى الأفراد من عمر عامين حتى الرشد.

وقد اشتمل المقياس على ١٥ اختباراً فرعياً تشكل نظاماً هرمياً يتكون من ثلاثة مستويات تمثل الذكاء العام (g) يليه ثلاثة عوامل عريضة تمثل القدرات المتبلورة ، والقدرات المائية - التحليلية ، والذاكرة قصيرة الأمد ، ويلي ذلك ثلاثة عوامل أقل اتساعاً هي الاستدلال اللفظي ، والاستدلال الكمي ، والاستدلال المجرد - البصري ، وهذه المستويات الهرمية الثلاثة موضحة بالشكل التخطيطي (٩ - ٩) التالي :

: (Thorndike , Hage & Sattler , 1986)

العامل العام		القدرات المتبولة	القدرات المائعة - التحليلية
		الاستدلال	الذكرة قصيرة الأمد
	المجرد - البصري	اللفظي - الكمي	
ذاكرة الصور	تحليل الأنماط	الكلمات	التعامل مع المقاييس
ذاكرة الجمل	مصفوفات	الكمية	الفهم
ذاكرة الأرقام	ثنى الأوراق	سلالس الأعداد	السخافات
ذاكرة الأشياء	وقطعها	تكوين المعادلات	العلاقات اللفظية

شكل (٩) يوضح النظام الهرمي لمقياس ستانفورد - بيئه للذكاء (الطبعة الرابعة)

وقد اشتملت عينة التقنين على أكثر من ٥٠٠٠٠ فرد من مختلف الولايات تم تصنيفهم وفقاً للمنطقة الجغرافية، وحجم المجتمع المحلي، والسلالة، والنوع. ووجد أن قيمة معامل استقرار الدرجات الكلية في المقياس تفوق .٩٠، وأكدت البيانات صدق التكوين الفرضي للمقياس ككل.

ويتميز المقياس بأنه يقيس نطاقاً واسعاً من عمليات تجهيز المعلومات لدى الفرد، ولكن لم تؤكد نتائج التحليل العاملى البنية الهرمية للتصنيف الموضحة بشكل (٩-٩).

لذلك ينبغي تفسير الدرجات الكلية فقط حيث إنها تدل على الذكاء العام. غير أن المقياس يتطلب مزيداً من البحث لتحديد جوانب القوة والضعف عند تطبيقه على أطفال ما قبل المدرسة، كما يتطلب أن تتضمن العينة أطفالاً من ذوى الاحتياجات الخاصة.

تعليق على مقياس ستانفورد - بيئه :

على الرغم من أن مقياس ستانفورد - بيئه يُعد من أكثر مقاييس الذكاء العام

استخداماً، حيث أشار ثورنديك (Thorndike, 1975) إلى أنه طُبِّقَ على حوالى ٨٠٠ فرد سنوياً ما بين عامي ١٩٦٠ و١٩٧٢ . ونال كثيراً من الدراسات والبحوث حيث أشار الكتاب السنوي للقياس العقلي عام ١٩٧٨ إلى ما يقرب من ١٥٩ مصدرًا يتعلق بالقياس (Buros, 1978).

وبذلك يمكن القول بأنه حق الغرض الذي بُنى من أجله، إلا أنه يُعاب عليه تركيزه الشديد على القدرة اللغوية. فتعليمات معظم أجزاءه تقدم شفوياً، وكثير من اختباراته الفرعية تتطلب استخدام الكلمات ، فإذا طبق المقياس على فرد لديه بعض الصعوبات اللغوية، فإنه ربما يصعب الحصول على صورة صادقة عن نسموه العقل، كما أنه يتطلب فاحصاً مدرياً تدريباً عالياً على استخدام المقياس لكي يتسعى له الحيطة من تدخل ظروف تؤدي إلى عدم دقة النتائج.

هذا بالإضافة إلى أن المقياس يعطي درجة كلية واحدة وبذلك لا يزودنا بمعلومات كافية عن القدرات المتعددة للفرد. وكذلك لا يناسب الكبار، حيث إن نوع الأداء العقلي الذي يمثله مفهوم العمر العقلي ينمو ببطء شديد بعد بلوغ السادسة عشر أو السابعة عشر من العمر مما يؤدى إلى مشكلات في حساب نسبة الذكاء IQ بالنسبة للأعمار التي تزيد عن ذلك كما سبق أن أوضحنا.

ولعل استخدام نسبة الذكاء الانحرافية فى طبعة عام ١٩٦٠ وما بعدها هدف للتغلب على هذه المشكلة كوسيلة لجعل نسبة الذكاء (IQ) قابلة للمقارنة في جميع المستويات العمرية.

ولعله يُعد من أفضل المقاييس السيكولوجية التي تنبأ بدراجة جيدة بالتحصيل الدراسي. كما أن المقياس يقدم للفاحص معلومات ثرية من خلال ملاحظته سلوك المفحوص أثناء حله للمشكلات وأدائه للمهام المتباعدة في درجة صعوبتها، كما يمكنه ملاحظة المدخل الذي يتوجه المفحوص في حل المشكلة مثل المحاولة والخطأ في مقابل أسلوب التفكير المنظم . لذلك يمكنه ملاحظة السلوك في موقف معنون إلى جانب الدرجة الكلية التي يحصل عليها المفحوص، ومقارنتها بالمعايير التي تم تحديدها.

مقاييس ويكسنر للذكاء : Wechsler Intelligence Scales

أعد ويكسنر Wechsler مجموعة من المقاييس الفردية للذكاء من منظور مختلف عن المنظور الذي استند إليه مقياس ستانفورد - بيئه ، حيث حاول تلافي أوجه قصور هذا المقياس التي أشرنا إليها فيما سبق.

ولعل أهم هذه الأوجه عدم ملاءمة مقياس ستانفورد - بينيه لقياس ذكاء الراشدين من حيث مفرداته ومعاييره، واقتصره على درجة كلية واحدة للفرد الذي يطبق عليه الاختبار تدل على الذكاء العام، واعتماد معاييره في الطبعات الأولى على مفهومي العمر العقلي ونسبة الذكاء اللذين كانا محل تساوق ونقد مستمر وبخاصة فيما يتعلق بذكاء الراشدين، وكذلك تأكيده لعامل السرعة في الإجابة عن مفرداته الموقوتة مما قد لا يناسب الراشدين.

لذلك اهتم ويكسنر ببناء مقياس لذكاء الراشدين في ضوء منظوره عن الذكاء كتجمع من القدرات المترابطة وليس كقدرة عامة واحدة. ونظرًا لأن مقياس ستانفورد - بينيه يعتمد اعتماداً أساسياً على الجانب اللغوي فيما عدا بعض المفردات في المستويات العمرية المبكرة، فقد عمل على تضمين مقياسه جانبين أحدهما لغوي والآخر عملي. والهدف من ذلك التغلب إلى حد ما على مشكلة اللغة والثقافة والمؤثرات التعليمية (Wechsler, 1939).

وبذلك يمكن قياس ذكاء الأفراد الذين لديهم صعوبات في اللغة أو خبراتهم اللغوية محدودة، ويحصل الفرد في المقياس على ثلث درجات كلية تمثل الأداء اللغوي، والأداء العملي، والذكاء العام. ففي هذا المقياس يمكن أن يحصل الفرد على نسبة الذكاء نفسها بطريق مختلفة بحسب المفردات التي يجتازها أو يُحقق في الإجابة عنها.

وبذلك يمكن الإفادة من مقياس ويكسنر في تشخيص أنماط تفكير الفرد بدرجة أفضل من مقياس ستانفورد - بينيه . كما يمكن الحصول على معلومات أكثر ثراء من خلال ملاحظة سلوك الفرد أثناء أداءه العملي . فإذا كان الأداء اللغوي يسمح للفاحص بملاحظة سلوك الفرد أثناء إجابته عن المفردات اللغوية ، فإن الأداء العملي بما يتطلبه من تركيز وانتباه ووقت أطول يمكن الفاحص من ملاحظة توجُّه الفرد ومدخله في حل المشكلات. لذلك فإن مقياس ويكسنر يستخدم كأدلة مناسبة في التشخيص الكلينيكي .

وقد صدر مقياس ويكسنر للراشدين أول مرة عام ١٩٣٩ ، وتشتمل على صورتين متكافتين ، وأطلق عليه «مقياس ويكسنر - بلفيو للذكاء» نسبة إلى مستشفى بلفيو للأمراض النفسية كما ذكرنا من قبل .

ولكن نظرًا لأن الجماعة المعيارية التي استخدمت في تقييم الاختبار وتحديد معاييره كانت غير ممثلة للمجتمع الأمريكي ، فقد أعاد ويكسنر تقييمه مرة أخرى عام ١٩٥٥ على عينة أكثر تمثيلاً ، وأطلق عليه «مقياس ويكسنر لذكاء الراشدين

«Wechsler Adult Intelligence Scale (WAIS) ثم أعيد تقييمه وُحدّدت له معايير جديدة عام ١٩٨١ (Wechsler, 1981)، وأطلق عليه (WAIS - R).»

وقد قام ويكسنر عام ١٩٤٩ ببناء مقاييس آخر لذكاء الأطفال، Wechsler Intelligence Scale for Children (WAIS)، وذلك للمدى العمري ٥ - ١٥ عاماً، كما قام عام ١٩٦٧ ببناء مقاييس ثالث لذكاء أطفال ما قبل المدرسة

Wechsler Preschool and Primary Scale of Intelligence (WPPSI)

وذلك للمدى العمري ٤ - ٦ أعوام. وسوف نوضح فيما يلى بإيجاز كل من هذه

المقاييس:

مقاييس ويكسنر لذكاء الراشدين (WAIS - R) :

صمم هذا المقاييس لقياس ذكاء الراشدين بخاصة وذلك في المدى العمري ١٦ - ٧٥ عاماً، لذلك راعى ويكسنر أن يكون محتوى المقاييس وطريقة تصحيحه مناسبة لهم، وأن تكون معاييره مستمدة من عينات من الراشدين. وقد تبنى تعريفاً للذكاء يؤكّد القدرة العامة كما أشرنا في مستهل هذا الفصل، حيث أوضح أن الذكاء هو قدرة الفرد الكلية على العمل الهايدف، والتفكير المنطقي، والتعامل مع بيئته بفاعلية، ولذلك اعتبر أن الاختبارات الفرعية التي اشتمل عليها المقاييس لا تقاس أنواعاً مختلفة من الذكاء.

ومع هذا فإنه أكّد الإلّافة من المقاييس كأداة تشخيص وذلك باستخدام أنماط أو بروفيل درجات الاختبارات الفرعية التي يشتمل عليها المقاييس كأساس للاستدلال على إمكانات الفرد العقلية وحالته الانفعالية.

ولكنه أشار إلى أن الذكاء لا يُعدُّ مجموع القدرات التي تكشف عنها الاختبارات الفرعية المختلفة، وإنما يُعدُّ تكوين متافق من القدرات إضافة إلى العوامل الانفعالية والشخصية التي تؤدي إلى السلوك الذكي (Wechsler, 1958).

ويشتمل المقاييس على مقاييسين فرعيين أحدهما المقاييس اللغطي Verbal Scale والأخر مقاييس الأداء العملي Performance Scale. ويكون المقاييس الأول من ستة اختبارات فرعية، والمقاييس الثاني من خمسة اختبارات فرعية. وهذه الاختبارات كالتالي:

المقياس اللغطي:

(١) المعلومات : أسئلة تتناول معلومات عامة في موضوعات مختلفة مرتبة

تصاعدياً من حيث درجة صعوبتها، وهي مصممة لقياس المعلومات العامة المتوافرة لدى معظم قطاعات المجتمع.

(٢) الفهم : أسئلة تقيس الحكم الواقعي والمعلومات.

(٣) الحساب : مسائل كلامية تتطلب قدرًا ضئيلاً من المهارات الحسابية، وتقيس التركيز.

(٤) المتشابهات : أسئلة تتضمن أزواجاً من الكلمات، وتتطلب توضيح أساس التشابه بين كل زوج منها، وتقيس التفكير المجرد.

(٥) وسع الذاكرة للأرقام : أسئلة تتطلب إعادة ذكر أرقام بترتيبها الأصلي أو بالترتيب العكسي.

(٦) معانى الكلمات : أسئلة تتطلب تعريف كلمات أو مصطلحات عامة، وتقيس مستوى فهم الكلمات.

مقياس الأداء العملى:

(١) رموز الأرقام : مهام تتطلب استبدال شيء برمز مجرد استرشاداً بقائمة توضح التمازج بينهما، وذلك في مدة زمنية تتراوح بين دقيقة ودقيقة ونصف، وتقيس الوظائف البصرية والحركية.

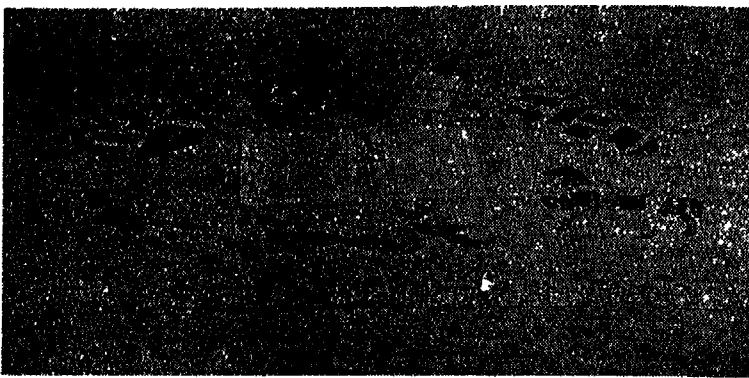
(٢) تكميل الصور: تتطلب تحديد الجزء المفتقد من صورة معينة، ويقيس إدراك التفاصيل.

(٣) تصميم المكعبات : يتطلب ترتيب مكعبات ذات ألوان معينة بحيث تكون ماثلة لتصميمات معطاة، وتقيس القدرة على التخطيط.

(٤) ترتيب الصور : تتطلب ترتيب بطاقات مختلفة ترتيباً متسلسلاً ، وتقيس الاستدلال غير اللفظي.

(٥) تجميع الأشياء : يتطلب تجميع أجزاء متاهات متباعدة في درجة تعقدتها بأسرع ما يمكن، ويقيس القدرة على تحليل علاقة الأجزاء بالكل.

ويوضح شكل (٩ - ١٠) بعض المواد الاختبارية التي يشتمل عليها مقياس ويكسنر للراشدين (Horrocks , 1964).



شكل (٩ - ١٠) يوضح بعض المواد الاختبارية في مقياس ويكسنر للمراشدين

وقد تم ترتيب مفردات كل من هذه الاختبارات الفرعية ترتيباً تصاعدياً بحسب مستوى صعوبتها، ويجري تطبيقها وفقاً لهذا الترتيب قدر الإمكان، ويتهى التطبيق إذا أخفق الفرد عدداً محدداً من المرات المتتالية. وبعض هذه المفردات ينبغي الإجابة عنها في زمن محدد، بينما البعض الآخر غير موقوت.

وينبغي على الفاحص الذي يستخدم المقياس أن يكون على درجة عالية من المهارة والتدريب، وأن يراعي اتباع التعليمات التفصيلية الموضحة في دليل المقياس بدقة. ويستغرق تطبيق المقياس ساعة تقريباً، وهو زمن أقل مما يتطلب تطبيق اختبار ستانفورد - بينيه.

ويتم تصحيح الاختبارات الفرعية بمقارنتها بقائمة الاستجابات المقبولة المحددة في دليل المقياس. وتقدر الدرجات أو النقاط على أساس درجة صحة الإجابة، وتراعي أيضاً سرعة الاستجابة في بعض الاختبارات الفرعية.

لذلك يعتمد مقياس ويكسنر على النقاط Point Scale وليس على العمر Age Scale كما في مقياس ستانفورد - بينيه .

وتحوّل الدرجات الخام الكلية في كل اختبار فرعى إلى درجات معيارية اعتدالية متوسطها ١٠ وانحرفتها المعياري 3 ، وتجمع الدرجات الكلية للإختبارات الفرعية، ثم تحول الدرجات المعيارية الاعتدالية لكل من المقياس اللغظى، ومقياس الأداء العملى،

والمقياس ككل إلى معيار «نسبة الذكاء الانحرافية DIQ» التي أوضحتناها في الفصل السادس (شكل ٦ - ١) وقد حدد ويكسنر متوسط هذه النسبة ١٠٠ وانحرافها المعياري ١٥ نقطة.

وهذا يعني أن الفرد الذي تتحرف درجته الخام في أحد المقياسين أو المقياس ككل وحده انحراف معياري واحد أعلى متوسط مجموعة عمرية تكون نسبة ذكائه ١١٥ مثل فرد في أي مجموعة عمرية أخرى تتحرف درجته الخام أيضاً وحدة انحراف معياري أعلى متوسط مجموعة عمرية. واستمدت معايير المقياس من عينة مستعرضة تمثل مجتمع الراشدين بالولايات المتحدة الأمريكية. وبإجراء هذا التحويل يمكن المقارنة بين الدرجات الكلية لكل من الاختبارات الفرعية على ميزان موحد سواء للفرد نفسه أو بين الأفراد.

ومما يميز نسبة الذكاء الانحرافية أنه يمكن حسابها لككل مجموعة عمرية على حدة. وهذا بالطبع يحدث في مقياس ذكاء الأطفال، ولكن يعد أمراً جديداً في حالة الراشدين، فمثلاً تبين أن المجموعة العمرية ٢٥ - ٢٩ وفقاً لهذا المعيار تؤدي أداء أفضل من أي مجموعة عمرية أخرى.

لذلك فإن أفراد هذه الأعمار ينبغي أن يحصلوا على درجات خام في المقياس أعلى من أفراد الأعمار الأخرى لكي يحصلوا على نسبة الذكاء الانحرافية نفسها.

وقد استمدت معايير الطبعة الأخيرة للمقياس من عينة مستعرضة قوامها ١٨٨٠ فرداً موزعة على تسعمجموعات عمرية تم اختبارهم في المدة بين عامي ١٩٧٧ ، ١٩٨٠.

وفيما يتعلّق بالخصائص السيكومترية للمقياس بَيَّنت الدراسات أن الدرجات الكلية للمقياس تميز بقدر مرتفع من الثبات، غير أن الدرجات الكلية للاختبارات الفرعية أقل ثباتاً نظراً لقلة عدد مفراداتها، وقيم معامل الاتساق الداخلي للمقياس اللغوي أعلى من قيم هذا المعامل لمقياس الأداء العملي. وعلى الرغم من أن استخدام درجات المقياسين كل على حدة يتطلب أن تكون قيم معامل الارتباط بينهما منخفضة ، إلا أن ويكسنر وجّد أن هذه القيم مرتفعة وتصل أحياناً إلى ٨١ ، للأفراد في المجموعة العمرية ٤٥ - ٥٤ عاماً مما ينفي افتراض أن كلاً من المقياسين مستقل عن الآخر.

ولهذه الأسباب أوضح ويكسنر أن تفسير بروفيل درجات الأفراد الأسوية في الاختبارات الفرعية يكون غير ذي معنى. فأى فرق بين درجات الفرد في أي اختبارين فرعيين ينبغي أن يكون كبيراً قبل أن نعتبره فرقاً دالاً، ومع هذا فإن الفروق الكبيرة بين

الدرجات المحولة للاختبارات الفرعية، وكذلك بين نسب الذكاء الانحرافية للمقياس اللغظى ومقاييس الأداء العملى يمكن الإفادة منها فى تشخيص بعض الاضطرابات العقلية.

أما فيما يتعلق بصدق المقياس ، فلا توجد أدلة كافية بهذا الشأن ، ولكن يوجد ارتباط مرتفع بين درجات مقياس ويكسler ومقاييس ستانفورد - بينيه بلغ حوالي ٨٥ ، ٨٠ ، ٧٩ ، ٧٦ ، ٧٤ ، ٧٣ ، ٧٢ ، ٧١ ، ٧٠ ، ٦٩ ، ٦٨ ، ٦٧ ، ٦٦ ، ٦٥ ، ٦٤ ، ٦٣ ، ٦٢ ، ٦١ ، ٥٩ ، ٥٨ ، ٥٧ ، ٥٦ ، ٥٥ ، ٥٤ ، ٥٣ ، ٥٢ ، ٥١ ، ٥٠ ، ٤٩ ، ٤٨ ، ٤٧ ، ٤٦ ، ٤٥ ، ٤٤ ، ٤٣ ، ٤٢ ، ٤١ ، ٤٠ ، ٣٩ ، ٣٨ ، ٣٧ ، ٣٦ ، ٣٥ ، ٣٤ ، ٣٣ ، ٣٢ ، ٣١ ، ٣٠ ، ٢٩ ، ٢٨ ، ٢٧ ، ٢٦ ، ٢٥ ، ٢٤ ، ٢٣ ، ٢٢ ، ٢١ ، ٢٠ ، ١٩ ، ١٨ ، ١٧ ، ١٦ ، ١٥ ، ١٤ ، ١٣ ، ١٢ ، ١١ ، ١٠ ، ٩ ، ٨ ، ٧ ، ٦ ، ٥ ، ٤ ، ٣ ، ٢ ، ١ ، ٠ .

وقد أشار ويكسler إلى صدق محتوى المقياس من حيث الوظائف التى تقيسها الاختبارات الفرعية وملاءمتها لتعريف الذكاء الذى تبناه . وبينت دراسات صدق التكوين الفرضى فيما يتعلق بالارتباطات بين المقاييس الفرعية ، وبين كل من المقياس اللغظى ومقاييس الأداء العملى أن مقياس ويكسler يقىس الذكاء العام ، وأكيد ذلك دراسات التحليل العاملى . غير أن التقييم الدقيق للخصائص السيكومترية للمقياس تتطلب مزيداً من البحث .

مقياس ويكسler لذكاء الأطفال (WISC):

يعد هذا المقياس امتداداً للصيغة الأولى لمقياس ويكسler - بلفيو لذكاء الراشدين ، بحيث يقيس ذكاء الأطفال فى المدى العمري ٥ - ١٥ عاماً ، ١١ شهراً ، والطبعة الأولى لهذا المقياس تم إصدارها عام ١٩٤٩ . واشتمل المقياس على خمسة اختبارات فرعية تقيس الأداء اللغظى ، وخمسة اختبارات أخرى تقيس الأداء العملى وإضافة اختبارين تكميليين لتطبيقاتهما إذا تبين صعوبة أحد الاختبارات الفرعية الأخرى بالنسبة للطفل المُختبر .

والاختبارات الفرعية التى يشتمل عليها مقياس الأداء اللغظى هي : المعلومات العامة ، والفهم العام ، والحساب ، والمتشابهات ، ومعانى الكلمات ، واختبار بدليل يقيس وسع ذاكرة الأرقام . أما الاختبارات الفرعية التى يشتمل عليها مقياس الأداء العملى فهو : تكميل الصور ، وترتيب الصور ، وتصميم المكعبات ، وتجميع الأشياء ، والترميز ، واختبار بدليل يشتمل على متاهات .

وهذه الاختبارات الفرعية تُناظر اختبارات الراشدين من حيث المحتوى والوظائف العقلية التى تقيسها ، ومفردات هذه الاختبارات مرتبة بحسب مستوى صعوبتها ، ومجمعة معًا وفقاً لمحتواها . ويُوقف تطبيق أي اختبار فرعى إذا أخطأ الطفل فى إجابته عدة مرات متتالية ، ولكن هذا العدد أقل مما هو فى حالة مقياس الراشدين ، وذلك

لتقليل شعور الطفل الصغير أو الأقل ذكاء بالإحباط أو الفشل، وتقليل زمن التطبيق الذي يستغرق ساعة واحدة تقريباً.

وقد أجريت بعض التعديلات على المقياس وصدرت طبعته الثانية عام ١٩٧٤ وأطلق عليه (WISC - R) ، ولا يختلف كثيراً عن طبعته الأولى فيما عدا تغيير ترتيب تطبيق الاختبارات الفرعية، وتوسيع مدى الأعمار عاماً واحداً ليشمل العمر (١٦ عاماً، ١١ شهراً ، ٣٠ يوماً) ، وأدى ذلك إلى بعض التداخل مع مقياس الراشدين الذي يبدأ بالمستوى العمري ١٦ عاماً حيث يتنهى مقياس الأطفال، وأعد دليل خاص بتفسير درجات المقياس.

وقد أعيد تقييم الاختبار في هذه الطبعة بحيث أصبحت العينة أكثر تمثيلاً لمجتمع الأطفال في الولايات المتحدة الأمريكية، واشتملت على ٢٢٠ طفل من مختلف المناطق الجغرافية، والمستويات الاجتماعية والاقتصادية، والعمر، والنوع.

وقد صدرت الطبعة المعدلة الثالثة للمقياس عام ١٩٩١ (Wechsler , 1991) تحت مسمى (WISC III) التي تعد أكثرطبعات استخداماً لقياس ذكاء الأطفال، وأجرى حولها كثيراً من البحوث في مجال التربية الخاصة. وقد أضيف في هذه الطبعة اختبار تكميلي في مقياس الأداء العملي هو اختبار البحث عن الرموز، حيث يُطلب من الطفل الفحص البصري لمجموعة من الرموز ويحدد وجود رمز معين، وبذلك أصبح عدد الاختبارات الفرعية عشرة اختبارات، وثلاثة اختبارات تكميلية.

كما أن الاختبارات الفرعية التي تقيس الأداء العملي موقوتة، وقد تم تحويل درجات كل اختبار فرعي إلى درجات معيارية متوسطها ١٠ وانحرافها المعياري ٣ كما في اختبارات مقياس الراشدين، وذلك لكي يتتسنى المقارنة المباشرة بين درجات الاختبارات الفرعية من خلال تحليل الصفحة النفسية (بروفيل) لدرجات الطفل على الرغم من النقد الذي وجه مؤخرًا إلى هذه الصفحات النفسية Profiles . ويمكن الحصول على نسب الذكاء الانحرافية لكل من الأداء اللغظي والأداء العملي والذكاء العام التي متوسطها ١٠٠ ، وانحرافها المعياري ١٥ ، وأعدت جداول خاصة يستخدمها الفاحص لتفسير انتشار درجات الاختبارات الفرعية، والفرق في الأداء اللغظي، ودالة الفرق بين الاختبارات الفرعية التي تختلف عن المتوسط العام لدرجات جميع هذه الاختبارات. كذلك تضمنت هذه الجداول مؤشرات لـ مكافئ الس درجات Score Equivalents وذلك لأربعة عوامل جديدة تم التوصل إليها عن طريق التحليل العائلي (Kaufman , 1979). وهذه العوامل هي :

(١) **الفهم اللغوي** : ويتضمن المعلومات ، والمتباينات ، ومعانى الكلمات ، والفهم العام .

(٢) **التنظيم الإدراكي** : ويتضمن تكملة الصور ، وترتيب الصور ، وتصميم المكعبات ، وتجميع الأشياء .

(٣) **التحرر من مؤثرات التشتت** : ويتضمن الحساب ، ووسع ذاكرة الأرقام .

(٤) **سرعة تجهيز المعلومات** : ويتضمن الترميز ، والبحث عن الرموز .

وقد أعدت معايير الاختبار استناداً إلى عينة مماثلة للعينة التي استخدمت في الطبعة الثانية للمقياس ، وروى تساوى عدد كل من الذكور والإإناث .

وأوضحت نتائج تقدير الثبات أن قيم معامل الاستقرار لجميع الاختبارات الفرعية فيما عدا الاختبارين التكميليين في مقياس الأداء العملي تراوحت بين ٦٠ ، ٩٥ ، ٠ ، ٩٥ ولنسبة الذكاء الانحرافية المناهضة لدرجات كل من المقياسين الفرعيين والمقياس ككل حوالي ٩٠ ، ٠ ، ٠ ، كما تراوحت قيم معامل الاستقرار للعوامل الأربع التي أشرنا إليها بين ٨٠ ، ٩٥ ، ٠ ، ٠ .

أما فيما يتعلق بصدق المقياس ، فقد قدمت نتائج التحليل العاملى بعض الأدلة عن صدق التكوين الفرضي للمقياس ، ولكن لا توجد مؤشرات عن الصدق المرتبط بمحك ، حيث إن الدراسات لا تزال جارية في هذا الشأن .

وينبغي عند تفسير درجات الطفل في الاختبارات الفرعية مراعاة المتغيرات التي يمكن أن تؤثر في هذه الدرجات ، فدرجات الطفل في الاختبار الفرعى المتعلق بالحساب ربما تتأثر بعض المتغيرات مثل المهارة الحسابية للطفل ، وقدرتة على الاستدلال العددى ، ومدى تشتت انتباذه ، والمعلومات المكتسبة من المدرسة ، ومستوى القلق لديه ، والقدرة على العمل في حدود زمن معين وغير ذلك .

مقياس ويكسنر لذكاء الأطفال قبل المدرسة (WPPSI):

أعد ويكسنر أيضاً هذا المقياس عام ١٩٦٧ ، بهدف قياس ذكاء الأطفال الذين تتراوح أعمارهم بين أربعة أعوام وستة أعوام ونصف ، ولا يختلف هذا المقياس كثيراً عن مقياس ذكاء الأطفال (WAIS - III) فيما عدا إضافة ثلاثة اختبارات فرعية تناسب طفل ما قبل المدرسة هي :

(١) **بيت الحيوان** : وهو اختبار موقت يتطلب من الطفل وضع أسطوانة ملونة في فتحة مناسبة أمام الحيوان .

(٢) تصميم هندسي : وهو اختبار يتطلب مهارة إدراكية حركية .

(٣) الجمل : وهو اختبار تكميلي لقياس التذكر الفوري .

كذلك لا يختلف هذا المقياس عن مقياس ذكاء الأطفال سواء في طريقة بنائه أو تطبيقه أو تصحيحه .

وقد استندت معايير هذا المقياس على عينة قومية ممثلة تمثيلاً جيداً للمناطق الجغرافية ومهنة الآباء ، والنوع ، والسلالة ، واشتملت على ٢٠ طفل من الذكور والإثاث بالتساوي في كل من المستويات العمرية الستة التي يفصل بين كل منها نصف عام . ونحصل عن طريق المقياس على نسبة الذكاء الانحرافية التي متوسطها ١٠٠ وانحرافها المعياري ١٥ لكل من المقياس اللغظي ، ومقياس الأداء العملي ، والمقياس ككل .

وقد بيّنت الدراسات أن قيم معامل الاتساق الداخلي لكل من المقياسين الفرعيين والمقياس ككل كانت في حدود ٩٠ ، أو أقل ، مما يدل على ارتفاع قيمة ثبات درجة كل منها ، كما تراوحت قيم معامل الاستقرار بفواصل زمني ثلاثة أشهر لعمر خمسة أعوام بين ٨٠ ، ٩٠ ، . كذلك وجّد أن قيمة معامل الارتباط بين الدرجات الكلية في المقياس والدرجات الكلية في مقياس ستانفورد - بيئي (طبعة عام ١٩٦٠) بلغت ٧٦ ، . مما يدل على أن المقياسين يتفقان إلى حد ما فيما يقيسانه .

مقارنة بين مقياس ويكسنر ومقياس ستانفورد - بيئي :

إن كثيراً من ميزات مقياس ستانفورد - بيئي توافر في مقياس ويكسنر ، وكما أوضحنا فإن المقياسين مرتبطان ارتباطاً كبيراً في معظم المستويات العمرية . وعلى الرغم من أن كلا من المقياسين ربما يستخدم بدرجة أفضل في أغراض معينة ، إلا أن اختيار أي منهما يعتمد على ما يفضله مستخدم المقياس ، لذلك ربما يكون من المفيد أن نوضح فيما يلى بعض الفروق الأساسية بين كل من المقياسين ليترشد بها في اختياره :

(١) إن الأساس المنطقي الذي بُني عليه مقياس ويكسنر هو الحصول على قياس للذكاء العام للراشدين وخاصة ، ومن ثم اعتمد على هذا المقياس في بناء مقياسيه الآخرين لذكاء الأطفال ، فالمقياسان مماثلان لمقياس ارشدين فيما عدا بعض الاختبارات الفرعية التي تناسب الأطفال ، وأحددهما مصمم للأطفال الذين تتراوح أعمارهم بين ٥ ، ١٧ عاماً ، والآخر لأطفال ما قبل المدرسة الذين تتراوح أعمارهم بين ٤ ، ٦، ٥ عاماً .

أما مقياس ستانفورد - بينيه فهو مصمم أساساً لقياس الذكاء العام للأطفال الذين تتراوح أعمارهم بين ٢ ، ١٨ عاماً ، ويمكن تطبيقه على الراشدين . لذلك فإنه يفضل عن مقياس ويكسنر في قياس ذكاء أطفال ما قبل المدرسة . وجميع هذه المقاييس تعد من المقاييس الفردية .

(٢) مقاييس ويكسنر تشتمل على اختبارات فرعية مرتبة وتطبق وفقاً لهاذا الترتيب ، بينما الاختبارات التي يشتمل عليها مقياس ستانفورد - بينيه مرتبة وفقاً للمستويات العمرية .

(٣) مقاييس ويكسنر تتضمن مهاماً لفظية وأخرى عملية ، بينما محتوى مقياس ستانفورد - بينيه مثقل بالجانب اللفظي .

(٤) مقاييس ويكسنر تزودنا بثلاث نسب للذكاء تتعلق بالجانب اللفظي ، والأداء العملي ، والمقاييس ككل ، وكذلك عشر درجات للاختبارات الفرعية ، بينما يقتصر مقياس ستانفورد - بينيه على نسبة ذكاء كلية واحدة إضافة إلى درجة تمثل العمر العقلي .

(٥) تُعد مقاييس ويكسنر من نوع الموازين التي تعتمد على النقاط Point Scale ، أما مقياس ستانفورد - بينيه فيُعد من الموازين التي تعتمد على العمر Age Scale .

(٦) يمكن استخدام مقاييس ويكسنر كأدوات تشخيصية في المواقف المدرسية والклиничية ، غير أن هذا يتطلب فاحصاً مدرباً تدربياً عالياً على كيفية استخدام هذه المقاييس وتفسير درجاتها .

(٧) على الرغم من عدم وجود أدلة كافية عن مدى تبنّؤ مقاييس ويكسنر بالتحصيل الدراسي والنجاح المهني ، إلا أن الأدلة المتاحة أوضحت أن هذه المقاييس لا تختلف في ذلك عن مقياس ستانفورد - بينيه .

(٨) يصعب باستخدام مقاييس ويكسنر قياس المستويات الدنيا للضعف العقلي وتؤدي إلى نسب ذكاء أكبر من حقيقتها عند هذه المستويات . ومع هذا فإن سقف المقاييس منخفض نسبياً في المستويات العليا ، وتوارد إلى نسب ذكاء أقل من حقيقتها في هذه الحالات . وعلى الرغم من اشتغال هذه المقاييس على مهارات لفظية وعملية ، إلا أنها مثل مقياس ستانفورد - بينيه تقيس مجموعة محدودة من العوامل أو القدرات التي تمثل السلوك الذكي .

ومن الجدير بالذكر أن مقياس ويكسنر - بلفيو للذكاء الراشدين والمهامين ومقاييسه للذكاء الأطفال ، وكذلك مقياس ستانفورد - بينيه أجرى تعريفها وتقنينها في البيئة المصرية .

بعض مقاييس الذكاء الفردية الأخرى:

توجد مقاييس فردية متعددة أخرى لقياس الذكاء، ولكنها ربما تكون أقل شيوعاً من مقاييس ويكسنر، ومقاييس ستانفورد - بينيه. وبعض هذه المقاييس تطبق على الأسواء، والبعض الآخر يناسب الأفراد من ذوى الإعاقات الحسية أو الجسمية أو اللغوية، فقد لاحظنا أن عينات تقدير المقاييس السابقة لم تتضمن أمثال هؤلاء الأفراد.

وعلى الرغم من أن معظم مقاييس الذكاء الفردية الأخرى لا ترقى في خصائصها السيكومترية، ونوعية عينيات تقديرها وكفايتها، واتمام أداتها وتعليماتها، إلى مستوى مقاييس ويكسنر وستانفورد - بينيه، إلا أنها تفي في أغراض معينة، أو لفئات خاصة. وكثير منها لا يعتمد بدرجة كبيرة على الاستجابات اللفظية التي اهتمت بها المقاييس السابقة، وإنما تتطلب أن يشير الفرد ، أو يستجيب بـ (نعم) أو (لا) دون الحاجة إلى إحداث تكامل بين الوظائف البصرية والحركية التي تتطلبها الاختبارات الفرعية لمقاييس ويكسنر .

كذلك يشتمل كثير منها على اختبارات لقياس الأداء العملي، ولعل النقص الملحوظ في هذا الجانب العملي في مقاييس ستانفورد - بينيه هو الذي شجع على بناء هذه المجموعات الأخرى من مقاييس الذكاء الفردية .

ونظراً لأن بعض هذه المقاييس صُمِّمت لأغراض معينة أو لفئات خاصة ، فإن استخداماتها تكون مقصورة على هذه الفئات ، وهذا يحد من تنوع الوظائف والإمكانات العقلية التي يمكن أن تقيسها ، ومع هذا فإنها تعد أكثر ملاءمة لهذه الفئات من المقاييس الرئيسة السابقة .

غير أنه في هذه الحالة لا نستطيع مقارنة نسب الذكاء التي نحصل عليها من هذه المقاييس الأخرى بحسب الذكاء التي نحصل عليها من المقاييس الرئيسة . ويمكن اعتبار نتائج هذه المقاييس الأخرى مكملاً لنتائج تطبيق مقاييس ويكسنر وستانفورد - بينيه، حيث تفي في أغراض الفحص والمتابعة وإعادة التقييم ، أو إذا لم يكن هناك وقت متاح لتطبيق أحد هذين المقاييس الرئيسين .

وفيما يلى أسماء بعض هذه المقاييس :

١) **بطارية التقييم للأطفال التي أعدها كوفمان عام ١٩٨٣ .**

Kaufman Assessment Battery for Children

وهي مجموعة اختبارات فردية لقياس كل من الذكاء والتحصيل وتناسب الأطفال الذين تتراوح أعمارهم بين ٢،٥ ، ١٢،٥ عاماً ، وتميز بتركيز مفرداتها على العمليات

العقلية التي يستخدمها الطفل في حل المشكلات بدلاً من تركيزها على المحتوى، ولا يعتمد قياس هذه العمليات على ما تعلمه الطفل في المواد الدراسية. وقد استند بناء الاختبارات الفرعية إلى النظريات السيكولوجية ونظريات سيكولوجية الأعصاب. وتتضمن الاختبارات مفردات لفظية وغير لفظية، وأعدت معايير متصلة للأطفال الأسياء والأطفال من ذوى الاحتياجات الخاصة في البيئة الأمريكية.

وتشتمل البطارية على ١٦ اختباراً فرعياً، عشرة منها تتعلق بالذكاء العام، والستة الباقية تتعلق بالتحصيل.

وتنقسم اختبارات الذكاء العام العشرة إلى مجموعتين تتطلب إدراهما اختبارات التجهيز المتتابع للمعلومات Sequential Processing، والأخرى تتطلب التجهيز الآنى للمعلومات Simultaneous Processing . وكلاهما يتطلب من الطفل توليف أو إحداث تكامل بين معلومات مكانية أو متشابهات.

أما الاختبارات الفرعية التي تتعلق بالتحصيل فتشتمل على مفردات تتضمن كلمات تعبيرية، وحساب ، وألغاز ، وقراءة ، وفهم المقروء .

ولا تطبق جميع الاختبارات الفرعية على الأطفال في المستويات العمرية المختلفة، وإنما تطبق سبعة اختبارات فرعية على الأطفال الذين تبلغ أعمارهم ٢،٥ عاماً، وثلاثة عشر اختباراً على الأطفال في المستويات العمرية ٧ - ١٢،٥ عاماً، ويمكن الحصول على درجتين إدراهما تشير إلى التجهيز المتتابع والأخرى تشير إلى التجهيز الآنى للمعلومات .

ونظراً لحداثة هذه البطارية ، فإنها تتطلب مزيداً من البحث ، فيما يتعلق بصدق قرارات التشخيص والتفسير في الإطار المدرسي استناداً إلى نتائجها.

(٢) **بطارية الاختبارات السيكولوجية التربوية التي أعدها وودكوك وجونسون عام ١٩٧٧ وأجري تعديليها في طبعتها الثانية عام ١٩٨٩ .**

Woodcock - Johnson Psychoeducational Battery (WJ - R)

وتشتمل على مجموعة متسعة من الاختبارات التي تطبق فردياً على الأفراد الذين تتراوح أعمارهم بين عامين ، ٩٥ عاماً، وتقيس الاختبارات القدرات المعرفية، والاستعداد الدراسي ، والتحصيل.

(٣) **مقاييس القدرات الفارقة التي أعدها إيليوت Elliot عام ١٩٩٠ .**

Differential Ability Scales (DAS)

وتشتمل على اختبارات معرفية واختبارات تحصيل للأطفال الذين تتراوح

أعمارهم بين ٢،٥ ، ١٧ عاماً، وتُعد هذه المقاييس تعديلاً لمقاييس الذكاء البريطانية British Ability Scales . وقد تم تنظيم مقاييس القدرات الفارقة في مجموعتين إحداهما البطارية المعرفية Cognitive Battery وتشتمل على مستويين أحدهما لأطفال ما قبل المدرسة والآخر لأطفال المدرسة الابتدائية ، والمجموعة الأخرى اختبارات تحصيل دراسي تشتمل على مهارات حسابية أساسية ، وهجاء ، وقراءة كلمات.

إلى جانب هذه المقاييس الفردية المهمة توجد مقاييس عديدة أخرى لذكاء الأطفال حديثي الولادة ، مثل مقياس برازيلتون (عام ١٩٧٣).

Brazelton Neonatal Assessment Scale (BNAS)

وكذلك مقياس بيلي Bayley (١٩٦٩) لنمو الأطفال الذين تترواح أعمارهم بين شهرين ، ٣٠ شهرًا Bayley Scale of Infant Development ، ويقيس الوظائف العقلية والحركية لهؤلاء الأطفال.

كما توجد مقاييس فردية للأطفال من ذوى الإعاقات وفئات خاصة ، مثل مقياس كولومبيا للنضج العقلى (عام ١٩٧٢) Columbia Mental Maturity Scale (CMMS) للأطفال الذين يعانون من إعاقات جسمية ، أو قصور لغوى ، أو فقدان سمع ، وتتراوح أعمارهم بين ٣ ، ١٢ عاماً، ولا يتطلب المقياس استجابات لفظية أو مهارات حركية دقيقة . ومقياس إيلينوى للقدرات السينكولوجية اللغوية Illinois Test of Psycholinguistic Abilities (ITPA) وهو من المقاييس الحديثة نسبياً التي تقيس العجز في التعلم Learning Disabilities لدى الأطفال الذين تترواح أعمارهم بين عامين وعشرة أعوام . واستند المقياس إلى المفاهيم المعاصرة في تجهيز المعلومات Human Information Processing حيث افترض أن إخفاق الطفل في الاستجابة الصحيحة لمثير معين ربما يكون ناتجاً ليس فقط من نظام مخرجات قاصر (أى استجابة) ، وإنما من مدخلات قاصرة ، ومن نظام تجهيز معلومات قاصر أو من أحدهما . وقد لاقى هذا المقياس اهتماماً كبيراً من جانب المربين والمهتمين بهذه الفئات الخاصة وكذلك الباحثين .

ويلاحظ أن معظم هذه المقاييس تقتصر على الأطفال الصغار ، وعددًا قليلاً منها يناسب الأطفال الأكبر سنًا والراهقين . وبعض هذه المقاييس يهتم بالمهارات غير اللفظية أو الذكاء غير اللفظي مثل المهارات الإدراكية الحركية . كما أنه يمكن أن نحصل باستخدام بعض هذه الاختبارات على درجة كلية واحدة كما في مقياس بيبيه ، والبعض الآخر يمكننا من الحصول على عدة درجات كما في مقياس ويكسنر . ويطلب تطبيق

بعض هذه الاختبارات فاحصاً مدرِّباً تدرِّباً عالِيَاً مثل مقاييس ويكسنر ويبينيه ، في حين يتطلب البعض الآخر قدر أقل من المهارة ، أو تطبق تحت إشراف أخصائي.

ثانياً: اختبارات الذكاء الجماعية :

سبق أن أوضحنا عند عرضنا لقياس الذكاء من منظور تاريخي نشأة حركة اختبارات الذكاء الجماعية بغرض انتقاء الجنود وتصنيفهم والتتوسيع بعد ذلك في استخدامها في مختلف مجالات الحياة المدنية . وتميز هذه الاختبارات بإمكانية تطبيقها على أعداد كبيرة من الأفراد في وقت واحد مما يُقلل زمن العملية الاختبارية وما تتطلبه من نفقات مالية وإمكانات فنية على عكس ما تتطلبه اختبارات الذكاء الفردية ، ولعل ذلك كان السبب الرئيسي لبناء اختبار الجيش ألفا Army Alpha Test ، واختبار الجيش بيتا Army Beta Test ، واختبار التصنيف العام لأفراد القوات المسلحة (AGCT) ، والاختبار البديل له (AFQT) ، وهي الاختبارات الجماعية التي أشرنا إليها من قبل .

وأختبارات الذكاء الجماعية تكون عادةً اختبارات ورقية وقلم ويشتمل معظمها على مفردات اختيار من متعدد مما ييسر عملية التصحيح . وبذلك لا تتطلب هذه الاختبارات مواد مرتفعة الكلفة أو فاحصين على درجة عالية من المهارة لتطبيق الاختبار أو تصحيحه بموضوعية وسرعة ، إذ يمكن تصحيحه باستخدام الحاسوب . غير أن استخدام هذا النوع من المفردات في كثير من هذه الاختبارات يجعل من غير الممكن اختبار الأطفال أو الأميين أو الذين لديهم إعاقات سمعية أو بصرية بها مما يتطلب اختبارات ذكاء جماعية غير لفظية ..

ويبينما تُستخدم اختبارات الذكاء الفردية في تقييم وتشخيص المشكلات النفسية في المواقف الكlinيكية ، فإن اختبارات الذكاء الجماعية تُستخدم في المدارس والجامعات والمصانع والمؤسسات العسكرية ، كما تُستخدم بكثرة في البحوث التربوية والنفسية . ويمكن الإفاده من نتائج هذه الاختبارات في اتخاذ قرارات انتقاء العاملين Personnel Selection ، وأغراض التصفية Screening ، والتوجيه المهني Vocational Guidance ، وذلك لتقييم القدرات العقلية ، والمهنية ، والاستعدادات الخاصة المتعلقة بمهن أو أعمال معينة .

غير أن هذه الاختبارات لا تسمح بتدخل القائمين بتطبيقها من حيث حفز المختبرين أو الثناء على استجاباتهم وهو ما يميز تطبيق اختبارات الذكاء الفردية . وهنا يفترض أن المختبرين لديهم دافعية مرتفعة وميل إلى التعاون دون حنهم على ذلك .

ونظراً لأن بعض المُختبرين ربما لا يكونون كذلك، أو تكون لديهم مشكلات صحية أو خوف من الفشل أو قلق ، فإنه يحتمل أن يحصلوا على درجات أقل مما يستحقون. والحقيقة أن الظروف الطارئة التي ربما تؤثر في أداء الفرد في الاختبار يصعب التعرف عليها في مواقف الاختبارات الجماعية. لذلك ينبغي الحيطة في تفسير درجات هذه الاختبارات، واعتبار هذه الدرجات تقديرات تقريرية وليس درجات مطلقة.

ويُفضل الحصول على أكثر من درجة واحدة للفرد قبل اتخاذ قرارات أو استدلالات تتعلق بذكائه أو إمكاناته العقلية، وقد سبق أن ناقشت ذلك في الفصل الثاني عندما تناولنا الشروط العلمية لتفسير درجات الاختبارات وتقديم نتائجها.

كما أن هذه الاختبارات أقل فائدة في التنبؤ بالنجاح الأكاديمي نظراً لأن نوع الذكاء المطلوب في الدراسة الأكademie متعلق بنوعية الدراسة. فهذا النوع يؤكّد بدرجة كبيرة الاستدلال اللفظي والعددي، بينما يؤكّد الجوانب الاجتماعية والعملية للذكاء بدرجة أقل. ففي المواقف الأكاديمية نهتم بقدرة الفرد في الأداء الأكاديمي ، فالقدرة المرتفعة على الاستدلال المجرد لدى الفرد ربما تختلف اختلافاً كبيراً عن القدرة أو الذكاء المطلوب للتعامل في المواقف الحياتية والتكيف مع هذه المواقف.

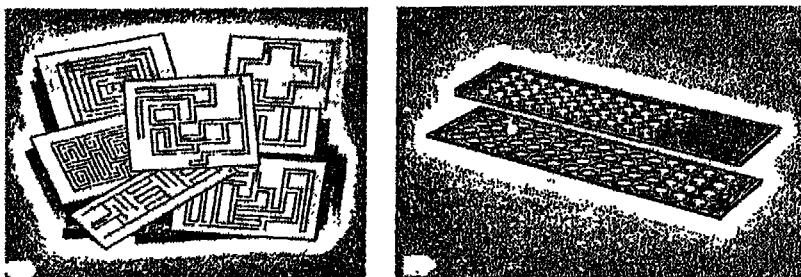
أنواع اختبارات الذكاء الجماعية:

تعدد أنواع اختبارات الذكاء الجماعية وتباين تباينًا ملحوظًا في محتواها وتكونيتها وكيفية الإجابة عن مفراداتها ، ولعل هذا يتضح بدرجة كبيرة عند إطلاعنا على المجلد الأخير لكتاب السنوى للفياس النفسي Mental Measurement Yearbook الذي سبق أن أشرنا إليه (Buros, 1998) .

محتوى الاختبارات : يمكن تصنيف محتوى اختبارات الذكاء الجماعية إلى نوعين : محتوى لفظي Verbal ، ومحظى غير لفظي Nonverbal ، فالمحظى اللفظي يتطلب القدرة على القراءة أو على الأقل فهم التعليمات التي يُقدمها القائم بعملية التطبيق سواء قام بقراءتها أو توضيحها.

أما المحظى غير اللفظي فلا يتطلب استخدام اللغة سواء في تطبيق الاختبار أو في الإجابة عنه ، ومثال ذلك اختبار الجيش بيتا Army Beta الذي أشرنا إليه من قبل. وهذا النوع من الاختبارات يكون مناسباً بل وضرورياً للأطفال الصغار والأمينين وضعاف العقول والفتات الأقل حظاً الذين لم ينالوا قدرًا مناسباً من اللغة.

ويشتمل محتوى هذه الاختبارات غير اللفظية على مواد مصورة ، وأشكال ، ومتاهات ، ومواد عيانية أخرى ، كما يمكن أن تتطلب نوعاً من الأداء العملي Preformance كما في مقاييس ويكسنر مثل بناء مكعبات ، ومتاهات ، والغاز مكانية . ويوضح شكل (١١ - ٩) التالي بعض هذه المواد غير اللفظية & (Noll Scannell, 1972)



شكل (٩ - ١١) يوضح بعض المفردات غير اللفظية لاختبار المتاهات

أما مفردات الجانب اللفظي في معظم اختبارات الذكاء الجماعية فتشتمل على: كلمات ، وإكمال جمل ، وحساب ، ومتباينات لفظية ، ومتضادات لفظية ، وسلسلات لفظية وعددية ، وتميز بين الفروق في تسلسل من الكلمات ، وأمثال ، وترتيب جمل ، واستدلال لفظي ، واسترجاع مؤجل ، أي إجابة سلسلة من الأسئلة بعد قراءة قصة معينة . أي أن محتوى هذه الاختبارات يتعلق بالكلمات ، والمعلومات العامة ، والحساب ، والاستدلال .

وعلى الرغم من أن الارتباط بين نوعي الاختبارات اللفظية وغير اللفظية ليس مرتفعاً بدرجة كبيرة ، إلا أن مقدار الارتباط يسمح بتبرير استخدام المحتوى غير اللفظي في الحالات التي تتطلب ذلك وبخاصة الأداء العملي . لذلك نجد أن كثيراً من اختبارات الذكاء التي تُستخدم في وقتنا الحاضر يشتمل على كل من المحتويين ، وأحياناً يكون محتوى الاختبار بطريقة تسمح بالحصول على نسبتي ذكاء إحداهما تتعلق بالجانب اللفظي والأخرى تتعلق بالجانب غير اللفظي ، وتمثلان معًا الذكاء العام مقاساً بطرقين مختلفتين .

تكوين الاختبارات :

على الرغم من أنها نحصل باستخدام بعض الاختبارات الجماعية على درجتين إحداهما للجانب اللفظي ، والأخرى للأداء العملي إلا أن هاتين الدرجتين لا تنفصل إحداهما عن الأخرى ، أي لا تدلان على قدرتين مختلفتين .

ولكن هناك اختبارات نحصل باستخدامها على مجموعة من الدرجات الفرعية التي تمثل جوانب مختلفة للذكاء، وتسمى بطاريات اختبارات متعددة العوامل Multifactor Aptitude Batteries سوف تتناولها في الفصل العاشر. وعادةً تُصنف الاختبارات بحسب الصنف الدراسي أو المستوى العمري. غير أنه توجد اختبارات ذات مستويات مختلفة يناسب كل مستوى منها صفوًا دراسية معينة، أي أنها تقيس الذكاء في مستويات الاختبار المتعددة. وتختار عينات التقنيين بحيث يمكن المقارنة بينها من مستوى إلى آخر، وهذا يسمح بمقارنة القياسات التي نحصل عليها عبر الزمن.

فلكي نحصل على قياسات النمو العقلي بمفردات متسعة المدى في صعوبتها تُبني سلسلة من بطاريات متداخلة متعددة المستويات Multilevel Batteries. ويختبر كل فرد بالمستوى المناسب له فقط، ويمكن إعادة اختباره في الأعوام التالية بالمستويات الأخرى، كما يمكن استخدام هذه المستويات في المقارنة بين مجموعات عمرية مختلفة، لذلك كان لابد من إحداث تداخل بين البطاريات المتسلسلة.

وتُصنف عادةً مفردات اختبارات الذكاء الجمعية، ويتم تجميع مفردات كل صنف في اختبار فرعي واحد، ومن ثم تُرتّب مفردات الاختبار الفرعي ترتيباً تصاعدياً بحسب مستوى صعوبتها. غير أن بعض الاختبارات تخلط أنواع المفردات بشكل حلووني شامل Spiral Omnibus Forms حيث تُقدم مفردة من كل نوع (لفظي بأنواعه، أو غير لفظي مثلاً)، وتكرر السلسلة ولكن بمفردات أكثر صعوبة. وهذا يسمح للفرد باستمرار الإجابة ولا تكون هناك ضرورة لتحديد زمن إجابة لكل اختبار فرعي.

كيفية الإجابة عن مفردات الاختبارات:

كثير من اختبارات الذكاء الجمعية تتطلب السرعة والدقة في الإجابة Speed Tests وهذه تكون اختبارات موقوتة، وتميز مفرداتها بالسهولة النسبية حيث يجب الفرد عن عدد كبير منها في وقت محدد. غير أن هناك بعض الاختبارات تعتمد على القوة Power Tests ، وهذه تكون مفرداتها متدرجة في صعوبتها وليس لها زمن محدد. وفي الحقيقة لا يستطيع الإجابة عن جميع مفردات الاختبار الذي يتطلب السرعة إلا الأفراد النابهين، أما درجة الفرد في اختبارات القوة فتعتمد على صعوبة المفردات التي يستطيع الإجابة عنها.

أمثلة لبعض اختبارات الذكاء الجمعية:

نظراً لعدد أنواع اختبارات الذكاء الجمعية وتنوعها حيث يوجد عدد كبير من هذه الاختبارات تناسب جميع المستويات العمرية من الحضانة إلى سن الرشد، فإنه يصعب حصر هذه الاختبارات أو تعریف القارئ بها، وإنما يمكن الرجوع إلى الكتاب

السنوى للقياس النفسي (1998 ، Buros) للاطلاع على أحدث هذه الاختبارات بالتفصيل ، وكذلك أدلة الاختبارات في البيئة العربية .

غير أننا سوف نقدم في الجزء السالى أمثلة لبعض أنواع الاختبارات الأكثر استخداماً في قياس الذكاء ، وعلى الرغم من أنها مقتنة في البيئة الأمريكية إلا أن بعضها تم تعربيه وتقنيته في البيئة المصرية وفي بيئات عربية أخرى .

اختبارات ذكاء جماعية لفظية متعددة المستويات :

(١) اختبار كاليفورنيا للنضج العقلي California Test of Mental Maturity (CTMM) عام ١٩٦٣ :

ويشتمل على سلسلة من الاختبارات الفرعية عددها ١٢ اختباراً تقيس عمليات عقلية تمثل في الذاكرة ، وال العلاقات المكانية ، والاستدلال المنطقى ، والاستدلال العددى ، ومفاهيم لفظية . والاختبار يتكون من عدة صيغ للفياس التابعى فى مستويات ستة مخصوص لكل منها كتيب اختبارى .

وهذه المستويات تبدأ من الحضانة إلى سن الرشد . ويمكن باستخدام الاختبار الحصول على ثلاثة درجات كلية للجزء اللفظى والجزء غير اللفظى ، والاختبار ككل . ويستغرق تطبيقه ما بين ٤٨ دقيقة ، ٨١ دقيقة للمستوى ، لذلك اجتُصر الاختبار في صيغة أقصر .

(٢) اختبار القدرات المعرفية Cognitive Abilities Test (CAT) ١٩٧٨ :

وهو اختبار متعدد المستويات حيث يشتمل على بطارية اختبارات للصفوف من الحضانة إلى الصف الثالث الابتدائى Primary Battery ، وبطارية متعددة المستويات Multilevel Battery للصفوف من الثالث إلى نهاية مرحلة التعليم الثانوى . وتكون البطارية الأولى من صيغتين لكل مستوى ، وكلاهما يشتمل على أربعة اختبارات فرعية هي : الكلمات الشفوية ، والمفاهيم العلاقية ، وعمليات عقلية متعددة ، ومفاهيم كمية . ويستغرق تطبيق كل اختبار فرعى ربع الساعة تقريباً في أربع جلسات منفصلة ، وتستخدم معاير نسبة الذكاء الانحرافية والمبنيات في تفسير الدرجات .

أما البطارية الثانية فتشتمل على ثلاثة بطاريات محتواها لفظى ، وكمي ، وغير لفظى على الترتيب . ويتضمن المحتوى اللفظى كلمات ، وإكمال جمل ، وتصنيف لفظى ، ومتشابهات لفظية ، ويتضمن المحتوى الكمى علاقات كمية ، وسلالس أعداد ، وتكوين معادلات ، وتعامل مع الأعداد وغيرها من الرموز الكمية ، ويتضمن المحتوى غير اللفظى أنماط أشكال هندسية ومكانية ، وهذه تتضمن تصنيف أشكال ، ومتشابهات

شكلية، وتكوين أشكال، وكل من هذه المحتويات تقيس باختبارات فرعية. وتعتمد معايير هذه البطارئ ذات المستويات المتعددة على الدرجات المكافئة للعمر، والرتب المئوية. والتساعيات المعيارية المكافئة للصفوف.

ومن الجدير بالذكر أن هذا الاختبار يعد تطويراً للاختبار الشائع الاستخدام الذي أُعده لورج، وثورنديك Thorndike Test - Lorge عام ١٩٦٣.

(٣) اختبارات أوتيس ولينون لقدرة العقلية العامة Otis - Lennon Mental Ability Tests عام ١٩٦٧ :

وتكون هذه الاختبارات من صيغة واحدة، وتشتمل على مفردات متعددة بشكل حلزوني شامل Spiral - Omnibus لقياس القدرة العقلية العامة. وتوجد ستة مستويات للاختبارات تمتد من الصف الأول الابتدائي إلى نهاية مرحلة التعليم الثانوى، ويستغرق تطبيق الاختبارات من ٤٥ - ٣٠ دقيقة بحسب المستوى. وتحصل باستخدام الاختبار على درجة واحدة للفرد. وتعتمد المعايير على نسب الذكاء الانحرافية، والمئويات، والتساعيات المعيارية المقابلة للعمر والصف الدراسي، وكذلك يمكن الحصول على الأعمار العقلية للمستويات الثلاثة الأولى.

(٤) اختبارات الذكاء التشاركي بين المدارس والكلليات Cooperative School and College Ability Tests , Series II (SCAT , Series II) عام ١٩٦٧ :

وهذه الاختبارات من إعداد مركز الخدمات الاختبارية ETS بالولايات المتحدة الأمريكية بالتعاون مع العديد من الخبراء التربويين، وهي سلسلة من الاختبارات متعددة المستويات تقيس القدرة العقلية العامة. وتُستخدم في أربعة مستويات تبدأ من الصف الرابع الابتدائي. وتوجد صيغتان متكافئتان أو ثلاث لكل اختبار في كل مستوى. ويشتمل كل منها على ٥٠ مفردة تتعلق بال مشابهات اللفظية، ٥٠ مفردة أخرى تتعلق بالمقارنات الكمية. ويمكن باستخدام هذه الاختبارات الحصول على ثلاثة درجات كلية للجوابات اللفظية والكمية وللاختبار عامنة، وتحوّل هذه الدرجات إلى رتب مئوية.

(٥) اختبار كالمان وأندرسون للذكاء Kuhlmann - Anderson Test - (KAT) عام ١٩٦٣ :

ويُعد من اختبارات الذكاء الجماعية، ويشتمل على ثمانية مستويات منفصلة لجميع الصفوف المدرسية، أي من الحضانة إلى نهاية مرحلة التعليم الثانوى، وتشتمل كل من هذه المستويات على اختبارات متعددة يحتوى كل منها على أنواع مختلفة من المفردات.

وهذه المجموعة من الاختبارات تُعد بمثابة بطارية متعددة المستويات Multilevel Battery نظراً لأنها تشمل مستويات متعددة من العمر أو الصف الدراسي. ومفردات الاختبارات في المستويات العمرية الدنيا محتواها لفظي ولا تتطلب إلا قدرًا ضئيلاً من مهارات القراءة، وكلما ارتفع المستوى يزداد المحتوى اللفظي ، ولكن يظل المحتوى غير اللفظي سائد. لذلك فإن هذه البطارية **تُناسب الأطفال الصغار ، وتُفيد أيضًا في قياس ذكاء من لديهم صعوبات في القراءة أو اتباع التعليمات اللفظية.**

ويُعد هذا الاختبار من أقدم وأفضل اختبارات الذكاء الجماعية للأطفال حيث اعتنى ببنائه عناية فائقة ، وأُجري تعديله عدة مرات منذ عام ١٩٢٧ حتى طبعته الأخيرة السابعة.

والدليل على تميز هذا الاختبار أن كثيراً من مفرداته استُخدمت في اختبارات جماعية متنوعة أخرى. ويمكن الحصول على ثلاثة درجات: لفظية، وكمية ، وكلية، وفي بعض المستويات يمكن التعبير عن هذه الدرجات بنسب ذكاء انحرافية ، وفي مستويات أخرى يعبر عنها بواسطة معايير المدى المئي Percentile Bands لتقدير فترات الثقة في الرتب المئوية، كما سبق أن أوضحنا في الفصل السادس. ويتراوح زمن تطبيق الاختبار بين ٥٠ - ٦٠ دقيقة.

(٦) اختبارات هنمنون، وتلسون للقدرات العقلية

: ١٩٧٣ Year Nelson Tests of Mental Ability (H - NT)

ويُعد هذا الاختبار أيضًا من الاختبارات الشائعة الاستخدام واعتُنى ببنائه بحيث يشمل أربعة مستويات بحسب الصفوف المدرسية أى الصفوف ٣ - ٦ ، ٦ - ٩ ، ٩ - ١٢ ، وهذه تمثل الصيغة الأولى ، والصفوف من الحضانة إلى نهاية الصف الثاني ، وهذه تمثل البطارية الأولية. وقد أعدت في طبعة حديثة اختبارات للمستوى الجامعي. وكل من المستويات الثلاثة للصيغة الأولى تشتمل على ٩٠ مفردة مرتبة ترتيباً حلزونياً شاملًا Spiral Omnibus ، وتتناول كلمات مختلطة ، ومتشابهات لفظية ، وتصنيف لفظي ، واستدلال لفظي ، وسلال عددية ، واستدلال حسابي ، ومتشابهات شكلية ، واتباع تعليمات. ويتراوح زمن تطبيق اختبارات هذه الصيغة ٣٠ دقيقة تقريباً.

أما البطارية الأولية فتشتمل على ثلاثة اختبارات فرعية هي: اختبار استماع إلى معلومات عامة ، واختبار كلمات مُصورة ، واختبار حجوم وأعداد. وهذه البطارية غير موقوتة ، ولكن تطبيقها يستغرق عادة حوالي ٣٠ دقيقة.

وتعتمد معايير جمجمة الاختبارات على نسب الذكاء الانحرافية، والتساعيات المعيارية، والرتب المئوية المقابلة للعمر والصف الدراسي. وعلى الرغم من أن الاختبار يؤدي إلى درجة كلية واحدة يفترض أنها تمثل الذكاء العام، وهو ما تهدف الاختبارات لقياسه، إلا أنه يمكن الحصول على مجموعتين من المعايير يعتمدان على توزيع الدرجات الخام وفقاً لكل من العمر والصف الدراسي على الترتيب.

اختبارات ذكاء جماعية غير لفظية:

ذكرنا فيما سبق أن اختبارات الذكاء الجماعية غير اللفظية NonVerbal تستخدمن في قياس ذكاء الأفراد الأميين، وبعض الفئات الخاصة التي لديها إعاقات ثقافية أو لغوية أو جسمية، والأفراد الذين تكون قدرتهم على القراءة محدودة، وكذلك الأطفال الصغار. ويمكن أن تكون هذه الاختبارات من النوع الذي يتطلب ورقة وقلم، أو يتطلب أداء مهام معينة أو يوضحها من قبل. وهذه الاختبارات لا تعتمد على اللغة، بل يمكن تطبيق بعضها دون الحاجة إلى تعليمات لفظية.

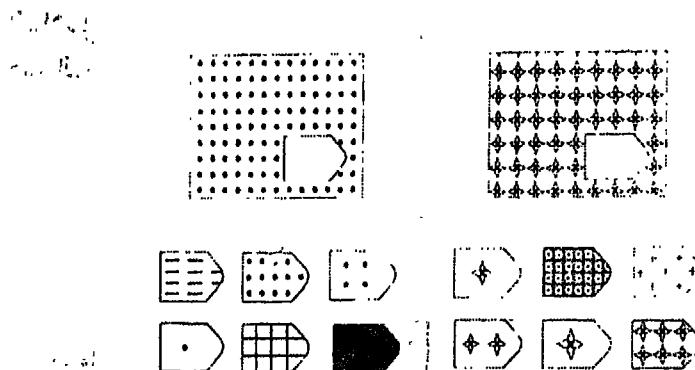
وفيما يلى أمثلة لبعض هذه الاختبارات الشائعة الاستخدام:

(١) اختبار المصفوفات المتتابعة لرافن

Raven Progressive Matrices عام ١٩٤٧ :

وقد أعد عالم النفس رافن . J . Raven . L. Penrose . هذا الاختبار في إنجلترا. ويُعد من أكثر مقاييس الذكاء الجماعية غير اللفظية شيوعاً واستخداماً في قياس القدرة العقلية العامة، وقد صُمم لتقييم ذكاء الأفراد المجندين دون أن تتأثر درجاتهم بالعوامل المتعلقة بالتعليم.

ويشتمل الاختبار على ٦٠ مصفوفة أو تصميم أحد أجزائه مقطوع، وعلى الفرد أن يختار الجزء المقطوع من بين بدائل معطاة عددها ستة، أو ثمانية. ومفردات الاختبار مصنفة في خمس مجموعات متسلسلة، كل منها تشتمل على ١٢ مصفوفة متزايدة الصعوبة. وتتطلب الإجابة إدراك المشابهات، وإجراء تبديلات على الأنماط وغير ذلك من العلاقات المنطقية. ويوضح شكل (٩-١٢) التالي مثلاً لهذه المصفوفات : (Jensen , 1980)



شكل (٩ - ١٢) يوضح مثلاً من مفردات اختبار المصفوفات المتتابعة

وللختبار صيغ ثلاثة إحداها صيغة معيارية، والثانية صيغة ملونة تناسب الأطفال في مرحلة ما قبل المدرسة، وضعاف العقول، والثالثة صيغة متقدمة أكثر صعوبة تناسب المستويات الأعلى من التعليم الابتدائي والمتوسط.

ويُعد الاختبار من نوع اختبارات القوة التي لا تتطلب زماناً محدوداً للإجابة، ولكنه يستغرق زمناً يتراوح بين ١٥ ، ٣٠ دقيقة، ويمكن تطبيقه فردياً أو جماعياً بتعليمات شفوية بسيطة للغاية. ومعاييره المئوية تم تحديدها باستخدام عينات من تلاميذ المدارس البريطانية والراشدين.

ويُعد هذا الاختبار من النوع الذي يطبق عليه الاختبارات غير المتحيز للثقافة Culture - Fair Tests ، والهدف من هذه الاختبارات إتاحة فرص متكافئة للأفراد من ثقافات مختلفة في إجابتهم عن الاختبار. غير أنه تبين أنه ربما يصعب بناء مفردات اختبار يكون محتواها مستقلاً عن الخبرات الممتداة للأفراد من ثقافات مختلفة. كما أن هذه الاختبارات لا تتبناً بدرجة جيدة بالتحصيل الأكاديمي أو الأداء في العمل، وصدقها في أغراض الانتقاء، والتسكنين يكون منخفضاً بالمقارنة بالاختبارات الجماعية المعتادة.

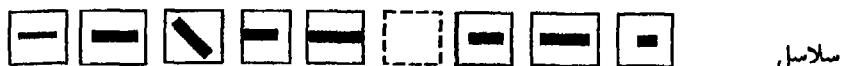
(٢) اختبار الذكاء غير المتحيز للثقافة

IPAT Culture Fair Intelligence Test : ١٩٥٧ عام

وهذا الاختبار شائع الاستخدام في الولايات المتحدة الأمريكية، وقد أعد تحت إشراف كاتل Cattell R. في معهد قياس الشخصية والقدرات. ويشتمل الاختبار الذي يتكون من صيغتين متكافئتين على ثلاثة مستويات لكل صيغة كالتالي: المستويات العمرية ٤ - ٨ أعوام، والراشدين من ضعاف العقول (٨ - ١٤) عاماً، والراشدين من متوسطي الذكاء، وتلاميذ المرحلة الثانوية والراشدين المتفوقين. ويشتمل كل اختبار على أربعة اختبارات فرعية تقيس قدرة الفرد على إدراك العلاقات بين الأشياء، مثل



السلسل ، والتصنيفات ، والمصفوفات . ويوضح شكل (٩ - ١٣) التالى مثلاً لبعض هذه المفردات :



سلسل



تصنيفات



مصفوفات

شكل (٩ - ١٣) يوضح عينة من مفردات اختبار الذكاء غير المتحيز ثقافياً

وبالإضافة إلى ذلك يوجد مقياس آخر ينقسم إلى أربعة اختبارات فرعية تقيس معلومات ثقافية وفهم لفظي . ويستغرق تطبيق الاختبار بعامة مدة زمنية تتراوح بين ٤٠ ، ٦٠ دقيقة . وقد وجد أن قيمة معامل ارتباط الاختبار بكل من مقياس ويكسنر ومقياس بيئية تتراوح بين ٥٦ ، ٨٥ ، ٠ . (Tannenbaum , 1968) .

وقد أمكن التتحقق من عدم تحيز الاختبار للثقافة بمقارنته معاييره التي تم إعدادها في كل من دول أوروبا الغربية والولايات المتحدة الأمريكية وأستراليا ، ووجد أن هناك تمازجاً فيما بينها .

(٣) اختبار الرسم : Goodenough - Harris Drawing Test (G - HDT)

يُعد من الاختبارات غير النظرية الجيدة ، ويمكن تطبيقه فردياً أو جماعياً ، ويطلب من الفرد رسم صورة رجل كامل ، وتُقدر الدرجة على أساس تفاصيل الجسم والملابس ، وتناسب الملامح والخصائص الأخرى ، ولا يهتم بالتوابع الجمالية في الصورة .

والاختبار غير موقوت ، ولكن الفرد يستغرق عادة ١٥ دقيقة تقريباً لرسم الصورة . وقد حددت خصائص صورة الرجل التي تميز بين مختلف الأعمار ، وتقدر لكل خاصية درجة واحدة ، فمثلاً إذا رسم الفرد رأس الرجل دون تحديد تفاصيل ملامحه تُقدر له درجة واحدة ، ومن ثم تُجمع الدرجات . وقد أعدت معايير الاختبار



للأطفال الذين تتراوح أعمارهم بين ٣ ، ١٥ عاماً متضمنة الدرجات المعيارية ، والرتب المئوية لكل من الذكور والإناث.

وعلى الرغم من أنه كان يفترض أن الاختبار غير متحيز للثقافة، إلا أنه قد تبين أن عملية رسم رجل تتأثر تأثيراً ملحوظاً بالثقافة، ومع هذا فإن الاختبار يتميز بسهولة التطبيق وسرعته ولا يتطلب إلا ورقة وقلم، ونحصل باستخدامه على تقدير تقريري سريع لذكاء الطفل. ويمكن الإفادة من هذا الاختبار بدمجه ضمن بطارية من اختبارات الذكاء لتأكيد المعلومات التي نحصل عليها نظراً لأن تطبيقه بمفرده ربما يؤدي إلى نتائج غير دقيقة.

وقد أُجري تعريب كثير من مقاييس الذكاء الجماعية والفردية وتقنيتها في البيئة المصرية وفي بيئات عربية أخرى، ويمكن لمستخدمي الاختبارات والباحثين الرجوع إلى أدلة الاختبارات المنصورة في الوطن العربي لاتقاء الاختبارات والمقاييس المناسبة للغرض المطلوب.

تحقيق عام على مقاييس الذكاء :

يتضح مما سبق أن اختبارات الذكاء يمكن أن يُختبر بها كل فرد على حدة، أو تُطبق على مجموعة من الأفراد في وقت واحد وذلك لقياس القدرة العقلية العامة أو الذكاء العام. ولعل التساؤل الذي يترتب على ذلك هو كيفية اختيار الاختبار المناسب من بين هذين التوقيعين الرئيسيين. والحقيقة أنه لا توجد إجابة محددة عن هذا التساؤل، وإنما يمكن مراعاة بعض الاعتبارات في هذا الشأن.

فلكل من نوعي الاختبارات استخداماته ومزاياه وعيوبه، فاختبارات الذكاء الفردية تناسب بدرجة أفضل الأطفال الصغار لكي يُختبر كل منهم على حدة، وذلك لأن كثيراً منهم لا يستطيع القراءة أو اتباع التعليمات التفصيلية، ومن السهل تشتيت انتباهم، مما يتطلب من الفاحص المتدرس توجيههم وحفظهم على الإجابة في إطار إجراءات مقتنة للحصول على قياسات دقيقة.

لذلك يُوصى بعدم استخدام اختبارات الذكاء الجماعية في قياس ذكاء الأطفال الذين تقل أعمارهم عن خمسة أعوام، أي أطفال ما قبل المدرسة نظراً ل حاجتهم إلى الإشراف الفردي. أما الأطفال الذين يبلغون من العمر خمسة أو ستة أعوام فيمكن تطبيق اختبارات الذكاء الجماعية عليهم لقياسهم قدرتهم العقلية العامة، ولكن في مجموعات صغيرة قدر الإمكان، أي لا تزيد عن ١٥ طفل مثلاً.

ولكن نظراً لأن قيم معامل استقرار درجات اختبارات الذكاء الجماعية بالنسبة لأطفال الصفوف الأولى من المرحلة الابتدائية عبر الزمن تكون منخفضة، فإن استخدام هذه الدرجات في اتخاذ قرارات تربوية مهمة تتعلق بهؤلاء الأطفال ربما لا يكون مناسباً.

لذلك إذا دعت الحاجة إلى قياس الذكاء العام لبعض الأطفال الصغار لأغراض معينة، فإنه يمكن استخدام اختبارات الذكاء الفردية، أو استخدام بعض الاختبارات الجماعية المناسبة لمستواهم النمائي والتي لا تتطلب إلا قدرًا ضئيلاً من القراءة أو الكتابة، وإنما يضع الطفل علامة على ورقة الأسئلة مباشرة حيث إن الاختبار في هذه الحالة يتضمن مواد غير لفظية.

لذلك يفضل استخدام اختبارات الذكاء الجماعية وبخاصة بطاريات الاختبارات متعددة العوامل Multi-Factor Batteries في قياس ذكاء المراهقين والراشدين العاديين حيث إن التنبؤ باستخدام هذه الاختبارات يكون جيداً بنفس القدر إذا استُخدمت الاختبارات الفردية. كما أن هؤلاء لا يتأثرون كثيراً بمواصفات الاختبارات الجماعية، ويعملون على توظيف أقصى إمكاناتهم في إجاباتهم.

لذلك تميز هذه الاختبارات الجماعية بأنها قليلة الكلفة، ولا يستغرق تطبيقها وقتاً طويلاً، ولكنها لا تناسب المواقف الكلينيكية التي تتطلب عمليات التشخيص، حيث ينبغي استخدام اختبارات الذكاء الفردية لهذا الغرض.

لذلك فإن الاختبارات الفردية تتطلب جهداً كبيراً في بنائها بواسطة اختصاصيين في القياس النفسي، بإعداد المواد المتعددة التي يشتمل عليها الاختبار، وإجراء عمليات تقديره وإعداد معايره، وتحديد وصياغة تعليماته، وطريقة تصحيحه، وتدريب الفاحصين يستغرق وقتاً طويلاً ونفقات مالية كبيرة.

وعموماً، فإن اختيار أيّ من نوعي الاختبارات يعتمد على الهدف الأساسي من استخدامهما، فإذا كان الهدف تقدير قدرة الأفراد الذين يكون الموقف الاختباري الجماعي مناسباً لهم من حيث التعليمات ومستوى الدافعية وما إلى ذلك فإنه يفضل استخدام اختبارات الذكاء الجماعية. أما المواقف التي تتطلب معلومات تفصيلية أكثر دقة، وكذلك ملاحظات الفاحص، فإنه يفضل اختيار أحد الاختبارات الفردية التي ناقشناها فيما سبق.

بعض القراءيات المتعلقة بقياس الذكاء:

(١) بناء اختبارات الذكاء:

عرضنا في الجزء السابق وجهات نظر متعددة تتعلق بمفهوم الذكاء ونماذجه ونظرياته وأدوات قياسه، ولعل من المناسب أن نوضح الأساس المنطقي الذي تستند إليه اختبارات الذكاء بأنواعها المختلفة التي أشرنا إليها.

فاختبارات الذكاء بعامة تُعد اختبارات مفنته ويُتبع في بنائها خطوات مماثلة لخطوات بناء الاختبارات التحصيلية المفنته أو الاختبارات التحصيلية مرجعية الجماعة أو المعيار التي ناقشناها بإسهاب في الفصل الثامن. إذ ينبغي أولاً تحديد النطاق السلوكي الذي يقيسه اختبار الذكاء ، وابتکار مفردات مناسبة لقياس مكونات هذا النطاق . وكتابة المفردات يُعد علماً وفنا له أصوله وقواعده، وينبغي كتابة عدد من المفردات أكبر مما هو مطلوب ، ويتختلف تفاصيل عملية بناء هذه الاختبارات باختلاف الجماعة المرجعية التي يطبق عليها الاختبار، أي ما إذا كانت هذه الجماعة من المراهقين والراشدين أم من الأطفال.

فبعد تطبيق الاختبار على العينة المستهدفة من المراهقين والراشدين يجري تحليل مفردات الاختبار كما سبق أن أوضحنا في الفصل السابع، حيث ينبغي تحديد النسبة المئوية من الأفراد الذين أجروا إجابة صحيحة عن كل مفردة.

ولا يجب أن تكون معظم المفردات غاية في السهولة أو الصعوبة، فإذا أجاب عدد محدود من الأفراد عن المفردات إجابة صحيحة أو إجابة خطأ، فإن الاختبار لن يُميز بين الأفراد تمييزاً كافياً. وإذا كان مستوى صعوبة كثير من المفردات متوسطاً، أي أن معامل الصعوبة 0.5 ، فإن الاختبار لن يُميز بين الحالات المتطرفة في توزيع الدرجات.

وكذلك ينبغي إيجاد معامل الارتباط بين درجات مفردات الاختبار والدرجات الكلية. فالمفردات التي أجبت عنها إجابة صحيحة يجب أن ترتبط ارتباطاً موجباً بالدرجات الكلية، كما يجب أن تكون الارتباطات بين المفردات موجبة أيضاً. أما إذا كانت العينة المستهدفة من الأطفال، فإنه إلى جانب تحليل المفردات السابق يكون بناء الاختبار أكثر صعوبة، وذلك لأن ذكاء الأطفال يتوقع أن يزداد بزيادة العمر، حيث يجب أن تزداد نسبة الأطفال الذين يجيبون عن المفردات إجابة صحيحة بزيادة العمر.

لذلك سوف تكون هناك كثير من المفردات إما سهلة أو صعبة للغاية بالنسبة لأطفال مستوى عمرى معين، ولكن لا يحيط الطفل فإنه عادة يُوقف تطبيق الاختبار إذا أخفق الطفل في إجابة عدد محدد من المفردات المتدرجة فى صعوبتها فى مرات متتالية. ويفترض عند تحديد هذه الخصائص السيكومترية للمفردات أن الذكاء عامل عام، غير أن هناك نماذج أخرى للذكاء كما رأينا فى مستهل هذا الفصل مثل نموذج ثيرستون أو جيلفورد يفترض عدداً من العوامل أو القدرات المستقلة. وفي هذه الحالة ينبغي مراعاة أن ترتبط المفردات المتعلقة بكل عامل أو قدرة فيما بينها ارتباطاً مرتفعاً نسبياً، وترتبط بغيرها من المفردات المتعلقة بالعوامل الأخرى ارتباطاً منخفضاً.

ومعظم اختبارات الذكاء تُحدد النسبة المئوية من المفردات التي يجب عنها الأفراد إجابة صحيحة، وهذا يُعد أحد محركات تقييم الذكاء. غير أن ثورنديك (Thorndike, 1968) قد أشار إلى أن قياس الذكاء لا يقتصر على قياس العلو أو الارتفاع Altitude. فإذا أجاب فرد عن عدد من المفردات المتكافئة في صعوبتها أكبر مما أجاب عنه فرد آخر في اختبار غير موقوت، فإن الفرد الأول يمكن اعتباره أكثر ذكاءً من الفرد الثاني، وعندئذ يقال أن لديه وسعة أو اتساع Breadth. وكذلك إذا استطاع فرد أن يجيب عن مفردات أكثر سهولة من فرد آخر في اختبار موقوت، فإن الفرد الأول يمكن اعتباره أكثر ذكاءً من الفرد الثاني نظراً لأنه أكثر سرعة في إجابته.

والحقيقة أن سرعة الإجابة تُعد من المتغيرات الشائكة، وذلك لأن الأفراد يختلفون في مقدار الزمن الذي يحتاجونه لحل مشكلات معينة، فالبعض ربما يترك المفردات الصعبة بالنسبة له دون إجابة بمجرد مواجهتها، وبعض آخر ربما يترك المشكلات التي كان من الممكن حلها بمزيد من المثابرة، ومع هذا، فإن السرعة كمؤشر للذكاء أصبحت تهتم بها الدراسات السيكولوجية في الآونة الأخيرة، فالأفراد الذين يمكنهم تجهيز المعلومات ومعالجتها بسرعة أكبر من غيرهم ربما يستطيعون بمرور الزمن تكون زخرفة من المعلومات الضرورية، وبالتالي يحصلون على درجات مرتفعة في اختبارات الذكاء.

ومن المشكلات المهمة الأخرى في عملية بناء الاختبارات هو أن تكون السمة المراد قياسها أحادية البُعد Unidimensional. فالبعد الأحادي يُمكّنا من تفسير معنى الدرجة الناتجة. فإذا كان الأداء في الاختبار يعتمد على بُعدين، فإنه من الممكن أن يحصل فردان على الدرجة نفسها بطريق مختلفة، إذ ربما يكون الفرد متميزاً في أحد الأبعاد وضعيفاً في البُعد الآخر، والعكس بالنسبة للفرد الثاني.

لذلك ينبغي التتحقق من فرضية أحادية البُعد باستخدام أساليب إحصائية مناسبة مثل التحليل العاملى الذى سبق أن أوضحناه، وذلك عند بناء اختبارات الذكاء أو غيرها من الاختبارات المقنتة.

(٢) بعض محددات الذكاء والعوامل المؤثرة فيه :

اتضح لنا مما سبق أن معظم اختبارات الذكاء روعى في بنائها أن يكون متوسط نسب الذكاء أو الدرجات المُحوَّلة ١٠٠ ، فنجد أن ثلثي أفراد الجماعة المرجعية تقريباً تحصل على درجات تتراوح بين ٨٥ ، ١١٥ (ذكاء متوسط أو قريب من المتوسط) ١١٪ تقريباً تكون درجاتهم أعلى من ١٢٠ (متفوقين أو موهوبين) ، ١٦٪ يحصلون على درجات أعلى من ١٤٠ (متفوقين جداً)، بينما يحصل ٨,١٪ تقريباً على درجات تنحصر بين ٣٠ ، ٧٩ (فئات الضعف العقلي) ، وذلك وفقاً للمتحنِي الاعتدالى الذى أوضحتناه في الفصلين الثالث والسادس.

غير أن توزيع الدرجات يختلف إلى حد ما من اختبار إلى آخر . ويعتمد هذا الاختلاف على ثبات درجات الاختبار، والسمات أو العوامل التي يقيسها، وطبيعة الجماعة المرجعية التي يُطبق عليها الاختبار وتُستمد منها معايره .

ويمكن أن يرجع تباين الأفراد والجماعات في الذكاء إلى مصادر متعددة، مثل العوامل الوراثية التي تبدو أهميتها في هذا المجال من البحوث المتعددة التي تناولت هذه العوامل بالدراسة وبخاصة البحوث التي قارنت بين ذكاء الآباء والأبناء، وبين التوائم المتطابقة والتوائم المتشابهة، وكذلك العوامل البيئية التي تلعب دوراً مهما في هذه الاختلافات في الذكاء وبخاصة في المراحل المبكرة من نمو الطفل .

تأثير العوامل الوراثية في الذكاء :

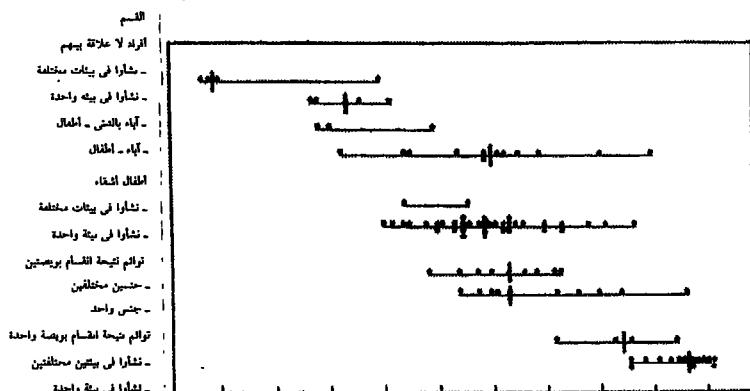
لقد بيَّن جنسن (Jensen, 1969) بجامعة كاليفورنيا في مقالته الشهيرة «إلى أي حد يمكننا رفع الذكاء والتحصيل الدراسي How Much Can We Boost IQ and Scholastic Achievement» ، أن السبب الرئيسي في عدم فاعلية البرامج التربوية التعويضية التي تقدم للأطفال الأقل حظاً في الولايات المتحدة الأمريكية أنها ركزت على الخبرات البيئية كمحدد للنمو العقلي .

وقدَّمَ بيانات تؤيد وجهة نظره مؤداها أن ٨٠٪ تقريباً من الفروق في نسب الذكاء تُعزى إلى الفروق الوراثية، في حين أن العوامل البيئية تسهم بنسبة ضئيلة في هذه الفروق، وذكر أن المستوى الأول للقدرة Level I Ability المتعلق بعمليات الترابط Associative يكون توزيعه متعدلاً بين مختلف الفئات الاجتماعية، بينما المستوى الثاني للقدرة Level II Ability المتعلق بالتفكير التصورى Conceptual يتوزع توزيعاً متبايناً بين الفئات الاجتماعية والاقتصادية المنخفضة والمتوسطة، حيث يتميز الأطفال من المستويات الاجتماعية، والاقتصادية المتوسطة عن الأطفال من المستويات المنخفضة في هذه القدرة الثانية .

أى أنه يرى أن الاختلاف بين هؤلاء الأطفال يرجع إلى اختلافهم في نمط قدرات التعلم، مما يتطلب تقديم أساليب تعليمية مختلفة لتنمية القدرات المتوفرة لدى كل من المجموعتين. وقد أكد أيزنك (Eysenck, 1971) في كتابه «الجدل حول نسبة الذكاء . Jensen ووجهة نظر جنسن The IQ Argument»

غير أن جنسن في كتابه «التحيز في القياس العقلى Bias in Mental Testing» الذي صدر عام ١٩٨٠ عاد وذكر أن أسباب هذه الفروق متعددة، وتحديد هذه الأسباب يعد من المهام المعقدة التي تتطلب بحوثاً متعددة التخصصات، وبخاصة في العلوم البيولوجية والسلوكية والقياس النفسي.

وفي دراسة أكثر اتساعاً قام إلينمير وكيملنجر (Elenmeyer - Kimling, 1963) بتلخيص نتائج ٥٢ دراسة أجريت في ثمانية دول، على مدى الصيف الأول من القرن الحالى حول مشكلة الوراثة في الذكاء الإنساني يوضحها شكل (٩ - ١٤) التالي:



شكل (٩ - ١٤) يوضح نتائج ٥٢ دراسة تتعلق بالوراثة في مجال الذكاء الإنساني

ويتبين من شكل (٩ - ١٤) أن هناك اتساقاً في النتائج على الرغم من اختلاف الدراسات في تكوين العينات من حيث المستوى الاجتماعي والاقتصادي والجنسية وأنواع الاختبارات المستخدمة. إذ يبدو أنه كلما اقتربت العلاقات الوراثية بين الأفراد كانوا أكثر تماثلاً في الذكاء.

ولعل قيم الارتباطات الموضحة تُعد الحدود الدنيا لهذه القيم، نظراً لأن الدراسات اعتمدت على بيانات مستمدة من الأطفال وليس من الراشدين، ولم تفصل بين الذكور والإإناث مما أدى إلى تقديرات لتأثير العوامل الوراثية أقل من حقيقتها،

فوسيط القيم المبنية في شكل (٩ - ١٤، ١) ويقرب كثيراً من قيمته المتوقعة. إذا اعتبرنا الوراثة هي المحدد الوحيد لدرجات اختبارات الذكاء . غير أن تأثير العوامل البيئية يتضمن من قيم الارتباط بالنسبة للأفراد الأقارب الذين نشأوا معًا التي تبدو أعلى من القيم بالنسبة للأفراد الأقارب الذين نشأوا في بيئتين مختلفتين على الرغم من ثبوت العوامل الوراثية .

كما يمكن ملاحظة الفروق بين الأفراد غير الأقارب الذين نشأوا معًا أو في بيئتين مختلفتين . وبذلك ربما يوضح شكل (٩ - ١٤) النزعة العامة لدور كل من الوراثة والبيئة كمحددات لدرجات اختبارات الذكاء .

وفي إطار الجدل حول رفع نسبة الذكاء ينبغي أن نميز بين ذكاء الفرد ، والدرجة التي يحصل عليها في اختبار للذكاء . ففي معظم الدراسات التي أجريت حول إمكانية رفع نسبة الذكاء اُتُّخذت الدرجة المرتفعة في الاختبار مؤشرًا على ارتفاع القدرة العقلية العامة ، على الرغم من أن هذا موضوع تساؤل .

فالتحاق الأطفال بدور الحضانة واكتسابهم بعض الخبرات يساعدهم في الإجابة عن مفردات اختبارات الذكاء بدرجة أفضل قليلاً ، ولكن إذا استُخدِمت مفردات مختلفة ربما لا يستطيعون الإجابة عن هذه المفردات مما قد يدل على عدم وجود تأثير متسبق لهذه الخبرات في القدرة العقلية العامة .

بعض علماء النفس عرَّف الذكاء بأنه ما تقيسه اختبارات الذكاء . والارتفاع الطارئ في نسبة الذكاء يمكن تفسيره إلى حد ما في ضوء هذا التعريف . فالطفل الذي يتحقق بدور الحضانة يكتسب اتجاهات ومهارات ربما تساعد في أن يصبح أكثر تمرساً في التعامل مع الاختبار ، وذلك لزيادة ثقته بالكبار المختلفين عن والديه ، مما يجعله أكثر تقبلاً للفاحص عندما يكرر تطبيق الاختبار عليه من أجل المتابعة .

كما أن التفات الطفل المستمر إلى المواد التي يشتمل عليها الاختبار ، مثل المكعبات والصور واللوحات ربما يزيد من ألفته بها ، مما يمكّنه من التعامل معها بدرجة أفضل .

وهذا العاملان ربما يُسْهِمان في رفع نسبة ذكائه عدة نقاط . ولكن هل يعني ذلك أن الطفل أصبح أكثر ذكاء؟ فلقد ذكرنا فيما سبق أن الذكاء لا يقاس قياساً مباشرأ . وقياسه بطريقة غير مباشرة يعتمد على تقييم استجابات الطفل لسلسلة من المواقف المقتننة ، أي مفردات الاختبار . والهدف من اختبار الذكاء هو تقدير درجة تدل على قدرة الطفل على التعامل مع هذه المواقف بحيث يمكنه توظيف هذه القدرة في التعامل مع جميع المواقف المماثلة .

ولكى يتحقق هذا الهدف يجب أن يشتمل الاختبار على عينة غير متحizable وكافية من المفردات. فإذا تدرب الطفل على كيفية إجابة مثل هذه المفردات، هل أداؤه عندئذ يتعلق بعينة غير متحizable من المفردات، أم أن هذه العينة تكون مصطنعة وتدل على أن الطفل ربما يؤدى أداء جيداً فقط إذا واجه مواقف مصطنعة مماثلة.

فمن المعلوم أن معظم الأفراد يُوظفون قدرًا ضئيلاً من قدراتهم العقلية. لذلك ربما يكون من المفيد تقديم خبرات إثرائية للطفل في حل المشكلات والفحص والبحث وما إلى ذلك. وعندئذ يمكن الإفاده من اقتراح جنسن Jensen بأن الذكاء تحدده العوامل الوراثية، وذلك بالعمل على توسيع الخبرات الإثرائية وجعلها ملائمة لنمط تعلم الطفل لكي يوظف إمكاناته العقلية الموروثة إلى أقصى حد ممكن.

تأثير العوامل البيئية في الذكاء :

إن السؤال حول أثر البيئة الاجتماعية في الإسراع بالنمو العقلى للأطفال كان ولا يزال مثار جدل وحوار وأبحاث ودراسات مستمرة. فقد تبين من كثير من البحوث أن هناك علاقة وثيقة بين ذكاء الأطفال والمستوى الاجتماعي والاقتصادي لأبائهم (McNemar , 1942 ; Harrel, 1945; Harrell, 1945 ; Eckland, 1965; Tyler, 1965 ; Ginitis 1973)

ولكن يجب عدم المبالغة في تأكيد هذه العلاقة، فالأطفال الذين تقع نسبة ذكائهم في المدى الممتد من التأخر العقلى إلى العبرية إنما هم لأباء من كل نمط مهنى. كما تبين من نتائج بعض الدراسات أن الأسر الذكية بحكم الوراثة (أى بيولوجياً) قد تستغل ذكائها فى توفير بيئات أكثر إثراء، وبذلك تعمل على رفع مستوى ذكاء أفرادها.

فمثلاً تبيّن من كثير من الدراسات أن الأطفال في المؤسسات الخاصة بضعف العقول تميل نسبة ذكائهم إلى الانخفاض بزيادة أعمارهم، أما إذا وضعوا في ملاجئ للأيتام لكي يتيسر لهم في هذه الملاجئ الاستثارة العقلية على الرغم من عدم تهيئة مناخ يسوده الأمان والثقة لا يُحدث أى انخفاض في نسبة ذكائهم، ولكن إذا انتقلوا إلى بيوت تبني طيبة ترتفع أحياناً نسبة ذكائهم، أو على الأقل يتوقف الانخفاض فيها.

ويرى البعض أنه يمكن رفع نسبة الذكاء عندما يُحول الأطفال المتخلعون عقلياً من الفصول العادية إلى الفصول الخاصة، إلا أن نتائج البحوث التي أجريت في هذا

المجال لم تؤكّد هذه الزيادة، في حين أن البرنامج التعليمي المعد إعداداً جيداً ربما يُسّرّ تحصيل هؤلاء الأطفال ويزيد من توافقهم الشخصي والاجتماعي. وعموماً لا تُوجَد إيجابية العوامل البيئية عن مدى تأثير العوامل البيئية في الإسراع في نمو الذكاء.

غير أن تحليل هذه العوامل المؤثرة في النمو العقلي يبيّن أن الذكاء هو نتاج العوامل الوراثية والبيئية، إذ بيّنت كثيرة من الدراسات أنه لا يمكن تنمية الذكاء أبعد من حدود الإمكانيات الفطرية البيولوجية. ومن بين هذه دراسة تيرمان Terman ما بين عامي ١٩٤٩ - ١٩٥٩ المتعلقة ببحث الانتقال الوراثي للذكاء، فقد وجد أن الأطفال يميلون إلى أن يشبهوا آباءهم في الذكاء. ولكته فسر هذه النتيجة بأن الآباء من ذوى الذكاء المرتفع ربما يوفرون بيئات أكثر إشارة لأبنائهم مما يؤدى إلى تفوقهم، أي أن الإسراع في النمو العقلي يكون نتيجة لتفاعل العوامل الوراثية والبيئية. وبالطبع من الصعب تحديد القدر الذي يسهم به كل من هذه العوامل، في النمو العقلي.

(٣) ثبوت نسبة الذكاء :

والحقيقة أنه لا توجد إجابة قاطعة لهذا السؤال، نظرًا لصعوبة تحديد نسبة ذكاء الفرد الواحد بشكل دقيق خلال مراحل حياته، ومع ذلك يمكن القول بأن نسبة ذكاء الفرد (متوسط الذكاء) تظل ثابتة نسبيًا في مختلف مراحل نموه بوجه عام مع تغير ربما پیتر او سون في مدى درجات.

ففي دراسة طولية أجراها هونزيك Honzik ، وماركفارلين Markfarlane ، والآن Allen (١٩٤٩) اعتبروا بتصنيعها لبحث استقرار نسب الذكاء ، حيث قاموا بتتبع ٢٥٢ طفلًا من الميلاد إلى الرشد. وقد اختبر الأطفال باختبار ستانفورد - بينيه في مختلف مراحلهم العمرية ، وقد تبين أنه كلما تقدم الطفل في العمر استقرت نسبة ذكائه. فمثلاً وُجد أن قيمة معامل الارتباط بين نسب الذكاء في المدى العمري ٤ - ٦ أعوام ، ٦٢ ، ٠، بينما قيمته في المدى العمري ٨ - ١٤ عاماً ، ٨٥ ، ٠ . أما درجات الاختبار في مرحلة ما قبل المدرسة كانت قليلة الفائدة في التنبؤ بالدرجات في المراحل التالية ، بل كانت عديمة الفائدة في ، مدة زمنية فاصلة قدرها عام واحد.

وفي خلال سن الدراسة في المدى العمري ٦ - ١٨ عاماً تذبذبت نسبة ذكاء حوالي ٦٠٪ من العينة ١٥ نقطة أو أكثر، ومن بين كل ثلاثة أطفال تذبذبت نسبة ذكاء

أحدهم ٢٠ نقطة أو أكثر، ومن بين كل عشرة أطفال تذبذبت نسبة ذكاء أحدهم ٣٠ نقطة أو أكثر، ١٥٪ من أفراد العينة تذبذبت نسبة ذكائهم أقل من ١٠ نقاط.

وعموماً، فإن تذبذب نسبة الذكاء من تطبيق إلى آخر للاختبار تقل كلما تقدم الطفل في العمر، وهذه النتيجة لها أهمية تربوية حيث تبين أن درجات اختبارات الذكاء العام في الصفوف المبكرة من المحتمل ألا تناظر الدرجات في الصفوف التالية، مما يتطلب تكرار تطبيق هذه الاختبارات في الصفوف المبكرة.

لذلك لا ينبغي استخدام نسبة الذكاء في وصف الطفل أو تصنيفه في نعط معين، لما لذلك من آثار نفسية وتعلمية واجتماعية على مستقبل حياته. فنسبة الذكاء تمثل أداء الطفل في وقت معين، وربما تبدو إمكانات الطفل أعلى أو أقل في الأعوام التالية.

غير أن تغير نسبة الذكاء يبدو أنه يرتبط بمستوى الأسرة من حيث تعليم الوالدين، والحالة الاجتماعية والاقتصادية. فالطفل الذي نسبة ذكاؤه مرتفعة ولكن مستوى تعليم والديه منخفضاً وفي مهن لا تتطلب مهارة، ربما تنخفض نسبة ذكائه كلما تقدم في العمر إلى أن يصل إلى مرحلة الرشد. وعلى العكس فإن الطفل الذي نسبة ذكاؤه منخفضة ولكن ثقافة والديه ومستوى تعليمهما متقدمة، فإنه ربما تزداد نسبة ذكائه في مراحل نموه التالية (Duncan, Featherman and Duncan, 1972).

غير أنه ينبغي أيضاً مراعاة تأثير العوامل المتعلقة بالشخصية والدافعية في الذكاء، وبخاصة أنه ربما يختلف هذا التأثير باختلاف نوع الطفل.

ولا يزال هناك أيضاً جدل وعدم اتفاق حول استقرار ذكاء الراشدين. فقد اختر Wechsler (1955) عينة مستعرضة من الراشدين ووجد أن الجوانب اللغوية للذكاء تزيد حتى سن ٣٠ عاماً ثم تبدأ في الانخفاض التدريجي، غير أن أسلوب المعاينات الذي استخدمه يشوه بعض القصور نظراً لاختلاف مستوى تعليم المجموعات العمرية المختلفة. فمستوى تعليم الأفراد الأصغر سنًا كان أعلى من مستوى تعليم الأفراد الأكبر سنًا، وربما أدى ذلك إلى هذه الفروق.

كما قام بيلي (Bayley, 1955) بدراسة طويلة استنتج منها أن نمو الذكاء يستمر حتى يبلغ الفرد ٥٠ عاماً. غير أن الاستنتاج الذي يمكن الاستناد إليه هو أن الوظائف العقلية العامة لا تنخفض انخفاضاً آلياً بتقدم العمر. فكما ذكرنا يمكن أن تعمل العوامل البيئية للراشدين أيضاً على رفع أو خفض أدائهم العقلى، ولكن إذا أخذنا بعين الاعتبار

الظروف الصحية لراشد عمره ٥٠ عاماً ، فإن قدرته العقلية ربما تكون مساوية لقدرته عندما كان عمره ٢٥ عاماً (Jarvik et al., 1973).

ويبقى الآن السؤال التالي :

ما السن التي يتوقف فيها النمو العقلى ؟ أى ما السن التي يكتمل فيها النضج العقلى للفرد ؟

فلقد حاول كثير من علماء النفس دراسة هذه المشكلة بالأساليب الإمبريقية، حيث انتقووا اختبارات ذكاء معينة، وقاموا بتطبيقها على أفراد من أعمار مختلفة، وقارنوا بين درجات الاختبارات لكل مجموعة متعددة في العمر الزمني والمجموعات الأخرى، ووجدوا استمراً في منحني النمو حتى أعمار معينة. والاتجاه العام في الدراسات المختلفة يذهب إلى أنه قلما يحدث تغير في النمو العقلى بعد سن الثامنة عشرة.

ولكن تظل الإجابة عن السؤال محاطة باعتبارات كثيرة في البحث عن أثر العوامل الاجتماعية والاقتصادية للعينة التي تطبق عليها اختبارات الذكاء، ومدى تمثيل العينة لجميع أفراد المجتمع، ومدى تنوع هذه الاختبارات بحيث تُظهر الاختلافات في القدرات أو الاستعدادات الخاصة. ولعل التمييز الذي اقترحه كاتل Cattell بين الذكاء المائع Fluid Intelligence ، والذكاء المتبلور Crystalized Intelligence ، والذي أشرنا إليه في مستهل هذا الفصل يلقي ضوءاً على هذه المشكلة.

خلاصة

من هذا يتضح أن مجال الذكاء الإنساني يُعد من المجالات الشائكة التي نالت قدراً كبيراً من اهتمام علماء النفس وخبراء القياس بل والعلماء من فروع العلم المختلفة، ولا يزال الحوار والجدل قائماً حول تعريف الذكاء، وأساليب ومنهجيات قياسه، وكيفية الإفادة من نتائج مقاييس الذكاء، والتغلب على المشكلات الناجمة عن سوء استخدام هذه النتائج في المجالات التربوية والاجتماعية.

لذلك اتجهت بحوث الذكاء في الآونة الأخيرة وجهة جديدة، حيث اعتبرت أن تقييم الذكاء بالأساليب التي تناولتها في هذا الفصل يُعد تقسيماً إستاتيكياً؛ وذلك لأن هذا التقييم لا يهتم كثيراً بامكانية إفادة الفرد من التعليم وبخاصة بالنسبة لحالات الضعف العقلي والأفراد الهاشمين، كما أن هذا التقييم يستند إلى مفهوماً ضيقاً للذكاء. ومن هنا حاولت بعض الدراسات الإفادة من التطورات التي حدثت في علم النفس المعرفي في تطوير أساليب تحليل العمليات المعرفية وتحديد مكوناتها وقياسها، كما حاولت تطوير أساليب جديدة لقياس إمكانات التعلم Learning Potential لدى الأفراد بحيث يكون التقييم ديناميكياً وليس إستاتيكياً كما هو الحال في القياس التقليدي للذكاء الإنساني.

وقد أوضح هيد Haywood، وبراؤن Brown، وينجنفلد Wingenfeld (١٩٩٥) أن التقييم الديناميكي للذكاء يؤكد تقييم العمليات وليس النواتج ، ومحاولات تحديد العقبات التي تحول دون التعلم الفاعل ، وكذلك محاولة تحديد الشروط التي تسهم في الأداء الجيد، والتمييز بين الأداء والإمكانات الفردية، كما أن هذا التقييم يؤكد تكامل العملية الاختبارية مع عملية التعليم.

لذلك ينبغي مراعاة أن الذكاء يُعد مفهوماً أكثر اتساعاً مما تشير إليه نسبة الذكاء، فهو يتضمن القدرة العامة على التكيف في مواقف الحياة الاجتماعية الفعلية، ويتميز الأفراد بالتنوع وتعدد القدرات، مما يتطلب تعدد جوانب القياس.

ولقد تناولنا فيما سبق بعض النماذج والنظريات التي اعتبرت الذكاء قدرات عقلية متعددة ، غير أن هذه القدرات ترتبط ارتباطاً وثيقاً بالدافعية والاتجاهات والأبعاد الأخرى ذات العلاقة.

لذلك لا يجب الارتكان إلى درجة كلية واحدة للذكاء في اتخاذ قرارات مهمة تتعلق بالفرد، حيث إن هذه الدرجة ينبغي النظر إليها على أنها مدى من الدرجات قد يكون متسعًا أو ضيقًا وفقاً لثبات درجات الاختبار.

وتعبر هذه الدرجة عن ذكاء الفرد في وقت تطبيق الاختبار، ولكنها ربما تتغير بمرور الزمن، وهذا التغير يمكن التحقق منه بتطبيق اختبارات أخرى، أو إعادة تطبيق الاختبار بفواصل زمنية معينة. وعند تفسير نسبة الذكاء ينبغي ربطها بخلفية الفرد من حيث الفرص المتاحة له للتعلم، ومهاراته اللغوية، وحفز الأسرة له على التعلم. كما أنه نظرًا لعدم توافر أدلة كافية حول التأثير النسبي لكل من الوراثة والبيئة في الذكاء حيث إن كلاهما يؤثر بلا شك بدرجة معينة، فإنه ينبغي العمل على إثراء البيئة التعليمية للأطفال والراهقين في المنزل والمدرسة ومواقف العمل، وتصميم أساليب تعليمية وتدريلية تتمكن الأفراد منها اختلاف مستويات قدراتهم وتبين سمات شخصياتهم من الإسهام الفاعل في إثراء حياتهم وتنمية مجتمعهم.

وهذا يتطلب التحول من القياس مرجعى الجماعة أو المعيار الذي يهتم بمقارنته أداء الفرد بأقرانه، إلى القياس مرجعى المحك الذي يُفيد في تشخيص مهارات واحتياجات الأفراد التربوية والمهنية والاجتماعية بهدف تحسين وتطوير هذه المهارات لديهم، وقد سبق أن ناقشنا هذا النوع من القياس بإسهاب في الفصل الثامن.

فالأفراد لا يختلفون فقط فيما بينهم في القدرة العقلية العامة، وإنما أيضًا في استعداداتهم الخاصة المتعلقة بمجالات تعليمية أو مهنية معينة، مما يدعو إلى الاهتمام بقياس وتقييم هذه الاستعدادات الخاصة، والعمل على الإفادة منها وتنميتها عن طريق تزويد الأفراد بالخبرات الثرية التي تُسهم في إبرازها بالقدر الذي تسمح به إمكانات كل منهم، وهو ما سوف نتناوله بالتفصيل في الفصل التالي.

الفصل العاشر

قياس الاستعدادات الخاصة

* مقدمة

* قياس الاستعدادات الخاصة من منظور تاريخى

* مفهوم الاستعداد وعلاقته بمفهوم القدرة

* تأثير العوامل الوراثية والبيئية في الاستعدادات الخاصة

* استخدامات اختبارات الاستعدادات الخاصة

* أنواع اختبارات الاستعدادات الخاصة

* بعض الافتراضات الأساسية في قياس الاستعدادات الخاصة

* أساليب بناء اختبارات الاستعدادات الخاصة

* إجراءات تحليل العمل

* اختبارات الاستعدادات الخاصة

* بطاريات الاستعدادات المتعددة

* اختبارات استعدادات تتعلق ب المجالات نوعية محددة

* الاستعداد الميكانيكي

* الاستعدادات النفسية الحركية

* الاستعداد الكتابي

* الاستعداد الموسيقي

* الاستعدادات الفنية

* الابتكارية

* الابتكارية ومفهوم التفكير التباعدي

* قياس قدرات التفكير الابتكاري

* اختبارات التفكير الابتكاري

* مشكلات قياس التفكير الابتكاري

* خلاصة

مقدمة :

عرضنا في الفصل السابق المفاهيم والنمذج النظرية المتعلقة بالذكاء الإنساني، ومنهجيات وأساليب قياسه، وبعض محدداته الوراثية والبيئية وما يرتبط بهما من مؤشرات متعددة، وقد أوضحنا أن الذكاء هو القدرة العقلية العامة حيث اعتبرها بعض العلماء قدرة عامة واحدة تنطوي تحتها كافة الأنشطة العقلية، في حين اعتبرها آخرون مجموعة من المكونات العاملية أو القدرات المتعددة التي يحدد تفاعلها مستوى القدرة العقلية العامة لدى الفرد، أي أن هذه القدرات اعتبرت بمثابة «عوامل الذكاء».

غير أن هناك تجمعات معينة لقدرات خاصة تبدو أهميتها في أداء مهام محددة ويمكن الإفادة منها في التنبؤ بالسلوك. وعادة يطلق على هذه التجمعات من القدرات الخاصة «الاستعدادات Aptitudes».

فالاستعداد يشير إلى استطاعة الفرد اكتساب سلوك معين أو مهارات معينة إذا أتيحت له ظروف ملائمة أو تدرب مناسب. والذكاء العام أو القدرة العامة وما تنطوي عليه من قدرات متعددة ليست كافية لتعرف إمكانات الفرد وطاقته العقلية. فاختبارات الذكاء العام التي قدمتها في الفصل السابق تبيّن فاعليتها وجدواها بدرجة أكبر في التنبؤ بالنجاح الأكاديمي بعامة، ولكن اختبارات الاستعدادات الخاصة Special Aptitudes تستخدم في التنبؤ بالنجاح في مجال دراسي أو تدريسي أو مهني أكثر تحديداً، مثل النجاح في دراسة اللغات الأجنبية أو الرياضيات العالية، أو التمكن من أداء أعمال ومهارات كتابية أو ميكانيكية أو فنية أو غير ذلك.

وقد سبق أن ميزنا في الفصل الثامن بين اختبارات الاستعدادات الخاصة واختبارات التحصيل الدراسي، واختبارات الذكاء العام، وأوضحنا أنها تقع على متصل واحد، ويتبين موقعها عليه وفقاً لنوعية وشمول الخبرات التي تتطلبها في إطار مفهوم القدرات النامية أو المستمرة.

فلكي يحدد الفرد إمكاناته وقابلياته بصورة أكثر شمولاً ينبغي أن يتعرف على استعداداته الخاصة إلى جانب قدرته العقلية العامة، مما يتطلب توافر اختبارات مناسبة تقيس هذه الاستعدادات، إذ ربما تكون القدرة العقلية العامة لفردين مختلفين متقاربة، ومع هذا فإن كليهما يختلفان اختلافاً ملحوظاً في نوعية استعداداتهما الخاصة المتعلقة ب المجالات دراسية متنوعة أو أعمال ومهام وظيفية ومهنية معينة.

لذلك سوف نتناول في هذا الفصل مفهوم الاستعداد والعوامل المؤثرة فيه، وأنواع اختبارات الاستعدادات الخاصة واستخداماتها وأساليب بنائها، وأمثلة متعددة لبطاريات الاستعدادات، واختبارات تتعلق ب مجالات نوعية محددة مثل الاستعداد الميكانيكي، والاستعدادات النفسية الحركية، والاستعداد الكتابي، والاستعدادات الفنية، ونوضح بمزيد من التفصيل مفهوم الابتكارية واختبارات التفكير الابتكاري، وما يتعلق بقياس الابتكارية من مشكلات.

قياس الاستعدادات الخارجية من منظور تاريخي :

إن تطور قياس الاستعدادات الخاصة من منظور تاريخي يرتبط ارتباطاً وثيقاً بتطور حركة القياس العقلى التي أوضحتها في الفصل السابق.

فقد ذكرنا أنه منذ قيام الحرب العالمية الأولى حتى عام ١٩٣٠ شهد مجال القياس العقلى نمواً متسارعاً حيث تزايد عدد الاختبارات الجماعية المقنتة، واتسع نطاق استخداماتها، وصدر العديد من الكتب والمقالات والبحوث المتعلقة بموضوعات القياس. غير أنه بدأت تبرر تساؤلات متعددة سواء من المستغلين ببناء الاختبارات أو من مستخدميها تتعلق بمحدودية هذه الاختبارات، وبخاصة اختبارات الذكاء وأوجه قصورها في قياس قدرات وإمكانات الأفراد.

وقد أدى ذلك إلى توسيع مجالات الاختبارات العقلية، وإجراء العديد من الدراسات والبحوث حول طبيعة القدرات العقلية والإفادة من نتائجها في تطوير هذه الاختبارات وغيرها من أدوات التقويم. ولعل قيام الحرب العالمية الثانية قد ساعد في تشجيع بناء اختبارات تستخدم في أغراض التوجيه التربوي والمهنى، وبخاصة اختبارات الاستعدادات.

كما ساعد في ذلك البرامج الحكومية في الولايات المتحدة الأمريكية المتعلقة بال المجالات المهنية المختلفة مثل الزراعة والتجارة والصناعة والتشييد، وكذلك التوسيع في برامج التعليم الثانوى وتنويعه مما تطلب تطوير برامج منظمة وفاعلة للإرشاد التربوى لطلاب هذا التعليم. وكذلك تطوير أساليب دقيقة لتصنيف وتسكين ملايين الأفراد المجندين.

وقد أسهم ذلك كله في دفع حركة بناء أنواع متعددة من اختبارات الاستعدادات للوفاء بهذه الاحتياجات المُلحّة، مثل اختبارات الاستعدادات الميكانيكية والكتابية والموسيقية والفنية.

وقد أدى زيادة التوجه نحو استخدام أساليب التحليل العاملی إلى تطوير طرق بناء اختبارات الاستعدادات الخاصة، حيث ساعدت هذه الأساليب في التوصل إلى عدد من العوامل أو القدرات النوعية المستقلة نسبياً حيث أمكن قياسها ببطاريات استعدادات متعددة منذ عام ١٩٤٥، وبخاصة في المجالات العسكرية لاستخدامها في انتقاء قادري الطائرات والمعدات العسكرية والفنين، وأُعدت بعد ذلك بطاريات أخرى للأغراض التربوية والمهنية والإرشاد وانتقاء الأفراد وتصنيفهم.

واستمر استخدام أساليب التحليل العاملی في دراسة التكوين العاملی للختبارات المتوفرة وتطوير اختبارات جديدة استُخدمت في البرامج الاختبارية واسعة النطاق التي انتشرت في الولايات المتحدة الأمريكية لأغراض متعددة، وكذلك اختبارات الاستعداد الدراسي، وختبارات القبول بالجامعات، وختبارات انتقاء الطلاب للالتحاق بالكلية الفنية والعملية والمهنية، وكذلك اختبارات انتقاء الطلاب للمنح الدراسية.

ومما ساعد بدرجة كبيرة في تطوير واستخدام كثير من هذه الاختبارات وغيرها تأسيس مركز الخدمات الاختبارية بولاية نيوجرسى الأمريكية Educational Testing Services (ETS) عام ١٩٤٧ ، وكذلك البرنامج الاختباري للكليات الأمريكية American College Testing Program الكبير الذي حدث في نظريات القياس ونماذجه منذ السبعينيات إلى توجه جديد في بناء أنواع متعددة من اختبارات الاستعدادات وإجراء عمليات تكافؤ درجاتها استناداً إلى النظرية المعاصرة في القياس التي أصبحت تعرف بنظرية الاستجابة للمفرددة الاختبارية Item Response Theory (IRT) كما ذكرنا في موضع آخر.

وهناك دراسات جادة يقوم بها علماء النفس بالقوات المسلحة الأمريكية في وقتنا الحاضر لبحث إمكانية استخدام التصنيف المتتابع للقدرات الخاصة بالاستعنة بالحاسوب Computer Sequential Ability Testing في مختلف أغراض الانتقاء والتدريب في ضوء أنسن ومبادئ هذه النظرية، مما يساعد في تشخيص وتوجيه هذه القدرات الخاصة أثناء عمليات التدريب.

ويعتمد هذا التصنيف على برنامج اختباري يُدار بواسطة الحاسوب تُحدَّد فيه مجموعة متسلسلة من المفردات الاختبارية لكل فرد، حيث يُختبر بمفردة تلو الأخرى، ويُتَّخذ قرار تصنيف مبدئي عقب إجابته عن كل مفردة. فإذا أُجري التصنيف في حدود قدر معين من الخطأ المحدد مسبقاً لا يستمر الفرد في الاختبار، أما إذا زاد مقدار خطأ التصنيف بما هو محدد يستمر الفرد في إجاباته عن المفردات التالية إلى أن يتم التصنيف في ضوء المحك المحدد للخطأ، أو إلى أن يُختبر الفرد بجميع المفردات المتوفرة.

ولعل هذا يمثل إجراءات تطبيق اختبار ستانفورد - بينيه ، فيما عدا أن إجراءات التطبيق والتصنيف تدار آلياً بواسطة الحاسوب ، ويتم تقدير استعداد الفرد باستخدام نظرية الاستجابة للمفردة الاختبارية المشار إليها .

كما أدى ذلك إلى التوجه نحو بناء نظرية علمية للقدرات المعرفية تحدّد وتصنّف جميع القدرات المعرفية الأساسية والفرعية بطريقة أكثر اتساقاً، وتقدم أساليب دقيقة لقياس هذه القدرات، وتوضح كيفية إسهامها في الأنماط المختلفة للسلوك المعرفي والتعلم والأداء .

مفهوم الاستعداد وعلاقته بمفهوم القدرة :

لعل من المناسب في مستهل هذا الفصل أن نلقى مزيداً من الضوء على مفهوم الاستعداد Aptitude وعلاقته بمفهوم القدرة Ability لكي يتسع استخدام كل من المفهومين استخداماً مناسباً.

فقد اتضح لنا من الفصل التاسع أن الذكاء أو القدرة العقلية العامة تنطوي على تجمُّع معين من العوامل أو المكونات . غير أن معظم مقاييس الذكاء تركز على نسبة الذكاء بعامة (IQ) وليس على تقييم الجوانب النسبية للقوة والضعف في هذه المكونات كل على حدة . ولعل هذا هو ما يميز بين نسبة الذكاء وتقييم الاستعدادات الخاصة .

وسوف يتضح عند مناقشتنا لاختبارات الاستعدادات الخاصة أنها تميّز بعمق محتواها وارتباطه ارتباطاً وثيقاً بالاستعداد المراد قياسه ، وهذا بالطبع لا يتحقق في اختبارات الذكاء . كما أن هناك استعدادات خاصة ، مثل الاستعدادات الفنية ، والاستعدادات الكتابية ، والمهارات الحركية أو العضلية ، وغيرها يندر أن تتناولها معظم أساليب تقييم القدرة العقلية العامة .

والاستعدادات الخاصة تتضمن تجمعاً من القدرات الخاصة التي تعد ضرورية لأداء مهمة معينة أو تسلسل من المهام ، وتجمّع هذه القدرات بالنسبة للإاستعدادات يماثل تجمُّع العوامل أو المكونات بالنسبة للذكاء العام . فالاستعداد يفسّر بواسطة ما تعلم الفرد أداءه من خلال توسط قدرته الخاصة .

ويمكن توضيح ذلك بالقدرة على التمييز بين طبقات الصوت Pitches ، فبعض الأفراد لديهم قدرة مرتفعة على التمييز بدقة بين طبقات الصوت المختلفة ، ويمكن قياس مدى دقة واتساق هذا التمييز ، وهنا ييدو دور التعلم في تمكين الفرد من تسمية أو تصنيف طبقة معينة من طبقات الصوت الذي يسمعه نظراً لأن هذه الإمكانيّة ليست موروثة .

وإذا بحثنا عن استعداد يتضمن هذه القدرة الخاصة المتعلقة بالتمييز بين طبقات الصوت نجد أنه الاستعداد الموسيقي. غير أن هذا الاستعداد يتضمن تجمعاً من القدرات الخاصة المتعلقة به، مثل إدراك أنماط الإيقاع، وذاكرة النغم وغيرها إلى جانب القدرة الخاصة السابقة. كما أن هناك مستوى آخر للاستعداد يسمى «الاستعداد من الدرجة الثانية JobA Second Order Aptitude» وهو استعداد متعلق بالعمل ptitude. وهذا الاستعداد أكثر شمولاً من تجمع هذه القدرات الخاصة التي تمثل استعداداً متفرداً ومتجانساً نسبياً مثل الاستعداد الموسيقي. فالاستعداد المتعلق بالعمل يمكن أن يشتمل على مجموعة متكاملة من الاستعدادات المترابطة التي تسهم في الأداء الموسيقي.

فإذا تحدثنا عن قدرة الفرد ليكون موسيقاراً، فإننا ننظر إلى جميع المتطلبات أو المهام التي ربما يقوم بها الفرد أو يعرفها أو يفضلها وما إلى ذلك، ونضئها معًا لكي تكون الاستعداد المتعلق بالعمل. فمستوى الاستعداد الموسيقي المرتفع لدى الفرد يعد مظهراً واحداً من مظاهر الاستعداد الأوسع المتعلق بالعمل كموسيقار. وتحديد المكونات المتعددة التي يشتمل عليها هذا الاستعداد يتطلب إجراء تحليل للعمل Job Analysis إلى جانب ملاحظة الأداء الفعلى في العمل.

من هذا يتضح أن الاستعداد الخاص لا يُعد عاماً تكاملياً أو وحيداً، وإنما يعد تنظيمياً معقداً لمكونات متعددة. فالفرد ليس لديه قدرة كامنة واحدة ، وإنما لديه قدرات كامنة متعددة، ويطلب أداء أي نشاط أو عمل معين كثيراً من هذه القدرات الكامنة.

لذلك فإن بناء اختبار الاستعداد يعتمد على تحليل دقيق لمكونات الأداء المراد قياسه أو التنبؤ به. وبعض القدرات الكامنة يمكن أن تتدخل فيما بينها أو تكون مترابطة، وبعض المظاهر التي تحدد القدرة الكامنة لفرد للنجاح في مهنة معينة ربما تحدد أيضاً إمكانية نجاحه في مهنة أخرى (Wesman , 1973 ,).

تأثير العوامل الوراثية والبيئية في الاستعدادات الخاصة :

لا يستطيع أحد أن ينكر أن العوامل الوراثية تلعب دوراً مهما في التعريف بحدود إمكانات الفرد. فقد اتضح لنا في الفصل السابق أنه في بداية القرن العشرين كان هناك تأييد من جانب كثير من علماء النفس لفكرة أن الذكاء وغيره من الاستعدادات موروث، وأن العوامل البيئية تلعب دوراً محدوداً في نموها. غير أن تطور نظريات التكوين العقلى والبحوث التي أجريت بعد ذلك جعلت العلماء يولون العوامل البيئية اهتماماً أكبر في تنمية وإثراء الاستعدادات. وقد أكدت الدراسات المعاصرة تفاعل كل من العوامل الوراثية والعوامل البيئية في الإسهام في نمو الاستعدادات.

لذلك فإن اختبارات الاستعدادات الخاصة تقيس الأداء بجانبيه المتفاعلين: الموروث والمكتسب. وبهذا تكون اختبارات الذكاء أو القدرة العقلية العامة مرتبطة باختبارات الاستعدادات أكثر من ارتباطها باختبارات التحصيل، إذ يمكن اعتبار اختبارات الذكاء نوعاً خاصاً من اختبارات الاستعدادات وتقيس القدرة العقلية العامة أو الطاقة العقلية. وعلى الرغم من أن هناك عناصر مشتركة بينهما، بل وتسمى اختبارات الذكاء عادة اختبارات الاستعداد الأكاديمي Scholastic Aptitude وبخاصة إذا استُخدمت هذه الاختبارات في التنبؤ بالتحصيل الأكاديمي، إلا أن لكل منها استخداماته الخاصة.

استخدامات اختبارات الاستعدادات الخاصة:

إن كثيراً من الاختبارات السيكلولوجية التي تستخدم في أغراض التنبؤ لا تفترض حصول الفرد على تدريب مناسب، كما أن بعضها لا يقيس الجوانب المعرفية. أما اختبار الاستعداد فينبع أن يقيس مهارة معرفية، ويستخدم أساساً في التنبؤ أو الانتقاء أو التسكين.

فاختبارات الذكاء العام ربما تكون مناسبة لتقدير القدرة العقلية العامة للطالب التي تُمكّنه من النجاح في كثير من البرامج الأكademie، أما اختبارات الاستعدادات الخاصة فإنها تستخدم في التنبؤ بالنجاح في تخصص أكاديمي أو مهني معين، مثل الهندسة أو الفنون أو الأداء العملية المتعددة أو المحاماة.

ويُعد انتقاء الأفراد Selection من أهم استخدامات اختبارات الاستعدادات. ويعتمد ذلك على تحديد درجة قطع Cutoff Score في الاختبار أو درجات قطع متعددة Multi Cutoff Scores إذا استخدمنا بطارية من الاختبارات. وينبغي أن ترتبط درجات الاختبار أو الاختبارات بمدى أداء مستقبلٍ متعلق بالاستعداد الذي يقيسه كل منها ارتباطاً مرتفعاً، وأن ترتبط فيما بينها ارتباطاً منخفضاً. وتتأثر درجات القطع بالسياسات التربوية أو المهنية، والاعتبارات الأخلاقية والاجتماعية. فتحديد درجة قطع مرتفعة للقبول بالكليات مثلاً سوف يقلل من الإنفاق الحكومي على التعليم العالي، ولكن هذا التوفير ينعكس أثراً سلباً على بعض فئات المجتمع، وبخاصة المستويات الاجتماعية والاقتصادية المنخفضة.

وقد سبق أن ناقشنا في الفصل الخامس الأخطاء الناجمة عن قرارات الانتقاء، وكيفية استخدام تحليل الانحدار في التنبؤ بدرجات المحك باستخدام الاختبارات التنبؤية، أي اختبارات الاستعدادات. وأوضحتنا في سياق عرضنا للصدق المرتبط بمحك وبخاصة الصدق التنبؤي بعض المفاهيم، مثل نسبة النجاح ونسبة الانتقاء وجدال التوقع، وكيفية توظيفها في أغراض الانتقاء.

وتُستخدم اختبارات الاستعدادات الأكاديمية Scholastic Aptitudes Tests بكثرة في أغراض الإرشاد والتوجيه التربوي. فالمرشد التربوي في المدرسة يهتم بمعرفة ما إذا كان التلميذ لديه صعوبات في القراءة أو الرياضيات أو اللغات الأجنبية، وهذا التشخيص يساعد المعلم في اقتراح خطة العلاج.

فمثلاً في المرحلة الابتدائية يمكن استخدام اختبارات التأهب Readiness Tests للتعرف على الأطفال الذين لديهم صعوبات في تعلم مواد دراسية معينة، وبخاصة القراءة والحساب. وفي المرحلة الثانوية يمكن استخدام اختبارات الاستعداد لتعلم اللغات الأجنبية في التعرف على الطالب الذين لديهم المهارات الضرورية لتعلم هذه اللغات، وتحليل أخطاء الإجابات في هذه الاختبارات يمكن أن يساعد في تحديد أفضل الأساليب العلاجية.

كما أنه يمكن باستخدام هذه الاختبارات التعرف على الطلاب الذين يتميزون باستعدادات مرتفعة في المجالات العلمية والفنية والموسيقية مثلاً ، بحيث يمكن تصميم برامج إثرائية لهؤلاء الطلاب في هذه المجالات.

أما اختبارات الاستعدادات المهنية Vocational Aptitudes Tests فتستخدم في التنبؤ بالنجاح في مهن أو أعمال معينة أو في برنامج تدريب على هذه الأعمال دون أن يتركز الاهتمام على الذكاء أو الاستعدادات الأكاديمية. وباستخدام هذه الاختبارات يمكن للمرشد أو المسئول عن التسكين أن يقدم معلومات للأفراد عن المهن أو برامج التدريب المناسبة لهم، أو أن يتخد قرارات تسريحهم في هذه المهن أو البرامج.

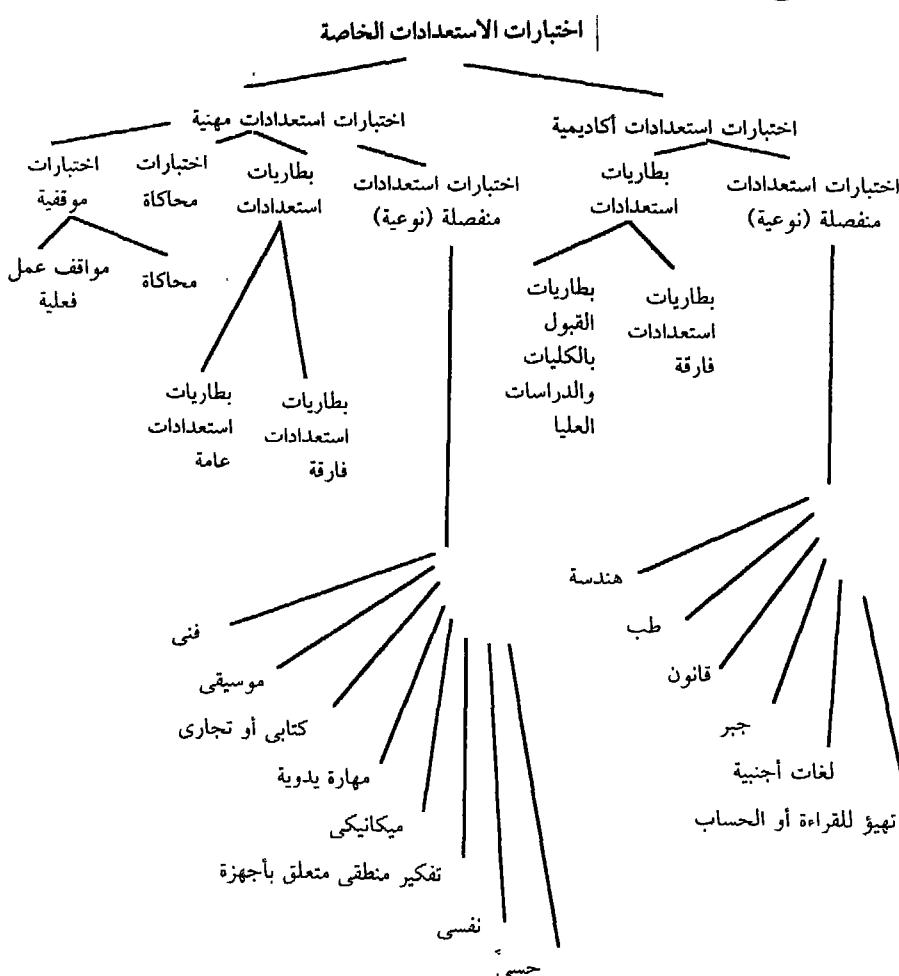
وكذلك تفيد هذه الاختبارات في إرشاد الطلاب، وبخاصة في مرحلة التعليم الثانوي عن نوع البرامج التعليمية المناسبة لاستعداداتهم والتي يمكنهم الالتحاق بها بعد انتهاءهم من هذه المرحلة، مثل البرامج الأكاديمية، والبرامج التطبيقية التجارية أو الصناعية أو الزراعية وغير ذلك، وتوجيههم التوجيه المهني المناسب.

ومن هذا يتضح أن بعض اختبارات الاستعدادات الخاصة يستخدم كأدوات وصفية، وبعضها يستخدم كأدوات تنبؤية. فاستخدام هذه الاختبارات في الإرشاد يهدف لوصف القدرات الخاصة للفرد أو الأفراد، وتعريف كل منهم بجوانب قوته وضعفه لكي يستخدم في ضوئها قرارات تتعلق بمستقبله الدراسي أو المهني. أما استخدام اختبارات الاستعدادات الخاصة في التنبؤ فالهدف منه تقديم معلومات تفيد المسؤولين عن انتقاء الأفراد سواء للالتحاق بالكليات أو بالمهن والأعمال المختلفة في اتخاذ قرارات تتعلق بهؤلاء الأفراد، مثل اختبارات القبول بالكليات أو المعاهد المهنية والفنية، واختبارات انتقاء المتقدمين للوظائف الشاغرة في مجالات العمل المختلفة.

أنواع اختبارات الاستعدادات الخاصة :

توجد أنواع متعددة من اختبارات الاستعدادات الخاصة في المجالات الأكاديمية والمهنية بعضها يقيس قدرة خاصة واحدة، وبعضها يقيس قدرات خاصة متعددة تشكل في مجملها استعداداً خاصاً أو أكثر، وتُعد بمثابة بطاقة من الاختبارات. وعلى الرغم من أن كثيراً من هذه الاختبارات تتطلب إجابتها ورقة وقلم، إلا أن بعضها منها يقيس أداء عملياً أو حركياً معيناً، وبعضاً آخر منها يتضمن عينات من المهام المتعلقة بعمل معين.

ويوضح الشكل التخطيطي (١٠ - ١) التالي تصنيفًا مقترحاً لهذه الأنواع :



شكل (١٠ - ١) يوضح تصنيفًا مقترحاً لأنواع اختبارات الاستعدادات الخاصة
(من تصميم المؤلف)

ويتضح من شكل (١٠ - ١) أنه يمكن تصنيف اختبارات الاستعدادات الخاصة في أربعة أنواع رئيسة نوجزها فيما يلى :

اختبارات استعدادات نوعية :

ويقىس كل منها قدرة خاصة أو استعداداً خاصاً واحداً في المجالات المهنية، مثل الاستعداد الحسى، أو النفسي الحركى، أو الميكانيكى، أو الفنى، أو مثل التهئي للقراءة، أو استعداد اللغات الأجنبية، أو الهندسة، أو القانون، وغير ذلك من المجالات الأكاديمية. ويمكن أن تكون هذه الاختبارات مستقلة بذاتها أو ضمن بطارية من اختبارات الاستعدادات. وتفيد هذه الاختبارات في أغراض الانتقاء والتسلكين، ولكنها محدودة الفائدة في أغراض الإرشاد. ويمكن أن تكون هذه الاختبارات من النوع الذي يتطلب ورقة وقلم، أو يتطلب أداء عملياً باستخدام أجهزة معينة.

لذلك فإن هذه الاختبارات ربما تطبق فردياً أو على جماعات محدودة العدد مما يجعل عملية التطبيق تستغرق كثيراً من الوقت.

بطاريات اختبارات الاستعدادات :

وتشتمل هذه البطاريات على عدد من الاختبارات المنفصلة التي تقيس مجموعة من القدرات الخاصة التي ينطوى عليها استعداد معين أو أكثر. ويمكن الحصول على درجة لكل من هذه القدرات أو المكونات، أو يمكن ضم جميع الدرجات للحصول على درجة كلية واحدة.

كما يمكن رسم صفحة نفسية (بروفيل) Profile للفرد باستخدام الدرجات التي حصل عليها في كل من الاختبارات التي تشتمل عليها البطارية. ولكن ينبغي مراعاة أن تكون معايير جميع الاختبارات مستمدة من جماعة مرجعية واحدة، وبذلك تُعطى قدرأً كبيراً من المعلومات في أقل زمن ممكن. وبطاريات اختبارات الاستعدادات يمكن أن تكون فارقة Differential، أو عامة General، وتستخدم البطاريات الفارقة في مرحلة التعليم الثانوى لكي يتعرف الطالب على جوانب القوة والضعف في استعداداته، ويترشد بذلك في تحضير مستقبله التعليمي والمهنى. غير أنه في مرحلة التعليم الابتدائى يمكن استخدام اختبارات الذكاء العام، حيث إن الاستعدادات الخاصة تفتح وتنمّى كلما ازداد نضج الطفل وخبرته.

وإذا استُخدمت بطارية اختبارات الاستعدادات الخاصة للتتبؤ بالكافيات المتعلقة بعمل معين يمكن مقارنة الصفحة النفسية للفرد بالصفحة البيانية العامة المتعلقة بمتطلبات العمل لمعرفة مدى التقارب بين الصفتين. ومن أمثلة هذه البطاريات فى

المجالات الأكاديمية والمهنية بطارية الاستعدادات الفارقة Differential Aptitude Tests (DAT) وبطاريات الاستعدادات الحسية الحركية Sensory - Motor Batteries وفي مجالات انتقاء الطلاب للقبول بالكليات أو الدراسات العليا أو التخصصات المهنية نجد أن هناك عدداً من البطاريات تمثل في اختبار الاستعداد الأكاديمي Scholastic Aptitude Test (SAT)، واختبار سجل الدراسات العليا Graduate Record Examination (GRE) وغيرها.

أما البطاريات العامة فتشتمل على اختبارات يمكن استخدامها في مدى واسع من الاستعدادات والمهن، ويمكن اختيار واحد أو أكثر من اختبارات البطارية لاستخدامه في أغراض مناسبة. كما أن هناك اتجاهًا نحو تضمين هذه البطاريات اختبارات لفظية، وتوسيع محتواها بحيث يمكن أن تقيس القدرة العقلية العامة وبعض الاستعدادات الخاصة، وهذا يفيد في زيادة فاعلية التطبيق والحصول على معايير قابلة للمقارنة في جميع اختبارات البطارية.

وهذه البطاريات العامة المتعددة الأغراض أكثر استخداماً في أغراض الإرشاد التربوي والمهني، ومن أمثلتها بطارية اختبارات الاستعدادات العامة General Aptitude Employee Aptitude Tests Battery (GATB) والبطارية المسحية للاستعدادات Flanagan Aptitude Survey (EAS) وبطارية فلانagan لتصنيف الاستعدادات Classification Tests (FACT).

وتميز بطاريات الاستعدادات الخاصة بأنه يمكن تطبيقها على مجموعة كبيرة من الأفراد في آن واحد. غير أن هذه البطاريات تقتصر عادة على الورقة والقلم حيث لا تستخدم أجهزة تتطلب أداء عملياً معيناً.

اختبارات محاكاة:

وتشتمل على الأنشطة الضرورية المتعلقة بعمل معين إما بتضمين الاختبار مواقف مصغرة لأنماط العمل الفعلى، أو بمحاكاة العمل دون تقديم مواقف فعلية.

وتستند هذه الاختبارات إلى فكرة أن العمل ينبغي أن يؤدي كوحدة متكاملة ترتبط مكوناتها فيما بينها ، وهذا الارتباط بين المكونات لا يقل أهمية عن المكونات ذاتها. لذلك فإن مثل هذه الاختبارات تكون أكثر صدقًا في التبؤ بالنجاح في العمل. ومن ميزات اختبارات المحاكاة Simulation Tests أنها لا تتطلب تحديد القدرات التي ينطوي عليها العمل، وإنما تعتمد على محاكاة العمل ولو بصورة جزئية، وبخاصة في المجالات الصناعية. ومن أمثلة ذلك اختبار محاكاة الاستعداد المتعلق بتشغيل المخارط

Miniature Test For Engine Lathe Aptitude اختبار محاكاة مكبس تثقيب المعادن Miniature Punch Press Test. ويطلب بناء مثل هذه الاختبارات فكراً مُبدعاً وقدرة عالية على التصور.

اختبارات موقفية:

تستخدم هذه الاختبارات في تقييم أداء الفرد المتعلق بعمل معين في ظروف اختبارية مقتنة Situational Assessment. وهنا يطلب من الفرد أداء جميع العمليات المتعلقة بعمل معين أو بعضها في ظروف مماثلة تماماً لظروف العمل الفعلية التي سيواجهها الفرد عندما يتحقق بالعمل.

فمثلاً في مجال إصلاح الأجهزة الكهربائية يمكن أن يقدم للفرد المُختبر جهاز به عطل معين ويطلب منه كشف العطل وإصلاحه. وفي الأعمال الإدارية أو الكتابية، أو الإرشادية يقدم للفرد المُختبر مواد متعددة تُزوده بخلفية من المعلومات، ثم يعطي أوراق ومعلومات تماثل ما يتعامل معه الإداري أو المرشد التربوي أو النفسي أثناء أداء كل منهم لمهامه الفعلية، وتسمى هذه المعلومات «معلومات في السلة In - Basket».

و يتم تقدير درجات هذه الاختبارات بفحص الاستجابات لتحديد أنماط اتخاذ القرارات والأحكام المهنية فيها. والاختبارات الموقفية Situational Tests لا تقتصر على تقييم الأفراد فقط، وإنما تفيد في أغراض التدريب، فتقدير الدرجات يعتمد أيضاً على معدل ومقدار التحسن في أداء الفرد المُختبر إذا أعطى قدرًا من التدرب.

لذلك يمكن أن يُعاد تطبيق هذه الاختبارات بعد التحاق الفرد بالعمل لإعطائه فرصة التحسن، وملاحظة مدى تكامل سلوكه الشخصي والاجتماعي، وكفايته المستقبلية في تعلم وأداء المهام التي يقيسها الاختبار. وتُستخدم اختبارات المحاكاة و الاختبارات موافق العمل Work-Sample Tests في انتقاء العاملين بالجيش والمصانع وتدريبهم، ولكنها ربما تكون قليلة الفائدة في المدرسة أو في أغراض الإرشاد التربوي.

بعض المؤشرات الأساسية في قياس الاستعدادات الخاصة:

(١) سبق أن أوضحنا أن الاستعدادات Aptitudes يمكن أن تضم مجموعة من القدرات الخاصة المترابطة التي يُعبر عنها باستجابات الفرد لمواصفات سلوكية معينة.

فمثلاً إذا قلنا أن لدى فرد استعداداً موسيقياً، فإن هذا يعني أنه يمتلك مجموعة مترابطة من المهارات والقدرات والخصائص التي تمكنه من تعلم

الموسيقى. أي أن الاستعداد يُعد تكويناً فرضياً يُشخص المهارات والقدرات والخصائص المتعلقة به والتي يمكن أن يمتلكها الفرد. لذلك فإن الاستعداد لا يقاس بطريقة مباشرة، وإنما يُستدل عليه من مكوناته السلوكية التي أشرنا إليها، والتي يهدف اختبار الاستعداد لقياسها.

(٢) يُعد قياس الاستعدادات الخاصة نوعاً من القياس مرجعى الجماعة أو المعيار Norm - Referenced ، فدرجة الفرد في اختبار الاستعداد تقارن بمعيار جماعة مرجعية يتمسّى إليها الفرد. فإذا قلنا أن الاستعداد الكتابي لفرد ما مرتفع، فإن هذا يعني أنه يمكنه تعلم المهارات الكتابية المتعلقة بالملفات والتقارير والمراسلات والطباعة والتدقّيق والاحتزال أسرع من أقرانه وبجهد أقل. أما إذا كان الاستعداد الكتابي لديه منخفضاً، فإن هذا لا يعني أنه لا يستطيع تعلم هذه المهارات، وإنما يحتاج لجهد ووقت أكبر مما يحتاجه معظم أقرانه.

(٣) يفترض في قياس السمات الإنسانية أنها مستقرة نسبياً ، ونظراً لأن الاستعدادات تفيد في التنبؤ المستقبلي بأداء الأفراد، فإنها ينبغي أيضاً أن تتسم بالاستقرار النسبي عبر الزمن. ولكن هذا لا يعني أن الاستعدادات تُعد سمات موروثة لا تتغير أو تتحسن طوال حياة الفرد، فالاستعدادات لها أساس وراثي بلا شك ، ولكنها تعكس أيضاً خبرات الفرد وتعلمه، أي أنها نتاج تفاعل العوامل الوراثية والعوامل البيئية كما ذكرنا في مستهل هذا الفصل. والاستقرار النسبي للدرجات اختبارات الاستعدادات يعكس اتساق نمط حياة معظم الأفراد.

(٤) إن الاستعداد المرتفع لدى فرد في مجال معين لا يعني أنه سوف يُنمّى قدرته في هذا المجال دون تدريب مناسب. أي أن التنبؤ استناداً إلى أحد اختبارات الاستعدادات يعني أن تجمع من المهارات والقدرات الخاصة التي تمثل هذا الاستعداد مع تدريب مناسب سوف يؤدى إلى سلوك معين.

ونظراً لأن التدريب الذي يُعد مناسباً لفرد ما قد لا يناسب فرداً آخرًا، فإن أساليب التدريب ينبغي أن تباين تباين الاستعدادات. لذلك فإن كرونياك Cronbach وسنو Snow (١٩٧٧) أكدوا أهمية التفاعل بين الاستعدادات وأساليب التدريب أي المعالجات (ATI) - Treatment Interaction (ATI) لكن يكون أسلوب التدريب فاعلاً مما يزيد من دافعية الفرد ويؤدي إلى تحسّن أدائه.

ولعل هذا ما دعا بعض علماء النفس والقياس النفسي إلى إعادة النظر في مفهوم الاستعداد Aptitude الذي أوضحته فيما سبق، حيث يرى بلوك Block ، ويسلوم Bloom ، وكارول Carroll وغيرهم أن أسلوب التدريب الجيد والمناسب لاستعدادات الفرد واهتماماته ويسعى دافعيته للتعلم يؤدى إلى تقليل الزمن الذي يحتاجه الفرد في إتقان المهارات المرجوة. أما إذا كان أسلوب التدريب غير مناسب للفرد، وبخاصة إذا كان استعداداته منخفضاً فسوف يحتاج إلى زمن أطول لتعلم المهارات المعينة.

لذلك رأى هؤلاء العلماء أن الاستعداد يمكن تعريفه من حيث الزمن الذي يحتاجه الفرد لتحقيق هدف معين أو التمكن من مهارة. فإذا تبيّنت المعالجة التعليمية أو التدريّبية بتباين استعدادات الأفراد بحيث تناسب المعالجات استعدادات الأفراد، وأتيح لكل منهم الوقت الذي يحتاجه للتعلم، فإن معظمهم يستطيع إتقان المهارة أو المطلب المرجو تحقيقه.

أي أن الاستعداد يتمثل في الزمن الذي يحتاجه الفرد للتعلم ولا يتمثل في مستوى صعوبة أو تعقد المادة المتعلمة. ولعل ذلك يرتبط بالقياس مرجعى المحك بدرجة أفضل من ارتباطه بالقياس مرجعى الجماعة أو المعيار. فمحك الأداء Performance Criterion هو الإطار المرجعي المناسب لقياس الاستعدادات، حيث إنه أكثر ملاءمة من معايير الجماعة Group Norms .

أساليب بناء اختبارات الاستعدادات الخاصة :

تختلف أساليب بناء اختبارات الاستعدادات الخاصة باختلاف نوع الاختبار وتكوينه والغرض الذي يستخدم من أجله، وسوف نلقى الضوء فيما يلى على بعض هذه الأساليب:

أساليب تستند إلى التحليل النظري لمكونات الاستعداد :

فعند بناء مقياس للاستعداد الموسيقى أو الفني مثلاً، يمكن تحديد الخصائص أو القدرات الأساسية التي تسهم في الأداء الفاعل في هذه المجالات. فالقدرات الخاصة المتعلقة بالاستعداد الموسيقي تتعلق بتميز النغم، والإيقاع، وشدة الصوت وغير ذلك من المكونات التي يمكن أن يقتربها المتمرسون في المجالات الموسيقية والتي تتأثر بنوعية التدريب.

أساليب اEmpirique :

فمعظم اختبارات الاستعدادات الخاصة تهدف للتنبؤ بمحك أداء مستقبلٍ، لذلك يُراعى في بناء هذه الاختبارات أن ترتبط درجاتها بمحك مناسب لكي يكون الصدق

التبؤى Predictive Validity مرتفعاً، حيث إن هذا يعد من الشروط الأساسية في هذا النوع من الاختبارات. فإذا اشتمل الاختبار على عدة اختبارات فرعية، فإنه ينبغي أن يكون الارتباط بين درجات هذه الاختبارات منخفضاً أو صفرياً، بينما يكون ارتباطها بدرجات المحك مرتفعاً، وبذلك يتم انتقاء المفردات على أساس إمبريقي. ويمكن الرجوع إلى الفصل الخامس لمراجعة تفاصيل ذلك.

أساليب قستندة إلى التحليل العاملى :

تعتمد كثير من بطاريات الاستعدادات الخاصة في بنائها وتكوينها على نتائج التحليل العاملى Factor Analysis، حيث يكون الهدف هو بناء بطاريات يشتمل كل منها على عدد محدود من الاختبارات الفرعية التي يمكن باستخدامها التنبؤ بدرجة أفضل بالنجاح في تخصصات دراسية معينة أو مهن مختلفة.

ونظراً لأن طبيعة المهن ومكوناتها ومواصفاتها دائمة التغير، فإن هذا يتطلب المراجعة المستمرة للإختبارات التي تشتمل عليها البطاريات استناداً إلى نتائج التطبيق الميداني المستمر لهذه البطاريات وما يُستمد من بيانات متعددة.

أساليب قستندة إلى نتائج تحليل العمل :

تصمم عادة اختبارات الاستعدادات الخاصة لقياس مهارات محددة و المعارف تتعلق بمجال مهنى معين. فاختبارات الاستعداد الميكانيكي أو الكتابي أو الفنى مثلاً تقيس المكونات السلوكية التي تتضمنها هذه الاستعدادات.

لذلك فإن بناء هذه الاختبارات يتطلب تحديداً دقيقاً لما تشتمل عليه هذه المكونات من مهارات و معارف استناداً إلى تحليل العمل Job Analysis. والهدف من هذا التحليل تحديد الأداء الفاعل في مهنة أو مجال معين. ونظراً لأن كثيراً من المؤسسات الصناعية والتقنية وغيرها من المؤسسات المهنية تود اتخاذ قرارات انتقاء الأفراد الفنيين العاملين بها على أساس علمية، فإنها تقوم بناء اختبارات استعدادات مناسبة للأعمال أو المهن المتعلقة بها، وبخاصة إذا لم تتوافر لديها اختبارات جاهزة.

ففي هذه الحالة ينبغي أن يستند بناء اختبارات الاستعدادات على نتائج تحليل العمل المتعلق بهذه المهن، مثل أعمال التجارة ، والكهرباء ، والخراءة ، وإصلاح الآلات ، وترويج السلع ، وما إلى ذلك.

إجراءات تحليل العمل :

يستند تحليل العمل Job Analysis على إجراءات منهجية منظمة نوضحها فيما يلى :

(١) تعريف المقصود بالعمل :

فالعمل Job هو تجمع متجلانس نسبياً من المهام الوظيفية التي يقوم بها الفرد لتحقيق غرض أساسي في مؤسسة معينة. والمتجلانس النسبي لهذه المهام قد يعني تماثلها في النوع أو المحتوى أو المتطلبات المشتركة، أو قد يعني أن الفرد الذي يستطيع أداء إحدى هذه المهام ربما يستطيع أداء مهمة أخرى. فتعريف المقصود بالعمل في ضوء المهام Work Tasks المتعلقة به يساعد في تركيز الاهتمام على السلوك الفعلى المرتبط بالعمل ومتطلباته، وهذا هو الهدف الأساسي من تحليل العمل. غير أنه ينبغي مراعاة احتمال تغير متطلبات العمل ومواصفاته بتغير مهارات الأفراد، أو طبيعة العمل والمشرفين عليه، أو بيئة العمل الفيزيائية والتفسية. أى أن ديناميكيات العمل يجب أن تؤخذ بعين الاعتبار عند تحديد المقصود بالعمل.

(٢) تحديد طبيعة العمل ومواصفاته :

إن تحليل العمل يصف الجوانب المهمة لعمل معين التي تميزه عن غيره من الأعمال. لذلك فإن تحديد طبيعة العمل ومواصفاته يزودنا بوصف ما يفعله الفرد في العمل لكي يتسمى بناء أدوات انتقاء الأفراد وتصنيفهم في ضوئه.

ويجب أن يتضمن هذا الوصف العمليات التي يؤديها الفرد، والأجهزة المستخدمة، وظروف العمل ومخاطرها، والتدريب الذي يتطلبها، وعلاقة العمل بغيره من الأعمال. وهذا الوصف أساسى حيث تُحدد في ضوئه مواصفات العمل Job Specifications ، أى قائمة الكفایات التي يستند إليها بناء أدوات القياس؛ لذلك يجب أن تكون هذه القائمة مترتبة بالعمل المحدد ومتطلباته التي تميزه عن غيره من الأعمال. كما يجب أن تُحدد القائمة جوانب الأداء التي تميز تميّزاً دقيقاً بين الأداء الجيد والأداء الضعيف.

فالأفراد من مستويات كفاية متباعدة ربما يختلفون في طريقة أدائهم معظم جوانب العمل، لذلك فإن هناك جوانب معينة من العمل هي التي توضح الفروق بين المتميزين والضعفاء، ويطلق على هذه الجوانب التمييزية «الأنشطة أو المتطلبات الحرجة Critical Incidents». وهذه الجوانب لا تفيد فقط في بناء اختبارات الانتقاء، وإنما أيضاً في تصميم برامج التدريب أو تعديل البرامج القائمة، وتطوير إجراءات العمل، وتقليل مخاطرها، وتطوير تصميم الآلات والأجهزة في المؤسسات الصناعية.

(٣) أساليب جمع المعلومات المتعلقة بالعمل :

إن تحديد المواقف التفصيلية للمهام في الخطوة السابقة يتطلب جمع معلومات تتعلق بالعمل من مصادر متعددة بأساليب منظمة لكي يتسع الإفادة منها. إذ يمكن الاطلاع على المطبوعات ذات العلاقة، مثل أدلة التشغيل والتدريب، والكتيبات الخاصة بأعمال مماثلة، وسجلات الأداء، وبخاصة تلك التي تتضمن مواصفات الأخطاء الشائعة في الأداء، وصعوبات التعلم، وأسباب إخفاق بعض الأفراد في العمل، كما يمكن فحص سجلات شكاوى العملاء، وبخاصة في الأعمال التجارية عدم تأثير الأداء بوجود القائم باللإلاحة. وللتغلب على ذلك يمكن استخدام أسلوب الملاحظة بالمشاركة Participant Observation، حيث يقوم الفرد الملاحظ بالمشاركة في العمل والوظائف كما لو كان واحداً من العاملين، ويسجل ملاحظاته عنهم أثناء مشاركته معهم.

كذلك تُعد المقابلة الشخصية مصدرًا مهمًا من مصادر جمع المعلومات المتعلقة بالأعمال المختلفة. ويمكن أن يُجرى المقابلة المشرفون على العمل، وأخصائيو التدريب والعاملين القدامى من ذوي الخبرة والأداء المتميز. غير أن التقارير التي يقدمها القائمون بالم مقابلة قد تنسى أحياناً بالذاتية أو تتضمن تعديلات أو مصطلحات غير متفق عليها، لذلك ينبغي أن تركز التقارير على الحقائق الموضوعية وما يؤديه الفرد العامل بالفعل.

ويمكن جعل هذه التقارير أكثر موضوعية باستخدام مذكرات قيد الأعمال أو الأنشطة Activity Logs بدلاً من الاعتماد على ذاكرة القائم بالم مقابلة. وكذلك يمكن استخدام «أسلوب الأنشطة الحرجة» (CIT) الذي Critical Incidents Technique اقتربه فلانagan (1954) والذي أشرنا إليه في الخطوة السابقة. ويتم إجراء هذا الأسلوب في مدة زمنية محددة، على أن يتضمن التقرير الأنشطة أو السلوك كما حدث، مع ضمان سرية البيانات التي يتضمنها التقرير.

ومن الأساليب الأخرى المهمة التي يمكن عن طريقها الحصول على معلومات منظمة وبيانات كمية تتعلق بتحليل العمل الاستبيانات أو القوائم Inventories التي يتم إعدادها مسبقاً، حيث تدون فيها المهام الأساسية المتعلقة بالعمل والتي أمكن الحصول عليها باستخدام بعض الأساليب السابقة، وكذلك فقرات تتعلق بظروف العمل وغير ذلك من الخصائص.

وتوجد أنواع متعددة من هذه الاستبيانات بعضها يتضمن ميزان تقدير رقمي متدرج Rating Scale يتطلب من الفرد وضع علامة على رقم التقدير الذي ينطبق على

ما يؤديه من عمل، أو على الزمن الذي يستغرقه في مهمة معينة بالنسبة للمهام الأخرى المتعلقة بعمل معين، وبعضاها الآخر يتضمن ميزاتٍ بيانيًا خطياً يُحدّد عليه الفرد نوعية الأداء الذي يتراوح بين أداء متميز وأداء ضعيف، ويضع علامة على الخط تناظر المستوى الذي يراه مناسباً له.

وهذه النوعان من الاستبيانات يتعلّقان بمهام العمل Job - Oriented ، لذلك فإن استخدامهما يقتصر على مجال مهني معين صمّمت من أجله . غير أن هناك أنواعاً أخرى من الاستبيانات تتعلّق بالفرد العامل نفسه، وخصائصه النفسية التي تسهم في رفع مستوى الأداء في العمل وتتناول سلوك الفرد في العمل الذي يتميز بدرجة أكبر من العمومية، ومثال ذلك استبيان تحليل العمل Position Analysis Questionnaire الذي أعده ماكورميك McCormic ، وجينرت Jeanneret ، وميكم Mecham . ويشتمل هذا الاستبيان على ١٩٤ عنصراً متعلّقاً بالفرد العامل وأدائه يمكن تصنيفها في عدة أقسام مثل: مصادر حصول العامل على معلوماته المتعلقة بالعمل ، والعمليات العقلية التي يوظفها في عمله كالاستدلال واتخاذ القرار والتخطيط ، والأنشطة البدنية ، والمهارات اليدوية والأدوات والأجهزة التي يستخدمها ، وعلاقة العامل بزملاهه ورؤسائه في العمل ، وبيئة العمل ، وغير ذلك من الخصائص .

أي أن هذه الاستبيانات تركز على مدخلات المعلومات لدى الفرد العامل ، وتجهيزه ومعالجته للمعلومات ، والفعل أو الاستجابة الناتجة . وتتطلّب من الفرد إما أن يضع علامة على العنصر الذي ينطبق عليه ، أو يحدد على ميزان التقدير الرقم الذي يدل على مدى الأهمية ، أو الزمن أو الصعوبة فيما يتعلّق بكل عنصر في هذه الأقسام .

وقد استُخدم هذا الاستبيان في تحليل ٥٣٦ نوعاً مختلفاً من الأعمال ، وأجرى تحليل عاملٍ للدرجات الناتجة ، حيث أمكن استخلاص خمسة عوامل تنطوي عليها استجابات الأفراد العاملين في الأنواع المختلفة من الأعمال تتلخص في : القدرة على اتخاذ القرارات والتواصل ، والمسؤولية الاجتماعية ، والقدرة على أداء أنشطة تتطلّب مهارات ، والنشاط الحركي البدني في بيئة عمل مناسبة ، والقدرة على تشغيل الأجهزة التي تتطلّب عمليات حسية وإدراكية ووظائف جسمانية ، وعمليات تجهيز ومعالجة المعلومات .

أي أن هذا النوع من الاستبيانات يساعد في الحصول على قياس كمي لكثير من المتغيرات النفسية والبدنية والاجتماعية والمعرفية التي تسهم في أداء المهام المتعلقة بالأنواع المختلفة من الأعمال ، وكذلك ظروف العمل وخصائصه الفيزيائية التي تتطلّب من الأفراد العاملين التكيف معها .

(٤) انتقاء أو بناء أدوات القياس :

بعد التحديد المبدئي للمهارات والمعرف والقدرات الخاصة التي تبدو أهميتها في أداء نوع معين من الأعمال، فإن الخطوة التالية تتعلق بترجمتها إلى إجراءات وموافق اختبارية.

ويمكن الاعتماد في ذلك على الاختبارات والمقاييس المتوفّرة التي يتبيّن ملاءمتها لقياس مهارات أو قدرات مشابهة لما أسفرت عنه نتائج تحليل العمل. أما إذا لم تكن هناك اختبارات أو مقاييس مناسبة، فإنه ينبغي ابتكار مهام اختبارية لقياس هذه القدرات والمعرف والمهارات المتعلقة بعمل معين. وهذا يتطلّب اتباع جميع الخطوات التي سبق أن أوضحتها في الفصل الثامن عند مناقشتنا لكيفية بناء الاختبارات المقننة. ويفضل تجريب الاختبار على عينة استطلاعية مناسبة للتحقق من وضوح مفرداته وتعليماته، وتحديد زمن إجابته إذا كان من نوع الاختبارات الموقوتة، وتحليل مفرداته والتحقق من ثبات درجاته. أما فيما يتعلّق بصدق الاختبار وما يتعلّق بذلك من اختيار المحك المناسب والتحقق من صدقه، وتحديد درجة القطع وما يتربّط عليها من أخطاء اتخاذ قرارات الانتقاء أو التصنيف، فقد أوضحتنا بالتفصيل في الفصل الخامس.

ومن الجدير بالذكر أنه ليس من الضروري الاعتماد على الاختبارات التي تتطلّب ورقة وقلم، وإنما يمكن استخدام اختبارات محاكاة، أو اختبارات تتضمّن عينات من السلوك أو الأداء الفعلي في العمل كما أوضحتنا في هذا الفصل. غير أنه في جميع الحالات ينبغي إجراء دراسات للتحقق من الصدق التنبؤي لهذه الأدوات.

اختبارات الاستعدادات الخاصة :

ذكرنا فيما سبق أنه توجد أنواع متعددة من اختبارات الاستعدادات الخاصة في المجالات الأكاديمية والمهنية أو أوضحتها بالشكل التخطيطي (١ - ١).

ففي أغراض التوجيه والإرشاد المهني والأكاديمي، وانتقاء الأفراد تبدو أهمية قياس قدرات ومهارات ومعرف الأفراد في هذه المجالات، مما قد يتطلّب تطبيق عدة اختبارات مختلفة. غير أن تطبيق هذه الاختبارات المنفصلة يستغرق وقتاً طويلاً، كما أن معايرها تكون مستمدّة من جماعات مرجعية مختلفة، وبذلك لا تستطيع المقارنة بين درجات الفرد في أحد هذه الاختبارات بدرجاته في اختبار آخر مما يجعلها محدودة الفائدة في أغراض الإرشاد.

لذلك يُفضّل في هذه الحالة استخدام بطاريات الاستعدادات المتعددة Multiple Aptitude Batteries، حيث يمكن تطبيقها على مجموعة كبيرة من الأفراد في آن

واحد، ورسم صفحة نفسية (بروفيل) للفرد باستخدام درجات الاختبارات الفرعية التي تشمل عليها البطارية.

وسوف نعرض فيما يلى بعض أمثلة من هذه البطاريات الفارقة والمعامة الشائعة الاستخدام في المجالات الأكاديمية والمهنية، ثم نعرض أمثلة أخرى لبعض اختبارات الاستعدادات المتعلقة بمحالات نوعية محددة.

بطاريات الاختبارات المترددة :

(١) بطارية اختبارات الفارقة

: Differential Aptitude Test Battery (DAT)

تعد من أكثر بطاريات الاستعدادات استخداماً وتميزاً، وقد أعدها بنيت Bennet، وسيشور Seashore، وزمان Wesman، وصدرت طبعتها الأولى عام ١٩٤٧، وأجريت عليها بعد ذلك عدة تعديلات في طبعاتها التالية حتى صدرت الطبعة الخامسة عام ١٩٧٤. ويتفق كثير من علماء النفس والتربية على أن هذه البطارية تتميز بوجودة بنائها وطريقة عرضها لسلسلة الاختبارات التي تشمل عليها، حيث استند بناؤها إلى إطار نظري وعملي متوازن، وروى أن يكون الارتباط بين اختباراتها الفرعية منخفضاً، وأن تقيس مجموعة من القدرات الخاصة التي يمكن أن تساعده في التوجيه التربوي والمهني لطلاب المرحلة الثانوية.

وعلى الرغم من أن مصممي البطارية لم يعتمدوا في بنائها على أسلوب التحليل العاملى، إلا أنهم استرشدوا في انتقاءهم للاختبارات التي تشمل عليها بنتائج دراسات التحليل العاملى للاختبارات المتاحة، وكذلك بالاحتياجات الفعلية لعملية الإرشاد التربوى والمهنى، أى أنهم حاولوا انتقاء اختبارات يمكن الإفاده منها عملياً وتتسم بالثبات والصدق.

كما تميز البطارية بسهولة التصحيح، وتعتمد على القوة. ولكن تطبيقها يستغرق ما يقرب من ثلاث ساعات، ويتراوح زمن تطبيق أيٌّ من اختباراتها الفرعية من ٦ إلى ٣٠ دقيقة ، ويمكن تطبيق أيٌّ من هذه الاختبارات على حدة. وت تكون البطارية في طبعتها الخامسة من صيغتين متكافئتين ، واستمدت معايرها في البيئة الأمريكية من عينة قومية اشتملت على ٦٤٠٠ طالب في مرحلة التعليم الثانوى. وأعدت معاير مستقلة لكل من البنين والبنات، ويمكن رسم صفحة نفسية (بروفيل) للرتب المئوية المقابلة للدرجات الخام لكل طالب بحيث يمكن أن يستند إليها المرشد التربوى في تعريف الفرد بمستوى استعداداته المتعددة لكي يتخذ الفرد قراره في صوتها. ويمكن الرجوع إلى شكل (٦ - ٤) في الفصل السادس للاطلاع على مثال لهذا البروفيل.

وتشتمل البطارية على ثمانية اختبارات فرعية هي : الاستدلال اللفظي، (المتشابهات)، والقدرة العددية (العمليات الحسابية)، والاستدلال المجرد، والعلاقات المكانية، والاستدلال الميكانيكي، والسرعة ، والدقة الكتابية، والهجاء، واستخدام اللغة (النحو ، والترقيم).

ويذلك يكون لكل فرد تسع درجات بما في ذلك الدرجة الكلية، ويمكن أيضاً الحصول على مجموع درجتي كل من اختبار الاستدلال اللفظي ، والقدرة العددية المؤشر للاستعداد الدراسي العام. وقد وجد أن هذا المؤشر يرتبط باختبارات القدرة العقلية العامة ارتباطاً تزيد قيمته عن ٧٠ ، غير أنه لوحظ أن قيم معامل الارتباط بين الاختبارات الفرعية التي تشتمل عليها البطارية تراوحت بين ٦٠ ، ٦٧ ، مما يدل على أن هذه الاختبارات لا تقىس «عوامل» مستقلة، وبذلك تحتاج البطارية لمزيد من الدراسات للتحقق من أنها بطارية فارقة بالفعل ، أي التحقق من قدرتها التمييزية. وقد أجريت دراسات متعددة حول القيمة التنبؤية للبطارية ، وأوضحت الارتباطات بين درجات اختباراتها الفرعية وكل من التقدير العام للمطلب فى المقررات المختلفة، واختبارات الذكاء، ودرجات بطاريات استعدادات أخرى ، واختبارات تحصيل موضوعية، وبيانات تتعلق بالمهن، أن هناك علاقة فارقة بين درجات اختبارات البطارية والتحصيل فى المجالات الدراسية المختلفة، ومستويات القدرة العقلية ، والتصنيفات المهنية ، وتقديرات المشرفين على العمل فى مجالات مهنية محددة.

وقد تراوحت قيم معامل ثبات درجات الاختبارات الفرعية باستخدام طريقة التجزئة النصفية بين ٨٧ ، ٩٧ ، للبنين، وبين ٨٣ ، ٩٦ ، للبنات فى الفرق المختلفة .

(٢) بطارية اختبارات الاستعدادات العامة

: General Aptitude Test Battery (GATB)

أعد هذه البطارية مكتب خدمات التوظيف بالولايات المتحدة الأمريكية، وتختلف عن البطارية السابقة فى أنها تُستخدم فى أغراض الانتقاء والتسلكين فى المجالات الصناعية والتجارية وخاصة . بطارية الاستعدادات الفارقة أجرى تقييدها فى الإطار المدرسى ، وستستخدم أساساً فى التنبؤ بالنجاح الأكاديمى .

وقد صُممت بطارية اختبارات الاستعدادات العامة استناداً إلى تحليل العمل، والتحليل العاملى لمجموعة من الاختبارات بلغ عددها ٥٩ اختباراً متعلقاً بأعمال مهنية مختلفة ، حيث تبين أن هذه الاختبارات تنطوى على عوامل أو قدرات مختلفة لا يزيد عددها عن عشرة، لذلك أعدت هذه البطارية لقياس معظم هذه العوامل .

فالطبيعة الحديثة لهذه البطارية اشتغلت على ١٢ اختباراً فرعياً يمكن باستخدامها الحصول على درجات لتسعة عوامل أو قدرات مختلفة كالتالى :

القدرة العقلية العامة (وتتضمن القدرة اللفظية ، والقدرة العددية ، والقدرة المكانية) ، وإدراك الصيغ ، والإدراك الكتابي ، والتآزر الحركي ، ومهارة الأصابع ، والمهارة اليدوية .

وتُقاس هذه العوامل أو القدرات بثمانية اختبارات ورقة وقلم ، وأربعة اختبارات تعتمد على أجهزة . وبعض اختبارات الورقة والقلم تشبه اختبارات بطارية الاستعدادات الفارقة .

وهذه البطارية تناسب طلاب المرحلة الثانوية والراشدين . وتحوّل الدرجات الخام في العوامل التسعة إلى رتب مئنية أو درجات محوّلة متوسطها ١٠٠ وانحرافها المعياري ٢٠ ، وبذلك يمكن مقارنة الدرجات المحوّلة للفرد بأنماط الاستعدادات المهنية التي يبلغ عددها حوالي ٣٦ نمطاً استمدت من تحليل درجات البطارية من عينة من الأفراد العاملين في أكثر من ٨٠٠ مهنة .

ولتيسير استخدام أنماط درجات الاختبارات التي تشتمل عليها البطارية في أغراض الإرشاد المهني ، تم تجميع المهن التي تتطلب استعدادات متماثلة في عدد قليل من المجموعات المهنية . وحددت درجات قطع للاستعدادات الثلاثة الأكثر أهمية لكل مجموعة مهنية بحيث تمثل الحد الأدنى للمتطلبات التي تُسهم في الأداء الفاعل في كل من هذه المجموعات .

ومما يؤخذ على هذه البطارية اعتماد الاختبارات على عامل السرعة في الإجابة . وعدد الاستعدادات التي تقيسها محدود ، حيث إنها لا تتضمن اختبارات تقيس الفهم الميكانيكي ، أو الاستدلال ، أو الابتكار ، وكذلك فإن الاستناد إلى درجات القطع المتعددة في الاختبارات في عمليات الانتقاء والتسيكين بدلاً من استخدام معادلات الانحدار يحتاج إلى مزيد من الدراسات .

(٣) بطارية اختبارات سجل الدراسات العليا

: Graduate Record Examination Aptitude Test (GRE)

تستخدم هذه البطارية من الاختبارات لانتقاء الطلاب الذين يودون الالتحاق بالدراسات العليا بالكلليات المختلفة في الولايات المتحدة الأمريكية ، وتطبق في مختلف دول العالم لانتقاء طلاب البعثات والمنح الدراسية .

وتنقسم البطارية إلى قسمين رئيسيين : أحدهما قسم عام يشتمل على اختبارات تقيس الاستعداد اللفظي ، والاستعداد الكمي ، والاستدلال ، وقسم متقدم يقيس

التحصيل في حوالي ٢٠ تخصصاً أكاديمياً رئيساً، مثل علم النفس، والتاريخ، والكيمياء وغيرها.

وأختبارات الاستعداد اللغطي تتضمن التعرف على المتضادات، واستخدام المتشابهات، وفهم الفقرات، وتتضمن اختبارات الاستعداد الكمي، الاستدلال الحسابي، والجبر، والهندسة.

وتحوّل الدرجات الخام في البطارية إلى درجات محولة متوسطتها ٥٠٠، وانحرافها المعياري ١٠٠. وتتطلب بعض الكلمات حدّاً أدنى للدرجات الكلية المحولة للقبول. وقد أعدت مطبوعات وكتيبات وأدلة ودورات تدريبية تُعين في التدرب على إجابة اختبارات هذه البطارية لكي تزيد فرص حصول الطالب على درجات مرتفعة، على الرغم من عدم وجود بيانات موثقة تُبين فاعلية هذه المطبوعات والتدريبات في تحقيق الغرض منها.

وقد أوضحت نتائج استخدام معامل التجانس الداخلي اتساق درجات الاختبارات الفرعية بدرجة مقبولة، غير أنه لا توجد أدلة كافية تُبيّن مدى صدق البطارية في التنبؤ بنجاح الطلاب بعد التحاقهم بالدراسات العليا بالكليات المختلفة.

أختبارات استعدادات تتعلق بمجالات نوعية محددة :

تهدف اختبارات الاستعدادات التي تتعلق بمجالات مهنية أو دراسية معينة التنبؤ باحتمال نجاح الأفراد في هذه المجالات. ومن بين الاستعدادات الخاصة التي سنوضحها فيما يلى: الاستعداد الميكانيكي، والاستعداد الكتابي، والاستعدادات الموسيقية، والفنية، والابتكارية.

وسوف نقدم أمثلة لبعض الاختبارات التي تقيس هذه الاستعدادات. ولعل اختيارنا لهذه المجالات يرجع إلى أن كثيراً من الأفراد يهتمون بالشخص في دراستها أو العمل في مهن متعلقة بها مما يتطلب توافر الاستعدادات المناسبة لديهم، وتهدف برامج الإرشاد والتوجيه التربوي والمهني للتحقق من ذلك.

الاستعداد الميكانيكي :

تعتمد كثير من المهن التي يشغلها عدد كبير من الأفراد على الاستعداد الميكانيكي Mechanical Aptitude، وتبين هذه المهن في مستوى المهارة التي ينبغي أن تكون متوافرة لدى الأفراد الذين يعملون فيها.

فأعمال ميكانيكا السيارات والآلات والأجهزة وبناء السفن، وأعمال الخراطة

والسباكه والتجارة تعد أمثلة لهذه المهن . والتنبؤ بدرجة كبيرة من الثقة بنجاح الأفراد الذين يرغبون في الالتحام إليها يتطلب تطبيق اختبارات متعددة لقياس الاستعداد الميكانيكي لديهم .

ويرى بعض علماء النفس أن هذا الاستعداد يُعد قدرة واحدة تختلف درجتها من فرد إلى آخر ، بينما يرى البعض الآخر أنه يشتمل على عدد من القدرات المستقلة نسبيا ، غير أن الدراسات العاملية المتعلقة بهذا الاستعداد أكدت تعدد القدرات الخاصة به مما يتطلب قياسه باختبارات متعددة أو بطارية من الاختبارات المناسبة .

فالمهن المختلفة في هذا المجال تتطلب مهاماً مختلفة مما يدعو إلى قياس القدرات الخاصة المطلوبة للأداء الجيد باستخدام مجموعات مختلفة من الاختبارات المتعلقة بالاستعداد الميكانيكي .

وهذا الاستعداد لا يشير إلى سلوك حركي أو عضلي أو يدوى كما يظن البعض ، وإنما يشير إلى وظائف عقلية رمزية ، فقد أوضحت الدراسات التي حاولت تحليل الاستعداد الميكانيكي أنه يشتمل على مكونتين رئيسيتين هما : الفهم الميكانيكي المتعلق بفهم العلاقات الميكانيكية والفيزيائية ، والإدراك أو التصور البصري المكاني المتعلق بالتعامل الرمزي مع الأشكال والصيغ في الفراغ .

لذلك نلاحظ أن اختبارات الاستعداد الميكانيكي اعتمدت في بداية تطورها على قياس الأداء المهارى الفعلى حيث يقدم للفرد المُختبر معدات ميكانيكية بسيطة مفككة ويطلب منه تجميعها . غير أنه لوحظ أن هذه الاختبارات تتأثر تأثراً واضحاً بخبرة الفرد ، حيث يمكن أن يحصل فرد لديه خبرة بهذه المعدات على درجة مرتفعة في الاختبار على الرغم من انخفاض استعداده الميكانيكي . وهذا يدل على أن هذه الاختبارات لا تصلح لقياس هذا الاستعداد ، علاوة على أنها تقيس متغيرات تتعلق بالمهارة اليدوية والتآزر البصري الحركي .

لذلك اتجه العلماء إلى قياس المكونتين الرئيسيتين اللتين ذكرناهما أعلاه المتعلقتين بالوظائف العقلية الرمزية .

(١) اختبار الفهم الميكانيكي

: Bennett Mechanical Comprehension Test (BMCT)

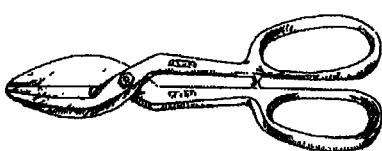
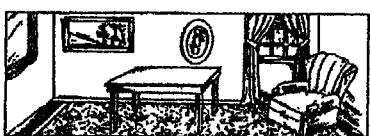
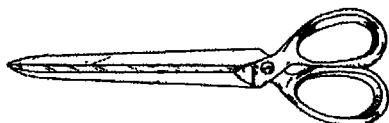
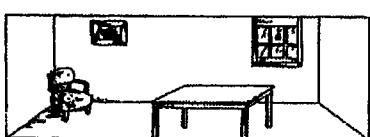
أعد هذا الاختبار بنيت Bennett ، وصدرت طبعته المعدلة عام ١٩٧٠ ، وهو من الاختبارات شائعة الاستخدام لقياس فهم المبادئ الميكانيكية المتعلقة بالأدوات التي تستخدم بكثرة في الحياة اليومية ، مثل المفك ، والشاوكش ، ومسنث الصمولة ، والمقص ، وغير ذلك .

ويُعد هذا الاختبار مماثلاً أيضاً لأحد الاختبارات الفرعية المتعلقة بالقدرة الميكانيكية في بطارية الاستعدادات الفارقة (DAT) التي أشرنا إليها فيما سبق. ويقيس هذا الاختبار فهم العلاقات الميكانيكية، وتطبيقات المبادئ الفيزيائية في مواقف عملية، والاختبار عبارة عن كتيب يتضمن ٦٠ رسمًا تخطيطيًا لمواصفات ميكانيكية لا تحتوى بالضرورة على أدوات يدوية، مثل العلاقات بين الجبال والبكرات، والعجلات، والقوى الفيزيائية، والتركيبيات الميكانيكية.

ولا يتطلب الاختبار أداء عملياً وإنما يتطلب تفكيراً رمزياً، ولا يعتمد على المعرف العامة أو تذكر الأشياء وإنما يعتمد على الفهم، والتحليل المنطقي، واستيعاب العلاقات بين الأشياء.

ممثلاً يقدم للفرد رسم تخطيطي لمقصين ويطلب منه تحديد أيهما يمكن استخدامه بدرجة أفضل في قطع المعادن، أو غرفتين تحتويان على قطع مختلفة من الآثار، ويطلب منه تحديد الغرفة التي يحدث فيها صدى أعلى للصوت.

ويوضح الشكل (١٠ - ٢) التالي هاتين المفردتين :



شكل (١٠ - ٢) يوضح بعض مفردات اختبار الفهم الميكانيكي

وقد أعددت ثلاثة صيغ للاختبار تتفاوت في درجة صعوبتها، حيث تُطبق صيغة معينة منها على المتقدمين للعمل في المستويات المهنية العليا، ويستغرق تطبيق الاختبار عادة حوالي نصف الساعة، ولكنه يعتمد على القراءة. وقد صُمم الاختبار لكي يناسب طلاب نهاية مرحلة التعليم الثانوي، والمتقدمين للالتحاق بالمهن الصناعية والعمالين فيها. وتُفسر درجات الاختبار في ضوء معايير مئوية خاصة مستمدة من طلاب نهاية المرحلة الثانوية العامة والصناعية وطلاب السنة الأولى بكليات الهندسة، ومعايير خاصة للمتدربين في المصانع، والمهندسين وغيرهم من العاملين، ومعايير خاصة بالإثاث.

ويشير دليل الاختبار إلى أن قيم معاملات صدق الاختبار في التنبؤ بالنجاح في المهن المختلفة التي تتطلب الاستعداد الميكانيكي تترواح بين ٣٠ ، ٦٠ ، . . . وقيم معامل ثبات درجات الاختبارات تترواح بين ٨٠ ، ٨٤ ، . . .

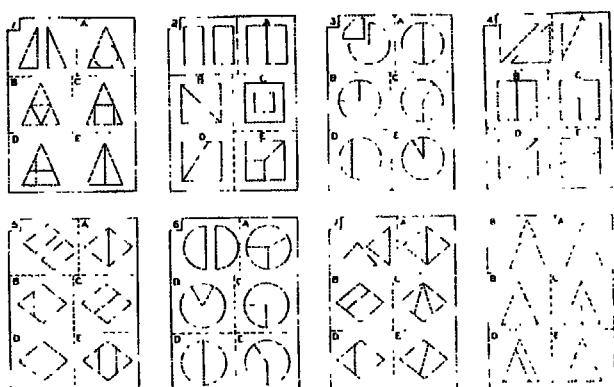
(٢) اختبار مينيسوتا للعلاقات المكانية

Minnesota Spatial Relations Test

أعد هذا الاختبار ترايبيو Trabue وزملاؤه. وصدرت طبعته الثانية المُعدلة عام ١٩٧٩ التي أعدها ليكرت Likert ، وكواشا Quasha. لقياس المكونة الرئيسة الثانية من مكونات الاستعداد الميكانيكي. فادراك العلاقات المكانية هو القدرة على الحكم والتعامل عقلياً مع الأشياء في الفراغ، ومعرفة تأثير إعادة تنظيمها ، والتوصل إلى العلاقات القائمة بينها في وضعها الجديد، أي أنه يتعلق بالتجهيز المكاني Spatial Orientation أو التصور البصري المكاني. والاختبار بعد تعديله يُطلق عليه اختبار لوحات الورق Revised Minnesota Paper Form Board Test (MPFB)

ويشتمل على ٦٤ رسمًا هندسياً لأجزاء غير مرتبة وُضعت كل مجموعة منها في ستة خلايا من لوحات مستطيلة، وبعض هذه الأجزاء مقلوبة أو تم تدويرها، ولللوحات متدرجة الصعوبة.

ويطلب من الفرد اختيار الشكل الهندسي الصحيح بعد تجميعه من بين بدائل أشكال معطاة كما هو موضح بشكل (١٠ - ٣) التالي (Guion , 1965) :



شكل (١٠ - ٣) يوضح بعض مفردات اختبار العلاقات المكانية

ومن الجدير بالذكر أن معظم بطاريات الاستعدادات الخاصة وكثيراً من اختبارات الذكاء تشتمل على اختبار فرعى يقيس العلاقات المكانية، نظراً لأن الدراسات أوضحت

أن هذه القدرة منبئ جيد بالنجاح في المهن وال المجالات الدراسية المختلفة. غير أن هذا الاختبار يُعد من الاختبارات العجيدة لقياس هذه العلاقات، حيث أشارت إليه معظم الدراسات التي تناولت الاستعداد الميكانيكي. فهو يفيد في التنبؤ بالنجاح في المجالات الهندسية والتصميم المعماري وتصميم الآلات وغيرها.

والاختبار يناسب طلاب الصنوف الأخيرة من التعليم الثانوي وكذلك الراشدين، وأعدَّ معايير مئينة خاصة بهؤلاء الطلاب، ومعايير أخرى للمتقدمين لهذه المهن والمتدربين والمهندسين.

وقد أعدَّ صيغتان متكافئتان للاختبار، وتراوح قيم ثبات درجات الاختبار بين .٨٠ ، .٨٥ ، .٩٠ ، .٩٢ ، إذا استُخدمت أي من الصيغتين، بينما تراوحت القيم بين .٨٩ ، .٩٠ ، .٩٢ ، إذا استُخدمت الصيغتان معاً؛ لأن ذلك يؤدي إلى إطالة الاختبار. وأجريت دراسات متعددة للتحقق من الصدق التنبؤى للاختبار في مهن مختلفة تتعلق بالمجالات الميكانيكية وغير الميكانيكية التي تتطلب إدراك العلاقات المكانية. وقد تراوحت قيم معامل الصدق التنبؤى للمجالات الميكانيكية بين .٣٨ ، .٥٩ ، .٥٧ ، .٢٨ ، كما وُجد أن ارتباط درجات هذا الاختبار بتقديرات الطلاب في كليات الهندسة بلغ .٣٥ ، وارتباطه بدرجات اختبار الفهم الميكانيكي تراوح بين .٥١ ، .٥٢ ، .٢٢ ، وباختبار تصميم المكعبات الذي يُعد أحد اختبارات مقياس ويكسنر - بلفيو .٤٠ ، .٤١ .

الاختبارات الاستعدادات النفسية الحركية :

يلاحظ أن هناك بعض اختبارات الورقة والقلم التي تقيس المعلومات والخبرات الميكانيكية، مثل اختبار أوروروك للمعلومات الميكانيكية O'Rourke Mechanical Information Test الذي يتطلب في جزءه الأول أن يزاوج الفرد بين صور الأدوات اليدوية وصور الأشياء التي تستخدم معها. وفي الجزء الثاني يشتمل على مفردات اختيار من متعدد تقيس معلومات تتعلق بالأدوات واستخداماتها في مجالات التجارة والكهرباء والطلاء والطباعة وغيرها.

كما أن هناك اختبارات الاستعدادات النفسية الحركية التي تقيس المهارات اليدوية المختلفة Manual Dexterity ، مثل اختبارات حركات الأطراف والأصابع والعضلات في أداء مهام تجميعية معينة بمهارة وسرعة ودقة. ولكن ينبغي التمييز بين الاستعداد الميكانيكي والاستعداد اليدوي، فعلى الرغم من أن كليهما يتطلب استخدام الأدوات والآلات والمعدات الميكانيكية، إلا أن الاستعداد الميكانيكي - كما سبق أن ذكرنا - يعتمد

على عمليات عقلية رمزية تؤدي إلى اتخاذ قرارات تتعلق بما يجب أن تفعله الأيدي والأدوات والآلات، وربما يجري الفرد هذه العمليات ببطء ومع هذا ينجح في مهنته الميكانيكية.

فما يهمنا هنا هو قدرة الفرد على تشخيص الأعطال، وكيفية إصلاحها. أما الاستعداد اليدوي فهو متطلب أساسي في خطوط التجميع الآلية التي لا تحتاج إلى اتخاذ قرارات معينة، حيث إن الفرد يعطي تعليمات محددة وواضحة بما يجب أن يفعله، وعليه أن يؤدي المهام بمهارة وسرعة ودقة، وهذا يعتمد على مهارته اليدوية وليس على فهمه الميكانيكي. وهناك اختبارات متعددة تقيس مدى سرعة ودقة حركات الذراع واليد والأصابع في التعامل مع الأدوات والأشياء والأجزاء الدقيقة، وتختلف هذه الاختبارات باختلاف درجة دقة الأجزاء المراد التعامل معها.

(١) اختبار مينيسوتا لمعدل الحركات

Minnesota Rate of manipulation Test (١٩٦٩):

وهو اختبار شائع الاستخدام يتطلب حركات اليد والذراع، وهو عبارة عن لوحة بها ٦٠ ثقباً في أربعة صفوف، ويوجد ٦٠ قرصاً صغيراً اسطوانيّاً متماثلاً إحدى قاعدتي كل منها ملون باللون الأحمر، والقاعدة الأخرى باللون الأخضر، كما هو مبين في الشكل (٤ - ١٠) التالي (Guion, 1965):



شكل (٤ - ١٠) يوضح بعض مفردات اختبار معدل الحركات

والمهام التي يتطلبها الاختبار هي . تدوير الأقراص أو تعديل وضعها وإدخالها في الثقوب، أو إدخال الأقراص في الثقوب من وضع جانبى بالنسبة لللوحة، وإخراج الأقراص من الثقوب، واستخدام يد واحدة في تعديل وضع الأقراص وإدخالها في الثقوب، ثم استخدام اليدين معاً في إجراء ذلك.

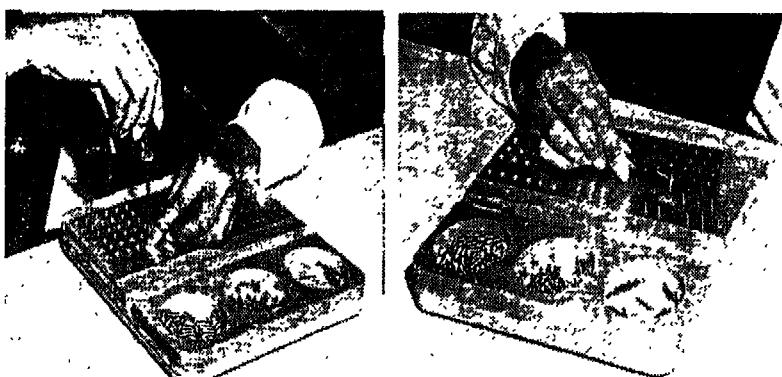
ومقدار الزمن الذي يستغرقه الفرد في إدخال جميع الأقراص في الثقوب يمثل درجته في الاختبار، وقد استمدت معايير الاختبار من تطبيقه على عينة اشتملت على ١١٠٠ من العاملين وغير العاملين من الشباب، وعينة أخرى من غير العاملين من الكبار اشتملت على ٣٠٠ فرد، ويقارن الزمن الذي استغرقه الفرد بهذه المعايير. وقدم دليل الاختبار بعض المؤشرات عن الصدق التنبؤي لل اختبار في موافق تطبيقية متعددة.

(٢) اختبار كراوفورد لمهارة التعامل مع الأجزاء الدقيقة

:Crawford Small Parts Dexterity Test

يقيس هذا الاختبار المهارة الدقيقة وتآزر العين واليد، ويشتمل الاختبار على جزئين: أحدهما يتطلب من الفرد استخدام ملقط في إدخال دبابيس في ثقوب ربعة ووضع غطاء فوقها، والأخر يتطلب وضع مسامير لولبية دقيقة في ثقوب بها خيوط وربطها بمنفك.

ويوضح شكل (١٠ - ٥) التالي صورة الجهاز المستخدم في ذلك (Guion , 1969)



شكل (١٠ - ٥) يوضح بعض مفردات اختبار مهارة التعامل مع الأجزاء الدقيقة

وتعتمد معايير الاختبار على الزمن الذي يستغرقه الفرد لإنجاز المطلوبين، واستمدت من عينات اشتملت على عشر مجموعات من العاملين والمتقدمين للالتحاق بالعمل. وقد تبين أن قيمة معامل الارتباط بين درجات جزئي الاختبار ،٤٠ ، تقريرياً، في حين أن ثبات درجات الاختبار بطريقة التجربة النصفية بلغت قيمته ،٨٥ ، تقريرياً.

والصدق التنبؤى للاختبار فى المجالات المهنية التى تتطلب مهارة يدوية دقيقة يحتاج إلى مزيد من الدراسات نظراً لأن ما أشار إليه دليل الاختبار فى هذا الشأن لا يمكننا من تقييم صدق الاختبار.

الاستعداد الكتابى :

يتضمن الاستعداد الكتابى Clerical Aptitude المهارات المكتبية المتعددة المتعلقة بأعمال الكتابة على الآلة الكاتبة أو الحاسوب، وحفظ السجلات وتحديثها، والاحتزال وتذقيق الحسابات وغير ذلك.

والاستعداد الكتابى ليس قدرة واحدة وإنما يتضمن قدرات متعددة. وعلى الرغم من أن هناك أفراداً يؤدون أحد هذه الأعمال وليس جميعها، إلا أن هناك اختبارات تقيس الاستعداد الكتابى العام، واختبارات أخرى تقيس قدرات خاصة نوعية. فالاستعداد الكتابى العام ينطوى على قدرات تمثل القدرات التى ينطوى عليها الذكاء العام.

لذلك تشمل الاختبارات التى تقيس هذا الاستعداد على مفردات مماثلة للمفردات التى تشمل عليها اختبارات الذكاء العام، مثل القدرات اللفظية والعددية، غير أنها تشمل أيضاً على مفردات تقيس المهارة السيدوية، وسرعة إدراك أوجه الشبه والاختلاف بين الكلمات والأعداد. كما أن بطارية اختبارات الاستعدادات الفارقة (DAT) التى أشرنا إليها فيما سبق تتضمن بعض الاختبارات الفرعية المتعلقة بالاستعداد الكتابى.

وسوف نقدم فيما يلى أمثلة لبعض اختبارات الاستعداد الكتابى العام، والاستعدادات النوعية شائعة الاستخدام المتعلقة به:

اختبارات تقيس الاستعداد الكتابى العام :

(١) اختبار الاستعداد الكتابى العام (١٩٧٢)

:General Clerical test (GCT)

يُعد هذا الاختبار بطارية من الاختبارات تشمل على تسعَة أجزاء يحصل الفرد فيها على ثلاثة درجات فرعية؛ إحداها تتعلق بسرعة ودقة أداء المهام الكتابية الروتينية، وتتطلب مراجعة الأعمال الورقية، وترتيب الحروف، والثانية تتعلق بالمهام الحسابية التي تُعد ضرورية لبعض الأعمال المكتبية، وتتطلب إجراء عمليات حسابية، وتحديد أخطاء الحسابات، والاستدلال الحسابي، والثالثة تتعلق بالجانب اللغوي أي المهارات اللغوية المهمة لأداء المهام الكتابية، وتتطلب مهارات الهجاء، ومهارات فهم المقروء، والكلمات، والنحو. ويمكن الحصول على درجة الاستعداد الكتابى العام بجمع الدرجات الثلاث.

وقد أعدَّت معايير مئينة مستقلة لكل من الأجزاء الثلاثة لكل من الذكور والإناث استناداً إلى جماعات مرئية متعددة تشمل المتقدمين للالتحاق بوظائف كتابية، وطلابات نهاية مرحلة التعليم الثانوي التجاري، والعاملين في الوظائف الكتابية المختلفة.

وتتميز البطارية بسهولة تطبيقها وتصحيحها ، ويستغرق تطبيقها ٤٧ دقيقة تقريباً، وحدَّد زمن مستقل لكل جزء من الأجزاء الثلاثة، وترواحت قيم معامل ثبات درجات كل من هذه الأجزاء بين ٠٨٢ ، ٠٩١ ، ٠٩٤ . وقيمة معامل ثبات الدرجات الكلية ٠٩٤ . واستُخدمت طرق مختلفة للتحقق من صدق البطارية في التنبؤ بالنجاح في الأعمال الكتابية بعامة وكذلك الصدق التلازمي ، وكانت النتائج إيجابية.

كما وجد أن درجات البطارية ترتبط بدرجات أحد اختبارات الذكاء العام ارتباطاً موجباً تراوحت قيمته بين ٥٤ ، ٨٣ ، ٠٠ ، مما يشير إلى العلاقة بين الاختبارين . غير أن هذه العلاقة تكون منخفضة في حالة اختبارات الاستعدادات الكتابية النوعية التي سنوضحها بعد قليل .

(٢) اختبار مينيسوتا للاستعداد الكتابي

؛Minnesota Clerical Test (MCT)

يُعد هذا الاختبار من أكثر اختبارات الاستعدادات الخاصة استخداماً حيث اعنى بنائه وتقنيته ومراجعته وأجريت حوله العديد من الدراسات، وقد أعده Andrew Paterson ، وباترسون ومتخصص في هذا الاختبار في التنبؤ بالنجاح في مهن كتابية مختلفة، مثل العمل المكتبي العادي، وأعمال الفحص والمراقبة الإدارية. فهذه الأعمال تتطلب سرعة الإدراك البصري، والسرعة والدقة في التعامل مع الأعداد والحوروف وغيرها من الرموز .

لذلك يهدف الاختبار لقياس سرعة القراءة، والفهم، واستخلاص معلومات من التقارير والمراسلات والأسماء والأعداد. ويكون الاختبار من جزئين: أحدهما يتعلق بمراجعة الأعداد والآخر يتعلق بمراجعة الأسماء ويشتمل كل من الجزئين على أزواج من الأعداد أو الأسماء (فمنا بتعريف الأسماء) كما يلى :

<u>المقارنة بين الأسماء</u>		<u>المقارنة بين الأعداد</u>
٣٤٥١٢	طارق شاهين	٣٤٥١٢
٤٧٣٠٦	حسن السيد	٤٧٣٠٦
٨٩٧٠٥	مدينة نصر ، القاهرة	٨٩٧٠٧
١٣٧٤٢٨	الشركة التجارية ش.م.م	١٧٣٤٢٨

ويُطلب من الفرد فحص هذه الأزواج ، ووضع علامة على الخط الفاصل بين مفردتي كل زوج منها إذا كان هناك تطابق بين المفردتين ولا توضع أى علامة إذا كانتا مختلفتين. وت تكون القائمة من ٢٠٠ زوج من الأعداد، ٢٠٠ زوج من الأسماء على الترتيب، وتقدر درجة لكل جزء، ولا تجمع درجات الجزئين.

ويتميز الاختبار بسهولة التطبيق والتصحيح، ويستغرق تطبيقه ١٥ دقيقة تقريباً حيث إنه يعتمد على السرعة، والاختبار يناسب طلاب المرحلة الثانوية. وتتعدد المعايير المئوية للاختبار ، إذ توجد معايير منفصلة للذكور والإناث ، وللفرق الدراسية بالمرحلة الثانوية ، وللمناطق الجغرافية ، وللمتقدمين للأعمال الكتابية ، ولبعض المهن الكتابية المختلفة. وقيم معاملات ثبات درجات الاختبار بطريقة إعادة تطبيق الاختبار تراوحت بين ٩١ ، ٧٧ ، ٠ . غير أن ارتباط درجات الاختبار بتقديرات المعلمين والمشرفين في مجالات العمل الكتابية تراوحت قيمه بين ٢٨ ، ٦٥ ، ٠ . والارتباط بين درجات الجزء المتعلق بمراجعة الأعداد ودرجات اختبار الاستعداد الأكاديمي بلغت قيمته ١٠ ، ٠ ، بينما بلغت قيمة الارتباط بين درجات الجزء المتعلق بمراجعة الأسماء، ودرجات اختبار الاستعداد الأكاديمي ٣٥ ، ٠ ، وكذلك وجد أن قيمة الارتباط بين هذا الاختبار وختبار الاستعداد الكتابي العام الذي أشرنا إليه فيما سبق ٣٦ ، ٠ .

ويوضح من ذلك أن السرعة الإدراكية تعد أحد العناصر الأساسية لكثير من الأعمال الكتابية ، ولكن ينبغي أيضاً قياس عناصر أخرى مثل الفهم اللغظى ، والعمليات الحسابية. وقد أعدت اختبارات أخرى لقياس هذه العناصر ، مثل الاختبار المسمى لاستعداد العاملين Employee Aptitude Survey الذى يقيس الفهم اللغظى المتمثل في القدرة على استخدام الكلمات فى التواصل والتفكير والتخطيط. ويتضمن الاختبار أيضاً مفردات لقياس القدرة العددية المتمثلة فى التعامل بسهولة مع الأعداد، وإجراء العمليات الحسابية البسيطة بسرعة ودقة . وكذلك اختبار فلانagan لتصنيف الاستعدادات Flanagan Aptitude Classification Test الذى يشتمل على مفردات لقياس مهارات التواصل، وبخاصة فيما يتعلق بتركيب الجمل ، والنحو .

الختبارات تقيس استعدادات الكتابية نوعية :

الختبارات الاستعداد المتعلق بالاختزال والألة الكاتبة:

يوجد العديد من الاستعدادات التى تتعلق بأعمال أو مهن خاصة ، مما يتطلب بناء اختبارات لقياس هذه الاستعدادات استناداً إلى إجراءات تحليل العمل ومتطلباته التي سبق أن أوضحتها بالتفصيل . فالأعمال الكتابية المتعلقة بالكتابة على الآلة الكاتبة ،

والاختزال Stenographic Jobs تتطلب قدرات خاصة ومهارات متعددة. فتعلم الاختزال يعتمد على اكتساب نوع جديد من التمثيل الرمزي اللغوي، وبذلك يتضمن المهارات اللفظية. وتعلم الآلة الكاتبة يعتمد على حدة البصر، ومهارة الأصابع، وتأثر العين والأصابع، وغير ذلك من المهارات. وتلعب السرعة دوراً مهماً في تقييم الأداء في هذه الأعمال.

ومن بين الاختبارات التي تقيس هذه الاستعدادات اختبار الكفاءة في أعمال الاختزال Seashore - Bennett Stenographic Profeciency Test (١٩٥٦).

وتُعرض المعلومات على الفرد المختبر عن طريق الحاسوب أو شرائط التسجيل، وهذه المعلومات عبارة عن خطابات تجارية غير فنية، وذلك لتقييم كفاءته في أعمال الاختزال. ويكون الاختبار من صيغتين متكافتين، وتتضمن كل منهما خمسة خطابات. وتتبادر في عدد فقراتها وسرعة عرضها. وبعد إملاء الخطابات في مدة ١٥ دقيقة يقوم الفرد بنسخها على الآلة الكاتبة في مدة ٣٠ دقيقة. ويُستخدم هذا الاختبار في انتقاء العاملين في مجال الاختزال وأعمال السكرتارية.

اختبارات الاستعداد المتعلق ببرمجة الحاسوب :

تطلب الأعمال المتعلقة ببرمجة الحاسوب قدرات خاصة أو مهارات معينة ينبغي توافرها في المتقدمين لهذه الأعمال ، غير أن الاختبارات التي تقيس هذه القدرات محدودة ، ومن أمثلتها بطارية اختبارات استعداد برمجة الحاسوب Computer Programming Aptitude Battery Palormo .

وتشتمل البطارية على خمسة اختبارات فرعية تتعلق بفهم الكلمات، وتقريب الأعداد، وسلسل الحروف، والاستدلال، وتصميم الأشكال التخطيطية. ويستغرق تطبيق البطارية ٧٥ دقيقة، وتقارن درجة الفرد في كل من الاختبارات الخمسة ودرجته الكلية بمعايير مئوية.

ونظراً لانتشار استخدام الحاسوب في الأعمال الكتابية فإن الاهتمام أصبح موجهاً لبناء اختبارات استعداد في هذا المجال بدليلاً عن الاختبارات المتعلقة بالآلة الكاتبة.

الاستعداد الموسيقي :

يُعد الاستعداد الموسيقي سمة مركبة من مجموعتين من المكونات تتعلق إداهما بالذوق والإلهام والانفعال والإبداع، وتعلق الأخرى بالجوانب الآلية والحسية والإدراكية. وكل من المجموعتين ربما يكون ضرورياً للأداء الموسيقي الجيد ولكنهما

غير كافيين للنجاح في المهن الموسيقية. ومع هذا فإن اختبارات الاستعداد الموسيقي المتوفّرة ترتكز على قياس المجموعة الثانية من المكونات، وذلك لأنّه يمكن تحديدها تحديداً دقيقاً وقياسها بسهولة.

غير أنّ كثيراً من هذه الاختبارات أُعدَتْ منذ مدة طويلة، والاهتمام ببناء اختبارات جديدة لقياس الاستعداد الموسيقي يبدو محدوداً. وتعتمد الإجابة عن مفردات معظم هذه الاختبارات على الحكم الشخصي للفرد المختبر، حيث لا يُطلب منه العزف الفعلي على آلة موسيقية معينة، ولكنه يستمع إلى لحن معين ويحكم على هذا اللحن من حيث النغمة والإيقاع وحدهما الصوت وغير ذلك من الخصائص، فالعزف على الآلات الموسيقية يتطلّب إلى جانب ذلك مهارات نفسية حركية.

وعلى الرغم من أنّ كثيراً من اختبارات الاستعداد الموسيقي تقيس المكونات الأساسية للاستعداد الموسيقي الناتجة عن تحليل هذا الاستعداد تحليلاً سيكولوجياً، إلا أن بعضها يقيس المحتوى الموسيقي في تكامله وشموله. وسوف نقدم فيما يلى أمثلة لبعض هذه الاختبارات :

(١) مقاييس سيشور للاستعداد الموسيقي

:Seashore Measures of Musical Talents

وتُعد هذه المقاييس أقدم مقاييس الاستعداد الموسيقي التي استندت إلى البحوث الرائدة التي أجرتها سيشور وزملاؤه بجامعة أيوا الأمريكية (Seashore , 1938)، وأعيد مراجعتها عامي ١٩٥٧ ، ١٩٦٠ ، حيث أجريت تعديلات في التعليمات، كما طرأت أجهزة التسجيل لكي تتماشى مع المستحدثات التكنولوجية.

واعتمد بناء المقاييس على عدد من العناصر المتعلقة بالموهبة الموسيقية، حيث يُقدم للفرد المختبر سلسلة من المهام ويطلب منه تمييز الفروق بين أزواج المثيرات، وتُقدم هذه المثيرات في ستة اختبارات منفصلة مسجلة على شريط ليزر. وتقيس هذه الاختبارات درجة الصوت، وعلوه، والإيقاع، والزمن، وسرعة النغمة، وذاكرة النغم.

ويمكن تطبيق هذه المقاييس على الأطفال من الصف الرابع الابتدائي إلى الرشد. ويستغرق تطبيق المقاييس ساعة واحدة تقريباً بمعدل عشر دقائق لكل اختبار. وقد أكد سيشور ضرورة تصحيح كل اختبار على حدة، ولا تُجمع درجات الاختبارات، وأعدت معايير مئنية لكل اختبار لفرق الدراسية المختلفة، وتراوحت قيم معامل ثبات

درجات الاختبارات الستة باستخدام طريقة التجزئة النصفية بين ٥٥ ، ٨٥ ، . . . وأوضحت نتائج بعض الدراسات أن ارتباط درجات الاختبارات بمحك التدريب الموسيقي تراوحت قيمته بين ٣٠ ، ٤٠ ، . غير أن بيانات الصدق التنبؤى للمقاييس غير كافية.

(٢) بطارية وينج المقتننة للذكاء الموسيقى

:Wing Standardized Tests of Musical Intelligence

أعد وينج Wing هذه البطارية في إنجلترا عام ١٩٤١ وأعيد مراجعتها وصدرت عام ١٩٦٢ بلغات ثلاث: الإنجليزية والفرنسية والألمانية . وتميز هذه البطارية عن مقاييس سيشور في أنها لم تعتمد على قياس مكونات الاستعداد الموسيقي ، ولكن اعتمدت على المحتوى الموسيقي ككل .

ويمكن تطبيق البطارية ابتداء من عمر ٨ سنوات ، ويُستخدم فيها موسيقى البيانو على شريط تسجيل ، وتكون البطارية من سبعة اختبارات فرعية هي : تحليل الأوتار من حيث عدد النغمات التي تُصدرها ، وتميز النغمات ، وذاكرة النغم ، وائلالنغم ، ووحدة الصوت ، والإيقاع ، والتعبير . وتخالف هذه البطارية أيضاً عن مقاييس سيشور في أن الاختبارات الأربع الأخيرة تتطلب من الفرد تقدير الجوانب الذوقية للنغمات المختلفة . ويستغرق تطبيق البطارية ساعة واحدة تقريباً ، ويمكن تطبيقها في جلستين .

وأعدت معايير منفصلة للأعمار من ٨ إلى ١٧ عاماً للدرجات الكلية ودرجات الاختبارات الفرعية الثلاثة الأولى التي اعتبرت بمثابة بطارية قصيرة تتناسب بالأطفال الصغار وكأداة فحص أولية للاستعداد الموسيقي . وهذه المعايير عبارة عن تقديرات بالحروف وكذلك الأعمار الموسيقية . وترواحت قيم معامل ثبات درجات الاختبارات الفرعية بين ٧٠ ، ٧٠ ، للأطفال الصغار ، و ٩٠ ، ٩٠ ، للأطفال الأكبر عمراً . كما أن صدف البطارية في التنبؤ بمحك الأداء الموسيقي يحتاج إلى مزيد من الدراسات ، على الرغم من أنه وجد ارتباط بين درجات البطارية وتقديرات معلمى الموسيقى لطلابهم بلغت قيمته ٦٠ ، أو أعلى . وقد نالت هذه البطارية اهتمام كثير من معلمى الموسيقى لأنهم يرون أن محتواها يتضمن المهارات ذات الأهمية في التدريبات الموسيقية .

لذلك يمكن استخدامها في التوجيه وانتقاء الطلاب الذين يرغبون في متابعة التدريبات الموسيقية المتقدمة . ومن الجدير بالذكر أن هناك بطارية أخرى تعرف باسم «بروفيل الاستعداد الموسيقي Musical Aptitude Profile» أعدتها جوردون عام ١٩٦٥ وأجرى مراجعتها عام ١٩٦٧ ، وتعتمد على آلات موسيقية مختلفة ، وتشتمل على سبعة

الاختبارات صُنفت في ثلاث مجموعات هي : تصور النغمات ، وتصور الإيقاع ، والحس الموسيقي .

ومن هذا يتضح أن الصدق التبؤى لبطاريات الاستعداد الموسيقى يتطلب مزيداً من البحث ، وربما يفضل استخدام اختبارات أخرى إلى جانب هذه البطاريات للحصول على تقدير أفضل لهذا الاستعداد .

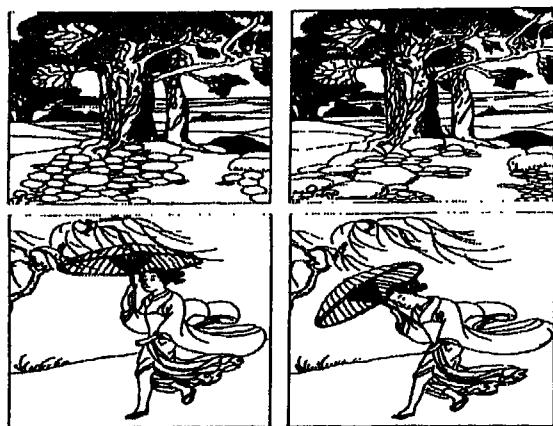
الاستعدادات الفنية :

إن قياس الاستعدادات الفنية Art Aptitudes يعد من المجالات التي لم تحظ بنفس القدر من الاهتمام الذي حظى به قياس الاستعدادات الأخرى التي أوضحتها فيما سبق .

ولعل هذا يرجع إلى صعوبة الفصل بين الاستعدادات الفنية والتحصيل ، حيث إن بعض اختبارات هذه الاستعدادات تقيس المعلومات والخبرات المستعملة في المجالات الفنية . كما أن الاهتمام بانتقاء وتصنيف وتسكين الأفراد في المدارس والمصانع والمجالات العسكرية نال معظم اهتمام علماء القياس مما أدى إلى قلة الاختبارات في مجال الفنون ، وساعد على ذلك أيضاً تعدد وتباطئ القدرات التي يتطلبه هذا المجال . كما أن الفرد الذي يستطيع تذوق الفن أو نقده ليس بالضرورة لديه الاستعداد الذي يمكنه من الإنتاج الفني .

لذلك ربما يكون من الضروري التمييز بين مقاييس الاستعدادات المتعلقة بالحكم الذوقى Aesthetic Judgment ، والإنتاج الفني Aesthetic Production التي تعتمد على عينات فعلية من الأعمال الفنية .

ويوضح شكل (٦ - ١٠) التالي بعض مفردات هذه الاختبارات : (Nunnally , 1975)



شكل (٦ - ١٠) يوضح بعض مفردات اختبارات التقييم الفني

وسوف نقدم فيما يلى أمثلة لبعض مقاييس كل من النوعين.

أولاً: اختبارات مايير للتقييم الفني

(١) اختبارات مايير للتقييم الفني Meier Art Judgment Tests

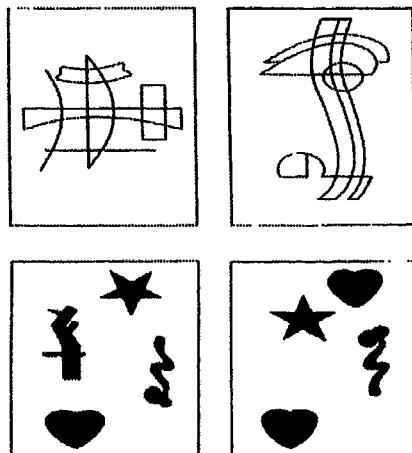
تُعد من الاختبارات المعروفة في هذا المجال، وقد أعدها مايير Meier في طبعتها الأولى عام ١٩٢٩ ، وأعيد طباعتها عام ١٩٦٣ ، وتشتمل الاختبارات على ١٠٠ زوج من اللوحات الفنية الشهيرة غير الملونة، وإحدى لوحات كل زوج تُعد لوحة أصلية واللوحة الأخرى بها بعض التعديلات البسيطة. ويُطلب من الفرد تحديد أوجه الاختلاف بينهما، ويقرر أيهما يفضل. فالاستعداد المتعلق بالحكم الذوقي ينبغي أن يتوافر لدى الناقد الفني، أما الاستعداد المتعلق بالإلتحاق الفني فينبغي أن يتوافر لدى الفنان نفسه. وهذا الإنتاج ربما يتطلب أيضًا الحكم الذوقي، ولكن العكس لا يكون صحيحاً.

وقد أعدت معايير مثينة لطلاب المرحلة الإعدادية والثانوية، وتراوحت قيم معامل ثبات درجات الاختبارات بطريقة التجزئة النصفية بين ٠٠,٧٠ ، ٠٠,٨٤ ، ٠٠,٤٠ ، ٠٠,٧٠ ، وهى قيم مقبولة. ومن الجدير بالذكر أن مايير Meier أعد اختباراً للإدراك الجمالى Aesthetic Perception عام ١٩٦٣ لقياس الفروق الفردية في إدراك القيمة الجمالية للطرق المختلفة في تكوين لوحة فنية ، ويشتمل الاختبار على ٥٠ لوحة فنية لكل منها أربعة بدائل ، ويُطلب من الفرد ترتيب اللوحات وفقاً لقيمتها الجمالية كما يدركها.

(٢) اختبارات جريفرز للحكم على التصميمات الفنية

؛ Graves Design Judgment Tests

أعد جريفز Graves هذا الاختبار عام ١٩٤٨ وأعيد طباعته عام ١٩٥١ . ويختلف عن اختبار ماير Meier في أنه لا يقدم للفرد بدائل أعمال فنية شهيرة ولكنه يعتمد على التصميمات المجردة ، حيث يتضمن على ٩ مجموعة من التصميمات المجردة مما يجعل الاختبار لا يعتمد على القيم الفنية المترافق عليها . وتكون كل مفردة من نسختين أو ثلاثة من التصميم نفسه ولكنها تختلف مبادئ الفن الذوقى المقبول من حيث الوحدة والتوازن والتماثل أو غير ذلك من مظاهر الذوق الفنى ، ويطلب من الفرد أن يقرر أيها أفضل . والاختبار غير موقوت ، ولكن تطبيقه يستغرق نصف الساعة تقريباً . ويوضح شكل (٧-١٠) التالي بعض مفردات هذا الاختبار (Nunnally , 1965) :



شكل (٧-١٠) يوضح بعض مفردات اختبار الحكم على التصميمات الفنية

وقد تراوحت قيم معامل ثبات درجات الاختبار بطريقة التجزئة النصفية بين ٨٠، ٩٣، ٠٠ ، ويطلب الاختبار مزيداً من دراسات الصدق التنبؤى.

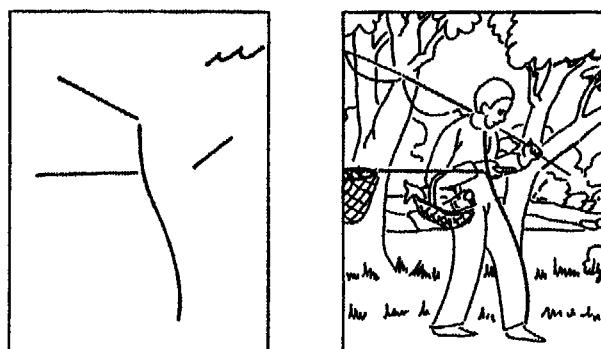
ثانية: اختبارات الاستعداد المتعلق بالإنتاج الفنى :

تعتمد معظم اختبارات الاستعداد المتعلق بالإنتاج الفنى على عينات من الأعمال الفنية مما يتطلب العناية بانتقاء هذه العينات ، وتأثير درجاتها بالتدريب المتظم ، ويوجد عدد محدود من هذه الاختبارات لعل أهمها:

قائمة هورن للاستعداد الفني :Horn Art Aptitude Inventory

أعد هورن Horn هذه القائمة التي تشتمل على عينات من الأعمال الفنية ، وت تكون من جزئين : أحدهما يتطلب من الفرد عمل رسم تصويري لعشرين شيئاً مائلواً مختلفاً مثل شجرة أو كتاب أو سكين ، على أن يتم العمل في مدة قصيرة لا تزيد عن عشر ثوان ، والآخر يتطلب التصور ، ويشتمل على 12 مستطيلًا يتضمن كل منها عدة خطوط يسترشد بها الفرد في رسم صورة مماثلة .

ويوضح شكل (١٠ - ٨) التالي أمثلة لبعض مفردات الجزء المتعلق بالتصور في هذه القائمة (Anastasi , 1976) :



شكل (١٠ - ٨) يوضح بعض مفردات قائمة الاستعداد الفني

وقد أشار دليل القائمة إلى أمثلة للرسم الممتاز والمتوسط والضعف ، وقدم بعض محكمات تقدير الدرجات . وقد وُجد أن معامل ارتباط درجات القائمة ومتوسط

تقديرات معلمى المجالات الفنية لعينة صغيرة من طلاب المرحلة الثانوية ٣٥ ، ، غير أن ثبات درجات القائمة وصدقها يحتاج إلى مزيد من الدراسات .

الابتكارية :

شهدت العقود التي تلت الحرب العالمية الثانية تطورات ملحوظة في كثير من المجالات المهنية العليا مثل المجالات التقنية، ومجالات البحث العلمي، والمجالات القيادية التنفيذية، ومجالات الفنون مما أدى إلى تنشيط بحوث القياس النفسي والتربوي لمواجهة المشكلات الجديدة الملحة التي نجمت عن هذه التطورات.

على الرغم من أن اختبارات الذكاء والاستعدادات الخاصة التي تناولناها فيما سبق يمكن باستخدامها التنبؤ إلى حد ما بأداء الأفراد في هذه المجالات المهمة، إلا أن نطاق هذه الاختبارات لم يمتد بحيث يشمل القدرات الأكثر أهمية المتعلقة بالتفكير الابتكاري Creative Thinking وما يتضمنه من أنشطة عقلية غير مألوفة. فاختبارات الذكاء التقليدية تتطلب من الفرد تطبيق ما تعلمه في موقف جديد، أو استنباط قاعدة معينة من أمثلة معطاة لفظية أو عددية أو شكلية، ويعيد تطبيق هذه القاعدة. وعادة تكون هناك إجابة واحدة لكل مفردة من مفردات الاختبار، وتحسب الدرجة الكلية في الاختبار بإيجاد مجموع درجات الإجابات الصحيحة.

لذلك فإن إجابة الفرد عن هذه الاختبارات تتطلب تفكيراً تقاريرياً محدداً Convergent Thinking. فمثلاً إذا طلب من الفرد أن يُحدد العدد التالي في السلسلة (٢ ، ٤ ، ٨ ، ١٦) فإن عليه أن يدرك أن كل عدد منها ضعف الذي يسبقه، وبذلك يستنتج أن العدد التالي (٣٢) ، غير أن الفرد ربما يتبع نمط تفكير مختلف، إذ ربما يلاحظ أن خانة العشرات في الأرقام الثلاثة الأولى (صفر) ، حيث يمكن كتابتها (٠٢ ، ٠٤ ، ٠٨) على الترتيب. وبعد إجراء هذا التحويل على المفردة الاختبارية، فإن الأعداد الثلاثة التالية يمكن أن تتبع قاعدة مماثلة لذلك .. ونظراً لأن خانة العشرات واحد صحيح في العدد التالي (١٦)، فإن العدد الذي يكمل السلسلة يمكن أن يكون (١٢)، وذلك لأن الرقم الأول (٢) هو رقم الآحاد في العدد (٣٢). وهذا النمط من التفكير على الرغم من صحته واختلافه عن نمط التفكير السابق، إلا أن الفرد يحصل على الدرجة صفر في هذه المفردة، ويعيد هذا النمط تفكيراً تباعدياً أو منطقاً Divergent Thinking. أي أنه في إطار اختبارات الذكاء التقليدية تُعد هذه الإجابة خطأ لأنها إجابة غير مألوفة، وذلك لأن مفردات هذه الاختبارات تُعد وحدات متكاملة في ذاتها،

ولا يجب على الفرد الذي يجب عنها إجراء أي تعديلات أو تحويلات عليه، وإنما يستخدم المنطق المألف.

ومثل هذه الاختبارات تناسب متطلبات النظم التربوية التقليدية التي تهتم اهتماماً أساسياً بالتفكير التقاري أو النمطي، ويمكن استخدامها في التنبؤ بأداء الطلاب في المجالات الدراسية التي تعتمد على قدرات الذاكرة. غير أن التوجهات التربوية المعاصرة تؤكد أهمية تنمية استراتيجيات التفكير التي تشجع الطلاب على إنتاج معلومات جديدة، وتوظيف إمكاناتهم العقلية في مواقف ثرية متعددة ، مما يتطلب بناء اختبارات من نوع مختلف لقياس التفكير التباعي وتقييم نواتجه .

الابتكارية ومفهوم التفكير التباعي :

تُعد الابتكارية Creativity من المفاهيم السيكولوجية التي نالت مزيداً من اهتمام علماء النفس والتربية في الآونة الأخيرة، غير أن هناك تبايناً ملحوظاً في تحديد وتعريف مفهوم الابتكارية نظراً لتعدد وجهات النظر حوله. إذ يرى بعض هؤلاء العلماء أن الابتكارية تتضمن عملية ممتدة عبر الزمن وتميز بالأصالة Originality ، والتوافق Adaptiveness ، والتحقيق Realization . وهذا يعني أنه ينبغي دراسة الأفراد الذين تميزوا بمستوى مرتفع في العمل الابتكاري. ويرى بعض آخر أن الابتكارية هي القدرة على كشف شيء جديد، بينما يرى آخرون أن الابتكارية ليست قدرة ، وإنما عملية أو عمليات سيكولوجية تسهم في تشكيل إنتاج يتميز بالجدة والقيمة ، في حين لا يركز البعض منهم في مناقشته لمفهوم الابتكارية على العملية وإنما على الإنتاج ذاته .

أى أن هذه التعريفات تتوارج بين النظر إلى الابتكارية على أنها مجرد حل للمشكلات يتميز بعنصر الجدة ، والنظر إليها على أنها التحقيق الكامل والتعبير عن جميع طاقات الفرد المتميزة .

ويرى ماكينون Mackinnon أنه ليس من المطلوب التوصل إلى أفضل تعريف من بين هذه التعريفات المتباعدة نظراً لأن الابتكارية تحمل جميع المعانى التى أشارت إليها وربما أكثر من ذلك. لذلك يفضل النظر إلى الابتكارية على أنها ظاهرة متعددة الأوجه بدلاً من اعتبارها مفهوماً نظرياً يلزم تعريفه دقيقاً، مما يعني أن الابتكارية Creativeness يمكن أن تضم عدداً من الأوجه المتراكبة . وأحد ميزات هذا التوجه وبخاصة في دراسة الابتكارية من منظور متسع، هو أنه يمكن تقسيم هذه الظاهرة التي تتسم بالتعقد إلى أوجه أو جوانب متمايزة يمكن التعامل معها بدرجة أفضل من مفهوم الابتكارية بعامة .

وفي إطار هذا التوجه توجد أربعة أوجه للابتكارية هي: الابتكارية كإنتاج، والابتكارية كعملية، والشخص المبتكر، والموقف الابتكاري.

وهذه الأوجه الأربع تتعلق بمجموعة من التساؤلات كالتالي:

(١) ما الإنتاج المبتكر؟ وما الخصائص التي تميزه عن غيره من الإنتاج المألوف؟

(٢) ما طبيعة العمليات الابتكارية؟ وما طبيعة ونوع العمليات السيكولوجية التي تسهم في الإنتاج المبتكر؟

(٣) ما السمات والخصائص التي تميز الأشخاص المبتكرين عن غيرهم من الأشخاص الأقل ابتكاراً؟

(٤) ما خصائص الموقف الابتكاري والظروف الاجتماعية والثقافية ومناخ العمل الذي يُسر ويُشجع على استثارة الفكر والعمل الابتكاري؟

وعلى الرغم من أهمية دراسة هذه الأوجه المتعددة للابتكارية، إلا أن ما يعنينا هنا هو القدرات أو الوظائف العقلية المعرفية وما يتعلّق بها من عمليات التفكير الابتكاري التي تسهم في إمكانية التوصل إلى إنتاج ابتكاري.

ولعل الدراسات الشريرة المتعلقة بالاستعدادات الابتكارية التي أجرتها جيلفورد Guilford ومعاونه في مشروع بحث الاستعدادات Aptitude Research Project (Guilford; 1967; Guilford & Hoepfner, 1971 ، دراسات Torrance وAssociates 1973) ، ودراسات توارنس ومعاونيه بجامعة مينيسوتا ، (Torrance et al. 1971) تعد من الدراسات الرائدة في هذا المجال. وقد أسهمت دراسات وبحوث جيلفورد ومعاونه في اقتراح نموذج بنية العقل Structure of Intellect (SOI) الذي أوضحته بالتفصيل في الفصل التاسع، حيث أشرنا فيه إلى الأبعاد أو الأوجه الثلاثة المتعلقة بالمحتوى والعمليات والتواتج التي تم التوصل إليها باستخدام منهج التحليل العاملی .

ومن بين هذه العمليات الرئيسة التفكير التقاري أو النمطي Convergent Thinking ، والتفكير التباعدي أو المنطلق Divergent Thinking . وقد سبق أن أشرنا أن محدودية اختبارات الذكاء وأوجه قصور مفهوم نسبة الذكاء IQ في النظر إلى القدرة العقلية أدى إلى زيادة الاهتمام بالتفكير التباعدي. وقد حاولت بعض الدراسات تعرف ما إذا كانت اختبارات هذا النمط من التفكير تقيس وظائف عقلية متفردة أم لا، أي ما إذا كان الذكاء العام يفسر الأداء في كل من نمطي التفكير، أم أن هناك عاملات خاصة يمكن أن يفسر ارتباط اختبارات التفكير التباعدي فيما بينها، غير أن كروبلی

(Cropley, 1966) قد اقترح في دراسته أن أفضل حل وسط بين طرفى النقاش هو أن الوظائف العقلية المتعلقة بنمط التفكير التباعدى ترتبط ارتباطاً وثيقاً بالوظائف العقلية المتعلقة بنمط التفكير التقاري. ودلل على ذلك بأن الأفراد الذين يستطيعون التوصل إلى حلول للمشكلات بتفكير منطلق يحصلون في غالبية الأحيان على درجات مرتفعة في اختبارات الذكاء التقليدية، ولكن أداؤهم في اختبارات التفكير التباعدى ربما يكون مرتفعاً أيضاً أو ربما لا يكون كذلك. بل إن بعض من يتميزون بتفكير تباعدى مرتفع ربما يكون أداؤهم في اختبارات الذكاء التقليدية منخفضاً. وقد استنتج كروبلى Cropley أن بعض الأفراد من ذوى القدرة العقلية المرتفعة ربما يفضل توظيف أحد نمطى التفكير، بينما يفضل بعضهم النمط الآخر.

قياس قدرات التفكير الابتكارى :

يعد كل من جيلفورد Guilford وتورانس Torrance من رواد حركة قياس قدرات التفكير الابتكارى، وقد أعدا اختبارات تقييم قدرات محددة تتعلق بهذا النمط من التفكير هى : **الطلاقة Fluency** ، والمرونة **Flexibility**، والأصالة **Originality**، وإدراك التفاصيل **Elaboration**، سوف نوضحها فيما يلى :

الطلاقة : هي القدرة على إنتاج عدد كبير من الأفكار والكلمات وأساليب التعبير عن الأشياء، أي طلاقة الأفكار **Ideational Fluency**، وطلاقة التداعى **Associational Fluency**، والطلاقة التعبيرية **Expressional Fluency**، فمثلاً يقدم للفرد كلمة أو فكرة ويطلب منه التفكير في أكبر عدد ممكن من الاستخدامات غير المألوفة للكلمة أو الشيء المعطى، أو مرادفاتها، أو تكوين جمل ذات خصائص معينة. وتقدر نقطة واحدة لكل فكرة جديدة، فكلما زاد عدد الأفكار التي يُتجهها الفرد حصل على درجة أعلى في القدرة على الطلاقة.

المرونة : هي القدرة على إنتاج أفكار تُبيّن انتقال الفرد من مستوى تفكير إلى مستوى آخر، أو تحول تفكير الفرد بالنسبة لمهمة معينة، ويقصد بذلك عدم افتقار الفرد على خط فكري واحد، أي مرونة التفكير. ونظرًا لاختلاف محتوى المثير الذي يقدم للفرد ونوع ناتج التفكير فإن المرونة تعتمد على هذا المحتوى، فإذا كان المحتوى يتعلق بالمعنى اللغوية فإنه يُطلق عليها المرونة التلقائية **Spontaneous Flexibility**، وإذا كان يتعلق بالأشكال فإنه يطلق عليها المرونة التوازمية **Adaptive Flexibility**. وتقدر نقطة واحدة لكل تحول في التفكير، ويشير مجموع النقاط إلى القدرة على المرونة.

الأصالة : هي القدرة على إنتاج أفكار نادرة بالمفهوم الإحصائي للندرة، أي أفكار غير مألوفة وتباعدية ومميزة. وإحدى الطرق الشائعة الاستخدام في تقدير

الدرجات هي إعطاء نقاط للأفكار التي يُتجهها خمسة أفراد من بين مائة فرد، وكذلك تقدر الدرجات على ميزان متدرج يشتمل على خمس نقاط، حيث تُقدر نقطة واحدة للأفكار التي ربما تكون معروفة ، وتدرج إلى خمس نقاط للأفكار التي تميز بدرجة عالية من السجدة. والأفكار نفسها التي يُتجهها أكثر من خمسة أفراد من بين مائة فرد يقدر لها صفرًا، ويشير المجموع الكلي للنقاط إلى الدرجة في هذه القدرة.

إدراك التفاصيل؛ هي القدرة على إضافة تفاصيل إلى فكرة رئيسية يُتجهها الفرد، ومثال ذلك أن يقدّم للفرد عدد من المربعات، ويُطلب منه أن يتوصّل إلى أفكار لا يستطيع أحد غيره أن يُتجهها، وذلك برسم صور باستخدام المربع المعطى، كأن يرسم باباً باستخدام المربع ويكتب عنواناً للصورة. وإذا قرر فرد أن يضيف تفاصيل أخرى إلى صورة الباب، مثل بعض الزخارفات، أو فتحة للرؤبة من الداخل، وهكذا، فإنه بذلك يضيف تفاصيلاً على الفكرة الرئيسية وهي الباب.

وتقدر نقطة واحدة لكل عنصر من العناصر التي أضيفت إلى الرسم، ويشير المجموع الكلي للنقاط إلى القدرة على إدراك التفاصيل.

وعلى الرغم من أن كلاً من جيلفورد Guilford وتوارنس Torrance استند إلى هذه القدرات الأربع في تقييم الابتكارية ، إلا أن جيلفورد حاول قياس كل منها بطريقة تتطلب من الفرد أداء مهام متعددة في ضوء المحتوى الشكلي بنوعيه، والرمزي، والسيمانتي، والسلوكي، ونماذج الوحدات، والفضاءات وال العلاقات، والنظم، والتحوليات، والتضميدات، أي في ضوء $5 \times 6 = 30$ قدرة من قدرات التفكير الباعدي كما سبق أن أوضحنا في شكل (٤ - ٩) .

ولكن حاول توارنس قياس القدرات الأربع الرئيسية نفسها بطريقة تتطلب من الفرد أداء عدد من المهام المعقدة التي تم تصميمها بحيث يوظف جميع هذه القدرات في آن واحد.

وقد أضاف توارنس مؤخراً عدداً آخر من القدرات، هي القدرة على التوليف Synthesis، والقدرة على التجريد Abstraction كما يتمثلان في إنتاج عناوين مميزة للرسوم التي يُتجهها الفرد، ودمج شكلين أو أكثر في صورة متكاملة متعلقة بها، وكذلك القدرة على الغلق Closure المتمثلة في تأجيل إتمام مهمة معينة لمدة زمنية تسمح بالتأمل من أجل إمكانية التوصل إلى إنتاج أفكار تميز بالأصالة . وفيما يلى توضيح لكل من هذين الاختبارين وغيرهما.

اختبارات التفكير الابتكاري :

(١) بطارية اختبارات الإنتاج التباعي لجليفورد

: Guilford Tests of Divergent Production

أعد جليفورد هذه البطارية في إطار برنامجه البحثي ونموذج بنية العقل الذي سبق أن أشرنا إليه، وتقيس قدرات محددة للتفكير التباعي، وتناسب بدرجة أفضل طلاب المرحلة الثانوية. ويمكن تصنيف اختبارات هذه البطارية على أنها اختبارات الابتكارия Creativity Tests. ومن أمثلة مفردات هذه الاختبارات: طلاقة الكلمات التي تتطلب سرعة كتابة قائمة بكلمات تشمل على حرف معين، والطلاقة التعبيرية التي تتطلب تسمية أشياء أو موضوعات تتمي إلى فئة معينة، وطلقة التداعي التي تتطلب إنتاج متزدفات، والاستخدامات المتعددة التي تتطلب كتابة قائمة بما يحتمل أن يترتب على فعل معين، وتتطلب جميع هذه الاختبارات استجابات لفظية. غير أن البطارية تشمل أيضاً على اختبارات تستخدم الأشكال، مثل اختبارات تكوين الأشياء الذي يتطلب رسم أشياء معينة استرشاداً بمجموعة محددة من التصنيمات، واختبارات مشكلات تتعلق بأعواد الكبريت، وتعتمد على تقديم تصميم معين بأعواد الكبريت، ويطلب من الفرد تحريك عدد معين من الأعواد لإنتاج تصميم آخر.

(٢) بطارية اختبارات تورانس للتفكير الابتكاري

: Torrance Tests of Creative Thinking

على الرغم من أن هذه البطارية تشبه بطارية جليفورد، إلا أن تورانس أعدها لتكون ملائمة للواقع التربوي الفعلى، ويفترض أنها تصلح للتطبيق من الحضانة إلى المراحل الأعلى. وقد عرف تورانس الابتكارية بأنها عملية استشعار الفجوات أو العناصر المفتقدة، وتكون فروض تتعلق بها، واختبار صحة هذه الفروض، وتقرير النتائج ، وربما أيضاً إجراء تعديلات على الفروض وإعادة التحقق من صحتها . (Torrance , 1962).

كما بينَ أنه يمكن تعريف الابتكارية من حيث الأصلة على أنها القدرة على التصور والابتعاد عن النمطية في الإدراك للتمكن من إعادة تكوين الصور الذهنية والأفكار المشاعر في روابط تتميز بالجدة (Khatena & Torrance , 1973). وتنقسم البطارية إلى قسمين رئيسين هما: التفكير الابتكاري بالكلمات، والتفكير الابتكاري بالصور ، وقد أضيف فيما بعد التفكير الابتكاري بالأصوات والكلمات.

وتم تصميم المفردات على شكل ألعاب أو صور. ويتضمن الجزء الخاص بالتفكير الابتكاري بالكلمات أنشطة متنوعة، كأن يعطي الفرد رسمًا تخطيطياً ويطلب

منه كتابة قائمة بالتساؤلات التي يراها تتعلق بما يجري فيما يمثله الرسم، أو تخمين الأسباب المحتملة لما يجرى، وكذلك تخمين ما يمكن أن يترب على ما يجرى . ومن الأنشطة الأخرى اقتراح تحسينات يمكن إدخالها على لعبة ، واقتراح استخدامات غير مألوفة لشيء مألوف ، واقتراح تساؤلات حول هذا الشيء ، وتخمين ما يترب على فعل غير محتمل حدوثه .

أما الجزء الخاص بالتفكير الابتكاري بالصور فيشتمل على ثلاثة اختبارات تتعلق باستخدام شكل ملون في تكوين صورة تميّز بالأصالة ، واقتراح عنوان لها ، ورسم أشياء باستخدام أزواج من الخطوط غير المنتظمة المعطاة ، وإنشاء تصميمات باستخدام أزواج من الخطوط المتوازية أو الدوائر .

والجزء الخاص بالأصوات والكلمات يشتمل على اختبارين يُطبّقان باستخدام أجهزة تسجيل حيث تقدّم التعليمات والمثيرات ، ويتعلّق أحد الاختبارين بالأصوات والصور ، والآخر بالأصوات التي تحاكي الكلمات المسموعة والصور ، ويستجيب الفرد لمفردات كل من الاختبارين لفظياً بعد سماعه كل مثير صوتي حيث يعبرّ عما يقترحه أو يتصوره بشأن هذه المثيرات .

وقد فضّل تورانس تسمية اختباراته «أنشطة» لكي يستبعد عوامل الخوف والقلق من الموقف الاختباري .

ويتم تقدير الدرجات وفقاً للمحکات التي سبق أن أوضحتها . وت تكون البطارية من صيغتين زمن التطبيق الفعلى لأى منها محدد وقدره ٤٥ دقيقة للجزء اللفظي ، ٣٠ دقيقة للجزء غير اللفظي .. وقد وجد أن معامل ثبات تقديرات اختبارات كل من صيغتي البطارية تراوحت قيمته بين ٥٠ ، ٥٠ ، ٩٣ ، . . ويشتمل دليل البطارية على أمثلة متنوعة تتعلق بكيفية تقدير درجات الاختبار والتدريب على ذلك ، كما تشتمل على معاير الاختبار لفرق الدراسية المختلفة المستمدّة من مختلف الأنظمة المدرسية .

وقد أجريت كثير من الدراسات حول هذه البطارية ، وبرزت تساؤلات متعددة تتعلق بما تقيسه الاختبارات ، وعلاقتها بالتحصيل الدراسي ، والذكاء ، ومحکات الإنجاز الابتكاري ، مما يتطلب مزيداً من البحث .

وعلى الرغم من ذلك فإن البطارية تعد جهداً رائداً في مجال بناء وتقنين اختبارات الابتكار .

(٣) اختبار التداعي المطلق : Remote Associates Test

أعد هذا الاختبار مدنick Mednick ، ومدنick Mednick (١٩٦٧) . ويقيس الطلقة في إنتاج تداعيات مألوفة لكنها صعبة الاستدعاء ، وفحص هذه التداعيات ، والانتقال

إلى تداعيات جديدة إلى أن يتوصل الفرد إلى أحد التداعيات المقبولة . فمثلاً يقدم للفرد مجموعات يشتمل كل منها على ثلاث كلمات مثل (مفاجأة ، خط ، عيد ميلاد) ويُطلب من الفرد التفكير في الكلمة رابعة تُستخدم في الحياة اليومية وترتبط بكل مجموعة من هذه الكلمات ، وتتطلب الإجابة التفكير في اتجاهات متعددة قبل أن يكتشف الفرد الكلمة التي في ذهن مؤلفي الاختبار .

ويبدو أن الاختبار يتطلب استجابات مألوفة أى لا تميز بالجدة ، غير أنه يفترض أن الأفراد الأكثر ابتكاريه يكون انتاجهم للتداعيات المتعلقة بكلمات شائعة ضعيفاً ، أما إذا كانت الكلمات غير شائعة ، فإن انتاجهم للتداعيات يكون أقوى ، أى أنه يمكنهم إنتاج تداعيات غير شائعة نسبياً .

وقد أوضحت الدراسات المتعددة التي تناولت العلاقة بين الاختبار ومحكمات مختلفة بالابتكاريه أن قيم معامل الارتباط تراوحت بين ٥٥ ، ٨٤ ، ٠ ، مما يشير إلى إمكانية الإفاده من الاختبار في التنبؤ بالابتكاريه .

ومن الجدير بالذكر أنه توجد اختبارات تقيس الابتكاريه في مجالات خاصة محددة ، مثل اختبار القدرة الابتكاريه AC Test of Creative Ability الذي أعدته شركة جنرال موتورز الأمريكية لقياس الابتكاريه لدى المهندسين والمسرفيين على العمل في قسم كهرباء السيارات بالشركة ، واختبار الابتكاريه في تصميم الآلات Owens Creativity Test for Machine Design الذي أعده أونز Owens عام ١٩٦٠ ، وهو اختبار أكثر صعوبة من الاختبار السابق ، ومحظوه يتناول أموراً فيه تتعلق بالآلات ، ويُطبق على المهندسين . هذا بالإضافة إلى الاختبارات التي تقيس الابتكاريه في المجالات الفنية والموسيقيه .

بعض المشكلات المتعلقة بقياس الابتكاريه :

اتضح لنا مما سبق أن مجال قياس الابتكاريه لا يزال في مرحلة المهد إذا ما قورن بمجال قياس الذكاء وقدرات التفكير التقاري حيث ترسخت كثيراً من الأسس والمبادئ التي يستند إليها المجال الأخير .

لذلك نجد أن قياس الابتكاريه محاط بالعديد من المشكلات المنهجيه والسيكومترية التي تستثير فكر علماء النفس والتربية وخبراء القياس في محاولة التغلب عليها ، وتطوير أساليب القياس في هذا المجال الذي يزداد أهمية في عالمنا المعاصر .

فالكشف عن الطاقات العقلية المُبدعة واستثمارها لخير الإنسان ومواجهتها التحديات في عصر دائم التغير والانفتاح يُعد ضرورة ملحة، وسوف نوضح فيما يلي بإيجاز بعض أهم هذه المشكلات:

مدى عمومية سمة الابتكارية،

تختلف اختبارات الابتكارية فيما بينها من حيث المحتوى، فبعضها يكون محتواه لفظياً أو غير لفظي مثل الصور والأشكال المجردة والآصوات، وبعضها يجمع بين هذه المحتويات كما في بطارية اختبارات تورانس. غير أن هناك اختبارات تقيس الابتكارية في مجالات محددة كالهندسة والفنون، والموسيقى، والتأليف، والبحث العلمي، ويكون محتواها أكثر تقييداً من الاختبارات التي تقيس الابتكارية بعامة.

وهذه الاختبارات النوعية لا يستطيع الإجابة عنها إلا المهتمون ب المجالاتها ولديهم خبرة وتدريب في هذه المجالات. وهنا يبرز تساؤل حول ما إذا كانت الابتكارية سمة نوعية بدرجة كبيرة أم أنها سمة عامة. فإذا كانت هذه السمة نوعية، فإن المهندس أو الفنان أو المؤلف المبتكر ربما لا يكون كذلك في مجالات أو مواقف تختلف عن مجال اهتمامه. أما إذا كانت السمة عامة فإن كلا من هؤلاء المبتكرين في مجالات نوعية محددة ربما تظهر بعض علامات ابتكارتهم في مجالات أو مواقف مختلفة.

غير أن فاعليتهم في ابتكار أفكار جديدة في هذه المجالات الأخرى سوف تكون محدودة نظراً لقلة تدريبهم في هذه المجالات. ومع هذا فإن نزعتهم الابتكارية يمكن قياسها بالاختبارات نفسها التي تستخدم في قياس الابتكارية في المجالات الأخرى.

لذلك فإن المشكلة التي تُعد مجالاً من مجالات البحث هي ما إذا كانت هناك بعض الخصائص المشتركة بين المبتكرين بغض النظر عن مجال اهتمامهم، وما إذا كانت هناك بعض الخصائص المتعلقة فقط بمجالات نوعية محددة.

علاقة اختبارات الابتكارية باختبارات الذكاء،

نالت هذه المشكلة الأساسية اهتماماً كبيراً من جانب علماء النفس والقياس، وقد أبرز ثورنديك (Thorndike, 1963)، ووالش وكوجان (Walsh & Kogan, 1965، 1962) وجيتزلس وجاكسون (Getzels & Jackson) هذه المشكلة في التساؤل التالي: هل اختبارات الابتكارية والذكاء تقيس مجالات أو نطاقات معرفية مختلفة؟ فلقد بينت نتائج دراسات والش Walsh وكوجان Kogan أن الأدلة لا تؤكد الرسم بأن اختبارات الابتكارية الشائعة الاستخدام تقيس نطاقاً موحداً من الوظائف المعرفية يختلف عن النطاق الذي تقيسه اختبارات الذكاء. وقد استند كل منهما في استنتاجه إلى أدلة ثلاثة:

(أ) الارتباطات بين الاختبارات الفرعية التي تشمل عليها مقاييس الذكاء كانت مرتفعة بعامة.

(ب) الارتباطات بين الاختبارات المختلفة التي تقيس الابتكارية كانت منخفضة بعامة.

(ج) الارتباطات بين اختبارات الذكاء واختبارات الابتكارية كانت مساوية تقريباً في مقدارها للارتباطات بين الاختبارات المختلفة للابتكارية ذاتها.

وقد أدى ذلك إلى استنتاج أن ما تقيسه اختبارات الابتكارية لا يعكس شيئاً أكثر مما تقيسه بالفعل اختبارات الذكاء، وأرجع الباحثان ذلك إلى قصور في بناء وتطبيق اختبارات الابتكارية.

ولمعالجة هذا القصور أجرى الباحثان دراسة على ١٥١ طفلاً من الصف الخامس الابتدائي ، واعتمدا في تقييم السلوك الابتكاري على أسلوب يسمح للأطفال بوقت كاف للإجابة عن مفردات الاختبار التي قدمت على شكل ألعاب بتعليمات تقلل من التناقض فيما بينهم ، وبذلك تصبح الظروف مناسبة لما يتطلبه السلوك الابتكاري . وقد وجدا أن متوسط قيم الارتباطات بين درجات كل من اختبارات الابتكارية والذكاء مساوياً ٩ ، ، ، ومتوسط قيم الارتباطات بين الاختبارات المختلفة للابتكارية ٤١ ، ، ، وبين الاختبارات المختلفة للذكاء ٥١ ، ، . وهذا ربما يكون دليلاً على أن هناك بعدين مستقلين للنطاق المعرفي الذي تقيسه كل من اختبارات الابتكارية واختبارات الذكاء .

لذلك اقترح الباحثان ضرورة الاعتماد في بناء اختبارات الابتكارية على نموذج يختلف عن نموذج بناء اختبارات الذكاء، أى يعتمد على فعل الابتكار ذاته كما يحدث بعيداً عن المواقف الاختبارية .

ويؤخذ على هذه الدراسة وغيرها من الدراسات التي تناولت العلاقة بين اختبارات الابتكارية واختبارات الذكاء اعتمادها على عينات من الأفراد تميز بذكاء يفوق المتوسط مما يجعل مدى الدرجات ضيقاً ، وبذلك تنخفض قيمة معامل الارتباط ، أما إذا اشتملت العينة على مدى متسع للذكاء فسوف يكون هناك ارتباط بين نوعي الاختبارات . غير أن قيمة الارتباط تتباين تبايناً ملحوظاً وفقاً للاحتجارات المستخدمة ، وطريقة تطبيقها ، وطبيعة عينة الأفراد المستخدمة .

ففى إحدى الدراسات طبق كروبل (Cropley 1966) اختبارات ابتكارية واختبارات ذكاء على عينة عشوائية من الأطفال ، ووجد أن الارتباط بين درجات نوعي الاختبارات ٥١ ، ، مما جعله يستنتج أن القدرات المتعلقة بكل من التفكير التقاري والتفكير التبعاعي متمايزة عالمياً ، ولكنها ليست مستقلة .

وقد توصلت دراسة جيتزلز Getzels ، وجاكسون Jackson إلى أن الذكاء فوق المتوسط ضروري للابتكارية، ولكن العكس غير صحيح، أي أن الذكاء المرتفع نسبياً ضروري للابتكارية ، ولكنه ليس كافياً .
وعومماً، فإن هذه المشكلة تعد ميداناً خصباً لمزيد من البحوث والدراسات استناداً إلى منهجيات وأساليب إحصائية مختلفة.

صدق اختبارات الابتكارية :

تُعد مشكلة صدق اختبارات الابتكارية من المشكلات الأساسية أيضاً، حيث إن رصيد المعلومات المتوافر حول هذه المشكلة محدوداً، فالدراسات التي تناولت السمات السيكولوجية المتعلقة بالابتكارية، وأساليب المناسبة للتنبؤ بها في مواقف الحياة الفعلية كانت نتائجها متعارضة.

فعند تقدير صدق اختبارات الابتكارية تُستخدم في معظم الأحيان تقديرات المعلمين كمحك على الرغم مما يشوبه من عيوب. فهذا المحك ربما يكون أكثر ملاءمة في مجال التحصيل الدراسي أو السلوك القيادي، ولكنه لا يناسب مجال الابتكارية، وكذلك تُستخدم أحكام الأقران كمحك لتقييم صدق هذه الاختبارات على الرغم من تأثير هذه الأحكام بمتغيرات متعددة مما يجعلها قليلة الفائدة (Torrance, 1959).

والحقيقة أن تحديد محكمات لتقدير صدق اختبارات الابتكارية أكثر صعوبة من تحديد محكمات لتقدير صدق اختبارات الذكاء. فالإنتاج الابتكاري يتميز بالندرة ولا يسمح بسهولة تصنيفه على ميزان منتظم، ومع هذا فإن الدراسات التي استخدمت اختبارات التفكير التباعدي توصلت إلى نتائج متباعدة ولكنها تشير إلى انخفاض قيم معامل الصدق التنبؤى لهذه الاختبارات (Wallach & Wing 1969).

وعلى الرغم من أن الأساليب التي اتبعها جيلفورد في بناء بطارية اختبارات الابتكارية تكفل الصدق العاملى والصدق التمايزى لهذه البطاريه، إلا أن الارتباط بين اختباراته ومحكمات مرتبطة بمواصفات الحياة الفعلية كان منخفضاً. كما أن نمط الارتباط بين درجات اختبارات تورنس ومحكمات خارجية متعددة لم يكن واضحاً.

لذلك فإن بطاريات اختبارات الابتكارية المتوافرة يمكن استخدامها لأغراض البحث، ولكنها لم تصل في تطورها إلى مرحلة تسمح بتفسير درجات اختباراتها تفسيراً يمكن الإفاده منه في الواقع الفعلى، مما يتطلب إجراء دراسات صدق شاملة ومنظمة لتوسيع العلاقة بين اختبارات الابتكارية، ومقاييس القدرات العقلية الأخرى، ومحكمات خارجية واقعية.

ثبات درجات اختبارات الابتكارية:

بيَّنت كثير من الدراسات أن ثبات درجات اختبارات الابتكارى منخفضاً بدرجة واضحة، ولعل هذا يرجع إلى الصعوبات المتعلقة بتقدير درجات هذه الاختبارات، حيث إن مفرداتها تتطلب إجابات مفتوحة في وقت محدد. ومما يزيد من حدة هذه المشكلة انخفاض الصدق نظراً لصعوبة التتحقق مما إذا كانت هذه الاختبارات تقيس الابتكارية بالفعل، وذلك لأن ما يُطلّق عليه ابتكارى يختلف اختلافاً كبيراً من شخص إلى آخر، ومن ثقافة إلى أخرى. وهذا يدعو إلى أهمية التوصل إلى أساليب لتقدير درجات مفردات هذه الاختبارات تساعده في الحكم الموضوعي على مدى تميز الإجابات بالجدة، ولعل مفهوم الأصالة يُعد مدخلاً لتحقيق ذلك.

فلكي نحافظ على ميزات المفردات التي تتطلب إجابة مفتوحة وفي الوقت نفسه يمكن تصحيح الاختبار بالطرق الآلية، يمكن أن تكون الاستجابة لفظية، ويفضل أن تكون كلمة واحدة.

فمثلاً يُقدم للفرد فقرات تتعلق بمجال معين يحدّد في نهاية كل منها المطلوب منه، ويلى كل فقرة مفردة اختيار من متعدد تتضمن عدداً من البديلات التي يشتمل كل منها على بعض الحروف المترفة يفصل بينها عدد محدد من المسافات ، مثل (ه---ل م --أ) وأحد هذه البديلات التي تبدأ بحرف معين يكون صحيحاً من وجهة نظر مؤلف الاختبار.

وقد استخدم فلانagan (1963 ، Flanagan) هذا الأسلوب في أحد اختباراته ، كما استخدمته بعض الاختبارات الأخرى. غير أنه يُعبّر على هذا النوع من المفردات أن الفرد الذي يتمسّ بمهارة في استخدام الكلمات ربما يتوصّل إلى الحل الصحيح ، دون تفكير منطلق ، وهذا بالطبع ليس الهدف من الاختبار. كما أن الفرد المبتكر ربما يفكّر في بديل متميز يختلف عن البديل المعطاة مما يجعل الفرد يشعر بأن الاختبار يحدّ من تفكيره المنطلق. ولعل اختبار التداعي الطليق الذي سبق أن أشرنا إليه يُعد من هذا النوع إلى حد ما.

ومن هذا يتضح أن قياس وتقدير الابتكارى بقدر كبير من الدقة والمهارة كما هو مُتبع في قياس الذكاء يتطلب مزيداً من الدراسات الإمبريقية ، والسيكومترية من منظور علم النفس المعرفي من أجل التعرف على العمليات المعرفية التي تُسهم في الإنتاج الابتكاري وما يرتبط بذلك من متغيرات غير معرفية ، مثل الدافعية ، والخصائص الشخصية، والبيئة المترتبة ، والمدرسيّة ، وغير ذلك من المتغيرات التي يمكن أن تلقى مزيداً من الضوء على ظاهرة الابتكارى .

خلاصة

إن اختبارات الاستعدادات الخاصة تعد مقاييس معرفية تُستخدم في التنبؤ بنجاح الفرد قبل التحاقه ببرنامج تعليمي أو تدريسي معين، كما تُستخدم في أغراض الانتقاء والتسكين والإرشاد والتوجيه التربوي والمهني. وتشتمل الاستعدادات على تجمعات من القدرات الخاصة المتعلقة بها ويمكن قياسها. وتتعدد أنواع هذه الاختبارات نظراً لتنوع الاستعدادات الخاصة.

فهناك بطاريات متعددة العوامل تشتمل على اختبارات فرعية منفصلة لقياس القدرات المتعلقة بعدد من الاستعدادات الأكademية أو المهنية التي تهدف البطارية لقياسها. ويمكن رسم صفحة نفسية أو برو菲ل للدرجات الفرد في الاختبارات الفرعية لتوضيح جوانب قوته وضعفه، ولكن ينبغي الحفطة عند تفسير دلالة الفروق بين الدرجات في هذه الاختبارات.

وهذا يعتمد على ثبات درجات الاختبارات، ودرجة ارتباط كل منها بالآخر، حيث ينبغي أن يكون الثبات مرتفعاً، والارتباط فيما بينها منخفضاً وهناك اختبارات استعدادات نوعية يقيس كل منها استعداداً خاصاً واحداً، مثل الاستعداد الميكانيكي، والاستعدادات النفسية الحركية، والاستعداد الكتابي، والاستعداد الموسيقي، والاستعدادات الفنية، والابتكارية، واستعدادات أخرى تُستخدم لأغراض خاصة يتطلب بناؤها أساليب منهجية منظمة لتحليل العمل.

ويستند قياس الاستعدادات الخاصة إلى مجموعة من الافتراضات الأساسية التي ينبغي مراعاتها. وتختلف أساليب بناء اختبارات الاستعدادات الخاصة باختلاف نوع الاختبار وتكوينه والغرض من استخدامه. وينبغي أن تميز جميع هذه الاختبارات بصدق تنبؤ مرتفع. وللحتحقق من ذلك يجب العناية باختيار المحك المناسب الذي توافر فيه الشروط المطلوبة، وكذلك تحديد درجة أو درجات القطع استناداً إلى الأساليب المنطقية، أو الإميريقية. غير أن اختبارات الابتكارية تختلف عن غيرها من الاختبارات في أسلوب بناء مفرداتها وتقدير درجاتها، وصعوبة تحديد محك مناسب لتقدير صدقها، كما أن ثبات درجاتها يكون منخفضاً نسبياً؛ لذلك فإن هذا النوع من الاختبارات يتطلب مزيداً من البحث والتطوير بفكر ابتكاري.



تطبيقات القياس والتقويم

التربوي والنفسى

(ثانياً) قياس وتقدير الجوانب غير المعرفية

(الوجهانية والشخصية)



الفصل الحادى عشر

قياس الميول

* مقدمة

- * قياس الميول من منظور تاريخى
- * الميول وعلاقتها بالاستعدادات الخاصة
- * الأنماط المختلفة للميول
- * أساليب بناء استبيانات الميول
- * بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية
- * بعض المشكلات المتعلقة بالاستبيانات المقتنة للميول
- * مشكلات تفسير درجات استبيانات الميول
- * أساسيات بناء استبيانات الميول المهنية
- * قياس الميول باستخدام الاستبيانات المقتنة
- * تعقيب عام على استبيانات الميول
- * استخدامات استبيانات الميول
- * تقويم ميول الطلبة في الصف الدراسي
- * أدوات قياس وتقويم ميول الطلبة في الصف الدراسي

مقدمة :

انصبَّ اهتمامنا في الباب الثاني على أساليب وأدوات قياس وتقدير الجوانب المعرفية للشخصية المتعلقة بالتحصيل الأكاديمي والكتابات والذكاء والاستعدادات الخاصة متضمنة الابتكارية ، وجميعها تهتم بقياس الأداء الأقصى للفرد Maximum Performance ، أي أنها تُوضع ما يستطيع الفرد إنجازه في ظروف معينة ، وتُعد منبهات جيدة بالنجاح الأكاديمي والمهني .

غير أن هناك جوانب أخرى للسلوك الإنساني لا تقل أهمية عن الجوانب المعرفية ، وينبغى قياسها وتقديرها والإفادة من نتائجها في اتخاذ القرارات المتعلقة بالأفراد ، وتعلق بالجوانب غير المعرفية أو الوجدانية أو الشخصية . وهذه الجوانب غير المعرفية Non - Cognitive لا تفصل عن الجوانب المعرفية Cognitive نظراً لتكامل السلوك الإنساني وتأثيره بهذه الجوانب مجتمعة ، فإذا اقتصرنا على قياس قدرات واستعدادات الفرد العقلية أو المعرفية فإن هذا لا يعني أن الجوانب المعرفية تؤثر وحدها في أدائه ، وكذلك إذا اقتصرنا على قياس الجوانب غير المعرفية ، فإن هذا لا يعني أن أداء الفرد لا يتاثر بالجوانب المعرفية ، فأثر كل من هذه الجوانب ينعكس في أداء الفرد أو سلوكه لأنها تمثل أوجه متعددة للشخصية الإنسانية تتفاعل فيما بينها . وتهتم أدوات قياس الجوانب غير المعرفية بتقييم الأداء المميز للفرد Typical Performance .

وعلى الرغم من أن إسهام هذه الأدوات في التنبؤ بالنجاح الأكاديمي والمهني أقل مما تسهم به أدوات قياس الجانب المعرفي ، إلا أنها تفيد بدرجة أكبر في الإرشاد والتوجيه ، حيث تضيف معلومات مهمة عن الفرد إلى جانب ما تقدمه الأدوات السابقة . وتتعدد الخصائص والمتغيرات التي تتضمنها الجوانب غير المعرفية أو الوجدانية Affective ، حيث تشمل الانفعالات والعواطف ، والميول ، والاتجاهات ، والمعتقدات ، والقيم ، والدوافع ، والسمات المزاجية ، والحسن الاجتماعي ، وغير ذلك من سمات الشخصية الإنسانية .

غير أننا سوف نركز في هذا الباب على بعض هذه المتغيرات المهمة وهي الميول ، والاتجاهات ، والشخصية . وعلى الرغم من أن المفهوم العام للشخصية Personality يتضمن عادة جميع الخصائص أو المتغيرات السابقة ، إلا أننا سوف نميز الميول والاتجاهات عن سمات الشخصية والسمات المزاجية ، وقد حظت هذه المتغيرات باهتمام كبير من جانب علماء النفس والتربية ، وتناولتها كثير من الدراسات

بالبحث والتطوير. وسوف نقتصر في هذا الفصل على مناقشة مفهوم الميول وما يتعلّق به من أدوات قياس، وكيفية بناء هذه الأدوات واستخداماتها والمشكلات الخاصة بها، وتناول الاتجاهات، والشخصية في الفصلين التاليين.

قياس الميول من منظور تاريخي:

لعل أول محاولة للبحث في مجال الميول هي الدراسة التي أجرتها ثورنديك E. Thorndike ونشرت في المجلة الشهرية للعلوم العامة Popular Science Monthly عام ١٩١٢ وتعلق باستقرار الميول وعلاقتها بالقدرات. واعتمدت هذه الدراسة على عينة تشمل على ١٠٠ طالب بالكليات حيث طلب من كل منهم ترتيب ميوله كما يتذكرها منذ كان في المرحلة الابتدائية، والثانوية، وعند التحاقه بالكلية، وكذلك ترتيب قدرته في هذه المراحل، وأُوجِدَ ارتباطات بين مختلف الترتيبات، واستنتج أهمية الميول المبكرة وعلاقتها الواضحة بالقدرة مما يبيّن أن كلاً منها يدل على الآخر، كما استنتج أن هذه العلاقة تميز بالتعقد. والمحاولة المبكرة الثانية تتعلق باستبيان أعده كيلي Kelley عام ١٩١٤ اشتتمل على أسئلة تستوضح تفضيلات الأفراد الترويجية، غير أن هذا الاستبيان لم ينل حقه من اهتمام الباحثين.

وقد بدأت النشأة الفعلية لاستبيانات الميول كما نعرفها الآن في معهد كارنجي التقني بولاية بنسينغ الأمريكية بعد الحرب العالمية الأولى بمدة قصيرة . فقد عقد هذا المعهد حلقة بحث تتعلق بالميول ، وقام الأساتذة وطلاب الدراسات العليا ببناء ما يقرب من ألف فقرة للتمييز بين ميول العاملين في المهن المختلفة، وأسهمت هذه الحلقة إسهاماً كبيراً في تشجيع الباحثين على بناء استبيانات مقتنة للميول على أساس علمية منظمة، وأُعدت بالفعل كثيراً من هذه الاستبيانات. فقد قامت مجموعة من علماء النفس الصناعي بقيادة بنجهام Bingham بدراسة المشكلات النفسية التطبيقية بما في ذلك مشكلة قياس الميول.

وفي عام ١٩٣١ أصدر فراير Fryer كتاباً بعنوان «قياس الميول» لخاص فيه التطورات الماضية التي حدثت في هذا المجال ، حيث أشار فيه إلى محاولة مور Moore عام ١٩٢١ قياس الميول الميكانيكية والاجتماعية للمهندسين، ومحاولات ريم Ream عام ١٩٢٤ التمييز بين الناجحين وغير الناجحين من العاملين بالمهن التجارية في ضوء ميولهم، وبحوث فرايد Freyd عام ١٩٢٤ التي أُجريت على مقاييس مور Moore في محاولة التمييز بين الميول الميكانيكية والاجتماعية لمجموعات من عامة الناس .

وقد شهد العقد الثاني من القرن العشرين بداية بناء الاستبيانات المفتوحة للميول بإشراف سترونج Strong الذي وافق تيرمان Terman على تعيينه أستاذًا بجامعة ستانفورد عام ١٩٢٤ ، وكان تيرمان يجري بحوثه في ذلك الوقت على الطلاب من ذوي الذكاء المرتفع . ونظراً لتأثير سترونج بنتائج حلقة البحث التي أشرنا إليها فقد قام بإعداد استبيان للميول المهنية عام ١٩٢٦ ، وتمت طبعته الأولى عام ١٩٢٧ ، ودليله ووثائقه المرجعية عام ١٩٢٨ ، وصدرت أول معلومات عن تصحيح هذا الاستبيان آلياً عام ١٩٣٠ . ونظراً لانتشار استخدام هذا الاستبيان، فقد تناولته العديد من البحوث والدراسات الطويلة، وأجريت عليه عدة تعديلات لتوسيع نطاقه خلال العقود التالية، وصدرت طبعته الحديثة التي يتم تصحيحها باستخدام الحاسوب عام ١٩٧٤ .

ومن التطورات التاريخية الأخرى المهمة المتعلقة بقياس الميول الاستبيان الذي أعده كيدور Kuder عام ١٩٣٩ والذي يُعرف بسجل كيدور للتفضيل Preference Record، ويختلف في أسلوب ومنهجية بنائه عن استبيان سترونج، وكذلك البحوث المتعلقة بالقياس الموضوعي للميول التي أجرتها علماء النفس بسلاج الطيران الأمريكي أثناء الحرب العالمية الثانية .

وفي عام ١٩٤٣ أعد ثورب Thorpe استبيان الميول المهنية Occupational Interest Inventory ، ويختلف أيضاً عن سابقيه في أسلوب ومنهجية بنائه . وفي عام ١٩٤٨ أعد جيلفورد Guilford وتعاونه استبيان مسح الميول Interest Survey Guilford's استندوا في بنائه على منهج التحليل العاملى ، ويمكن باستخدامه الحصول على درجات متضمنة للميول المهنية والميول غير المهنية .

ونظراً لاعتماد معظم مقاييس الميول على الجانب اللغظى ، فقد أعد جايس استبيان الصور Geist Picture Inventory ، غير أنه توجد أدلة محدودة عن صدق مثل هذه الاستبيانات غير اللغظية . وقد ازداد بعد ذلك الاهتمام بقياس الميول غير المهنية أي المتعلقة بالأنشطة الحياتية المختلفة ، مثل استبيان كيدور لمسح الميول بعامة الذي نشر عام ١٩٦٤ ، وكذلك بناء استبيانات لخريجي المدارس الثانوية الذين لا يستمرون في الدراسة وإنما يلتحقون بالعمل ، ومثال ذلك استبيان مينيسوتا للميول المهنية . وتتوالى الدراسات والبحوث حتى وقتنا الحاضر في هذا المجال ، وبخاصة علاقة الميول بجوانب السلوك المختلفة المعرفية وغير المعرفية .

الميول وعلاقتها بالاستعدادات الخاصة :

لعل من المناسب في مستهل هذا الفصل أن نوضح المقصود بالميول وعلاقتها بالاستعدادات الخاصة، وما إذا كانت تؤدي في النهاية بهذه الاستعدادات أو التسبّب بالنجاح في العمل، فالميول يمكن اعتباره تهيئاً أو قابلية للمشاركة في نشاط معين، وبتحديد أكثر فإن الميول شرط إدراكي **Perceptual Condition** يتضمن الوعي المعرفي والشعور الوجوداني معاً في نمط معين من المشاعر. ومن الممكن أن تكون هذه المشاعر إما مؤقتة أو دائمة، وهذا يعتمد على كم وكيف الخبرة التي أدت إلى هذه المشاعر، وعادة يشار إلى الميول بأنه قابلية الفرد لأن ينتقى انتقاء حراً أنشطة محددة تستثير لديه مشاعر سارة مثل القراءة في المجالات العلمية، أو تنسيق الزهور، أو ممارسة لعبة رياضية معينة، أو تفضيل العمل في مجالات مهنية محددة.

ومصطلح «الميول» يُعد نوعاً من المتغيرات المتصلة التي ربما تتغير في المشاعر تغييراً متطرضاً على حسب الظاهرة أو الشيء المرتبط به، غير أنه يمكن التمييز بين التعبير اللغظي للفرد عن ميله تجاه نشاط معين، ومشاركته الفعلية في هذا النشاط. إذ يمكن أن يعبر الفرد عن عدم ميله للموسيقى، ومع هذا ربما يقوم عادة بالعزف الجيد على إحدى الآلات الموسيقية، وبالمثل في كثير من الأنشطة. لذلك ينبغي أن يكشف الفرد عن ميوله في نشاط معين يقوم به، فالميول يتطلب نشاطاً معيناً ومشاركة في هذا النشاط.

فالفرد الذي يميل إلى إصلاح الآلات الميكانيكية يُعد لديه ميل من مستوى معين، ربما المستوى المهاري. أما إذا كان الفرد لا يميل إلى التعامل المباشر مع هذه الآلات، وإنما يميل إلى التصميمات المجردة المتعلقة بها، فإن ميله يكون من مستوى آخر، ربما المستوى المهني المتخصص. لذلك فإن اعتبار الميول في كل من الحالتين يمثل الميول الميكانيكي ربما يحدث بعض الخلط.

فعدم وضوح تعريف هذا الميول له انعكاساته على ما إذا كانت توجد علاقة بين الميول والقدرة أم لا توجد. وعموماً ينبغي أن يوصف الميول بعبارات أنشطة محددة، كأن نقول أن شخصاً يميل إلى التحدث عن السيارات، ويقوم بإصلاحها، والإطلاع على مستحدثاتها، أو يميل إلى رسم تصميمات هندسية تتعلق بها. لذلك فإن كثيراً من تعرفيات الميول تُركز على الأنشطة التي يُفضلها الفرد أو لا يُفضلها أو لديه تزعة ل القيام بأنشطة محددة، أو يحاول تجنبها. ومن الناحية الإجرائية يشار إلى الميول بأنها مهنية أو تربوية، على الرغم من أن الفقرات أو العبارات التي تشير إلى الأنشطة المحددة نفسها يمكن أن تستخدم في كل من الحالتين.

وعادة يتم التمييز بين الميول والقدرات المستمرة، أي التحصيل الدراسي والذكاء والاستعدادات الخاصة التي تشير إلى مهارات أو كفايات معينة. ويوضح هذا التمييز إذا أردنا تقييم مهارات أو كفايات الطلاب في العلوم مثلاً في مقابل تقييم تفضيلاتهم أو مدى إقبالهم على إجراء التجارب العلمية.

ومفهوم الميول يماثل مفهوم الذكاء أو القدرات في أنه تجريد للسلوك Abstraction، أي أن مجمل الأنشطة التي يشارك فيها الفرد بمحض اختياره والتي يبدو أنه يجمعها شيء مشترك، مثل تفضيل الفرد للعمل الميكانيكي، يمكن تفسيرها بأنها تُعبر عن ميل الفرد للقيام بأعمال أو أنشطة ميكانيكية. غير أنه قد تبين عمامة أن علاقة الميول بالقدرات العقلية منخفضة، إذ يمكن أن يكون لدى الفرد ميل مرتفع نحو عمل معين، ولكن أداؤه المتعلق بهذا العمل يكون عاديًا، أو ربما تكون قدرته مرتفعة على أداء العمل ولكن لا يكون لديه ميل نحوه.

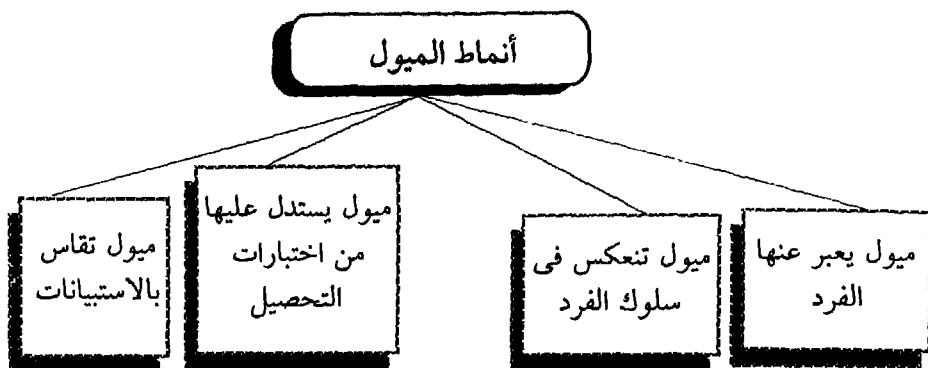
فالميول تشير إلى نطاق من التفضيلات مثل الطب، أو الهندسة، أو الأعمال التجارية، بينما تشير القدرات إلى استطاعة الفرد أداء مستوى معين من مستويات مهارة أو إنجاز معين، إذ ربما يستطيع أن يكون متخصصاً في مهنة معينة أو ماهراً أو شبه ماهر في المهنة نفسها.

ومع هذا فإن الميول تتفاعل بلا شك مع القدرات، فالقدرات ليست ثابتة دوماً، وإنما تتأثر بالتعلم المستمر والخبرة وتغير متطلبات المهن والأعمال. فالفرد يمكن أن يغير مهنته أو تخصصه داخل المهنة الواحدة. غير أن دراسات سترونج (Strong, 1955)، وكامبل (Campbell, 1971) أوضحت أنه في إطار الأقسام العريضة للمهن مثل الطب، والهندسة، والأعمال التجارية، يكون هناك استقرار نسبي للميول في مرحلة الشباب، ولا ينزع معظمهم إلى تغيير مهنته تغييرًا ملحوظًا بقية حياته.

وعموماً فإن اتخاذ القرارات التربوية والمهنية المتعلقة بالفرد يتطلب الاستناد إلى معلومات متعددة من مصادر مختلفة.

الأنماط المختلفة للميول :

يمكن تصنيف الميول في أربع مجموعات استناداً إلى الأسلوب الذي يُتبع في الحصول على المعلومات، وقد حدد سوبر وكرايتس (Super & Crites, 1962) هذه المجموعات التي يوضحها الشكل التخطيطي (١١ - ١) التالي:



شكل (١١ - ١) يوضح الأنماط المختلفة للميول

وفيما يلى توضيح لكل من هذه الأنماط:

(١) الميول التي يُعبر عنها الفرد : يمكن تحديد هذه الميول بأن يطلب من الفرد التعبير شفاهة أو كتابة عن الأنشطة المهنية أو غير المهنية التي يفضلها عن غيرها ويستمتع أو لا يستمتع بها. وعلى الرغم من أن هذه الطريقة تُعد أبسط طرق الحصول على معلومات تتعلق بميول الفرد، إلا أنه يُعبّر عليها أنها لا تميّز بالثبات وتأثر بخلفية الفرد وخبراته وخيالاته، وما يمكن أن يكون مقبولاً اجتماعياً، كأن يُعبر عن ميله للموسيقى الكلاسيكية أو العمل في مجال الفضاء. كما أنه ربما يحاول عدم التعبير عن ميله نحو الأعمال الحرفية، ولكنه يوضح ميله نحو الهندسة بدلاً عنها على الرغم من نفوره من الرياضيات.

ومع هذا فإن المعلمين يستندون إلى الميول التي يُعبر بها الطلاب عن أنفسهم Expressed Interests في معاونة الطلاب اختيار موضوعات القراءة والتغيير، والوسائل التعليمية المناسبة، واستشارة المناقشات الجماعية.

(٢) الميول التي تتعكس في سلوك الفرد : فالميول تتكتشف من خلال الأنشطة الحرة التي يقبل الفرد عليها ويمارسها أو يشارك فيها Manifested Interests فالفرد الذي يسهم في العمل التطوعي، والطالب الذي يشارك بانتظام في الجمعية العلمية بالمدرسة يعكس بذلك ميله نحو هذه الأنشطة.

ويمكن تعرُّف ذلك عن طريق الملاحظة المباشرة للفرد أو محاولة معرفة هواياته وأنشطته التي يمارسها بالفعل، فالفرد الذي ينسق الزهور بانتظام في حديقة منزله، أو يقتني التحف، أو يقوم بإصلاح الأعطال في أجهزته المنزلية، فإنه يكشف بذلك عن

ميله نحو نشاط معين. غير أن هذه الميول ربما تكون انتقالية، أى ربما يتتحول اهتمام الفرد من نشاط إلى آخر لأسباب متعددة. فالفرد الذى يذهب إلى النادى ربما يجعلنا نظن أن لديه ميل للمشاركة فى الألعاب الرياضية، فى حين أنه يرى أن هذا المجال مناسب لتكوين علاقات اجتماعية، أو أنه يذهب لمجرد قضاء وقت الفراغ.

فالميول التى يعبر عنها الفرد لا تطابق بالضرورة الميول التى تتعكس فى سلوكه، ولكن كثيراً من الأفراد الذين يشاركون بانتظام فى أنشطة حرة يعبرون بالفعل عن ميلهم الإيجابى نحو هذه الأنشطة.

(٣) الميول التى يستدل عليها من اختبارات التحصيل : يمكن الاستدلال على ميول الطالب من درجاته فى اختبارات التحصيل فى المجالات الدراسية المختلفة Inferred Interests. وهنا يفترض أن الدرجات المرتفعة فى أحد هذه المجالات توافق مع ميل الطالب نحو هذا المجال.

غير أن كثيراً من الباحثين - كما سبق أن ذكرنا - وجدوا أن الارتباط بين الميول المقاسة والقدرات منخفضة. ومع هذا فإنه إذا قارنا ترتيب درجات الطالب فى استبيان الميول بدرجاته فى اختبارات التحصيل المتعلقة بها، فإنه ربما نلاحظ أن هناك تطابقاً بين درجاته المرتفعة فى مجال دراسى معين وميله المرتفع نحو هذا المجال.

(٤) الميول التى تقادس بالاستبيانات : يمكن تحديد ميول الفرد عن طريق قياسها باستبيانات مقتنة حيث يقارن بين ميوله فى أنشطة متعددة Inventorized Interests. وهذه الاستبيانات تتطلب من الفرد اختيار الشاط الذى يفضله أو لا يفضله من بين مجموعة كبيرة من الأنشطة المتعلقة بالمهن والمجالات الدراسية والتربوية والأنشطة المصاحبة والهوايات.

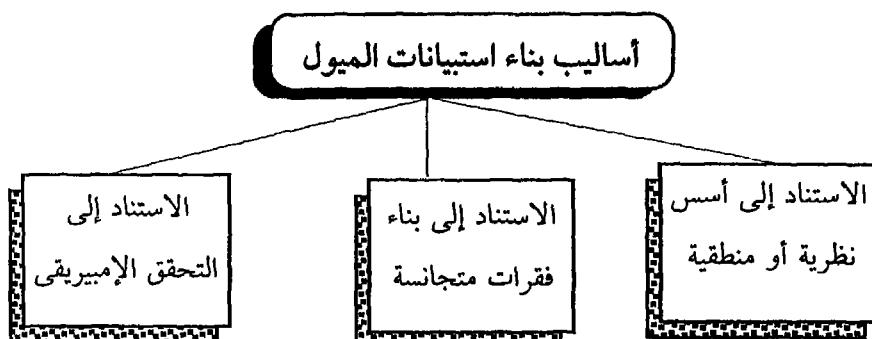
كما تتطلب منه ترتيب ظروف العمل المختلفة من حيث أهميتها بالنسبة له، واختيار السمات الشخصية والاجتماعية التى يرى أنها ضرورية. وتستند بعض هذه الاستبيانات وموازين التقدير إلى أن هناك أنماط ميول متماثلة تميز الأفراد الناجحين فى مهنة معينة أو عمل محدد كالتدريس مثلاً، كما أن هناك أنماط ميول فارقة بين الأفراد الناجحين فى مهن مختلفة كالهندسة والطب. وبالطبع تقيس هذه الاستبيانات عينة محددة من ميول الفرد، حيث إن كثيراً من الاستبيانات المتوفرة فى الوقت الحاضر تفيد فى التوجيه المهىنى، بينما عدد قليل منها يمكن الإفاده منه فى التوجيه التربوى.

أساليب بناء استبيانات الميول :

تختلف أساليب بناء استبيانات الميول Interest Inventories وتقدير درجاتها Keying عن الأساليب المتعارف عليها في قياس الجوانب المعرفية، وذلك لأن مفردات الاختبارات المتعلقة بهذه الجوانب يكون لها إجابة صحيحة واحدة مقبولة، بينما مفردات مقاييس الجوانب غير المعرفية أو الوجданية بعامة ليس لها إجابة صحيحة محددة.

إذ ربما تُعد استجابة الفرد لإحدى فقرات استبيان للميول صحيحة بالنسبة لأعضاء مهنة معينة بينما تُعد الاستجابة خطأ بالنسبة لأعضاء مهن أخرى. ويعتمد تفسير درجات هذه الاستبيانات على الأسلوب الذي يتبع في بنائها.

ويوضح الشكل التخطيطي (١١ - ٢) التالي أهم هذه الأساليب :



شكل (١١ - ٢) يوضح أهم أساليب بناء استبيانات الميول

وسوف نوضح فيما يلى كلا من هذه الأساليب بمزيد من التفصيل :

أسلوب يستند إلى أساس نظرية أو منطقية : استُخدم هذا الأسلوب في مستهل نشأة بناء استبيانات الميول والشخصية، حيث تُتحقق فقرات أو مفردات الاستبيان استناداً إلى التحليل النظري أو المنطقي للسمة أو الخاصة المراد قياسها، وهذا يماثل الأسلوب الذي يتبع في تقدير صدق محتوى الاختبارات الذي أوضحته في الفصل الخامس. وهذا يتطلب من القائم بناء استبيان الميول أن يحدد الأساس المنطقي الذي يُستند إليه في تصنيف الميول وبناء الفقرات التي تبدو أنها تقيس المجالات المختلفة.

وقد استُخدم لى وثروب Lee - Thrope هذا الأسلوب في بناء استبيان الميول المهنية الذي سوف نوضحه فيما بعد. ويتم تقدير درجات الفقرات على أساس منطقي. وعلى الرغم من أنه ربما لا يكون هناك اتفاق عام على تقدير درجة فقرة ما، إلا أنه عند

بناء المفردات ينبغي الاهتمام بانتقاء الفقرات التي يسهل تقدير درجاتها. فمثلاً تقدر درجة الفقرة «أميل إلى قراءة المقالات الاجتماعية» ربما يكون أيسر من تقدير درجة الفقرة «أفضل المشاركة في الحفلات الاجتماعية مع الأصدقاء، عن المشاركة في الحفلات الجماعية الكبيرة». وعموماً فإن بناء الاستبيان استناداً إلى الأسلوب المنطقي يُعد خطوة أولية يمكن أن يتبعها خطوات تحليلية إحصائية أو إمبريقية أخرى لتنقية الفقرات، وذلك لأن الفقرات التي تبدو أنها تقيس سمة أو مجالاً معيناً ربما لا تقيس بالفعل. لذلك لا يوصى بالاقتصار على هذا الأسلوب في بناء الاستبيانات.

أسلوب يستند إلى بناء فقرات متجانسة: يُعد هذا الأسلوب Homogeneous Keying امتداداً واستكمالاً للأسلوب السابق. فبعد انتقاء فقرات الاستبيان على أساس منطقي وتصنيف هذه الفقرات في مجموعات متجانسة تقيس مجالات مختلفة مثل الميل الخلوي، والميل إلى مهنة معينة، والميل إلى الأعمال الإدارية مثلاً، يمكن باستخدام الأساليب الإحصائية التي أوضحتها في الفصل الرابع التحقق من الاتساق الداخلي أو تجانس الفقرات.

إذ يرى كثير من علماء القياس أن المقاييس المستسقة داخلياً أو متجانسة هي التي تُضفي معنى على المقياس وتُمكّناً من تفسير درجاته. فالفقرات التي يكون ارتباطها بغیرها من الفقرات التي يشتمل عليها المقياس منخفضاً بدرجة واضحة يجب حذفها لأنها لا تتنسمى إلى مجموعة الفقرات التي تقيس سمة أو مفهوماً معيناً. وهذا يماثل أسلوب التتحقق من صدق التكوين الفرضي الذي أوضحته في الفصل الخامس.

ويمكن أيضاً إخضاع مصفوفة الارتباطات بين الفقرات المختلفة التي يشتمل عليها الاستبيان للتحليل العاملى Factor Analysis للتوصل إلى أقل عدد من العوامل التي تفسر هذه الارتباطات. وبذلك نحصل على تجمعات من الفقرات التي يقيس كل منها عاملًا معيناً يمكن تسميته بحسب محتوى فقرات كل تجمع، وهذه العوامل أو السمات يمكن قياسها. ففي مجال الميول ربما ترتبط مجموعة من الفقرات فيما بينها ارتباطاً مرتفعاً، وبذلك تقيس عاملًا أحادي البعد يمكن أن يطلق عليه مثلاً الميل الميكانيكي.

وينطبق ذلك أيضاً على مقاييس الاتجاهات واستبيانات الشخصية كما سيتضح في الفصلين التاليين. وقد استخدم جيلفورد ومعاونوه Guilford et al. هذا الأسلوب في بناء استبيان مسح الميول، كما استخدمه كيودر Kuder في بناء استبيان المسح العام للميول.

أسلوب يستند إلى التتحقق الإمبريقى : يعتمد هذا الأسلوب في انتقاء فقرات استبيانات الميول على درجة تمييزها بين مجموعة من الأفراد تنتمي إلى مهنة

معينة ومجموعة من عامة الناس أو مجموعات من مهن مختلفة. فالمجموعة الأولى تكون استجاباتها لفقرات الاستبيان مختلفة عن استجابات عامة الناس، أو أفراد المهن الأخرى للفقرات نفسها. فعندئذ يمكن تضمين هذه الفقرات في استبيان الميول، ويتم تقدير درجات فارقة تتناسب مع اختلاف استجابات هذه المجموعات.

ويتطلب هذا الأسلوب أيضاً تجميع عدد كبير من الفقرات وتجريبيها ميدانياً على مجموعة من الأفراد الناجحين في مهنة معينة مثلاً ومجموعة أخرى من عامة الناس. ويمكن تعريف الفرد الناجح بأنه الذي قضى في مجال عمله عدداً معيناً من السنوات على الأقل، ويواظب على هذا العمل، ويفضله عن غيره من الأعمال، ويحصل على تقديرات متميزة من مشرفيه، ولا يقل عمره أو يزيد عن عمر معين. أى أنه ينبغي تعريف محك النجاح الذي سيتم التمييز في ضوئه نظراً لأن هذا الأسلوب لا يستند إلى افتراضات معينة تتعلق بسمة أو خاصية من خصائص أفراد المجموعات المختلفة، وإنما يتم انتقاء الفقرات التي تميز بين هذه المجموعات، فالانتفاء إلى مجموعة معينة هو المحك الذي يقدر في ضوئه صدق الاستبيان. ويستخدم هذا الأسلوب أيضاً في بعض استبيانات الشخصية كما سيوضح في الفصل الثالث عشر.

ولعل هذا الأسلوب يتميز بأنه لا يمكن المستجيبين من تزوير استجاباتهم لأن استجابات المجموعة أو المجموعات التي يُستند إليها كمحك لا تكون معلومة لهم. غير أن اختيار الفرد الناجح في مهنة معينة يعتمد على تعريف معين ربما لا يكون كافياً أو شاملاً، فعندئذ لا يكون الاستبيان مناسباً، كما أن هذا الأسلوب في بناء الاستبيانات ربما يتحقق الصدق التلازمي. ولكن يبرز تساؤلان حول إمكانية الإفادة من الاستبيان في التنبؤ بمحك مستقبلي وعدم الاقتصر على تقسيم انتماء الفرد لمجموعة مهنية معينة، وكذلك مدى تعميم نتائج الاستبيان، أى التوصل إلى فهم متكامل للسمة المقاسة في ضوء إطار نظري معين استناداً إلى نتائجه. فالسؤال الأول يتعلق بالصدق التنبؤى الذي يمكن التتحقق منه إEmpirically، أما السؤال الثاني فيتعلق بصدق التكوين الفرضى الذى يتطلب أدلة متعددة.

لذلك يتضح أنه على الرغم من تعدد أساليب بناء استبيانات الميول والشخصية، إلا أن هناك تكاملاً فيما بينها، فالأسلوب المنطقي Rational Keying يعد خطوة أولية لانتقاء مجموعة من الفقرات. ويتم تجريب هذه الفقرات على جماعة مرجعية مناسبة ممثلة للمجتمع الذي سيطبق عليه الاستبيان فيما بعد، وتكون مصفوفة الارتباطات بين درجات الفقرات، وإجراء التحليل العاملى من أجل التوصل إلى مجموعات متGANSA

من الفقرات وبعد ذلك يمكن التتحقق من الصدق المرتبط بمحك بالأساليب الإمبريقية Emperical Keying، وبذلك يكون للاستبيان أساس نظري ومنطقى، وفي الوقت نفسه تتأكد فائدته العملية نتيجة استخدام الأسلوب الإمبريقى.

بعض المشكلات المتباقة بقياس الجوانب غير المعرفية :

إن قياس الجوانب غير المعرفية أو الوجданية للسلوك يُعد أكثر صعوبة من قياس الجوانب المعرفية لأسباب كثيرة نذكر منها:

(١) عدم الاتفاق التام على تعريف مكونات الجوانب الوجданية للسلوك، كما أن هذه المكونات تصعب ملاحظتها مباشرة في سلوك الأفراد، إذ كثيراً ما يجد المعلم أو المرشد التربوي أو المهني صعوبة في تحديد ما يود ملاحظته في سلوك الفرد الذي لديه ميل نحو نشاط أو عمل ما، أو اتجاه معين. فمما لا شك فيه أن هذه المكونات الوجданية تحدث داخل الفرد. وقياس وتقويم هذه المكونات غير المنظورة يحتاج إلى تحليل سلوكي بحيث يمكن ملاحظة السلوك ملاحظة مباشرة تعكس تأثير هذه المكونات. وتبين وجهات النظر تبايناً ملحوظاً حول إجراءات ونواتج هذا التحليل المتعلق بالأداء المميز للفرد.

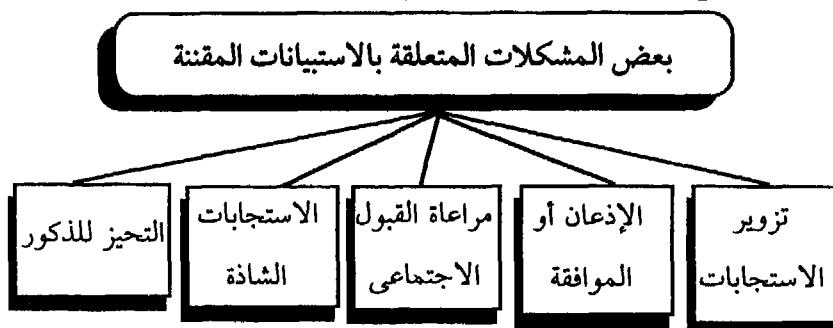
(٢) يصعب قياس بعض الجوانب الوجданية ، حيث إنها ربما تتطلب ملاحظة الفرد ملاحظة متكررة في مواقف مختلفة يمكن أن تظهر فيها الاستجابات التي نهتم بملاحظتها ، نظراً لأن ملاحظته في موقف واحد لا يكفي للحكم على أدائه المميز. فالسلوك الوجданى يعتمد على الحالة المزاجية للفرد ، وأفته بال موقف ، وظروف البيئة المحيطة به ، وغير ذلك من المتغيرات . فالفرد الذى يبدو عدوانياً في بعض المواقف ربما يجعلنا نحكم بأنه شخص عدواني ، ولكن ربما يكون في حالة من الضيق نتيجة قلقه المستمر على شيء ما ، ويبدو في حالة تكيف تام بعد انقضاء عدة أسابيع أو عدة شهور ، وبذلك فإن ما نلاحظه في مواقف قليلة عن فرد معين ربما يكون عرضياً وقتياً يميز الفرد في هذا الوقت .

(٣) نظراً لصعوبة التحديد الفعلى الدقيق لكثير من المفاهيم والمكونات الوجدانية ، فإننا لا نستطيع الاعتماد على أداة قياس واحدة . إذ من الصعب احتواء هذه المكونات جميعاً وقياسها مباشرة بأداء قياس واحدة؛ لذلك ربما يفضل اتباع استراتيجية الالتفاف حول هذه المكونات ، وذلك باستخدام مقاييس متعددة مناسبة ، وبذلك يمكن أن نحصل على صورة أفضل عن المكونة المراد قياسها .

(٤) إن قياس الجوانب غير المعرفية أو الوجدانية يشمل الفروق بين الأفراد وكذلك التغيرات التي تحدث في السلوك عبر الزمن، وهذا يُلقي عبئاً كبيراً على القائم ببناء مقاييس هذه الجوانب في التتحقق من صدق وثبات البيانات المستمدة من هذه المقاييس.

بحرين المشكلات المتعلقة بالاستبيانات المقننة للميلو :

سوف نناقش فيما يلى بعض أهم المشكلات المتعلقة بالاستبيانات المقننة لقياس الميلو، وتنطبق هذه المشكلات أيضاً على استبيانات الاتجاهات والشخصية، ويمكن توضيحها بالشكل التخطيطى (١١ - ٣) التالى :



شكل (١١ - ٣) يوضح بعض المشكلات المتعلقة بالاستبيانات المقننة

تزيف أو تحيز الاستجابات : تُعد مشكلة تزيف أو تحيز الاستجابات Faking في استبيانات الميلو وغيرها من استبيانات الجوانب غير المعرفية من المشكلات التي اهتمت كثير من البحوث بدراستها ، وبخاصة بعد دراسة كرونباك Cronbach التي قدم فيها مراجعة شاملة لهذه المشكلة (Cronbach , 1950 ،). ففي اختبارات الذكاء والاستعدادات الخاصة لا يلجأ الفرد إلى تزيف استجاباته ، وإنما يمكن أن يُخْمَن الإجابة الصحيحة ، أو يختار إجابات خطأ في مفردات الاختيار من متعدد ، أو يترك المفردة دون إجابة. ولكن في استبيانات الجوانب غير المعرفية يستطيع الفرد تزيف إجابته بطريق متعدد .

لذلك اهتم الباحثون بدراسة هذه المشكلة المعقدة. فعلى الرغم من أن الاستبيانات التي تقيس هذه الجوانب تهدف لمساعدة الفرد في تعرف ميوله واتجاهاته وسمات شخصيته وتوجيهه تربوياً ومهنياً استناداً إلى نتائجها، إلا أن بعض الأفراد يلجأون إلى تزيف استجاباتهم لصالحهم، وبذلك لا تُعبر هذه الاستجابات عن أدائهم المميز الفعلى. ولعل هذا يعتمد على الغرض من استخدام استبيان الميلو أو

الشخصية، فإذا كان الغرض انتقاء الأفراد لوظيفة أو مهنة شاغرة، فإن احتمال تزيف فرد معين لاستجاباته لكي يعطي انطباعاً مقبولاً عن نفسه يكون أكبر مما لو كان الغرض من الاستبيان التوجيه المهني. بل ربما أيضاً في التوجيه التربوي يلتجأ الطالب إلى الكذب في إجاباته على استبيان الميول لكي يبين أنه يميل إلى دراسة الطب أو الهندسة مثلاً.

وعلى العكس من ذلك، فإن هناك بعض الأفراد يعتمدون إظهار أنفسهم بمظاهر عدم التكيف أو أنهم غير أسواء للتهرب من الالتحاق بالجيش أو الحجز بمستشفيات الأمراض العقلية، أو مرتكبي الجرائم، ولكن ليس معنى ذلك أن تزيف الاستجابات يكون مُعمداً في جميع الأحوال، فبعض علماء النفس يرون أن الفرد ربما يكون لديه دوافع أو نزعات شبه شعورية Subconscious للاستجابة بطريقة أو نمط معين. فالمرء الذي يرفض سلطة الوالدين ويؤود والده إلحاقه بكلية معينة ربما يلتجأ إلى الكذب في استبيان الميول وبخاصة في المقياس المتعلق بالمييل نحو المهنة التي تُعد هذه الكلية لها. لذلك ينبغي تعرُّف دوافع الفرد ونزعاته عند تفسير الدرجات في استبيانات الميول وغيرها من الاستبيانات.

الاتجاه العقلي في الاستجابة : إن الاستبيانات أو التقارير الذاتية ليست عُرضة فقط لتزيف الاستجابات أو تحيزها، وإنما أيضاً الدرجات المستمدّة منها عرضة للتأثر باتجاهات عقلية معينة في استجابات الفرد، وبخاصة إذا كانت تشتمل على فقرات أو عبارات غامضة. وهذه الاتجاهات العقلية ربما ترجع إلى نزعات شبه شعورية لدى الفرد للاستجابة بطريقة أو توجُّه معين بغض النظر عن محتوى فقرات أو عبارات الاستبيان. ويُطلق على هذه النزعات الاتجاه العقلي المعين في الاستجابة Response Sets. وهذا الاتجاه العقلي يجعل الفرد يستجيب لفقرات الاستبيان بطريقة تختلف عن استجاباته لو قدمت له الفقرات بشكل آخر أو في مضمون مختلف. فالاستبيان الذي يفترض أنه يقيس سمة أو خاصية معينة لدى الفرد مثل الميول ربما يقيس أيضاً خاصية أخرى أي الاتجاه العقلي، وهذه الخاصية الأخرى ربما لا ينعكس تأثيرها في درجته لو أنها أحدثنا تغييراً في الموقف الاختباري بعامة، مثل نمط أو شكل فقرات الاستبيان، وأسلوب صياغتها، وعدد وطبيعة البذائل التي تشتمل عليها الفقرات، وتعليمات الاستبيان، والغرض الذي سوف تُستخدم فيه درجات الاختبار، أو القرارات التي سوف تستند إلى هذه الدرجات.

ويوجد نمطان رئيسيان لتبنيات الاتجاه العقلى: أحدهما انتقالى ويقتصر على استبيان معين أو موقف اختبارى معين، والآخر مستقر ومتسع ويتصف بالعمومية إلى حد ما، أى يؤثر فى استجابات الفرد في الاستبيانات والمواقف الاختبارية المختلفة. ويوجد أيضاً نمطان لهذا الاتجاه العقلى المستقر: أحدهما يعكس فروقاً فردية ثابتة نسبياً ولكنها غير ذات أهمية مثل استخدام اللغة أو العادات التي تبدو من تعابيرات الفرد، والآخر يعكس جوانب مهمة للشخصية.

لذلك فإن التباين الذى يُعزى إلى هذا النمط الثانى ربما يُعد مؤشراً لخصائص الشخصية؛ ولذلك يطلق عليه أسلوب أو نسق الاستجابة Response Style. وهذا الأسلوب أو النسق يعنى اتساق طريقة الاستجابة لشكل ما من أشكال فقرات الاستبيانات بغض النظر عن محتوى نوعى معين.

ويتميز هذا الأسلوب أو النسق بالاستقرار النسبي عبر الزمن ويُسمى فى تعميم الاستجابات سواء فى المواقف الاختبارية أو غير الاختبارية، وهذا الأسلوب يُنبع دراسته فى حد ذاته (Jackson & Messick , 1958).

وسوف نعرض فيما يلى الأنواع الرئيسية للاتجاهات العقلية Response Sets التي أشارت إليها معظم أدبيات القياس التربوى والنفسى:

الإذعان أو الموافقة : اهتمت كثير من الدراسات بهذا النوع من تحيز الاستجابات أو الاتجاه العقلى. ويُقصد بالإذعان أو الموافقة Acquiescence نزعة الفرد فى استجابته على فقرات استبيانات الميل أو الشخصية إلى الموافقة بغض النظر عما تشير إليه الفقرات. أى أن الفرد ينزع إلى الاستجابة دوماً إما بـ «نعم» أو «لا»، أو «موافق» أو «غير موافق». فإذا ما طبق استبيان للميل أو الشخصية على هؤلاء الأفراد، فإن الدرجات التى يحصلون عليها لا تُعبر عن السمة المقاومة لديهم بقدر ما تعكس هذا الاتجاه العقلى، وبذلك لا تستطيع الاستناد إلى هذه الدرجات فى اتخاذ قرارات مهمة بشأنهم.

وهناك أدلة بحثية متعددة أشارت إلى أن الإذعان أو الموافقة تُفسّر جزءاً كبيراً من تباين استبيانات الجوانب غير المعرفية، وبخاصة استبيانات الشخصية. وقد سبق أن أوضحنا أن بعض علماء النفس يرون أن هذا الاتجاه العقلى ربما يُعد سمة شخصية عند

الأفراد الذين لديهم نزعة شبه شعورية للاستجابة في معظم الأحيان لفقرات استبيانات الميول والشخصية بالموافقة، فهم يرون أن مثل هؤلاء الأفراد يكونون منقادين، وممثلين، دائمًا وينسجون من التحديات في حياتهم اليومية.

ويقترح سميث (Smith , 1961) أنه يمكن التحكم جزئياً في هذا الاتجاه العقلي في الاستجابة بالعناية بصياغة فقرات الاستبيان، فالفقرات التي تشتمل على عبارات متطرفة ربما لا تؤدي إلى استجابات الموافقة أو الإذعان بنفس قدر العبارات المعتدلة. كما يرى إيليوت (Elliott , 1961) أن هناك ثلاثة محددات للإذعان أو الموافقة هي: نوع أو شكل الفقرة (عبارات موجبة، وعبارات سالبة، و اختيار من متعدد)، ومدى تعلق محتوى الفقرة بسمة شخصية، والمستوى العقلي للفرد المستجيب، ويرى أن هذه المتغيرات الثلاثة متفاعلة.

مراجعة القبول الاجتماعي : يُعد هذا الاتجاه العقلي في الاستجابة Social Desirability نوعاً من التحيز، حيث يحاول الفرد إظهار نفسه بصورة مقبولة اجتماعياً في استجاباته لفقرات معينة في استبيانات الميول والشخصية، أي أن الفرد يحاول أن يظهر نفسه مثلاً كإنسان متكيف ومنفتح ومتفاعل مع الآخرين. فكثير من الأفراد يوافقون على عبارة مثل «أحاول قدر استطاعتي تجنب إشارة المشاكل»، أو «أفضل الأعمال التي تقدم خدمات إنسانية»، أو «أميل إلى مساعدة الآخرين». وهذه الموافقة تتم بطريقة شبه شعورية، حيث إن الفرد يحاول أن يكون مقبولاً اجتماعياً، وبذلك لا تُعد استجاباته نوعاً من الكذب أو الخداع لأن الفرد لا يتعمد ذلك.

ويعتمد هذا الاتجاه العقلي على محتوى فقرات مقاييس التقدير الذاتي مثل الاستبيانات. فالفرد لا يستطيع تزييف استجابته الحقيقة لكي تكون مقبولة اجتماعياً ما لم يدرك العلاقة بين محتوى الفقرة التي يستجيب لها والطريقة المحتملة لتقدير درجتها.

ولعل تبرير هذا الاتجاه العقلي أكثر وضوحاً من الاتجاهات العقلية الأخرى. فهنا ربما يكون التبرير هو تجنب الفرد للنقد أو التهديد الخارجي أو أن يتضح سوء تكيفه.

وقد أوضحت دراسة كراون ، ومارلو (Crown & Marlowe , 1964) ودراسة فريديريكسن (Fredrikson , 1965) أن شدة الاتجاه العقلى المتعلق بمراعاة القبول الاجتماعى يرتبط بحاجة الفرد إلى حماية نفسه ، والامتثال لما هو سائد في المجتمع . ولقد كان إدواردز (Edwards , 1957) أول من اهتم ببحث متغير «مراعاة القبول الاجتماعى» ، وبين أن هذا المتغير أحادى البعد . فإذا طلبنا من عدد من المحكمين تقدير درجة القبول الاجتماعى لكل عبارة من العبارات التسعة يشتمل عليها استبيان معين ، فإنه يمكن الحصول على ميزان أو مقاييس متدرج أحادى البعد لهذا الاتجاه العقلى .

كما أجرى إدواردز (Edwards , 1957 ، Marlowe, 1960, 1964) بحوثاً متعددة حول تأثير «مراعاة القبول الاجتماعى» في درجات استبيانات التقرير الذاتى ، وأعدوا موازين يمكن استخدامها في تقييم التزعة العامة للفرد لإظهار نفسه بصورة مقبولة اجتماعياً . كما يمكن استخدام هذه الموازن أثناء بناء استبيانات الجوانب الوجدانية للتحكم في هذا الاتجاه العقلى في الاستجابات الأفراد ، وذلك بحذف الفقرات التي ترتبط بمقاييس مراعاة القبول الاجتماعى .

ومن أجل تقليل تأثير هذا الاتجاه العقلى في الاستجابات اقترح إدواردز Edwards استخدام استبيانات تتطلب الاختيار الجبرى Forced-Choice Inventories ففي هذه الاستبيانات تُقدم أزواج من العبارات المقبولة اجتماعياً بدرجة متكافئة ومتباعدة في محتواها وصدقها ، أي أن كل من العبارتين ترتبط ارتباطاً متماثلاً بمحكات إيميريقية مناسبة . فإذا اختار الفرد إحدى العبارتين التي يرى أنها تنطبق عليه بدرجة أفضل يحصل على (درجة واحدة) إذا ارتبطت العبارة بالمحك ، بينما يحصل على (صفر) إذا اختار عبارة لا ترتبط بالمحك .

ويمكن أن يُقدم أي عدد من بدائل العبارات لكل فقرة من فقرات الاستبيان بدلاً من زوج واحد . وسوف يتضح ذلك في عرضنا لأمثلة من استبيانات الميل ، وبخاصة استبيان كيودر حيث استخدام الفقرات جبرية الاختيار . وعلى الرغم من أن هذا النوع من الفقرات يزودنا بمعلومات أكثر تفصيلاً من الفقرات المألوفة ، إلا أنه يمكن تزييف استجاباتها أيضاً . إذ من الصعب الحصول على زوج من العبارات متماثل تماماً ، والفرد الذي يود إعطاء صورة أفضل عن نفسه يمكن أن يعرف أحد البديلين الذي يتحمل بدرجة أكبر أن يكون مقبولاً .

كما أن تطبيق هذا النوع من الفقرات يتطلب وقتاً أطول للحصول على عدد من الاستجابات مساوياً للعدد الذي نحصل عليه إذا استخدمت فقرات غير جبرية الاختيار .

الاستجابات الشاذة : يميل بعض الأفراد إلى الاستجابة الشاذة أو غير الشائعة Deviant (Atypical) Responses على فقرات استبيانات الجوانب الوجданية، وبخاصة مقاييس الشخصية، وهذه التزعة تكون متسقة عبر فقرات الاستبيان وبمرور الزمن، فالأشخاص غير الأسواء يستجيبون لهذه الفقرات استجابة شاذة بغض النظر عن محظواها.

ومعظم الدراسات التي اهتمت بدراسة هذا الاتجاه العقلي طلت من هؤلاء الأفراد أن يختاروا استجاباتهم لفقرات استبيان معين على ميزان تقدير يتراوح بين: أهتم جداً أو أهتم جداً، ولا أهتم ، أو لا أهتم على الإطلاق. وهنا يلجم هؤلاء الأفراد إلى اختيار استجابات تختلف عن الاستجابات الشائعة لجماعة مجتمعية معينة. وقد قدم سيشرست وجاكسون (Sechrest & Jackson, 1963) تحليلًا تفصيليًّا لتزعمات الاستجابات الشاذة وكيفية قياسها وتفسيرها.

التحيز للذكور: وُجِّهَ كثير من النقد لاستبيانات الميول في مستهل نشأتها بحجة أنها تتحيز للذكور، وُتُسْهَمُ في توجيه الإناث إلى أعمال ومهن تتفق مع الأدوار التقليدية للمرأة، مثل أعمال السكرتارية، والتمريض، والتدرис للأطفال، والحرف اليدوية البسيطة. لذلك فإن استبيان سترونج للميول المهنية في طبعته الأولى كانت له معايير منفصلة لكل من الذكور والإناث. وقد طالبت العديد من المؤسسات الأمريكية التي تدافع عن حقوق المرأة بإعادة النظر في هذه الاستبيانات بحيث تكشف عن ميول المرأة نحو المهن العليا التي ترفع من مستواها الاجتماعي والاقتصادي.

لذلك أُعيد النظر في استبيانات الميول بحيث لا يكون هناك استبيان خاص بالذكور وأخر خاص بالإناث، وإنما يُستخدم مقاييس واحد ومعايير واحدة لكل منها. غير أن هذا التطوير قد أسهم في تقليل صدق استبيانات الميول، نظرًا لأن الواقع الاجتماعي وال النفسي يشير إلى اختلاف ميول الذكور عن الإناث (Campbell , 1977 ,).

لذلك يُلاحظ أن بعض الاستبيانات عادت مرة أخرى إلى التمييز بين معايير الذكور والإناث في المقاييس المتعلقة بالميول المهنية، ولكنها دمجت معايير كل منها في مقاييس الميول العامة، وعمومًا فإن هذا ربما يُقلل من التحيز لأحد الجنسين ولكنه لا يستبعده تماماً.

تحقيق على مشكلة تحيز الاستجابات :

لاحظنا تعدد مصادر الاتجاهات العقلية Response Sets التي يمكن أن تسهم في تحيز استجابات الأفراد في مقاييس التقرير الذاتي مثل الاستبيانات. غير أنه لا يوجد

اتفاق حول مدى استقلالية كل من هذه المصادر عن الآخر. وعلى الرغم من أنه يمكن التبؤ بالاتجاه العقلي المتعلق بالإذعان أو الموافقة من معرفتنا بالاتجاه العقلي المتعلق بمراعاة القبول الاجتماعي، مما يعني أن المصادرين يعبران عن ظاهرة واحدة، إلا أنه لا يوجد اتفاق أيضاً بهذا الشأن.

فالدراسة التي أجرتها كاوش وكينيستون (Couch & Keniston 1960) حول «الإذعان أو الموافقة» كمتغير من متغيرات الشخصية انتقدها كل من إدواردز ، ووالكر (Edwards & Walker , 1961) ، وتايلر (Tyler 1961) على أساس أن نتائجها يمكن تفسيرها في ضوء «مراعاة القبول الاجتماعي» بدلاً من «الإذعان»، ولكن عندما أجرى الباحثان الأصليان تحليلًا عاملياً لبيانات دراستهما توصلوا إلى أن مراعاة القبول الاجتماعي، والإذعان يُعدان عاملين مختلفين ، بينما توصل إدواردز ، ووالكر من نتائج تحليل عاملٍ آخر أن المصادرين يجمعهما عامل مشترك .

وإذا سلّمنا بأن «الإذعان أو الموافقة» يُعد مصدراً مستقلاً، فإن هناك أيضاً عدم اتفاق حول أهميته النسبية مقارنة بمصدر «مراعاة القبول الاجتماعي»، لذلك فإن الدراسات لم تتفق على أي المصادر تُسهم في التباين الكلى لدرجات الاستبيانات إسهاماً أكبر . ولكن أوضحت بعضها أن مصادر الاتجاهات العقلية في الاستجابات تُسهم في التباين الكلى لدرجات أكثر مما تُسهم به السمات التي يفترض قياسها بهذه الاستبيانات ، (Keniston & Messick , 1961) .

غير أن رورار (Rorrer , 1965) قام بمراجعة شاملة للدراسات في هذا المجال واستنتاج أن أساليب أو أنساق الاستجابات Response Styles نادرًا ما تفسر أكثر من ٢٠٪ من تباين درجات الاستبيانات ، مما يجعل اعتقادنا بأن الأفراد يعتمدون في استجاباتهم لاستبيانات الميول والاتجاهات والشخصية على هذه الأساليب أو الأنماط غير صحيح .

وعلى أي حال فإنه ينبغي أن نحتاط للأثار المحتملة لهذه التزععات في بناء وتفسير نتائج هذه الاستبيانات في ضوء المقتراحات التالية :

(أ) تكوين علاقات ودية مع المختبرين لإقناعهم بالفائدة التي سوف تعود عليهم مستقبلاً نتيجة تحري الصدق في استجاباتهم وتشجيعهم على ذلك .

(ب) محاولة إخفاء الهدف من الاستبيان عند بنائه بحيث يبدو أن محتوى الاستبيان يقيس شيئاً مختلفاً عما يراد قياسه . غير أن هذا الاقتراح يناسب بعض الاستبيانات دون الآخر ، كما أنه يؤدي إلى مشكلات عملية وأخلاقية ،

وربما يؤدي كذلك إلى عدم ثقة المستجيب في المرشد التربوي أو النفسي مما يسء إلى علاقة المودة بينهما .

(ج) صياغة الفقرات صياغة واضحة ومحايدة، وإعداد مفاتيح تصحيح خاصة لكشف هذه الاتجاهات العقلية في الاستجابات، كما في مقاييس كيودر للميل وكما سيتضح في الفصل الثالث عشر، والعناية بإعداد تعليمات الاستبيانات.

(د) استخدام فقرات من نوع الاختيار الجيري كما أشرنا منذ قليل، غير أن هذا لا يكشف عن الاتجاهات العقلية وإنما يقلل من أثرها إلى حد ما.

مشكلات تفسير درجات استبيانات الميل:

تعتمد الاختبارات التي تقيس الجوانب المعرفية في تفسير درجاتها على المعايير أو محركات الأداء. وعلى الرغم من أن مقاييس الجوانب غير المعرفية تعتمد أيضًا على الإطار المرجعى النسبي (المعايير)، والإطار المرجعى المطلق (محركات أو مستويات الأداء) في تفسير درجاتها، إلا أن بعض هذه المقاييس، وبخاصة استبيانات الميل والشخصية، تعتمد في ذلك على الإطار المرجعى الفردي Ipsative Scales.

فالدرجة في ضوء هذا الإطار المرجعى تصف مستوى الفرد في أحد المتغيرات أو السمات بالنسبة لمستواه في متغير آخر أو سمة أخرى، أي أن تفسير الدرجات في هذه الحالة يُركز على مقارنة الفرد بنفسه في أكثر من متغير. وبالطبع يختلف ذلك عن تفسير الدرجات في ضوء الإطار المرجعى النسبي الذي يقارن درجة الفرد بدرجات أقرانه في متغير أو أكثر، أو تفسير الدرجات في ضوء الإطار المرجعى المطلق الذي يقارن درجة الفرد بمحرك أو مستوى أداء محدد مسبقًا، كما سبق أن أوضحنا في الفصول السابقة.

ففي ضوء الإطار المرجعى الفردي يمكننا تفسير درجات الفرد في استبيان للميل بأن نقول أن درجة فرد في الميل العلمي ٧٠ ، ودرجه في الميل الكتابي ٣٠، لذلك فإن ميله العلمي أقوى. أما التفسير في ضوء الإطار المرجعى النسبي فيكون مختلفاً، فإذا حصل فرد على الدرجة ٧٠ في اختبار ما وحصل فرد آخر على الدرجة ٣٠ في الاختبار نفسه ، فإن هذا يعني أن الفرد الأول أفضل من الفرد الثاني فيما يقيسه الاختبار.

وتميز درجات المقاييس التي تُناسب إلى هذا الإطار المرجعى الفردي Ipsative Scales بأن قيم متوسطات الدرجات في جميع المقاييس تكون متساوية لكل فرد، فمثلاً يمكن أن تكون درجة فرد ما في مقاييس الميل العلمي والكتابي ٧٠ ، ٣٠ بينما

درجة فرد آخر في المقاييسين ٩٠ ، ١٠ على الترتيب، حيث يظل المتوسط ٥٠ لكل منها. ويمكن الحصول على الدرجات التي تُنفي بهذه الخاصية من الاختبارات والمقاييس التي يتبع في بناها أسلوب معين يؤدي مباشرة إلى مثل هذه الدرجات فردية المرجع.

كما يمكن الحصول عليها باستخدام طرق إحصائية لتحويل الدرجات المعيارية أو المئويات وغيرها من الدرجات المُحوَلة إلى درجات تُنْسِب إلى الإطار المرجعي الفردي. فالدرجات الخام في المقاييس الفرعية لاستبيانات الميول تحول أولاً إلى درجات معيارية ، وبذلك يتساوى متوسطها وانحرافها المعياري ، وبعد ذلك يضاف مقدار ثابت إلى درجات كل فرد لكي يصبح مجموع درجات كل منهم متساوياً. فمثلاً إذا كان لدينا أربعة متغيرات ، وكان المجموع الكلى للدرجات فردين في هذه المتغيرات ٢٠٠ ، ٢٨٠ ، فإنه ربما تُضَيِّفَ مقداراً ثابتاً لكل درجة من درجات الفرد الأول في هذه المتغيرات ول يكن ٢٥ ، وتُضَيِّفَ مقداراً ثابتاً آخر لكل درجة من درجات الفرد الثاني في هذه المتغيرات أيضاً ول يكن ٥ بحيث يصبح مجموع الدرجات الكلية لكل منهما في هذه المتغيرات ٣٠٠.

وبذلك تم المقارنة داخل الأفراد، فكل فرد ربما يحصل على درجات مرتفعة في بعض المتغيرات ، ودرجات منخفضة في المتغيرات الأخرى ، أو ربما يحصل على درجات متساوية لقيمة المتوسط في جميع المتغيرات. أي أن الفرد لا يمكن أن يحصل على درجات مرتفعة أو درجات منخفضة في جميع المتغيرات.

وهذا التفسير الذي يُنْسِب إلى الفرد ذاته يمكن الإفادة منه في مواقف متعددة، بل أحياناً يكون ضرورياً، ففى مواقف الإرشاد والتوجيه وما شابه ذلك، حيث يتذكر الاهتمام على الفرد وتحديد ميوله الفعلية والمقارنة بين جوانب قوته وضعفه فى التحصيل والقدرات، تُعد الدرجات التي تُنْسِب إلى الإطار المرجعي الفردى Ipsative Scores أبسط الطرق التي تُروُدنا بهذه المعلومات مباشرة. كما أن هذه الدرجات تُفيد فى أساليب التقدير الذاتى التى تستند إلى استبيانات الاختيار الجبرى Forced - Choice Items من أجل تقليل أثر الاتجاه العقلى المتعلق «بمراجعة القبول الاجتماعى» وغيره من الاتجاهات العقلية فى الاستجابات ، وبذلك تُسهم فى زيادة صدق الاستبيان بدرجة أكبر مما لو استندت الدرجات إلى معايير جماعة مرجعية معينة.

غير أنه ينبغي ملاحظة أن تفسير درجات استبيانات الميول استناداً إلى الإطار المرجعي الفردى لا يعكس مدى عمق تفضيلات الفرد أو مشاعره وهو أمر ضروري فى تفسير استجاباته. فإذا كان فرداً يميلان إلى لعبة رياضية معينة، فإننا نود معرفة أيهما أكثر ميلاً لهذه اللعبة، والاستناد إلى الإطار المرجعي الفردى لا يمكننا من التتحقق من ذلك ؛ لأن هذا الإطار يُفيد فقط فى مقارنة الفرد بنفسه.

لذلك يعترض البعض على التفسيرات التي تستند إلى هذا الإطار، وبخاصة إذا حولنا درجات الفرد إلى معايير مئنية مثلاً، أى إذا دمجنا تفسير الدرجات استناداً إلى كل من الإطارات الفردية والنسبية، وهذا ما يحدث بالفعل في كثير من استبيانات الميول مثل استبيان كيودر للميول المهنية كما سنوضح بعد قليل.

فنقول مثلاً أن الفرد حصل على درجة تناظر الرتبة المئنية ٨٠ في مقاييس الميل الكتابي، أى أنه إذا واجهه مواقف متعددة تمثل مواقف الحياة الواقعية ، فإنه يفضل الأنشطة الكتابية بدرجة تفوق ٨٠٪ من أقرانه ، ولكن إذا حصل فرداً على درجة في الميل الكتابي تناظر الرتبة المئنية ٨٠ فإننا لا نستطيع تفسير ذلك بأنهما متساويان في ميلهما ، وذلك لأن الإطار المرجعي الذي تُسَبِّبُ إليه درجة كل منهما فردي *Ipsative*.

ويرى بعض علماء القياس أن الدرجات التي تُسْتَمدُ من الاستبيانات جبرية الاختيار *Forced - Choice* تكون أكثر تمثيلاً لواقع الحياة من الاستبيانات التي تكون الاختيارات فيها مفتوحة ، فالفرد في حياته يكون مضطراً في غالبية الأحيان إلى الاختيار بين بدائلين من الأنشطة ، كأن يختار أحد نشاطين ترويجيين. غير أن البعض الآخر يرى أن هذا التقييد يعمل على إحباط الفرد عند الاختيار ، وبخاصة إذا كانت جميع الاختيارات إما مرغوبة أو غير مرغوبة.

أساسيات بناء استبيانات الميول المهنية :

يتطلب بناء استبيان الميول المهنية *Interest Inventory* اتخاذ بعض القرارات الأساسية قبل البدء في عملية البناء نوجزها فيما يلى :

(١) ينبغي تحديد الهدف من الاستبيان وتبرير بنائه واستخدامه ، حيث إننا قد ذكرنا فيما سبق أن هناك أنماطاً مختلفة للميول هي : الميول التي يعبر عنها الفرد ، والميول التي تتعكس في سلوك الفرد ، والميول التي يُستدل عليها من اختبارات التحصيل ، والميول التي تُقاس بالاستبيانات . وكل من هذه الأنماط يمكن قياسها بطريقة مختلفة ، مما يدعو إلى ضرورة تبرير اختيار نمط الميول التي تُقاس بالاستبيانات مع ملاحظة أن جميع أدوات قياس الميول المهنية تتسمى إلى النمط الأخير.

(٢) ينبغي اتخاذ قرار يتعلق بمستويات ومجالات الميول التي يهدف الاستبيان لقياسها ، حيث إن المهن والوظائف لها مستويات *Levels* و المجالات *Areas* ، فالمهن *Occupations* مثل المحاماة والهندسة والطب والفنون والتجارة تُعد مجالات أو أنشطة متعددة يندرج تحتها العديد من الأعمال . وبعض هذه الأعمال ربما يتطلب الحصول على درجة جامعية أو تدريب متخصص ، أى تكون من مستويات مهنية أو

تخصصية Professional . أما المستويات شبه المهنية Semiprofessional ، فإنها تتطلب عادة دراسة جامعية أو تدريبياً متقدماً، والمستويات المهارية Skilled تتطلب عادة تدريبياً مهنياً، والمستويات غير المهارية لا تتطلب مهارات معينة أو ربما تتطلب قدرًا محدوداً من المهارات .

لذلك ينبغي مراعاة التمييز بين مجالات ومستويات الميول المهنية عند بناء الاستبيانات من أجل اتخاذ قرارات صائبة تتعلق بميول الفرد . فالطلب مثلاً يُعد مجالاً مهنياً يتدرج تحته مستويات مهنية متباينة في متطلباتها التخصصية والمهارية ، مثل الأعمال الفنية الطيبة ، والتمريض وغير ذلك . ويلاحظ أن بعض استبيانات الميول المهنية اعتمدت على تجمعات من المهن الأساسية وضمت مجالات ومستويات هذه المهن في معظم المقاييس المهنية التي اشتملت عليها كما في استبيان سترونج Strong ، وبعضها اعتمد على قياس الميول في المجالات المهنية فقط مثل استبيان كيدر Kuder .

(٣) ينبغي اتخاذ قرار بشأن الأسلوب الذي يتبع في بناء الاستبيان ، فقد ذكرنا فيما سبق أن هناك ثلاثة أساليب : أحدها يستند إلى أسس نظرية ومنطقية ، والثانى يستند إلى بناء فقرات متجانسة ، والثالث يستند إلى التحقق الإмирقي ، وتناولنا كلاً منها بالمناقشة . ونلاحظ أن عدداً محدوداً من استبيانات الميول المقننة استندت في بنائها على الأسلوبين الأول والثانى ، بينما يبدو أن كثيراً من الاستبيانات استندت إلى الأسلوب الثالث ، حيث اختيرت فقرات الاستبيان على أساس محك إميريقي مناسب كما سيتضح بعد قليل .

قياس الميول باستخدام الاستبيانات المقننة :

تُعد الاستبيانات المقننة Inventories أكثر الأدوات استخداماً في قياس الميول ، فمحتوى هذه الاستبيانات يتضمن مجموعات متكاملة من الأنشطة في مجالات متعددة يعتنى باختيارها . وهذا يُمكّن من التصنيف المنظم للفقرات في مجالات ميول متعددة مما يضمن عدم إغفال أي مجال ربما يختاره الفرد ، وبذلك تستند درجات التفضيل إلى أسلوب كمي منظم . وتقدر أوزان لكل فقرة بحيث تكون مميزة لمجال ميل معين عن غيره من المجالات ، ومميزة لمهنة معينة عن غيرها من المهن . كما أنه يمكن تطبيق استبيانات الميول على جماعة مرجعية مماثلة تقارن في ضوئها درجة الفرد في مجالات متعددة ، وكذلك مقارنة درجته بدرجات الأفراد الذين يعملون بالفعل في مهن معينة متنوعة . وبذلك يستطيع الفرد تعرف ما إذا كان ميله مرتفعاً بدرجة كافية تشجعه على الالتحاق بمهنة معينة أو الإعداد لها .

وعلى الرغم من توافر استبيانات الميول المقتنة، إلا أن عددها ليس كبيراً كما في اختبارات الذكاء والاستعدادات والمقاييس الأخرى الشخصية. فقد أشار الكتاب السنوي السادس للقياس العقلي (Buros ، 1974) إلى ٥٥ استبياناً مقتناً للميول تم نشره ومعظمها يقيس الميول المهنية، ومعظم هذه الاستبيانات من نوع الورقة والقلم، وتتطلب من الفرد تقدير أنشطة معينة أو الاختيار بين بدائل أنشطة معطاة.

وسوف نقدم فيما يلى بعض الأمثلة لهذه الاستبيانات الشائعة الاستخدام، والتي أجري حولها كثير من الدراسات، واستند بناء كل منها إلى أحد الأساليب الثلاثة التي سبق أن أوضحتها.

(١) استبيان سترونج - كامبل للميول

The Strong - Campbell Interest Inventory (SCII)

يُعد هذا الاستبيان من أكثر أدوات قياس الميول شيوعاً واستخداماً، وأجرى حوله كثير من الدراسات والبحوث ما يزيد عن نصف قرن من الزمان، وقد سبق أن أوضحتنا في مستهل هذا الفصل التطور التاريخي لهذا الاستبيان . ويعُد سترونج Strong رائداً لحركة قياس الميول ، كما كان بيته Binet رائداً لحركة القياس العقلي (Cronbach ، 1965).

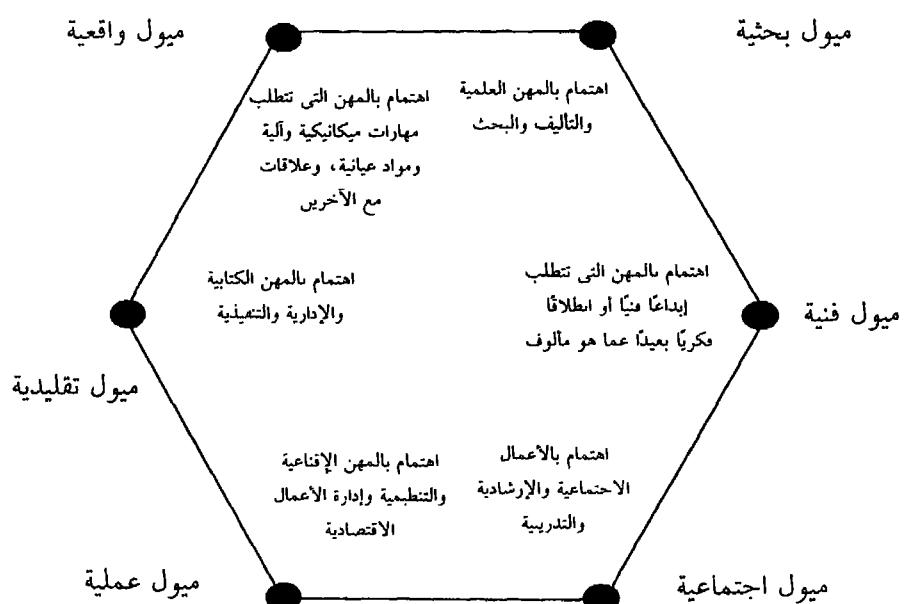
فقد قام سترونج Strong وبعضاً زملائه بفحص دقيق للأنشطة التي يحبها الأفراد الذين يتمون إلى مهن مختلفة، وكذلك الأنشطة التي لا يحبونها. ولا حظوا أن أفراد المهن المختلفة يتباينون في أنماط ميولهم، وربما يرجع ذلك إلى أن الأفراد يختارون المهن التي يميلون إليها. فالرسامون يميلون إلى اللوحات الزيتية ، والذين يعملون في حرفة النجارة يميلون إلى أعمال الأخشاب بدرجة أكبر من الذين يعملون في حرفة الطباعة مثلاً، كما لا يحظوا أن أفراد المهنة الواحدة ميولهم متماثلة ويفضلون أنشطة ترويحية واحدة، بل ويميلون إلى قراءة نوع معين من الكتب والمجلات . واستناداً إلى نتائج هذا الفحص بدأ سترونج في إعداد استبيان لمزاجة ميول الفرد بميول مجموعة محكمة من الأفراد Criterion Group الذين يتمون إلى مهن مختلفة يحبونها وناجحون فيها.

لذلك استند سترونج في بنائه للاستبيان إلى أسلوب التحقق الإميريقي Empirical Keying ، أي الأسلوب الثالث الذي سبق أن أوضحتناه. وقد ظهرت الطبعة الأولى من هذا الاستبيان الذي أطلق عليه استبيان سترونج للميول المهنية (SVIB) عام ١٩٢٧ ، وهذه الطبعة كانت خاصة بالذكور ، وصدرت الطبعة الأولى من الاستبيان الخاصة الإناث عام ١٩٣٣ . وتمت مراجعة الطبعة الخاصة بالذكور عامى

١٩٣٨ ، ١٩٦٦ ، والطبعة الخاصة بالإناث عامي ١٩٤٦ ، ١٩٧٩ ، ولم تُنزل الطبعة الخاصة بالإناث اهتماماً مماثلاً لما نالته الطبعة الخاصة بالذكر، وذلك لأن الإناث في ذلك الوقت لم يكونوا جزءاً فاعلاً من القوة البشرية العاملة.

وقد قام كامبل Campbell وهو أحد علماء النفس بمراجعة الاستبيان ومحاولة التغلب على أوجه القصور فيه، ودمج فقرات الطبعتين معاً في طبعة واحدة محايدة يمكن تطبيقها على كل من الذكور والإناث، وأطلق عليه استبيان ستروننج - كامبل للميول (SCII) عام ١٩٧٤، وصدرت طبعة حديثة من الاستبيان عام ١٩٨١، وبذلك تغلب على مشكلة التحيز للذكور. وقد بيّنت البحوث أن الميول المهنية لا تستقر قبل عمر ١٧ أو ١٨ عاماً أو ربما بعد ذلك.

لذلك فإن هذا الاستبيان يُطبق عادة على الراشدين، ولكن يمكن تطبيقه على الشباب صغار السن حيث إن صياغته بسيطة، وهو غير موقوت، ولكن تطبيقه يستغرق ساعة واحدة تقريباً. ونظراً لأنه كان يُعاب على الاستبيان عدم استناده إلى إطار نظري يفسر تفضيلات الأفراد الذين يتمون إلى مهن معينة، فقد حاول كامبل الاسترشاد بالنظام التصنيفي للميول المهنية للأفراد الذي اقترحه هولاند Holland عام ١٩٧٥. ويرى هولاند أن الميول تُعبّر عن شخصية الفرد المميزة، واقتصر تصنيف الميول المهنية في ستة أقسام يوضحها الشكل التخطيطي (١١ - ٤) التالي (Holland, 1973).



شكل (١١ - ٤) يوضح تصنيف هولاند للميول المهنية

وقد أوضحت الدراسات أن هذا النظام التصنيفي أفضل من غيره من الأنظمة التصنيفية للمهن في وصف الأنشطة المتعلقة بالأعمال المختلفة، والمتطلبات العامة للتدريب، والحوافز المهنية. ولعل اتفاق هذا التصنيف مع أنماط الميول المهنية التي اهتم بها سترونج في استبيانه الأصلي هو الذي شجع كامبل على الاسترشاد به في الاستبيان.

ويشتمل الاستبيان في طبعته الحديثة على ٣٢٥ فقرة تم تجميعها في سبعة أقسام من حيث المحتوى كالتالي (Campbell & Holland 1981) :

(١) المهن (١٣١ فقرة)، مثل : المحامي الجنائي، الكاتب الحر، الممثل، فيأشعة.

(٢) المجالات الدراسية (٣٦ فقرة)، مثل : الأدب، الاقتصاد، الفنون.

(٣) الأنشطة (٥١ فقرة)، مثل : تججير العظام، الطهي، مشاهدة فيلم طبي.

(٤) الأنشطة الترويحية (٣٩ فقرة)، مثل : الألعاب الرياضية، الاستماع إلى الموسيقى الكلاسيكية، صيد السمك.

(٥) أنماط من الناس (٣٤ فقرة)، مثل : كبار السن، أفراد انفعاليين، مغايرين.

(٦) التفضيل بين نشاطين (٣٠ فقرة)، مثل : التفضيل بين العمل الحكومي والعمل الحر، أو بين قراءة كتاب ومشاهدة فيلم.

(٧) الخصائص المميزة للفرد (١٤ فقرة)، مثل : استطاعة الفرد تصميم إعلان متميز، تكوين صداقات بسهولة.

ويجب الفرد بوضع علامة مقابلة لكل فقرة توضح ما إذا كان يحب النشاط المحدد، أو لا يحبه، أو لا يهتم به.

وفيما يلى مثال من كل قسم من الأقسام السبعة السابقة :

<u>لا أحب</u>	<u>لا يهتم</u>	<u>أحب</u>	
...	مهندس معماري
...	الاقتصاد
...	مشاهدة فيلم طبي
...	صيد السمك
...	الرحلة

<u>أفضل</u>	<u>محابٍ</u>	<u>لا أفضّل</u>
أشاهد مباراة كرة قدم مقابل الذهاب إلى متحف
يمكنتني الترويج عن الآخرين	نعم	لا
...

ومن الجدير بالذكر أن طبعة عام ١٩٨١ أضافت عدداً من المهن حيث بلغ عددها ١٦٢ مهنة، كما أن الجماعة المرجعية المحكمة لكل مهنة أصبحت ممثلاً بعينة قومية، وروعيت بعض الاحتياطات التي تُيسّر تفسير الدرجات. فقد كان يعبّر على الاستبيان أن أفراد المجموعات المحكمة أكبر سنًا من الأفراد المبتدئين في المهنة. لذلك روعي في هذه الطبعة المقارنة بين صغار السن والكبار في كل مجموعة محكمة لتحديد ما إذا كانت الميول والقيم للمبتدئين تختلف عن ميول وقيم الذين قضوا في العمل عدداً من السنوات. كذلك اشتملت هذه الطبعة على معايير خاصة بالذكر والإثاث لكل من هذه المهن. ويتم تصحيح الاستبيان وتقييم درجاته وإعداد البروفيل باستخدام حاسوب مركزي، وتحول درجات المقاييس الفرعية التي يشتمل عليها الاستبيان إلى درجات معيارية.

ويتضمن تقرير النتائج ما يلى:

(١) ستة توجهات مهنية عامة: تصف النمط أو التوجه العام للميول المهنية للفرد كما هو موضح بالشكل التخطيطي (١١ - ٤)، فالتوجهات المتقاربة معًا في الشكل التخطيطي المذكور تكون أكثر تماثلاً من التوجهات المتباينة، وتقدر درجة معيارية لكل نمط أو توجه.

(٢) الميول الأساسية: تصف مجموعات من الأنشطة الأكثر تحديداً والمرتبطة بالتوجهات المهنية العامة الستة، وعدد مقاييسها الفرعية ٢٣. فمثلاً ترتبط أنشطة التدريس، والخدمة الاجتماعية، والرياضيات، والفنون الشعبية بالتوجه أو الميول الاجتماعية. وتتوسط الميول الأساسية بين التوجهات المهنية والميول المهنية الأكثر نوعية.

فهي تصف تجمعات أو مجالات من الأنشطة التي يحبها الفرد ، وهذه المجالات ربما تشتراك فيها كثير من المهن النوعية .

(٣) الميول المهنية : تصف مدى تماثل نمط ميول الفرد مع ميول الأفراد الناجحين في مهن معينة عددها ١٦٢ مهنة، مثل : مزارع ، أو ضابط طيار ، أو مسئول إعلانات ، أو فنان ، أو معلم رياضيات وعلوم . ويحصل الفرد على درجة معيارية واحدة لكل مهنة.

(٤) ميول غير مهنية : يوجد مقاييسان فرعيان للميول غير المهنية Nonoccupational بما : مقاييس التوجه الأكاديمي ، ومقاييس الانطواء - الانبساط . ويقيس الأول مدى حب الفرد للأنشطة العقلية ، ويقيس الثاني مدى حب الفرد للتفاعل مع الآخرين والعمل معهم في مقابل حبه للعمل منفردًا . وقد استند في بناء هذين المقاييس على محركات غير مهنية . فقد قورن في المقاييس الأول بين النسبة المئوية المناظرة لاستجابات طلاب الكليات المتفوقين في دراستهم والنسبة المئوية المناظرة للطلاب الضعفاء . وقورن في المقاييس الثاني بين الطلاب الذين أمكن تحديدهم بأنهم منطعون أو منبسطون باستخدام استبيان الشخصية متعدد الأوجه (MMPI) . وقد بيّنت الدراسات أن هناك فروقاً كبيرة بين متosteطات الدرجات في كل من المقاييس في الاتجاه المطلوب بين المجموعات التي تتسمى إلى مهن مختلفة .

ويُستخدم هذا الاستبيان في معظم الأحيان في أغراض الإرشاد حيث يتطلب متخصصاً لديه خبرة في تفسير نتائجه تفسيراً مستثيراً ، وبخاصة أن نتائجه تكون في شكل بروفيل أو صفحة نفسية متعددة الجوانب يُعدّها الحاسوب . ولتسهيل تفسير النتائج أوكل إلى عدد من المؤسسات المتخصصة بالولايات المتحدة الأمريكية ، مثل نظم الحاسوب القومية National Computer Systems ، وشركة Mincomp لإعداد تقارير متعددة الصفحات تتعلق بتفسير نتائج الاستبيان المتمثلة في البروفيل وتقديم المقترنات للعميل والمرشد .

وتكون مهمة المرشد عندئذ هي معاونة العميل في تحديد الاتساق بين الدرجات في مختلف أقسام المقاييس ، أو عدم الاتساق بين درجات كل من مقاييس الميول المهنية ، ومقاييس التوجهات المهنية العامة . ولعل من الأفضل استخدام هذا الاستبيان إلى جانب المعلومات التي يكشف فيها الفرد عن ميوله لفظياً (الميول التي يعبر عنها الفرد) ، وكذلك مؤشرات عن قدراته باستخدام اختبارات القدرات العقلية التي اعنى بتقنيتها .

(٢) استبيانات كيودر للميول : The Kuder Interest Inventories

أعد كيودر Kuder مجموعة من الاستبيانات التي تقيس الميول بتوجهات تختلف عن التوجهات التي استند إليها استبيان سترونج - كامبل ، فقد أعد كيودر استبيان

التفضيلات المهنية Kuder Vocational Preference Record عام ١٩٣٩ ، وظهرت طبعته الثانية عام ١٩٦٤ ، وأطلق عليه استبيان مسح الميول العامة . Kuder Form E General Interest Survey (GIS)

ويُطبق على طلاب المدارس الثانوية والكلية والراغبين من الجنسين . واستند في بنائه إلى انتقاء فقرات متتجانسة Homogenous Keying (الأسلوب الثاني) حيث تم تجميع فقرات كل مجال عريض من مجالات الميول ، مثل الميول الأدبية ، أو الميول العلمية ، وغيرها في مقياس خاص به ، وتطبيق الفقرات على عينة تجريبية من الأفراد ، وإيجاد ارتباط درجات كل فقرة بالدرجات الكلية في المقياس الذي تتنمي إليه ، واستبعاد الفقرات التي تكون ارتباطاتها منخفضة .

وبذلك أمكن الحصول على عشرة مقاييس متعددة اتساعاً داخلياً تتعلق بالميول الخلوي ، والميكانيكي ، والحساسي ، والعلمي ، والإقناعي ، والفنى ، والأدبى ، والموسيقى ، والخدمة الاجتماعية ، والكتابى . وتحوّل درجات الفرد في كل من هذه المقاييس إلى رتب مئانية بحيث يمكن مقارنة نمط ميوله بنمط ميول الأفراد الذين يحبون المجال الذي يعملون فيه . واستخدم كيودر فقرات جبرية الاختيار يشتمل كل منها على ثلاثة بدائل لأنشطة Choice Triad Items - Forced ، حيث يقدم ثلاثة أنشطة يختار الفرد من بينها النشاط الذي يحبه بدرجة أكبر ، والنشاط الذي يحبه بدرجة أقل ، مثل :

(أ) أكتب خطاباً لأحد الأصدقاء .

(ب) أقوم بإعداد ترتيبات حفل .

(ج) أقوم بتحصيل الفواتير المتأخرة السداد من العملاء .

وهذا يختلف عن استبيان سترونج - كامبل الذي استخدم فقرات تتطلب تحديد ما إذا كان الفرد يحب نشاطاً معيناً أم لا يحبه ، أي أن الاختيار هنا مطلق ، بينما في استبيان كيودر يكون التفضيل نسبياً ، وبذلك نحصل في النهاية على تراتيب للتفضيلات ، وهو ما أطلقنا عليه فيما سبق القياس الذي يُناسب إلى إطار مرجعي فردي Ipsative Measurement ، أي أن درجة الفرد في أحد المقاييس التي يشتمل عليها الاستبيان تتأثر بدرجاته في المقاييس الأخرى . فإجراءات التصحيح وتقدير الدرجات في هذه الحالة ترتب الميول من حيث شدتها النسبية داخل الفرد نفسه ، ولا تُعطى أي مؤشرات عن الشدة المطلقة لميوله ، فالفرد الذي لا يهتم بعامة بمجالات معينة ولكن لديه ميل قوى نحو مجال واحد سوف يحصل على درجة مرتفعة في هذا المجال .

لذلك فإنه على الرغم من ميزات انتقاء فقرات الاستبيان على أساس منطق واتساق داخلي، إلا أنه لا ينبغي إغفال المشكلات المتعلقة بتفسير الدرجات في هذه الحالة. كما أن استخدام البداول الثلاثة في الاختيار تؤدي أيضاً إلى مشكلات أخرى، أحياها ربما لا يفضل الفرد الأنشطة الثلاثة التي تمثلها كل من هذه البداول، وأحياناً أخرى ربما يفضل هذه الأنشطة جميعها، مما يؤدي إلى إرباك الفرد واستغراف وقت كبير في الاستجابة.

وقد أعد كيودر بعد ذلك استبياناً آخر أطلق عليه «استبيان كيودر لمسح الميول المهنية» (KOIS) The Kuder Occupational Interest Survey في أواخر السبعينيات وتمت مراجعته وأعيد طباعته عام ١٩٨٥ (Kuder & Diamond, 1985)، ويستخدم مع الأفراد من سن أصغر حيث توجد معايير للسن من ٦ - ١٢. وقد استند في بنائه إلى أسلوب التتحقق الإمبريقي مثل استبيان ستروننج - كامبل بدلاً من أسلوب انتقاء الفقرات المتتجانسة الذي استند إليه في استبيانه السابق. ولكن استبقى على نظام الفقرات جبرية الاختيار ذات البداول الثلاثة كما سبق.

ويختلف هذا الاستبيان عن استبيان ستروننج - كامبل في أن الدرجة في كل مقياس يقيس ميلاً مهنياً معيناً تمثل بالارتباط بين نمط ميل الفرد ونمط أفراد جماعة مرجعية في مهنة معينة، فقد أوجد كيودر معامل الاقتران بين استجابات الفرد لفقرات الاستبيان ونسبة عدد الأفراد في مجموعة مهنية معينة الذين استجابوا للفقرات. فكلما زادت قيمة معامل الاقتران دل ذلك على اقتران أو تمايز درجة الفرد بنمط ميول هذه المجموعة.

وقد تم تقيين الاستبيان على عينة قوامها ١٢٠٠ طالب بالسنة الأولى بالكلية المختلفة في الولايات المتحدة الأمريكية، واشتمل الاستبيان في طبعته الأخيرة على ٧٧ مقياساً للميول المهنية، ٤٩ مقياساً للميول نحو تخصصات أكاديمية للذكور، ٥٧ مقياساً للميول المهنية، ٢٧ مقياساً للميول نحو تخصصات أكاديمية للإناث، والاستبيان غير موقوت، ويستغرق تطبيقه ساعة واحدة تقريباً. ولتصحيح الاستبيان وتقدير درجاته ينبغي إعادة الاستبيان بعد استيفائه إلى الناشر لإجراء ذلك باستخدام الحاسوب.

وقد بيّنت الدراسات أن درجات هذا الاستبيان أكثر استقراراً من درجات استبيان ستروننج - كامبل في أغراض إرشاد الطلاب للمهن غير المتخصصة، وتوارحت قيم ثبات درجات استبيانات كيودر بين .٧٠، .٩٠، .١٠، وعلى الرغم من أن الثبات المتعلق بتقديرات الاتساق الداخلي يشير إلى اتساق بروفيل الميول في مدى زمني قصير، إلا أنه ربما لا نستطيع توقع اتساقها على مدى زمني طويل ما لم تبلور ميول الفرد.

لذلك يلاحظ تذبذب برو菲ل ميول الطلاب الذين تقل أعمارهم عن ١٧ أو ١٨ عام في كل من استبيان كيودر واستبيان سترونج (1969 , Crites) . و تستقر الميول نسبياً لدى طلاب الكليات ، ويزداد استقرارها في نهاية المرحلة الجامعية (Campbell , 1966)

(٣) استبيان لـ- ثروب للميول المهنية

؛ Lee - Thrope Interest Inventory

على الرغم من أن هذا الاستبيان يقتصر استخدامه في البحوث نظراً لعدم وجود أدلة كافية تتعلق بثبات درجاته وصدقه، إلا أنها آثروا أن نشير إليه بـيجار لأنها استندت في بنائه إلى الأسلوب المنطقي (الأسلوب الأول) وبذلك تكون قد قدمتنا أمثلة لاستبيانات استندت في بنائها إلى كل من الأساليب الثلاثة التي سبق أن أوضحتها. ومن الجدير بالذكر أن هذا الأسلوب المنطقي محدود الاستخدام في مجال بناء المقاييس بعامة.

وقد أعد هذا الاستبيان لي Lee ، وثروب Thrope وأعيد طباعته بعد مراجعته عام ١٩٥٦ ، ويتضمن مستويين : أحدهما يناسب طلاب المرحلة الثانوية ، والآخر مستوى متقدم يناسب الراشدين ، وهو استبيان غير موقوت ويستغرق تطبيقه ٤٥ دقيقة تقريباً.

وقد قام المؤلفان بفحص دليل عناوين المهن الذي أعدَّ مكتب خدمات التوظيف بالولايات المتحدة الأمريكية ، وقاموا بتصنيف المهن في ستة أقسام عريضة أو مجالات ميول وهى : شخصي - اجتماعي ، طبيعي ، ميكانيكي ، تجاري ، فنون ، عا.م. وأعدوا استبياناً يمكن باستخدامه الحصول على درجات يمكن تصنيفها في مجموعات ثلاثة هى : مجالات الميول ، وأنواع الميول ، مثل الميل اللغظى والميل اليدوى والميل الحسابى ، ومستوى الميول . وقاموا ببناء فقرات لكل من هذه الأقسام الستة تمثل المستويات الثلاثة للميول ، أي ميل مرتفع ومنخفض ومتوسط . واشتمل الاستبيان على ١٢ زوجاً من الفقرات من نوع الاختيار الجبرى ، وكما ذكرنا فإن هذا الاستبيان يحتاج إلى دراسات للتحقق من خصائصه السيكومترية .

وبالطبع توجد استبيانات أخرى كثيرة لقياس الميول ، وسوف نشير إلى بعض منها فقط فيما يلى :

(٤) استبيان جاكسون لمسمح الميول المهنية

؛ The Jackson Vocational Interest Survey (JVIS)

ويُعد استبياناً حديثاً نسبياً حيث صدرت طبعته الثانية عام ١٩٨٢ ، وأشار مؤلفه (Jackson , 1982) إلى أنه أفاد من التطورات التقنية لإحداث تكامل بين الاختبارات والأساليب المتقدمة في بنائها .

واستند في بنائه إلى تصنیف هولاند الموضّع بشكل (٤ - ١١) . ويشتمل الاستیان على فقرات من نوع الاختیار الجبّری ، حيث يختار الفرد بين بدیلین يتعلقان بأدوار عمل Work roles ، أو أسلیب عمل Work Styles ، ونتیجة ضمّ البیانات يمكن الحصول على درجات معيارية لعدد يبلغ ٣٤ مقیاساً يتصلق ٢٦ منها بأدوار العمل ، و ٨ بأساليب العمل ، وقد صُممَت هذه المقایيس لمقارنة المیول المهنية للفرد ب Miyol جماعات مرجعية من الذكور والإناث ، وإعطاء صورة ممثّلة للتوجهات المهنية العامة للفرد في ضوء تصنیف هولاند ، ويُعد تقریر بروفیل النتائج لكل فرد باستخدام الحاسوب . وما يميز هذا الاستیان أن مؤلفه قد العدید من الأدلة عن صدق التکوين الفرضی ، وبذلك ربما يصبح هذا الاستیان من الأدوات المهمة في قیاس المیول المهنية .

(٢) استیان میتسوتا للمیول المهنية

: The Minnesota Vocational Interest Inventory (MVII)

أعدَّ هذا الاستیان Clark ونشر عام ١٩٦٥ ويمكن باستخدامه الحصول على معلومات منظمة تتعلق بأنماط المیول المتعلقة بالذكور في المهن غير التخصصية أو الحرف التي تتطلب مهارة . وقد أُعدَّ هذا الاستیان في ضوء نتائج البحوث التي أجريت لتوسيع نطاق استیان سترونچ واستیان کیودر ، حيث إنها لا يناسبان المهن ذات المستويات الأدنى . وهذا الاستیان يناسب الذكور الذين تبلغ أعمارهم ١٥ عاماً أو أكثر ، ويستغرق تطبيقه ٤٥ دقيقة . ويمكن الحصول على درجات تتعلق بحرف مختلفة يبلغ عددها ٢١ حرفة أو صنعة مثل: في طباعة ، خباز ، نقاش ، كهربائي وغير ذلك . واستند في بنائه إلى الأسلوب الإمبريقي ، وذلك بمقارنة استجابات الأفراد الذكور الذين يتمون إلى كل من هذه الحرف بالعاملين الحرفيين بعامة .

بالإضافة إلى أن الاستیان يتضمن تسعة مقایيس للمیول العامة الميكانيكية ، والخدمات الصحية ، والأعمال المكتبة ، والإلكترونيات ، والخدمات الغذائية ، وأعمال التجارة ، والمبيعات ، والأعمال الخلوية وغيرها .

وهذه المقایيس تشتمل على مجموعات من الفقرات المتراقبة ، وتعتمد معايير الاختبار على الدرجات التالية حيث تُعَدُّ الدرجة ٥٠ عن متوسط درجات المجموعة المحكمة في كل من المقایيس المتعلقة بالحرف المختلفة . وتتراوح قيم ثبات درجات هذا الاستیان بطريقة الإعادة بمدة زمنية فاصلة تبلغ شهراً بين ٦٢ ، ٨٨ ، ..

(٣) استبيان أوهايو لمسح الميول المهنية

: The Ohio Vocational Interest Survey (OVIS)

أعدَّ هذا الاستبيان دكوستا D'Costa وواينفوردنر Winefordner بجامعة أوهايو عام ١٩٦٦ ، وتم تقنيته عام ١٩٧٩ ، وقد صُمم لكي يستخدم في إرشاد طلاب المرحلة الثانوية في اتخاذ قرارتهم المهنية، ويستغرق تطبيقه ما بين ساعة وساعة ونصف. وقد استند الاستبيان إلى دليل عناوين المهن الذي أشرنا إليه سابقاً لتصنيف المهن. وأعد المؤلفان في ضوء ذلك ٢٤ مقياساً فرعياً للميول عين لكل منها رمز من ثلاثة أرقام يشير الرقم الأول إلى «بيانات Data» ، والرقم الثاني إلى أفراد People ، والرقم الثالث إلى «أشياء Things» كما وضحة الدليل المذكور في شكل مكعب . فمثلاً العمل المتعلق بالأعمال الفنية الطبية يرمز له بالرمز (٢٢٢) حيث تشير هذه الأرقام إلى أن العمل يتطلب التعامل مع البيانات والأفراد والأشياء بدرجة كبيرة ، نظراً لأن الأرقام المناظرة لكل بُعد من الأبعاد الثلاثة للمكعب (بيانات ، وأفراد، وأشياء) تتراوح بين صفر (لا يوجد تعامل) ، ٢ (تعامل بدرجة كبيرة).

ويشتمل كل من هذه المقاييس على ١١ فقرة تشير إلى أنشطة يستجيب لها الفرد. ولكن تظهر الفقرات منفردة ، وليس ثلاثيات كما في استبيان كيودر ، ولكل فقرة ميزان متدرج مكون من خمس نقاط (أحب جداً، أحب، محابيد، لا أحب، لا أحب على الإطلاق). ومن أمثلة هذه الفقرات : (أزرع الخضرروات، أرسم رسوماً للأطفال، أكتب برامج للحاسوب ، أصلاح المناضد والمقادع)، وتصحيح الاستبيان مماثل لتصحيح استبيان كيودر.

وتعكس نتائج الاستبيان الميول النسبية للفرد في مختلف الحرف ، كما تعكس ميله لكل حرف مقارنة بميل العاملين فيها من الجماعة المرجعية مما يساعد في إلحااق الأفراد ببرامج التوعية المهنية. وقد استمدت معايير الاختبار من عينة قومية تشمل على ... ٤٥ طالب في المرحلة الثانوية بالولايات المتحدة الأمريكية.

(٤) استبيان جايسن المصور للميول

: The Geist Picture Interest Inventory (GPII)

يختلف هذا الاستبيان عن الاستبيانات السابقة في أنه لا يعتمد على الجوانب اللفظية ، وإنما يعتمد على الصور ، وقد نُشر أول مرة عام ١٩٥٨ . والاستبيان يناسب الأفراد الذين لديهم صعوبات في القراءة ، ويقتصر استخدامه على الذكور ، ويمكن تطبيقه فردياً أو جماعياً. ويشتمل على ٤٤ فقرة مصورة يتضمن كل منها بدائل ثلاثة من الصور المتعلقة بمختلف المهن والهوايات . وبلي كل فقرة ثلاثة سؤال يجيب عنه الفرد باختيار البديل الذي يفضل له.

ويذكر جايسن (Geist 1959) مؤلف الاستبيان، أنه يمكن استخدامه في المواقف التي تتطلب تخطيئاً و اختياراً مهنياً، و دراسة أنماط القوى الدافعية المرتبطة بالميول، وكذلك دراسة الميول ذاتها. ويقترح أنه يمكن الإفاده من الاستبيان في معاونة المعلمين اختيار أنشطة تعليمية مفيدة، ورثما أيضاً في دراسة أسباب اختيار الأطفال هوايات محددة، و اختيار الكبار مهناً معينة. ويمكن باستخدام الاستبيان الحصول على ١١ درجة للميول، ٨ درجات للدافعية. وتشير درجات الميول إلى: الميل الإقناعي، والكتابي، والميكانيكي، والدرامي، والموسيقي، والعلمى، والخلوى، والأدبى، والحسابى، والفنى، والخدمة الاجتماعية. أما درجات الدافعية فتشير إلى : الأسرة، المركز الاجتماعى، العائد المادى، الدافع الذاتى، الشخصية، الوسط البيئى، الخبرة السابقة، لا أعرف.

وقد أُجرى تقييم الاستبيان على ١٢٠٠٠ فرد من جميع المستويات العمرية في الولايات المتحدة الأمريكية. وهاراي، وبورتوريكو. وتعتمد المعايير على الدرجات التالية لكل مجال من مجالات الميول المذكورة، و تمثل هذه الدرجات في بروفييل لكل فرد يوضح مستويات الميول وقوه الدافعية. وأوضحت الدراسات أن قيمة معامل ثبات درجات الاستبيان بطريقة الإعادة بمدة زمنية فاصلة بلغت ستة أشهر لعينات من طلاب المدارس الثانوية والكليات قد تراوحت بين ٣٧، ٩٤، . لجميع المقاييس الفرعية، أما صدق الاستبيان فيتطلب مزيداً من الدراسات.

تحقيق عالم على استبيانات الميول :

اتضح لنا من الأمثلة السابقة لاستبيانات الميول أنها استندت في بنائها إلى أساليب متباعدة و اشتملت على فقرات مختلفة، واعتمد بعضها مثل استبيان كيدور على أنواع الفقرات التي تتطلب من الفرد اختيار الجبرى من بين بديلين أو ثلاثة، أى أن الفرد يقوم بترتيب البديل بحسب درجة تفضيله لها، وهذا يعني أنه كلما قبيل الفرد مجال ميل معين ينبغي عليه رفض مجال ميل آخر، إذ لا يمكن أن يقبل جميع المجالات.

واعتمد البعض الآخر مثل اختبار سترونج على فقرات تتطلب استجابة غير مقيّدة، وذلك بأن يوضح الفرد جبه لنشاط معين أو عدم جبه له. غير أن هذا النوع من الفقرات أكثر تأثراً بالاتجاهات العقلية في الاستجابات من النوع السابق على الرغم من أن تفسير درجات الاستبيان الذي يشتمل على فقرات جبرية الاختيار يستند إلى الإطار المرجعي الفردي Ipsative Scores، مما يؤدي إلى صعوبات في تفسير هذه الدرجات كما سبق أن أوضحنا.

وبعض الاستبيانات مثل استبيان سترونج، واستبيان مينيسوتا، واستبيان كيودر في طبعها الحديثة استندت في بنائها إلى أسلوب التحقق الإمبريقي، حيث استخدم كل من الاستبيان الأول والثاني الناس بعامة كمجموعة متحكمة.

بينما قارن الاستبيان الثالث الأفراد الذين يتمون إلى مهنة معينة بغيرهم في المهن الأخرى. أما استبيان سترونج لمسح الميول العامة فقد استند إلى أسلوب انتقاء فقرات متتجانسة، واستند استبيان لي - ثروب للميول المهنية في بنائه إلى الأسلوب المنطقي. وبعض الاستبيانات مثل استبيان سترونج واستبيان كيودر اشتتملا على أنشطة وميول تتعلق بالمهن والحرف التخصصية.

بينما تركز استبيان مينيسوتا على الأنشطة والميول المتعلقة بالمهن غير التخصصية أو الحرف التي تتطلب مهارة، أما استبيان أوهايو لمسح الميول المهنية فقد اعتمد على دليل عناوين المهن، و Ashton على العديد من الأنشطة التي ربما يكون لدى طلاب المرحلة الثانوية معلومات أو خبرة محدودة حولها. وجميع استبيانات الميول غير موقته، ويستغرق تطبيقها ساعة واحدة تقريباً بطريقة جماعية، ويمكن تطبيقها فردياً. وكثير من هذه الاستبيانات وبخاصة الاستبيانين الرئيسيين لسترونج وكيودر يتم تصحيحهما وتقدير درجاتها باستخدام حاسوب مركزي، وتُعاد النتائج للجهة المرسلة.

وتتبادر استبيانات الميول في مدى سهولة تصحيحها وتقدير درجاتها، - حيث يصعب إجراء ذلك يدوياً في بعض هذه الاستبيانات وبخاصة استبيان سترونج وكيودر. وفيما يتعلق بصدق استبيانات الميول وثبات درجاتها فإن استبيان سترونج يعد أكثر الاستبيانات التي نالت دراسات صدق وثبات متعددة، بيليه استبيان كيودر، وأقلها استبيان جايست المصور للميول. أما فيما يتعلق بتفسير درجات هذه الاستبيانات فقد اتضح لنا أن بعضها يستند إلى الإطار المرجعي الفردي Ipsative نظراً لأنها تشتمل على فقرات جبرية الاختيار مثل استبياني كيودر ومينيسوتا، والبعض الآخر يستند إلى إطار مرجعي نسبي أي إلى معايير الجماعة؛ وعدداً محدوداً استند إلى كل من الإطارات.

ولعل استبيان كيودر هو الاستبيان الوحيد الذي اشتمل على مفاتيح لكشف عدم فهم التعليمات أو الفقرات، وعدم تحري الدقة والأمانة في الاستجابات.

وعلى الرغم من هذا الاختلاف في أساليب البناء، وتقدير الدرجات، وتقدير نتائج مقاييس فرعية متعددة بسميات مختلفة ، إلا أن كول وهانسون (Cole & Hanson , 1961) قد أوضحوا أن معظم هذه الاستبيانات وما تشتمل عليها

من مقاييس فرعية يمكن تصنيفها بسهولة في إطار نموذج هولاند المبين بشكل (٤-١١) الذي يشتمل على ستة عوامل تتعلق بالشخصية الإنسانية والمهن المختلفة . ويرى جيلفورد Guilford ومعاونوه أن استبيانات الميول ربما يمكن استخدامها كمقاييس للشخصية .

واقترحوا أنه لا مجال للشك في أن الميول ينبغي أن ترتبط بالدوافع والطموحات وال الحاجة نظراً لأنها لا تكون أو تنمو في فراغ ، ولعل الدراسات الحديثة المتعلقة بمجال النمو المهني أكدت صدق هذا الاقتراح ; Holland 1973 ; Super : 1972 . Tayler , 1983) .

استخدامات استبيانات الميول :

يمكن استخدام استبيانات الميول استخدامات متعددة في المجالات التربوية والمهنية ، غير أن الإفادة منها في هذه المجالات يعتمد على الغرض من استخدامها والقرارات التي تتخذ في ضوئها . فاستبيانات الميول تعد أحد أدوات قياس خصائص الأفراد وسلوكهم المميز ، لذلك ينبغي مراعاة حدود استخدام هذه الاستبيانات وعلاقتها بأدوات قياس الجوانب المعرفية المتعلقة بالسلوك الأقصى للأفراد . وسوف نوضح فيما يلى أهم أغراض استخدام هذه الاستبيانات ومدى الإفادة منها في كل من هذه الأغراض :

أغراض التوجيه التربوي والمهني : تُستخدم استبيانات الميول بكثرة في أغراض التوجيه والإرشاد التربوي والمهني إلى جانب اختبارات التحصيل والاستعدادات العقلية . إرشاد أو توجيه الفرد إلى الدراسة أو المهنة المناسبة له يتطلب بيانات متعددة لا تحصل عليها من استبيانات الميول بمفردها . غير أن هذه الاستبيانات إذا استُخدمت استخداماً واعياً يمكن أن تزود المرشد التربوي أو المهني بمعلومات أساسية تتعلق بميول الفرد أو تفضيلاته الأكademie والمهنية .

ومع هذا فإن هناك أدلة متعددة تشير إلى أن ما يقرره الأفراد من تفضيلات مهنية ربما يكون غير واقعي ، وبخاصة في بدء المرحلة الثانوية وربما بعد ذلك ، نظراً لأن هؤلاء لا يكرونون على دراية تامة بالأنشطة المتعلقة بالمهن المختلفة ومتطلباتها ، بل إن كثيراً من الشباب ربما لا يكون لديهم خبرة كافية للاختيار من بين مدى واسع من الأنشطة التي تتضمنها الاستبيانات . وتزداد حدة المشكلة إذا استُخدمت هذه الاستبيانات في المرحلة المتوسطة حيث يجد بعض الطلاب صعوبات في فهم المصطلحات المهنية التي يقررون في ضوئها اختيارهم المناسب .

لذلك يمكن أن يستخدم المرشد نتائج هذه الاستبيانات في إلقاء الضوء على بعض المشكلات التي يمكن أن يُقدّم فيها العون للفرد ، مثل عدم اهتمامه بالعمل

الجماعي، أو العلاقات الشخصية، أو الصعوبات الدراسية وغير ذلك. كما يمكنه استخدامها في تعرف مدى اتساق نمط الميول التي عبر عنها الفرد في استجاباته، وما إذا كان البرنامج الإرشادي يناسب بروfil ميوله واستعداداته.

وينبغي عدم اتخاذ قرارات فاصلة استناداً إلى نتائج استبيانات الميول. إذ لا ينبغي أن تكون هذه النتائج هي المحك الأساسي الذي يستند إليه إرشاد الفرد إلى تخصص أكاديمي معين أو مهنة محددة. لذلك يجب على المرشد عند تفسيره لنتائج استبيانات الميول أن يوضح للفرد الفرق بين الميول والاستعدادات.

فكثير من الأفراد يستنتجون أن لديهم القدرات أو الاستعدادات التي تؤهلهم للنجاح في مهنة معينة وذلك لأنهم يميلون إليها، في حين أن الاستبيان لا يوضح سوى أن ميول الفرد تمثل ميول جماعة مجتمعية من العاملين في هذه المهنة، وقد سبق أن ذكرنا أن الارتباط بين الميول والاستعدادات يميل إلى الانخفاض، مما يدل على أن الأفراد في غالبية الأحيان لا يمتلكون القدرات أو الاستعدادات التي تتطلبها الأنشطة المعينة التي يميلون إليها ، وربما يكون العكس صحيحاً . لذلك ينبغي تفسير نتائج استبيانات الميول في إطار معلومات متكاملة عن قدرات الفرد واستعداداته الخاصة وداعيته وتحصيله وسمات شخصيته .

أغراض الانتقاء والتسكين : يختلف استخدام استبيانات الميول في أغراض الانتقاء والتسكين عنه في أغراض الإرشاد والتوجيه، ففي أغراض الانتقاء يتربّط على نتائج الاستبيانات قرارات تمس مستقبل الفرد مباشرة .

فالفرد في هذه الحالة يبغى الالتحاق بمهنة أو وظيفة شاغرة، لذلك فإنه ربما يلتجأ إلى تزييف استجاباته أو جعلها متحيزة للمهنة أو الوظيفة التي تقدم إليها، فقد يبيّن معظم الدراسات أن الاتجاهات العقلية في الاستجابات Response Sets التي سبق أن أوضحتها ربما تلعب دوراً كبيراً في هذه المواقف الانتقائية ويصعب التغلب عليها (Kuder , 1966). فهناك أفراد يحاولون إعطاء انطباع جيد عن أنفسهم باختيارهم نسبة كبيرة من الفقرات المرغوبة اجتماعياً .

وعلى الرغم من أن بعض استبيانات الميول كما رأينا تستخدم فقرات جبرية الاختيار أو تشتمل على فقرات تساعد في التحقق من صدق الاستجابات، إلا أنه لا توجد أدلة بحثية تؤكد فاعلية ذلك في التغلب على مشكلات الاتجاهات العقلية في الاستجابات؛ لذلك لا يوصى خبراء القياس باستخدام استبيانات الميول في أغراض الانتقاء أو على الأقل عدم الاقتصار عليها كأداة قياس وحيدة .

وإذا استُخدمت في أغراض الانتقاء أو التسكين ينبغي مراعاة ما سبق أن ذكرناه في الإرشاد والتوجيه، إذ ينبغي توقيع أن يكون هناك اختلاف بين أنماط الميول الأربع التي أشرنا إليها من قبل، وهي الميول التي يُعبر عنها الفرد، والميول التي تقيسها الاستبيانات، والميول التي يُستدل عليها من الاختبارات الأخرى، والميول الفعلية التي تتعكس في سلوك الفرد. فالفرد يمكن أن يؤدي أنشطة معينة كهواية، ولكنه لا يميل إليها كمهنة. كما أن المصطلحات المتعلقة بالمهن أو تسمياتها ، أو مواصفاتها تختلف باختلاف الثقافات بل وباختلاف الاستبيانات، فالمقياس الفرعى المتعلق بمهنة المزارع في الطبيعة الأولى في مقياس سترونج استُمدت معايره من أفراد تخرج معظمهم من كلية الزراعة في ولايتنين مختلفتين، بينما استُمدت معاير المقياس نفسه في استبيان كيودر من أفراد تخرج نصفهم فقط في كلية الزراعة من ولايات متعددة.

أى أن كلاً منها استند إلى عينة متسقة من خريجي كلية الزراعة، وبخاصة استبيان سترونج. كما أن مواصفات هذه المهنة وغيرها تختلف من ثقافة إلى أخرى. لذلك واجه خبراء القياس في إنجلترا مشكلات متعددة عند استخدام هذين الاستبيانين، نظراً لاختلاف المهن ومواصفاتها عن الولايات المتحدة الأمريكية إضافة إلى المشكلات الأخرى، وللتغلب على ذلك أعدَّ في إنجلترا عدة مقاييس بديلة، وفي استراليا أعدَّ مقياس واحد أطلق عليه مقياس روثيريل - ميلر Rothwell - Miller Interest Test حيث اشتمل على 12 مجموعة من المهن يقوم الفرد بترتيب ميله إليها. كما أن اختبار مهنة معينة يتأثر بتفاعل كثير من المتغيرات والظروف مثل: نوع الفرد وأدواره، وضغوط الأسرة والأصدقاء، والدافعية، وكلفة التدرب على المهنة، والدخل، والقيم الشخصية.

أغراض تربوية: يمكن الإفادة من نتائج استبيانات الميول ، وبخاصة المقاييس البسيطة التي يدها المعلم لطلابه كما سنوضح بعد قليل ، في أغراض تربية تشخيصية وعلاجية ، كما يمكن الإفادة منها كمؤشر في تقويم فاعلية البرامج التعليمية ، فإذا بینت نتائج تطبيق استبيان الميول في مستهل العام الدراسي أن بعض الطلبة لا يميلون إلى مجال دراسي معين أو ينفرون منه ، فإن هذا يمكن أن يكون مؤشراً لبعض الصعوبات التي ربما تكون سبباً في ذلك . وعندئذ يستطيع المعلم أن يعاون هؤلاء الطلبة في التغلب على هذه الصعوبات بتقديم أنشطة محببة لهم في المجال الدراسي الذي لا يميلون إليه .

فمثلاً إذا كان الطالب ينفر من الرياضيات ، ولكنه يميل بدرجة كبيرة إلى الدراسات الاجتماعية ، فيمكن أن يقدم المعلم أنشطة اجتماعية تتطلب قدرًا من

العمليات الرياضية، ويسعى الطالب على القراءة فيها لكي يجري النشاط المطلوب. كما يمكن توظيف ميول الطالب الأكاديمية في أنشطة عملية تتعلق بمهن مرتبطة بميوله في إطار المحتوى الدراسي.

كذلك يمكن في تقويم البرامج التعليمية تطبيق استبيانات ميول مناسبة للمستويات العمرية للطلبة إلى جانب اختبارات الجوانب المعرفية، وذلك في بداية العام الدراسي وقرب منتصفه أو في نهايته. فإذا حدث تحسن ملحوظ في ميول الطلبة، وارتفع مستوى تحصيلهم، فإن هذا ربما يكون مؤشراً على فاعلية البرنامج التعليمي، وهذا يؤكد أهمية الجوانب المعرفية والجوانب غير المعرفية، وتفاعلها وانعكاس أثرها في أداء الطلبة وسلوكياتهم.

تقويم ميول الطلبة في التحفظ الدراسي:

إن استبيانات الميول المقننة التي أوضحتناها فيما سبق تفيد في التنبؤ بالنجاح في مجال مهني أو عمل معين والرضا عن العمل وذلك لتوافر معايير لهذه الاستبيانات يمكن في ضوئها تفسير الدرجات، غير أن المعلم ربما يجد من الأفضل أن يقوم ببناء استبيانات لقياس ميول طلابه إذا لم تكن الاستبيانات المقننة متواقة وغير ملائمة للطلاب. ولعل الجهد الذي يبذل المعلم في بناء استبيانات الميول يمكن الإفاده منه في تقويم الطلبة والتحقق من فاعلية البرنامج التعليمي في التأثير في الجوانب الوجدانية لديهم.

فقياس ميول الطلبة يساعد المعلم في تعرُّف هذه الميول وحفظ الميول المرغوبة وتنميتها، وتقديم مثيرات تساعده في اكتساب ميول جديدة، ومحاولة استبعاد الميول غير المرغوبة، وجعل مجالات الميول جزءاً متكاملاً من المحتوى الدراسي ، وإرشاد الطلبة تربوياً ومهنياً .

لذلك فإن المعلم ينبغي أن يهتم بتقويم ميول طلابه، وأن يدرِّبهم على كيفية إجراء ذلك بأنفسهم. وأن يضع خطة منتظمة لإجراء مسح لميولهم في مستهل العام الدراسي لكي يتعرَّف توجُّه الميول بعامة. وجمع أدلة تتعلق بالعمليات التي تحدث عند محاولة تعزيز الميول المرغوبة والتخلص من الميول غير المرغوبة، وتعرف التغيرات التي تحدث في أنماط الميول لدى الطلبة، وإعادة مسح الميول في نهاية العام الدراسي لتحديد التغيرات المرغوبة التي حدثت، وجمع معلومات عن ميول كل طالب للإفاده منها في التوجيه التربوي والمهني.

ولكى يتحقق المعلم ذلك عليه أن يجمع بيانات دقيقة وموضوعية تتعلق بميل الطلاب للإفادة منها فى أغراض التقويم. وعلى الرغم من أنه يستطيع استخدام بعض استبيانات الميل المقننة المتوفرة مثل التى أوضحتها، إلا أنه ينبغى أن يعتمد بدرجة كبيرة على أدوات القياس التى يُعدها بنفسه لطلابه أو بالمشاركة مع زملائه المهتمين بذلك.

أدوات قياس وتقدير ميل الطلبة في الصف الدراسي:

تتعدد أدوات قياس وتقدير الميل التى يمكن أن يقوم المعلم بتصميمها وبنائها، واستخدامها في الصحف المدرسية، ويمكن أن تكون هذه الأدوات بسيطة مثل مناقشة الطلبة في الفصل حول اهتماماتهم وفضائلهم، والمقابلات الشخصية، وملحوظة سلوكهم، ويمكن أن تكون أكثر تنظيماً ودقة، مثل قوائم المراجعة، والاستبيانات، وترتيب الأنشطة، وموازين التقدير، والموازنات الثنائية، ومؤشرات الأداء، واختبارات المواقف وغيرها من أدوات القياس. وسوف نوضح فيما يلى بعض هذه الأدوات:

قوائم المراجعة: يمكن استخدام هذه القوائم Checklists في تسجيل ميل الطالب وفضائلاته، وكثير من المعلومات المتعلقة بأنشطة معينة؛ لذلك ينبغي اختيار مجالات الميل المراد قياسها وما تتضمنه من أنشطة. ويمكن أن تكون هذه المجالات عريضة بحيث تشمل على أنشطة من الجوانب المختلفة للمنهج الدراسي، أو من جوانب مقرر دراسي معين.

ويبدون اسم الطالب في أعلى القائمة ، والسلوك أو النشاط المطلوب إلى الجانب الأيمن من القائمة، وترك أماكن خالية إلى الجانب الأيسر لكى يتسرى للمعلم وضع علامات أمام كل من هذه الأنشطة بما يراه مناسباً للطالب . ويمكن أن يكون السلوك لفظياً مثل اشتراك الطالب في المناقشة أثناء الدرس ، أو غير لفظي مثل اشتراكه في نشاطات الجمعيات المدرسية ، أو الأنشطة الرياضية .

موازين التقدير: تختلف موازين التقدير Rating Scales عن الاختبارات التي تقيس الجوانب المعرفية في عدم وجود إجابة صحيحة أو خطأ لكل فقرة من الفقرات، ويحدد الطالب ما يحبه أو يفضلها من أنشطة، وما لا يحبه أو لا يفضلها منها، أو يحدد درجة تفضيله لكل منها على ميزان تقييم متدرج يشتمل على خمس نقاط مثلاً كالتالى:



أحب جداً أحب غير متأكد لا أحب لا أحب مطلقاً

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------|
| <input type="checkbox"/> | قراءة كتب علمية |
| <input type="checkbox"/> | حضور محاضرات علمية |
| <input type="checkbox"/> | مشاهدة أفلام علمية |
| <input type="checkbox"/> | إجراء تجارب علمية |

وتقدر درجة لكل تفضيل مقابل لكل فقرة، ومتوسط مجموع الدرجات يدل على التفضيل العام للطالب لمجال ميل معين، وهذا يتطلب أن تكون جميع الفقرات متعلقة بهذا المجال. أما إذا كانت الفقرات تقسيس مجالات ميول مختلفة، فلا يجوز إيجاد المتوسط العام لجميع الفقرات، وإنما يمكن إيجاد متوسط تقديرات جميع طلاب فصل ما في كل فقرة من الفقرات التي يشتمل عليها ميزان التقدير. ويمكن استخدام أساليب أكثر دقة في بناء موازين التقدير سوف نوضحها في الفصل القادم.

المقابلات الشخصية : يستطيع المعلم إجراء المقابلات الشخصية Interviews في الحصول على معلومات تتصف بالسرية والثقة بينه وبين الطالب ويمكن أن تُعد خطة المقابلات الشخصية مسبقاً أو أن تكون غير سابقة الإعداد، وعادة تُجرى المقابلة الشخصية وجهاً لوجه بين المعلم والطالب، وبذلك تفي في الحصول على معلومات عن الطالب تساعد المعلم في إرشاد الطالب وتوجيهه. ولكن يجب التأكد من أن الطالب لا يجيب عن الأسئلة التي يوجهها إليه المعلم بما يعتقد أنه إرضاعه، والمقابلات الشخصية تناسب بدرجة أكبر التلاميذ الصغار الذين لم يتمكنوا بعد من مهارات القراءة والكتابة.

اللحظة السلوكية : يمكن أن يستعين المعلم بطرق الملاحظة Observation للحصول على معلومات صادقة عن ميول الطالب وتفضيلاته لأنشطة معينة. لذلك يجب أن تكون هذه الملاحظات واقعية وكافية قدر الإمكان لكي يستطيع المعلم الإفاده من نتائجها. ويُجدر ملاحظة السلوك التلقائي للطالب في مواقف طبيعية لكي يستجيب الطالب استجابة تعبّر عن ميوله بالفعل. كما يستطيع المعلم أن يستخدم مؤشرات الأداء Performance Indicators في الموقف الذي يود أن يُظهر فيها الطالب ميوله من خلال ما يختاره ويؤديه من أنشطة حرة أو محددة.

ترتيب الأنشطة : يمكن أن يُعد المعلم قائمة بأنشطة متنوعة ويطلب من الطالب ترتيبها Ranking بحسب درجة ميله لكل منها. وينبغي أن يكون عدد الأنشطة التي تتضمنها القائمة مناسباً للمستوى العمري للطالب، فاللهم يُمكنه يُستطيع ترتيب عدد

محدود من الأنشطة، بينما يستطيع الطالب في المرحلة الثانوية مثلاً ترتيب عدد أكبر من الأنشطة المتتجانسة نسبياً. ويمكن أن يُعين الطالب الرقم (١) للنشاط الذي يُفضله بدرجة كبيرة، والرقم (٢) للنشاط الذي يُفضله بدرجة أقل، وهكذا.

ويمكن أن يوجد المعلم متوسط رتب الفصل في كل نشاط متضمن في القائمة، وبذلك يستطيع تقويم ميول طلبة الفصل بعامة وتفضيلاتهم، وينبغى مراعاة أن هذه الأرقام تدل على التفضيل النسبي لكل طالب.

الاستجابات الحرة: يمكن أن يعبر الطالب عن ميوله من خلال استجاباته الحرة Free - Responses لمجموعة من الأسئلة التي تتطلب منه كتابة مقال مثلاً عن هواياته أو الأنشطة التي يفضل ممارستها، أو يمكن تحديد مجال الأسئلة قليلاً بحيث يكمل الطالب جملة ناقصة أو سؤالاً يتطلب إجابة قصيرة مثل:

أفضل قضاء وقت فراغي في

أميل إلى مادة

أحب الأنشطة التي تتطلب

ما الأنشطة التي تناول اهتمامك غالباً؟

ويتطلب بناء معظم الأدوات السابقة تحديد مجالات الميول أولاً، وتحليلها إلى مكوناتها السلوكية بحيث يمكن قياسها وتسجيل نتائجها بحيث يُستدل منها على إيجابية الميل أو سلبيته أو درجته. فمثلاً إذا كان المعلم مهتماً بتقويم ميل الطالب نحو دراسة العلوم مثلاً، فإنه يمكنه تحليل هذا الميل إلى مجموعة من الأنشطة التي يمارسها الطالب الذي لديه الميل ، مثل:

(١) يحرص على حضور ومتابعة درس العلوم.

(٢) يتلقى الكتب التي تعالج موضوعات علمية عند زيارته لمكتبة المدرسة.

(٣) يقضي وقت فراغه في قراءة المجالات التي تتناول موضوعات علمية.

(٤) يحرص على المشاركة في الرحلات العلمية التي تنظمها المدرسة.

(٥) يمارس هوايات علمية.

(٦) يشتراك في الجمعية العلمية بالمدرسة.

(٧) يشارك بانتظام في أنشطة الجمعية العلمية.

ويمكن أن يسجل المعلم ملاحظاته عن كل طالب في قوائم أو بطاقات ملاحظة تضم لهذا الغرض.

وفي ختام هذا الجزء ينبغي الإشارة إلى أن مقاييس الميول تُستخدم أيضاً في البرامج التربوية التي تهدف لتنمية الوعي المهني لدى الطلبة في المرحلة الثانوية، وربما قبل ذلك. وتهتم الدول المتقدمة بتصميم برامج منظمة للتوجيه التربوي والمهني تستند إلى نظم الحاسوب. وتُعد مقاييس الميول أحد مكونات هذه البرامج، وتهدف هذه البرامج لتنمية مهارات الطلبة في اتخاذ القرارات المهنية، والتضييق المهني، والتأكيد من مناسبة الاختيار المهني.

ويُعد ذلك من المجالات البحثية النامية حيث تتناول مجالات قياس إدراك الطالب لحاجاته وعلاقته بعملية اتخاذ القرارات المهنية، ومدى تأكده من اختياره المهني، والمشاعر المتعلقة بكفاية الطالب وفاعليته، وثقته في المهام والسلوك المرتبط باتخاذ القرارات المهنية، وتصميم أدوات قياس لهذه المجالات.

خلاصة

تُعد الميول من الجوانب غير المعرفية للشخصية، والتعرف على ميول الأفراد يُعد من الأمور المهمة التي تُوجه في ضوئها البرامج التعليمية والتدريبية لتنمية الميول المرغوبة، وبخاصة الميول المهنية. وتتنوع أنماط الميول وأساليب قياسها وتقديرها. فالميول يمكن أن يعبر عنها الفرد شفاهة أثناء المقابلة الشخصية أو خلال مواقف الحوار، ويمكن أن تعكس في سلوكه بحيث يمكن ملاحظة السلوك وتسجيله، ويمكن أن يستدل عليها من اختبارات التحصيل بافتراض أن التحصيل المرتفع في مجال دراسي معين يرتبط بميل الطالب نحو هذا المجال، وإن كانت الدراسات لم تؤكِد صحة هذا الافتراض، نظراً لتأثير التحصيل بمتغيرات أخرى متعددة. كما تقيس الميول باستخدام الاستبيانات المقنتة التي يمكن بناؤها وفق أساليب متعددة تستند إلى أسس نظرية أو منطقية، أو بناء فقرات متجانسة، أو إلى التحقق الإميريقي. وقد قدمنا أمثلة متعددة لاستبيانات مقننة استند بناؤها إلى كل من هذه الأساليب، ومن أهمها استبيان سترونج - كامبل، واستبيان كيدر وغيرها.

وعلى الرغم من المشكلات التي تتعلق بهذه الاستبيانات، وبخاصة مشكلات الاتجاهات العقلية في الاستجابات مثل التزيف، والإذعان أو الموافقة، ومراعاة القبول الاجتماعي، والاستجابات الشاذة، ومشكلة تحيز الاستبيانات للذكر، وغير ذلك، فإن هذه الاستبيانات المقنتة تُعد أدوات قياس أساسية في برامج الإرشاد والتوجيه التربوي والمهني، ويُستفاد منها إلى حد ما في برامج الانتقاء والتسكين.

وتهتم كثير من الاستبيانات المقنتة بقياس الميول المهنية بدرجة أساسية، وبخاصة المجالات التخصصية، غير أن بعض الاستبيانات اهتمت بقياس الميول في مجالات العمل التي تتطلب مهارة أو شبه مهارة مثل الحرف المختلفة. ونظراً لأهمية تعرُّف ميول الطلبة لتزويدهم بصورة واضحة عن مجالات اهتماماتهم الأكاديمية وتوجهاتهم المهنية، فإن المعلم يستطيع تصميم وبناء أنواع متعددة من مقاييس الميول المناسبة لطلابه مثل قوائم المراجعة، وموازين التقدير، والم مقابلات الشخصية، وملاحظة السلوك، وترتيب الأنشطة، والاستجابات الحرة، كما أن قياس الميول يمكن أن يمثل جزءاً أساسياً من تقويم البرامج التعليمية لأسباب متعددة. إذ يمكن اعتبار

الميول متغيراً تابعاً، وذلك عندما نود معرفة ما إذا كان للبرنامج التعليمي تأثير في ميول الطلبة، كما يمكن اعتبارها متغيراً تفاعلياً، وذلك عندما نود معرفة ما إذا كان البرنامج فاعلاً بالنسبة للطلبة من ذوى الميول المرتفعة في مجال معين عن أقرانهم من ذوى الميول في مجالات أخرى، لذلك ينبغي قياس الميول قبل بدء البرنامج التعليمي.

كذلك ربما نود قياس الميول لمعرفة تأثيراتها الجانبية، فالبرنامج التعليمي لا يهدف فقط لتنمية الكفايات المعرفية لدى الطلبة، وإنما أيضاً زيادة اهتمامهم وميلهم نحو مجال دراسى معين وما يرتبط به من المهن.

وعموماً فإن أي أداة قياس منفردة لا تكون كافية لتقييم الميول باعتبارها ظاهرة مركبة. فكل أداة قياس اعتنى ببنائها وتقنيتها تصيف معلومات مهمة يمكن الاسترشاد بها في عمليات التقييم والإرشاد والتوجيه التربوى والمهنى، فهذه المعلومات تزداد فائدتها إذا رُوعى أيضاً خلفية الفرد ودافعيته وظروفه الحالية.

الفصل الثاني عشر

قياس الاتجاهات

* مقدمة

* قياس الاتجاه من منظور تاريخي

* مفهوم الاتجاه وأهميته

* علاقة مفهوم الاتجاهات بغيره من المفاهيم الوجودانية

* مكونات الاتجاهات

* خصائص الاتجاهات

* الأسس المنطقية لقياس الاتجاهات

* تصنیف مقاييس الاتجاهات

* أساليب بناء موازين الاتجاهات

* أسلوب الفترات المتزاوية ظاهريًا (ثيرستون)

* أسلوب التقدير الجمعي (ليكرت)

* الأسلوب التراكمي (جتمان)

* بعض الأساليب الخاصة

* أسلوب تمایز معانی المفاهیم

* أسلوب التصنيف الترتیبی

* أسلوب البسط

* أساليب قياس العلاقات الاجتماعية

* أساليب بناء موازين الاتجاهات متعددة الأبعاد

* تعقیب عام على أساليب بناء موازين الاتجاهات

* إرشادات عامة لبناء موازين الاتجاهات

* أنواع أخرى من مقاييس الاتجاهات

* بعض مشكلات قياس الاتجاهات

* بعض استخدامات مقاييس الاتجاهات

مقدمة :

أوضحنا في الفصل السابق أن الجوانب غير المعرفية أو الوجدانية لشخصية الفرد لا تقل أهمية عن الجوانب المعرفية؛ نظراً لأنها ترتبط بحاجاته ورغباته ومشاعره ودواجهه النفسية، وتُفهم في نجاحه الأكاديمي والمهني، وتناولنا مفهوم الميول الذي يمثل أحد الجوانب الوجدانية للسلوك الإنساني، حيث يشير إلى تفضيلات الفرد لانشطة معينة ينعكس أثراً لها في سلوكه، وناقشتُنا أساليب ومنهجيات قياس الميول وما يتعلق بها من مشكلات، وكيفية الاسترشاد بنتائج القياس في مجالات الإرشاد والتوجيه التربوي والمهني، وفي الأغراض التربوية المختلفة. وأحد الجوانب الوجدانية الأخرى المهمة التي توجه سلوك الفرد وعلاقاته الشخصية والاجتماعية تتعلق بمفهوم الاتجاهات Attitudes.

فالاتجاهات تُعد بمثابة مؤشرات متوقعة في صورها سلوكاً معيناً مميزاً للفرد في موقف لاحقة، فاتجاه الطلاب نحو الكتب المدرسية ربما يؤثر في قدرتهم على تعلم القراءة، واتجاههم نحو المدرسة وبرامجها ربما يؤثر في سلوكهم وقدرتهم على التعلم في المدرسة، واتجاه الفرد نحو المؤسسة التي يعمل بها، والمشرفين على العمل، والآلات والأجهزة التي يستخدمها يؤثر في جودة أدائه. فكما أن الفرد ربما يميل إلى أنشطة معينة، أي يفضل ممارستها، كذلك ربما يكون لديه اتجاه إيجابي أو سلبي بدرجة ما نحو مؤسسات أو مجتمعات من الأفراد، أو شخصيات، أو موضوعات، أو ممارسات، أو عادات، أو مفاهيم، أو أفكار معينة.

لذلك يمكن التمييز بين مفهومي الميول والاتجاهات من حيث الإطار المرجعي الذي يشير إليه كل منها على الرغم من أنهما ليسا مستقلين استقلالاً تاماً. إذ ربما يكون الفرد لديه ميل لتنسيق الزهور (نشاط) واتجاه إيجابي نحو الزهور بعامة (أشياء)، وعلى العكس من ذلك ربما يكون لديه ميل للتدرис (نشاط) واتجاه سلبي نحو الأطفال (جماعة من الأفراد). فالاتجاه Attitude كما يعرّفه أللبروت (Allport, 1954) يعد استعداداً عصبياً نفسياً تنظمه الخبرة ويؤثر في استجابات الفرد لجميع المواقف والموضوعات المتعلقة به، ولكن نظراً لأنه لا يوجد اتجاه واحد يشمل جميع المواقف أو الموضوعات أو الجماعات أو الأشخاص، بل إن كل متغير معين ربما يؤدي إلى انفعالات خاصة به لدى الفرد، وربما تباين شدة إيجابية أو سلبية هذه الانفعالات، فإن السلوك الوجداني الذي يعكس اتجاهًا معيناً لدى الفرد يتحدد جزئياً بالشروط المتوازنة وقت ظهور المثير، وفي إطار الظروف التي ظهر فيها. وسوف نفرد هذا الفصل لمناقشة الاتجاهات وطبيعتها وأبعادها، وعلاقتها بغيرها من السمات أو الخصائص الوجدانية،

والأساليب والطرق المختلفة لقياسها، والمشكلات المتعلقة بذلك، واستخدامات أدواتها في مجالات تقويم الأفراد والبرامج التربوية. وسوف نستهل هذا الفصل بتوضيح التطور التاريخي لقياس الاتجاهات.

قياس الاتجاهات من منظور تاريخي :

إن تطور قياس الاتجاهات لم يحدث بمعزل عن علم النفس الاجتماعي، فحتى أواخر العقد الثاني من القرن العشرين لم يكن هناك سوى محاولات متفرقة لقياس في هذا العلم. فعلى سبيل المثال، أعد بوجاردوس Bogardus عام ١٩٢٥ مقياساً للمسافة الاجتماعية Social Distance Scale لتقدير درجات التقبل للأفراد من جنسيات مختلفة في الولايات المتحدة الأمريكية، ويطلب من المستجيب أن يُبيّن ما إذا كان يتقبل هؤلاء الأفراد كمواطني عاديين، أو كجيران، أو كاصدقاء، أو كأزواج. وفي عام ١٩٢٥ قام ألبرت وهارتمان Allport & Hartman بدراسة لتعرف اتجاهات الطلاب نحو مجالات سياسية مختلفة، وأوضحا الحاجة إلى القياس المقنن للاتجاهات. لذلك أعدّ مجموعة من الفقرات المتعلقة بكل من هذه المجالات، وطلبا من ستة محكمين تقدير رتب لهذه الفقرات، وأوجداً متوسط رتب كل فقرة لتمثيل القيم المناظرة لكل منها، وتطلبت استجابة الطلاب اختيار الفقرة التي تعبر عن رأيهم بدرجة أفضل حيث قدّمت لهم الفقرات مرتبة وفقاً لدرجة أفضليتها، وقد لفت هذه الدراسة نظر ثيرستون L. Thurstone الذي كان يُعد في ذلك الوقت أسلوب الحكم المقارن Comparative Judgment وغيره من الأساليب المتعلقة بالتحليل السيكوفيزياي الذي استند إليه في اشتاقاق موازين سيكولوجية تتعلق بخصائص المثيرات التي لا يناظرها خصائص فيزيائية يمكن قياسها مثل الخصائص الحسية والجمالية. وقد طور هذا الأسلوب وطبقه على بعض فقرات مقياس ألبرت وهارتمان، وبذلك توصل إلى أساس منطقي لتوزيع آراء الطلاب في المقياس بحيث تكون الفترات المتتساوية على المقياس تمثل فروقاً ظاهرية متتساوية في الآراء.

وقد أدرك ثيرستون أهمية هذا الأسلوب في قياس الاتجاهات الاجتماعية، ووضح ذلك في مقالته الرائدة بعنوان «الاتجاهات يمكن قياسها» التي نُشرت عام ١٩٢٨، حيث تناول فيها الأساس المنطقي لأسلوبه الجديد الذي أطلق عليه «الفترات المتتساوية ظاهرياً Equal Appearing Intervals». ويطلب هذا الأسلوب أن يقوم المستجيب بتصنيف الفقرات في مجموعة من الأقسام المرتبة التي يبلغ عددها عادة ٩ أو ١١ قسمًا، بحيث إن الأقسام تبدو أنها تمثل تدريجياً للآراء متتساوي الفترات. ومن ثم يستخدم أسلوب الحكم المقارن أو أي أسلوب رياضي مماثل للتوصيل إلى قيم الميزان المتدرج Scale Values تمثل الأفضلية النسبية لكل فقرة تجاه الموضوع أو

القضية التي تتضمنها الفقرة. وقد أسهم هذا الأسلوب في تطوير البحث في مجال الاتجاهات الاجتماعية Social Attitudes، كما أسهم في بناء العديد من مقاييس الاتجاهات في مجالات متعددة، وطبقت هذه المقاييس على جماعات متنوعة، ومن أهمها مقاييس ريمرز ومعاونيه Remmers et al. عام ١٩٣٢ مدخلًا مختلفًا لقياس الاتجاهات والجماعات العرقية والأنشطة المتعلقة بالأسرة، والممارسات الاجتماعية والأخلاقية، والمهن المختلفة، والإعلانات، حيث بدأ إعدادها عام ١٩٣٠. واستند في ذلك إلى أساليب تحليل المفردات المستخدمة في مجالات الاختبارات العقلية. وتطبق هذه الأساليب مباشرة على استجابات الأفراد المتعلقة بالموافقة أو عدم الموافقة على كل فقرة من فقرات مقاييس الاتجاهات دون الحاجة إلى استخدام أحکام الخبراء في الحكم على درجة أفضليّة الفقرات. ويُطلق على هذا الأسلوب «التقديرات الجمعية Summated Ratings»، حيث يُطلب من المستجيب أن يُحدد درجة موافقته أو عدم موافقته لكل فقرة من فقرات المقياس على ميزان متدرج يتكون من خمسة أرقام لكل منها وزن يبدأ من نقطة واحدة إلى خمس نقاط، مثل : أوفق بشدة، أوفق، غير متأكد، لا أوفق، لا أوفق على الإطلاق. وتُجمع النتائج المناظرة لاستجابات الفرد على الفقرات للحصول على درجته في المقياس.

وفي عام ١٩٤٤ قدم جتمان L. Guttman أسلوبًا جديداً ووضَّحه في مقال بعنوان «أسس تدريج البيانات الكيفية» ويتصل بتحليل مصفوفات بيانات الموارين المدرجة Scalogram Analysis ، حيث يتم ترتيب المفردات والأفراد وفقًا لاستجاباتهم، فالفرد الذي يوافق على فقرة معينة يوافق أيضًا على جميع الفقرات الأدنى منها في الرتبة. وتلا ذلك الأسلوب الذي اقترحه لازارسفeld Lazarsfeld عام ١٩٥٠ المتعلق بتحليل البنية الكامنة للمتغيرات Latent Structure Analysis. ويُعد هذا الأسلوب نموذجًا رياضيًّا عامًّا يربط احتمال الاستجابة لفقرة معينة بموقعها على متغير كامن ينطوي عليها مثل الاتجاه Attitude. والحقيقة أن أساليب ليكرت، وجتمان، والتحليل العاملى يمكن أن تعد حالات خاصة من نموذج لازارسفeld إذا تحققت شروط معينة في توزيع متغير الاتجاه، وشكل الدالة التي تربط احتمال الاستجابة بالمتغير الكامن (الاتجاه)، ومع هذا فإن أسلوب لازارسفeld لا يستخدم إلا في الأغراض البحثية في مجال الاتجاهات. كما قدم كومبس Coombs عام ١٩٥٠ نظرية البيانات Data Preferences Scaling Theory of البيانات التي تُعد نظرية عامة لتدريج استجابات التفضيل Preferences، وفي عام ١٩٥٠ أيضًا نشرت دراسة أدورنو ومعاونيه Adorno et al. المتعلقة بالشخصية السلطوية The Authoritarian Personality تناولت منهجهية جديدة تضم استخدام أسئلة حقائق، ومقاييس تجمع الرأى والاتجاه في أسئلة إسقاطية مفتوحة النهاية.

ومقابلات كلينيكية، حيث يرون أن الأساليب المتعددة غير المباشرة Disguised مهمة للتمييز الصادق بين اتجاهات الأفراد.

مفهوم الاتجاه وأهميته :

نال مفهوم «الاتجاه Attitude» اهتمام علماء النفس الاجتماعي وعلماء القياس نظراً لتأثير سلوك الأفراد تأثيراً ملحوظاً بالاتجاهات، مما يؤثر بدوره في العلاقات الإنسانية التفاعلية بين أفراد المجتمع الواحد والمجتمعات المختلفة. ويعتمد استقرار هذه العلاقات على مدى تأثير أنماط الاتجاهات السائدة في المجتمع في حياة الأفراد دون ضغوط أو توترات. لذلك اهتم خبراء التربية أيضاً باتجاهات الطلاب والتحقق من فاعلية العملية التربوية في تكوين أو تنمية أنماط اتجاهات تُسهم في إحداث تفاعل إيجابي وتكامل بين القوى المختلفة في المجتمع. فإذا كان المعلم يعمل على تنمية وإثراء تفكير طلابه، فإن عليه أن يدرك أن التفكير لا يحدث في فراغ، وإنما يوظف أنماط اتجاهات الطالب وخبرته، فليس المهم فقط كيف يفكر الطالب، وإنما المهم هو فيم يفكر. كما أن التكيف الاجتماعي الانفعالي للطالب يتأثر بتقييمه لسلوكه ورغباته ومشاعره في علاقتها بالقيم الاجتماعية كما يدركها بنفسه. فاتجاهه نحو البيئة ونحو زملائه ومعلميه والمؤسسات المختلفة في مجتمعه يؤثر تأثيراً أساسياً في توجهات تفكيره إما بالإيجاب أو بالسلب. لذلك فإن تكامل الجوانب العقلية والوجدانية للطلاب يتعلق باتجاهاتهم بقدر تعلقه بالتكامل الاجتماعي وتماسك المجتمع. ولهذا فإن مفهوم الاتجاه يُعد من المفاهيم السيكولوجية الاجتماعية التربوية المهمة، حيث إنه يُعد من المكونات الوجدانية الأساسية.

وقد تبانت تعريفات الاتجاه تباعاً ملحوظاً، فقد ذكر ألبرت Allport منذ أكثر من نصف قرن من الزمان سبعة عشر تعريفاً مختلفاً للاتجاه، كما وجد ماكجواير McGuire في مسحه لتعريفات الاتجاه عام ١٩٦٩ أن هناك ما يقرب من ثلاثة تعاريفاً. وحاول ألبرت استخلاص العناصر الأساسية للاتجاه من التعريفات التي ذكرها، وتوصل إلى ثلاثة عناصر هي: التهيئ أو الاستعداد لاستجابات القبول أو الرفض، والاستجابات التي تُنظمها الخبرة، وتحضر الاستجابات في وجود جميع الأشياء والموافق التي يتعلق الاتجاه بها.

وعلى الرغم من تعدد التعريفات بمرور الزمن، إلا أن معظمها يجمعها العناصر الأساسية الثلاثة السابقة. ولعل الاهتمام بتعريف الاتجاه تعريفاً إجرائياً يساعد في تحديد أساليب القياس المناسبة.

وخلاله التعريفات التي احتوت العناصر الثلاثة السابقة يمكن صياغتها في أن الاتجاه تكوين افتراضي يتضمن استجابة محفزة عندما يواجه الفرد مثيرات اجتماعية بارزة، وتتميز هذه الاستجابة بخصائص تقيمية. ويحتاج هذا التعريف إلى مزيد من التوضيح كما يلى :

(١) الاتجاه متغير مثل غيره من المتغيرات السيكولوجية، أي أنه متغير أو سمة كامنة أو تكوين فرضي، وليس متغيراً يمكن ملاحظته مباشرة، وبهذا لا يشير مفهوم الاتجاه إلى فعل واحد أو استجابة واحدة للفرد، وإنما يُعد تجريداً مشتقاً من مجموعة كبيرة من الأفعال أو الاستجابات المترابطة، أي نستدل عليه من هذه المجموعة من الاستجابات التي ينبغي أن تكون متسقة. وهذا يعني أنه إذا كان الأفراد الذين يعبرون عن رفضهم لقضية معينة يُحتمل أيضاً أن يعارضوا ما يتعلق بهذه القضية وما يترتب عليها، فإنه يمكننا عندئذ أن نستدل على اتجاههم نحو هذه القضية.

فالاتجاه إذن كمتغير كامن يستخدم لوصف اتساق أو تغاير عدد من الاستجابات المختلفة لمثيرات تنتهي إلى مجموعة عامة واحدة، أي أن المتغير الكامن (الاتجاه) يتوسط الاستجابات والمثيرات؛ لذلك تغيير الاستجابات لتعلقها بالمتغير الكامن نفسه. فأهمية هذا المتغير الكامن أو الفرضي ترجع إلى أنه توجد مجموعة من البيانات المتعلقة بالاستجابات الملاحظة. ومثل هذا المتغير يطلق عليه «سمة Trait» أو «متغير وسيط Factor» أو «عامل Intervening Variable»، أما الاستجابات الملاحظة فيُطلق عليها متغيرات ظاهرة أو صريحة أو مؤشرات Indicants.

(٢) الاتجاه الذي لدى الفرد يجعله يلاحظ ويستجيب بطريقة انتقائية، فإذا كان الطالب مثلاً لديه اتجاه إيجابي نحو معلميه، فإنه ربما يلاحظ مدى معاونتهم لطلابهم، وربما يُقرربقاء بعد انتهاء اليوم المدرسي، لمتابعة المناقشة معهم، بينما إذا كان اتجاهه سلبياً نحوهم، فإنه ربما يلاحظ مدى سيطرتهم على الطلاب، وربما يفتلع الأعذار للتغيب عن المدرسة، لذلك فإن الملاحظة الدقيقة للقرائن التي يلمسها الطالب والأفعال التي يقوم بها بعد ذلك ربما تفيد في الاستدلال على اتجاهاته.

(٣) الاستجابات التي يتضمنها الاتجاه تسمح بالاستدلال عليه منها عندما يواجه الفرد مثيرات مناسبة، وهذه المثيرات يمكن أن تكون لفظية يعبر عنها الفرد في استجاباته على مقاييس تتطلب الورقة والقلم، أو غير لفظية يمكن ملاحظتها من سلوك الفرد في موقف فعليّة أو محاكاة، وفي جميع الأحوال يستدل على الاتجاهات من نمط استجابات الفرد لفترة أو مجموعة من المثيرات.

(٤) الاتجاه يُعد تقنياً شخصياً لموافق أو أشخاص أو مؤسسات أو أفكار، حيث يحتوى على عناصر موجبة مثل الموافقة أو الاقتراب أو التفضيل، وعنابر سالبة مثل عدم الموافقة أو محاولة الابتعاد أو الكراهة.

والخلاصة أنه يمكن اعتبار الاتجاه انفعالاً معتدل الشدة يهبيء الفرد أو يجعله مستعداً للاستجابة الممتنعة التي تدل على الموافقة (التأييد) أو عدم الموافقة (عدم التأييد) عندما يواجه موضوع الاتجاه. وهذا التعريف يساعد في تمييز مفهوم الاتجاه عن غيره من المفاهيم المتعلقة به والتي سنوضحها فيما يلى:

علاقة مفهوم الاتجاه بغيره من المفاهيم المتعلقة به :

لكل يتضح مفهوم الاتجاه بدرجة أفضل، ربما يكون من المناسب أن نوضح علاقته بغيره من المفاهيم المتعلقة به. وسوف نقتصر على أهم هذه المفاهيم وبخاصة التي يحدث خلط بينها وبين مفهوم الاتجاه، وهي: الميل، والأراء، والمعتقدات، والقيم، والدعاوى، والعادات والسمات، والروح المعنية.

الميل : تعكس الميل Interests المتمثلة في السلوك ما يجذب الفرد أو يبعد اهتمامه عن أنشطة معينة، أي أن الميل تعبّر عن تفضيلات الفرد للقيام بأنشطة دون سواها، في حين أن الاتجاهات لا تقتصر على المشاعر، حيث إنها تتضمن أحکاماً قيمة، أي استجابات القبول أو الرفض.

وعلى الرغم من أنه لا ينبغي النظر إلى الميل على أنها منفصلة عن الاتجاهات، إلا أنه ربما يميل فرد بدرجة كبيرة إلى نشاط معين على الرغم من أن حكمه القيمي على ما يتضمنه هذا النشاط يكون سليماً.

الأراء : هناك أوجه تشابه بين الأراء Opinions والاتجاهات، غير أن مفهوم «الرأى» يشير عادة إلى الاقتراع Opinion Polls ، حيث يطلب من الفرد التعبير عن مشاعر التفضيل تجاه كل فقرة على حدة، وليس على مجموعة متجانسة من فقرات استبيان أو مقياس معين، مثل ما تقديرك لنظام اليوم الدراسي الكامل؟ (ممثاز، جيد، مناسب، غير مناسب). والأراء تتضمن مشاعر أقل مما تتضمنه الاتجاهات، وأقل منها موضوعية، وتعتمد بدرجة أكبر على حقائق أو معلومات لدى الفرد فيما يتعلق بالموضوع الذي يُقترن عليه، كما أن الأراء لفظية، بينما الاتجاهات تتوسط أحياناً بعمليات غير لفظية أو لا شعورية، أي أن الأراء تتعكس في استجابات لفظية بينما الاتجاهات تُعد استعدادات سابقة أو تهيئاً . Predispositions

المعتقدات : تتعلق بمستوى قبول معين لقضية مرتبطة بخاصيص شيء أو حدث معين. ويعرف إنجليش وإنجليش (1958) Beliefs (English & English) المعتقدات بأنها القبول الانفعالي لقضية أو مبدأ يعتبره الفرد أساسياً. وتشير معظم التعريفات إلى أن المعتقد يصبح اتجاهًا إذا صاحبته مشاعر معينة تعكس تقييم تفضيل الفرد لخاصيص الشيء أو الحدث المعين، وعندئذ يعبر عن الاتجاه بمجموع هذه المعتقدات. غير أن

هناك من يرون أن المعتقدات أقل قابلية للتغير من الآراء بل والاتجاهات، ومن أمثلة ذلك الاعتقاد بأن الأرض كروية، أو الاعتقاد بالشخصية الوطنية وغير ذلك.

القيم : يرى سميث (Smith, 1963) أن القيم Values هي تصورات فكرية شخصية لما هو مرغوب ومتصل بسلوك انتقائي، مثل الحكم على العمل بما يدره من دخل، أو مراعاة الأخلاق في جميع التصرفات، أو احترام العلم والمتعلمين، أى أن القيم تتعلق بأشياء أو مفاهيم مهمة في الحياة. ويمكن أن تكون القيم موجبة (التسامح الاجتماعي، والتطلع العلمي، والمعايير الخلقية) أو سالبة (الجريمة، والتحيز العرقي، وعدم التحمل، والفقر، والجهل). وتكون القيم من تجمع ميول الفرد وتذوقاته وتفضيلاته واتجاهاته المتعلقة بظواهر مختلفة، وتتضمن مشاعر وانفعالات أقوى من الاتجاهات. وتحتاج القيم إلى وقت طويل لكي تتكون وتشكل، ومن الناحية التربوية، فإن القيم تنمو وتتكون وتُتعلم عن طريق ما نقوله ونلاحظه ونفعله في البيئة المحيطة بنا.

الدّوافع : تتضمن الاتجاهات الدوافع Motives، فإذا كان لدى الطالب اتجاه إيجابي نحو معلم معين، فإن هذا ربما يكون دافعاً له للالقاء به. وتوكيد بعض التعريفات أن الاتجاهات استعدادات سابقة أو تهيئ لدافع الاستشارة، فالدافعة إذن مؤقتة، وأكثر نوعية حيث تستهدف تحقيق هدف معين، بينما الاتجاه يداوم تأثيره من موقف إلى آخر، ويحفز للسلوك.

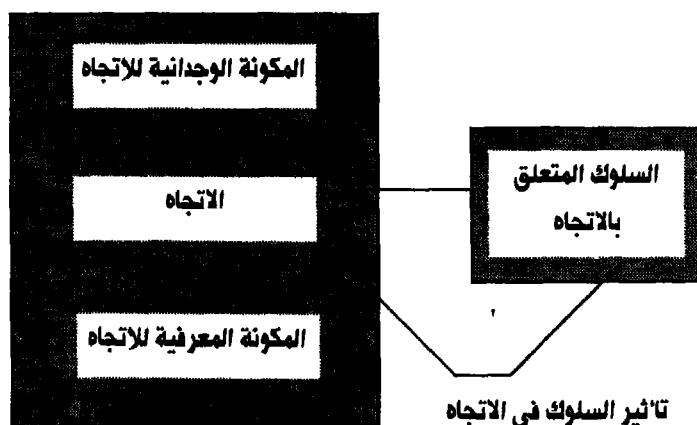
العادات والسمات : تختلف الاتجاهات عن العادات Habits والسمات Traits، فالعادات تشير إلى نزعة الفرد لعمل معين، وتميز ببنية معقدة مستديمة. والعادات كالاتجاهات تكون مكتسبة، ولكنها لا تتضمن حكمًا تقيميًّا على العكس من الاتجاهات، أما السمات فهي استعدادات متسلقة، وأقل استقراراً لاستجابة الفرد بطريقة ما تميزه عن غيره من الأفراد، وتختلف الاتجاهات عن السمات في كونها نوعية التوجّه، بينما السمات تميز بعمومية توجّه الفرد نحو أشياء متنوعة ومتعددة.

الروح المعنوية : تشير إلى التكامل الانفعالي لجماعة معينة، فالروح المعنوية Morale تعد نمطاً متكاملاً من الاتجاهات التي تشير إلى أشياء معينة ذات أهمية حيوية للجماعة، ومن أمثلة ذلك الروح المعنوية لطلاب صف معين، أو الروح المعنوية في مؤسسة صناعية، أو عسكرية، أو الروح المعنوية للأفراد في أوقات العروض. وارتفاع الروح المعنوية يسهم في تحقيق أهداف معينة، وانخفاضها ربما يهدد مصالح حيوية للجماعة.

ومن هذا يتضح تعدد المفاهيم المتعلقة بمفهوم الاتجاه، ويرجع ذلك إلى عدم الاتفاق على معنى ودلالة كثير من المفاهيم الوجدانية، واختلاف الإطار النظري الذي يستند إليه الباحثون في هذا المجال، ونواتج التتحقق الإمبريقي الذي يهدف لتحديد أوجه التشابه والاختلاف بين هذه المفاهيم المتعددة. ونظرًا لاختلاف الأسس والمحendas التي يستند إليها في تعريف الاتجاه Attitude بمرور الزمن، لذلك ربما لا تتوصل إلى تعريف نهائي واحد لهذا المفهوم في الوقت الحاضر. ويُفضل اعتبار جميع هذه المفاهيم أبعادًا متعددة يمكن تعريفها وقياسها بدرجات متفاوتة. ولعل تمييز مفهوم الاتجاه عن غيره من المفاهيم السابقة يعتمد في وقتنا الحاضر على ما ينال اتفاق الباحثين.

مكونات الاتجاهات :

اتضح لنا مما سبق أن الاتجاهات Attitudes تُعد من خصائص الأفراد، حيث تصف مشاعرهم الموجبة والسلبية تجاه موقف أو أفكار أو أشخاص أو مؤسسات أو موضوعات معينة. ويرى واجنر (Wagner 1969)، وزيم وإيسون (Zimbardo & Ebbeson, 1970) أن الاتجاهات لا تقتصر على المكونة الوجدانية أي المشاعر نحو موضوع ما، وإنما تتضمن أيضًا مكونة معرفية تتعلق بمعارف الفرد أو أفكاره أو إدراكاته أو آرائه أو معتقداته حول الحقائق المرتبطة بموضوع الاتجاه، كما تتضمن مكونة سلوكية تتعلق بأفعال الفرد واستجاباته وسلوكه الملاحظ نحو الموضوع المعين. ويمكن توضيح هذه المكونات الثلاث للاتجاه بالشكل التخطيطي (١-١٢) التالي :



شكل (١-١٢) يوضح المكونات الوجدانية والمعرفية والسلوكية للاتجاهات وتفاعلها

فمعظم الاتجاهات يتعلّمها الأفراد من والديهم ومن المحيطين بهم. ولعل الآراء التي يتم تعزيزها بانتظام من جانب الوالدين والأصدقاء تدعم المكونة المعرفية للاتجاهات Cognitive Component. وبالمثل فإن افتتان المعلومات المتعلقة بموضوعات الاتجاهات بالاستجابات الانفعالية للأفراد ربما يؤدي إلى دعم المكونة الوجدانية Affective Component. كما أن اتجاهات الأفراد تعتمد على اتجاهات أفراد الجماعات الأخرى التي يتمون إليها، وهذه الجماعات تُسهم في دعم مكونتي الاتجاه، وتسمى الجماعات المرجعية Reference Groups. وربما تكون كل جماعة مرجعية مناسبة لاتجاهات معينة بالنسبة للفرد، إذ ربما ينسب الفرد اتجاهه نحو موضوع معين، إلى إحدى الجماعات أو المؤسسات المرجعية، بينما ينسب اتجاهًا آخر إلى جماعة أو مؤسسة مرجعية أخرى. والحقيقة أن تحليل الاتجاهات إلى المكونات الثلاث الموضحة بشكل (١-١٢) يشير إلى أن اكتساب الاتجاهات يُعد من العمليات المعقدة. فالاتجاهات تعد بمثابة متغيرات وسيطة Intervening Variables. والسهوم المتوجه من السلوك إلى الاتجاه كما يوضحه شكل (١-١٢) يُعد بمثابة تغذية راجعة للاتجاه مما يشير إلى حقيقة أن الاتجاهات يمكن أن تتغيّر بتأثير المكونة السلوكيّة Behavioral Component.

وذلك السهم المتوجه من المكونة المعرفية يوضح تأثير الاتجاه بها أيضًا، وذلك لأن إدراك الفرد لموضوع الاتجاه ومعلوماته عنه تؤثر في درجة مشاعره نحوه مما يؤثر بدوره في سلوكه الملاحظ.

وعلى الرغم من أن هذا النموذج الذي يربط بين مكونات الاتجاه يُعد من النماذج الكلاسيكية ، إلا أن الدراسات الإمبريالية قدمت أدلة تؤيده Kahn , 1951; Vidulich (1967) Woodrnansee & Cook , 1966 ; Krevanick (1967) Shaw & Wright, أوضحوا أن المكونة المعرفية تُعد الأساس الذي يستند إليه الفرد في تقييمه لموضوع الاتجاه ، وبالتالي لاتجاه ذاته الذي يهوي الفرد للاستجابة بطريقة معينة . ويرى شو ، ورأيت أن الاتجاهات يمكن اعتبارها مجموعة من ردود الأفعال الوجدانية نحو موضوع الاتجاه استناداً إلى مفاهيم ومعلومات ومعتقدات الفرد حول هذا الموضوع ، وتجعله أكثر تهيئاً لكي يسلك بطريقة معينة نحوه . وتعُد المكونة المعرفية الأساس الذي يستند إليه تقييم الفرد لموضوع الاتجاه .

خصائص الاتجاهات :

تتميز الاتجاهات ببعض الخصائص العامة عن غيرها من المفاهيم الوجدانية التي سبق أن أوضحناها. وتفيد هذه الخصائص في إلقاء مزيد من الضوء على الأبعاد المتعددة لهذا المفهوم. وسوف نوضح فيما يلى أهم هذه الخصائص:

(١) التوجُّه : يُعد من الخصائص الأساسية للاتجاهات، فالتجوُّه Direction يتعلق بما إذا كانت مشاعر أو انفعالات الفرد تتجاه موضوع أو قضية معينة موجبة أم سالبة. وهذه الخاصة تُعد من الخصائص المعقّدة؛ نظراً لأن الاتجاهات تتضمن استجابات ظاهيرية متعدّة، مثل الثقة، والحب، والتعاون، والموافقة، والقبول، في مقابل الكراهية، والشك، والانسحاب، والاعتراض، والهاجمة. ويمكن أن تدرج جميع هذه المشاعر تحت مفهوم عام يتعلق بالمسافة السيكولوجية Psychological Distance بين الفرد وموضوع الاتجاه. فالطالب الذي لديه اتجاه موجب نحو المدرسة يكون موجّهاً توجيهياً إيجابياً نحو بعض الأشياء المتعلقة بالمدرسة مثل الطلاب والمعلمين، والجمعيات المدرسية، والمنهج والمواد الدراسية، أما إذا كان لديه اتجاه سالب نحو المدرسة فإنه يحاول الابتعاد عن المدرسة وما يرتبط بها من أنشطة.

غير أن هذه الخاصة تُبرز التساؤل المتعلق بالاتجاهات المحايدة، وما إذا كانت تقع بين الاتجاهات الموجبة والاتجاهات السالبة. فتصورنا للتوجُّه الموجب والسلب على أنهما توجّهان متضادان، يعني أن اتجاه الفرد لا يمكن أن يقع على طرف النقيض في آن واحد. لذلك ربما يكون من المناسب معالجة درجات التفضيل ودرجات عدم التفضيل على أنهما متمايزان من الناحية النظرية، على الرغم من أنهما بلا شك يكونان مرتبطين من الناحية الإمبريقية، أي أنه يمكن من الناحية النظرية التصورية اعتبار أن الاتجاه خاصية ثنائية القطب Bipolar Attribute .

(٢) المقدار والشدة : يشير مقدار الاتجاه Magnitude إلى درجة الاتجاه الموجب أو السالب. وعادة يحدث خلط بين مقدار الاتجاه وشدته Intensity ، فشدة الاتجاه تشير إلى أهمية أو قوة مشاعر الفرد تجاه موضوع معين، بعض الأفراد تكون لديهم مشاعر أو انفعالات أكثر شدة من غيرهم من الأفراد، وكذلك المشاعر ذاتها يمكن أن تتبادر في درجة شدتها، فالكراهية تعد من المشاعر الأكثر شدة من عدم الحب، ويمكن أن يكون لدى أحد الأفراد اتجاه موجب ولدى فرد آخر اتجاه سالب نحو موضوع ما، ومع هذا تكون شدة اتجاه كل منها متساوية. وكلما زادت درجة التفضيل أو عدم التفضيل زادت شدة الاتجاه، أما الأفراد المحايدون فإنه يكون لديهم

اتجاه أقل شدة، والحقيقة أن مقدار الاتجاه وشدته يُعدان من الخصائص الأساسية التي تميز الاتجاه عن غيره من الجوانب الوجدانية، لذلك تهتم مقاييس الاتجاهات بقياس هاتين الخاصيتين.

(٣) ثنائية المشاعر: تشير إلى مدى تناقض مشاعر الفرد نحو جوانب مختلفة لموضع الاتجاه نفسه، وتتأتي ثنائية المشاعر Ambivalence من تصور التوجه ثنائي القطب للاتجاه الذي أشرنا إليه، فكلما تساوت مشاعر التفضيل ومشاعر عدم التفضيل ارتفعت درجة تناقض المشاعر.

(٤) المركزية: تشير هذه الخاصية Salience إلى التهيز لاستثارة الاتجاه، أي مدى اقتراب الاتجاه من التفكير المباشر للفرد، فمثلاً إذا سألنا بعض الأفراد عن أهم مشكلة تواجه البشرية اليوم، فإنهم ربما يفكرون على الفور في منع الحروب، بينما إذا سألنا بعضًا آخر ربما يفكر على الفور في محاربة الإدمان، ويفكر غيرهم مباشرة في حقوق الأطفال وهكذا، أي أن هذه الخاصية تتعلق بما إذا كان اتجاه معين يعد مركزيًا لتفسير اتجاهات واسعة ومتعددة.

والاتجاهات المركزية هي تلك التي يهتم بها الفرد، ويكون لديه قدر كبير من المعلومات حول موضوعها. ويمكن قياس هذه الخاصية إذا أتيحت الفرصة للفرد للتغيير الحر عن اتجاهه دون سبر. ونظراً لأن معظم مقاييس الاتجاهات تشتمل على فقرات تتطلب استجابات محددة مثل «موافق» أو «غير موافق» فإننا لا نستطيع باستخدامها قياس خاصة «المركزية»، وإنما يمكن قياسها باستخدام أساليب الملاحظة والمقابلات الشخصية حيث يمكن أن يعبر الفرد عن اتجاهاته دون قيد.

(٥) المركزية الوجدانية: تشير إلى الدرجة التي يصبح فيها الفرد انفعالاً بدرجة كبيرة في تعسّره عن اتجاهاته نحو موضوع معين. أي أن خاصية المركزية الوجدانية Affective Salience تتعلق بدرجة تقدير الفرد أو انفعاله المتمرّك حول موضوع الاتجاه.

(٦) المرونة: تشير إلى مدى قابلية الفرد للتغيير أو تعديل اتجاهاته نحو موضوعات معينة، وعادة ينظر إلى خاصية المرونة Flexibility على أنها سمة عامة للأفراد بدلاً من اعتبارها تشير إلى اتجاه محدد.

(٧) التضمن: فالاتجاه يمكن أن يكون منفصلاً عن الجوانب الأخرى، مثل المعتقدات، أو القيم، أو غيره من الاتجاهات، أو يمكن أن يرتبط بهذه الجوانب من خلال الاقتران والتصنيف والتبرير وهكذا. فمفهوم الأم يمكن أن يكون مُتضمناً أو مطموراً Embedded في شبكة من الترابطات بمقاييس متمايزة مثل الحماية، ورعاية الطفولة، وغير ذلك.

(٨) **الشمول أو المدى** : فالطالب يمكن أن يكره بشدة مظهراً أو مظاهرين متعلقين بالمدرسة مثلاً، بينما يمكن أن يكره طالب آخر كل ما يتعلق بالتعليم، أو ربما يكون لدى أحد الأفراد اتجاه سالب نحو الشباب الذي يتميز بخصائص معينة فقط، أو ربما يكون اتجاهه السالب نحو الشباب بعامة الذين تختلف خصائصهم عنه. أى أن الشمول أو المدى Pervasiveness يتعلّق بمدى تعميم الاتجاه أو مدى شموله لعناصر موضوع الاتجاه، فإذا علمنا أن فرداً ما متحفظ، فإنه يُمكّنا التنبؤ باتجاهاته نحو موضوعات متنوعة واسعة، وكثير من الاتجاهات تتميز بإمكانية تعميمها بدرجة كبيرة.

(٩) **التعقد المعرفي** : يشير إلى تفاصيل المكونة المعرفية للاتجاهات، مثل ثراء المحتوى، أو عدد الأفكار التي لدى الفرد عن موضوع الاتجاه، وأحياناً يطلق على التعقد المعرفي Cognitive Complexity تسميات متعددة مثل: التمايز، والتعدد. وتعتمد درجة التعقد المعرفي إلى حد ما على منظور السفرد عن موضوع الاتجاه، فمنظور الفرد عن الناس بعامة ربما يختلف عن منظوره لفرد معين واحد. فإذا وصف الفرد أحد الأشخاص بأنه حسن الخلق، ومهذب، فإن استخدام هذين الوصفين المترادفين معًا للتعبير عن اتجاه الفرد، ربما يشير إلى اتجاه أكثر تعقداً مما لو استخدم أحد هذين الوصفين فقط. غير أن هذا التفسير لا يزال موضع خلاف في بحوث الاتجاهات.

(١٠) **الاتساق** : تُميز خاصية الاتساق Consistency بين الخصائص الوجданية ورد الفعل الوجданى الذى تُحدّثه مواقف أو أحداث معينة، فبعض الأفراد تكون استجاباتهم على مقاييس الاتجاهات متسبة اتساقاً ملحوظاً، بينما يستجيب البعض الآخر استجابات متناقبة لموضوع واحد من حيث التفضيل أو عدم التفضيل. لذلك فإنه ينبغي أن تكون هناك درجة ما من اتساق الاستجابات لكي نتمكن من الاستدلال على اتجاه الفرد، وإذا لم تتوافر هذه الدرجة فإن استجابة الفرد ربما تُعزى إلى عوامل خارجة عنه، ولا تُعزى إلى اتجاهه الفعلى. والحقيقة أن خاصية الاتساق تُعد من الخصائص الأساسية للاتجاهات، بالإضافة إلى خاصية مقدار أو شدة الاتجاهات وتوجهها، ويمكن ملاحظة ذلك في تعريف مفهوم الاتجاه الذي سبق أن قدمناه. لذلك تركز موازين الاتجاهات Attitude Scales على قياس هذه الخصائص الثلاثة كما سيتضح لنا فيما بعد.

الأسس المنطقية لقياس الاتجاهات :

- سبق أن أوضحنا في الفصل الأول أن القياس النفسي والتربوي يتعلق بجمع ملاحظات أو بيانات عن سلوك الأفراد، وتعيين أرقام تمثل هذه الملاحظات وفقاً لقواعد محددة، وينطبق ذلك على قياس الاتجاهات. وتعتمد إجراءات القياس على الافتراضات النظرية المتعلقة بطبيعة الاتجاه الذي تود قياسه، وعلى العلاقات القائمة بينه وبين السلوك، وكذلك على علاقته بالقواعد المستخدمة في تعين أرقام للملاحظات السلوكية. فالنماذج المختلفة لبناء الموازين Scaling Models المستخدمة في قياس الاتجاهات تهدف لتحويل مجموعات من الملاحظات أو استجابات الأفراد إلى مؤشرات Indices للاتجاه الذي تطوى عليه هذه الاستجابات، غير أن تفسير البيانات المستمدة من هذه الموازين يعتمد على العلاقة بين مفهوم الاتجاه الذي يتباين من يقوم بإعداد الميزان Scale والأسلوب الذي يستخدمه في قياس هذا الاتجاه. ولعل وضوح هذه العلاقة يعد أمراً ضرورياً لفهم النماذج والأساليب المختلفة لقياس الاتجاهات. وتتم عملية قياس الاتجاهات بعدة مراحل متالية نوجزها فيما يلى:

(١) تصور مبدئي لطبيعة الاتجاه المراد قياسه، ولعل التعريفات التي قدمناها فيما سبق توضح ذلك.

(٢) تحديد الأبعاد المتعلقة بالاتجاه للاستناد إليها في القياس، وعلى الرغم من عدم وجود اتفاق كبير على ماهية الأبعاد المهمة للاتجاهات، إلا أن هناك عدداً من الأبعاد التي اتضحت أهميتها أشرنا إليها فيما سبق، كما أن هناك جهوداً متزايدة تهدف للتوصيل إلى اتفاق حول الأبعاد الأساسية للاتجاهات.

(٣) ترجمة هذه الأبعاد النظرية وما تتضمنه من عناصر إلى مواقف يمكن ملاحظتها بحيث يستند إليها في التوصل إلى أدلة Indicators تتعلق بهذه الأبعاد. وعلى الرغم من اختلاف الإستراتيجيات والأساليب التي تتبع في إجراء ذلك، إلا أن الاستراتيجية أو الأسلوب المستخدم يُعد بمثابة تحديد إجرائي للاتجاه الذي تمثله الملاحظات الإيميريقية التي تُسند منها أدلة الاتجاه. وهذا التحديد الإجرائي للاتجاه ربما يتفق أو يختلف إلى حد ما عن التصور المبدئي لطبيعة الاتجاه المراد قياسه، غير أن هذا يعتمد على كفاية الأدلة التي تم اختيارها، فكلما كانت الأدلة تمثل أبعاد الاتجاه تمثيلاً جيداً يكون الاتفاق بينهما أفضل. ولكن نظراً لأن هذه الأدلة تكون عادة بمثابة عينات من السلوك الذي ينطوى عليه الاتجاه، فإنه ينبغي توقيع بعض الاختلاف بينهما، ويمكن العمل على جعل هذا الاختلاف أقل مما يمكن بمراعاة أن تكون هذه العينات مماثلة للنطاق السلوكي لهذه الأبعاد.

(٤) ضم الدرجات المستمدة من هذه الأدلة في مؤشرات Indices تُمثل الاتجاه الذى تتطوى عليه. وتُعد هذه المرحلة من المراحل التى تتعلق بدرجة أساسية بمفهوم القياس، أى تعيين أرقام تدل على الملاحظات وفقاً لقواعد محددة. فاختيار أدلة معينة Indicators تُميز الاتجاه المراد قياسه يتطلب تحديد السلوك الملاحظ المتعلق بهذه الأدلة. أى ينبغي تحديد طبيعة الموقف الاختبارى الذى تجمع فيه الملاحظات، وما إذا كان سيقصر على مواقف اصطناعية يجبر فيها الفرد لفظياً عن مجموعة من الأسئلة المتعلقة بالاتجاه المعين، أو يجب تحريرياً عن فقرات استبيان، أو تجمع الملاحظات فى مواقف حياتية فعلية بعد تحديد جوانب الأدلة التى ينبغي ملاحظتها فى ضوء طبيعة الاتجاه وأبعاده. ويطلب أيضاً تحويل الملاحظات إلى مقايير كمية وضم هذه المقايير معًا لتُصبح درجات على ميزان، واقتراح عمليات تنظيم وتصنيف لهذه الملاحظات بحيث يعبر عنها النظام التصنيفى المقترن، وربما تُبعد بعض الملاحظات التي لا تطابق هذا النظام. لذلك يعد النظام التصنيفى مكملاً للتصور البذئى لطبيعة الاتجاه الذى تحدّد في المرحلة الأولى. وتحقيق ذلك في قياس الاتجاهات يعتمد على نماذج فى بناء الموازين Scaling Models تتطلب تحقق فروض معينة في الملاحظات أو البيانات لكي تستطيع التوصل إلى خصائص قياسية يفضل توافرها. ولكن هذه الخصائص ربما لا تناظر التصور البذئى لطبيعة الاتجاه المراد قياسه. ويرى كومبس Coombs أنه ربما يصعب تتحقق فروض هذه النماذج في البيانات المتعلقة بالاتجاهات وغيرها من الجوانب الوجدانية.

لذلك يقترح التحرر من فرض هذه النماذج إذا لم تتحتمل البيانات هذه الفرض، وفي هذه الحالة ينبغي أن نسمح لخصائص البيانات أن تبرز أثناء عملية التحليل. والحقيقة أنه لا يوجد تعارض بين التوجهين حيث إنهما يختلفان فقط في القدر من التباين الذي يمكن إعزاوه إلى الأخطاء العشوائية، والأخطاء التي يمكن احتواوها في إطار النموذج المستخدم. لذلك تنقسم هذه النماذج إلى نماذج صارمة فيما تتطلبه من تحقيق فروض معينة في البيانات، ونماذج لا تتطلب إلا قدرًا ضئيلاً من الفرض. وبالطبع تؤدي النماذج الثانية إلى مؤشرات للاتجاهات أقل دقة من الناحية السيكومترية، ولكن تفيد من جميع البيانات أو الملاحظات التي تم جمعها.

وعموماً، فإن استخدام أى من هذه النماذج يعتمد بدرجة أساسية على طبيعة البيانات ومستوى قياسها، أى ما إذا كان مستوى القياس اسمياً أو رتيبياً أو فترياً كما أوضحتنا في الفصل الأول. فالإعداد في المستوى الاسمي تُصنف البيانات أو الملاحظات المتعلقة بالاتجاه المطلوب إلى أقسام منفصلة، ولذلك فإن هذه الأعداد لا تكون مرتبة وبالتالي يصعب قياس الاتجاهات بمثل هذه الأعداد، أما في المستوى

الرتبي، فإن الفرد الذي يحصل درجة ولتكن ٤ يفترض أن اتجاهه أكثر إيجابية من الفرد الذي حصل على الدرجة ٣ مثلاً. فالأعداد في هذا المستوى تكون مرتبة، ولكن المسافات فيما بينها غير متساوية، فالفرق بين الدرجة ٤ والدرجة ٣ لا تساوي في إيجابية الاتجاه الفرق بين الدرجتين ٦ ، ٥ مثلاً وذلك لعدم توافر صفر مطلق. أى أن الميزان الرتبي يتعلق فقط بترتيب الأفراد وفقاً لدرجة إيجابية الاتجاه لديهم، ولكن الفروق بين الرتب المختلفة لا تكون متساوية، بينما يحدث ذلك في المستوى الفتري حيث تكون الفترات التي تفصل بين الأعداد المتتالية متساوية. فالفرق بين الدرجتين ٤ ، ٣ في إيجابية الاتجاه تساوى الفرق بين الدرجتين ٦ ، ٥ في الإيجابية، ولكن لا نستطيع اعتبار الدرجة ٦ ضعف الدرجة ٣ في الإيجابية نظراً لعدم وجود صفر مطلق أيضاً، حيث إن هذا يقتصر على المستوى النسبي. غير أنها ذكرنا فيما سبق أن الاتجاهات تراوح في شدتها بين الإيجابية المرتفعة والسلبية المرتفعة. وهذا يعني أن هناك نقطة محايضة تناظر الصفر على متصل الاتجاه، يتغير عندها الاتجاه من الإيجابية إلى السلبية أو العكس، وتحديد موقع هذه النقطة يعد أمراً ضرورياً في قياس الاتجاهات. إذ ربما نود معرفة ما إذا كان تعديل اتجاه الفرد من موقع معتدل إلى موقع إيجابي أو سلبي متطرف أيسراً من تعديل اتجاهه من موقع يدل على اتجاه إيجابي منخفض إلى حد ما إلى موقع يدل على اتجاه سلبي منخفض، وهل الموقف المحايض يعني التطابق مع الجماعة. والحقيقة أن نقطة الصفر على متصل الاتجاه تعد من مشكلات قياس الاتجاهات. ويرى بعض علماء النفس أن الاتجاهات تكون عادة إما موجبة أو سالبة، فإذا لم تكن كذلك، أى إذا كانت محايضة أو تقع عند نقطة الصفر، لا نستطيع أن نطلق عليها مصطلح الاتجاهات . Attitudes

ذلك فإن تساوى وحدات متصل الاتجاهات الممثل في المستوى الفتري يُعد أمراً مطلوبًا، وذلك لكي نتمكن من إيجاد متوسط درجات فرد في موازين مختلفة، ومقارنة التغيرات التي تحدث في اتجاهات الأفراد على هذا المتصل، وكذلك مقارنة درجات أحد الأفراد بدرجات فرد آخر في ميزان معين. غير أن الموازين التي تُستخدم في قياس الاتجاهات، والتي سنوضحها بعد قليل ربما لا تحقق ذلك، وإن كانت وحدات بعض هذه الموازين تقى بذلك إلى حد ما.

وكثير من الموازين التي تقيس الاتجاهات تكون أحادية البعد Unidimensional، أى أنها تقيس متغيراً واحداً أو اتجاهًا واحدًا حيث يمكن تمثيل درجة الفرد بنقطة على متصل الاتجاه. فإذا حصل فرداً أو أكثر على درجة متساوية في الاتجاه الذي يقيسه الميزان Attitude Scale ، فإن درجة كل منهم يكون لها نفس الموضع على متصل

الاتجاه، أى يكون لهم نفس الاتجاه تقريباً. ولكن إذا كان الميزان يقيس أكثر من اتجاه واحد، أى متعدد الأبعاد Multidimensional، فإن الدرجة نفسها يمكن أن تمثل أنماط اتجاهات مختلفة. فمثلاً إذا كان ميزان اتجاه نحو أفراد المهن المختلفة يشتمل على ستراتيفيات تتعلق بالاتجاه نحو كل مجموعة، فإن هذا الميزان لا يكون أحادى البعض، أما إذا كانت موازين الاتجاهات نحو المجموعات المختلفة مرتبطة فيما بينها ارتباطاً مرتفعاً، فإنه يمكن اعتبارها تمثل ميزاناً أحادى البعض.

وتعد خاصية البعض الأحادي Unidimensionality من الخصائص المهمة التي ينبغي توافرها في الموازين التي تقيس الاتجاهات نظراً لسهولة تفسير نتائجها، ويعتمد ذلك اعتماداً كبيراً على الأسلوب الذى يستخدم فى بناء الميزان. فإذا كان لدينا أكثر من ميزان واحد يشتمل كل منها على عينة من الفقرات المختلفة المستمددة من نطاق سلوكي واحد للاتجاه المراد قياسه، فإنه ينبغي أن تكون هذه الموازين مرتبطة ارتباطاً مرتفعاً، بل وربما نود أيضاً أن يكون توزيع درجات كل منها يتبع نفس الشكل العام، وكذلك نفس نمط الارتباطات الداخلية بين فقرات كل منها، ونفس مقدار الثبات.

تعريف مقاييس الاتجاهات :

تهدف جميع مقاييس الاتجاهات للتوصل إلى استدلالات تتعلق بالاتجاهات استناداً إلى أدلة يمكن ملاحظتها. وعلى الرغم من تعدد طرق تصنيف هذه المقاييس، إلا أنه تجمعها عناصر مشتركة، إذ يمكن التمييز بين ثلاثة أنواع من النطاقات السلوكية للاتجاه Attitude Universes التي يُطلب فيها من الفرد أن يستجيب استجابة لفظية، وهذه الاستجابة تعبر عادة عن رأى الفرد أو تصف مشاعره. ونطاقات هذه الاستجابات يمكن تسميتها الاتجاهات اللفظية المستدعاة Elicited Verbal Attitudes. ولكن تعبر الفرد عن آرائه في بعض الأحيان في محادثته المعتادة مع أصدقائه يُمثل الاتجاهات اللفظية التقائية Spontaneous Verbal Attitude، أما القسم الثالث من النطاقات السلوكية فيتعلق بالسلوك اللفظي أو غير اللفظي أي الفعل الموجه نحو موضوع الاتجاه، ويمكن تسميتها الاتجاهات المتمثلة في الفعل Action Attitudes. وتتمثل هذه الأقسام الثلاثة في أن يُطلب من الفرد إبداء رأيه في قضية معينة، أو أن يوضح تلقائياً لأصدقائه رأيه في القضية، أو أن يسلك بطريقة معينة تجاهها.

ويمكن تعين أرقام تناظر الاستجابات في كل من هذه الأقسام وفقاً لقواعد معينة. وعلى الرغم من أن معظم الموازين المستخدمة تقيس الاتجاهات اللفظية المستدعاة، أى القسم الأول، إلا أنه يمكن بناء أدوات لقياس الاتجاهات اللفظية في

القسمين الآخرين. إذا يمكن عن طريق أساليب الملاحظة والمقابلة الشخصية التي يعنى بتصميمها إعداد موازين لقياس الاتجاهات اللغوية التلقائية، والاتجاهات المتمثلة في الفعل. وإلى جانب هذه الأساليب المباشرة في قياس الاتجاهات، توجد أيضًا أساليب غير مباشرة أو متنكرة Disguised Techniques، مثل الأساليب الإسقاطية بأنواعها المختلفة، كما توجد أيضًا أساليب فسيولوجية.

لذلك فإن مقاييس الاتجاهات يتم تصنيفها عادة على أساس نوع الأدلة Indicators التي يستدل منها على الاتجاه المطلوب كالتالي:

القسم الأول: يشتمل على الطرق التي تمكنا من الاستدلال استناداً إلى استجابات الفرد وفقاً لمشاعره الحقيقة لمجموعة من الفقرات أو الصفات، وتُعرف هذه الطرق بأساليب بناء الموازين Scaling Techniques، والأدوات التي يتم بناؤها تسمى «موازين Scales». ومن هذه الأساليب الشائعة الاستخدام موازين ثيرستون Thurstone، وموازين ليكرت Likert Scales، وموازين جتمنان Guttman Scales، وبعض الأساليب الخاصة الأخرى مثل أسلوب التمايز اللغوي Semantic Differential Unfolding Scales الذي اقترحه أورجود ومعاونوه Osgood et al.، وأسلوب الفرد Technique الذي اقترحه كومبس Coombs، وأسلوب Q الذي اقترحه ستيفنسون Stephenson، والموازين متعددة الأبعاد Multidimensional وغيرها. وتختلف جميع هذه الموازين عن الاختبارات التي تقيس الجوانب المعرفية في أن نتائجها لا تستخدم في تحديد نجاح الفرد أو إخفاقه، أو قوته أو ضعفه، وإنما تُستخدم في قياس درجة ما يمتلكه الفرد من الخاصية الوجدانية، أي الاتجاه نحو موضوع معين، مثل اتجاه طيبة المدارس الثانوية نحو التعلم الذاتي، أو اتجاه طلبة الجامعات نحو التعليم المستمر. لذلك لا توجد إجابة صحيحة وإجابة خطأ للفقرات التي يشتمل عليها ميزان الاتجاهات Attitude Scale.

القسم الثاني: يشتمل على الطرق التي تمكنا من الاستدلال استناداً إلى سلوك ظاهر Overt. فملاحظة السلوك في مواقف حياتية فعلية يزيد من صدق هذا الاستدلال. وهذه المواقف لا تتطلب تعاون المستجيب، كما لا يجعله يشعر بأنه في موقف اختباري أو أنه يُعامل بطريقة خاصة، وكذلك لا تؤدي إلى تغير في الظاهرة التي يجري قياسها. ويرى سشرست (Sechrest, 1969) أنه لا ينبغي الاستناد إلى موازين الاتجاهات استناداً تاماً، وأوصى باستخدام أساليب ملاحظة السلوك إلى جانب هذه الموازين، فقد بيّنت كثير من الدراسات التي قام ويكر (Wiker, 1969) بمراجعتها أن

هناك احتمالاً كبيراً أن ترتبط الاتجاهات التي تُقاس بهذه المعايير بالسلوك الظاهر ارتباطاً منخفضاً. واستخلص من هذه الدراسات أن ١٠٪ على الأكثر من تباين درجات مقاييس السلوك الظاهر يمكن تفسيرها باستخدام البيانات المتعلقة بالاتجاهات. ومن أمثلة أساليب ملاحظة السلوك: لعب الأدوار Role Playing ، والاختيارات السوسيومترية Sociometric Choices وغيرها.

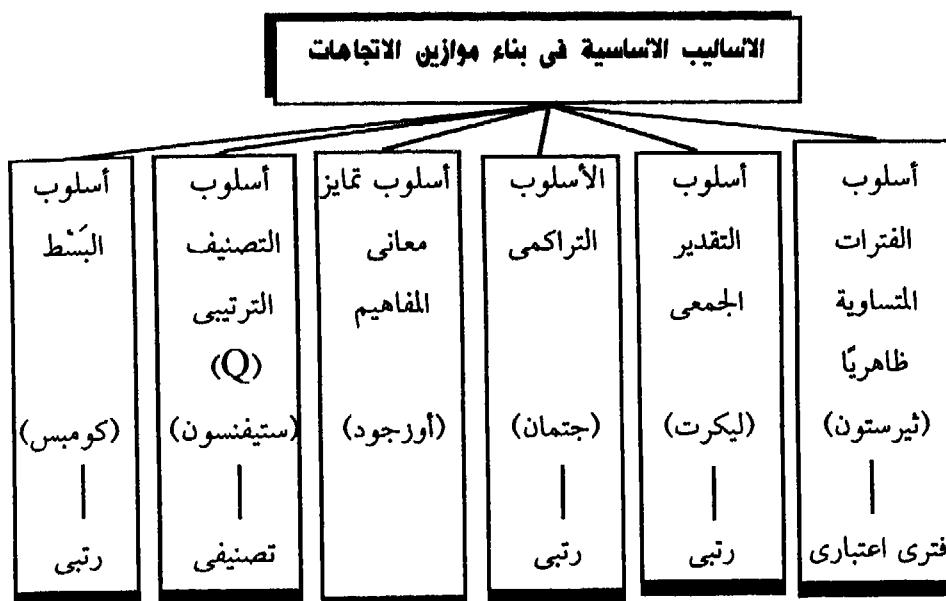
القسم الثالث: يشتمل على الطرق التي تُمكّننا من الاستدلال استناداً إلى مستويات استجابة أكثر عمقاً من الاستجابة لفقرات معايير الاتجاهات. وتقدّم المثيرات بحيث يفهمها الفرد المستجيب بطريقة تختلف عما يقصده الفاحص، أي أن هذه الطرق غير المباشرة لقياس الاتجاهات تحمل معنى معيناً بالنسبة للمستجيب، بينما يستند تفسير الاستجابات إلى نظام مختلف في المعنى، ومن أمثلتها الطرق الإسقاطية كما سبق أن أشرنا.

القسم الرابع: يشتمل على الطرق التي تُمكّننا من الاستدلال استناداً إلى مقاييس نفسولوجية للاتجاهات. وقد اقترح عدد من هذه المقاييس، مثل الاستجابة الجلفانية للجلد، واتساع أو ضيق حدق العين، ومعدل نبض القلب عند عرض تمثيل مصور لمواقف تتعلق بموضوع الاتجاه. وعلى الرغم من أن هذه المقاييس تكشف عن الاستشارة أو الشدة، إلا أن المعلومات التي تستمدّ منها لا تتعلق تعلقاً تاماً بقياس الاتجاهات، حيث إن توجّه الاستشارة وهدفها يظل غير معلوم. وعلى الرغم من أن فيشباین ، وأجزن (Fishbein & Ajzen , 1965) اقترحا أهمية الحصول على قياس غير لفظي للاتجاهات يكون خارجاً عن تحكم المستجيب، إلا أنهما علقاً على ذلك بأنه ليس من المحتمل أن أيّاً من مقاييس رد الفعل النفسيولوجي يمكن أن تحقق هذا الاقتراح.

ونظراً لأن الطرق التي تتّمّي إلى القسم الأول المتعلقة بأساليب بناء معايير الاتجاهات تعد أكثر استخداماً من الطرق التي تتّمّي إلى القسمين الآخرين، كما يسهل تفسير نتائجها، فسوف نناقشها بالتفصيل فيما يلي، ونوضح بعض الطرق الأخرى بإيجاز.

أساليب بناء موازين الاتجاهات :

أشرنا فيما سبق إلى أنه توجد أساليب متعددة لبناء موازين الاتجاهات Attitude Scales ذكرنا بعضاً منها. وسوف نوضح فيما يلى كيفية استخدام كل من هذه الأساليب، وميزات وعيوب كل منها، ونقارن بينها من حيث الأسس المنهجية التي استندت إليها، ويوضح الشكل التخطيطي التالي (٢-١٢) أهم هذه الأساليب:



شكل (٢-١٢) يوضح أساليب بناء موازين الاتجاهات الشائعة الاستخدام

أسلوب الفترات المتساوية ظاهرياً (ثيرستون) :

يعد هذا الأسلوب من المحاولات المبكرة التي قام بها ثيرستون (Therstone, 1931) 1929 لبناء موازين الاتجاهات، والحقيقة أنه قدم أساليب مختلفة استناداً إلى تجارب علم النفس الفيزيائى Psychophysics تُفيد في بناء موازين لأحكام الأفراد حول خصائص فизيائية مختلفة مثل الوزن، فكلما كان الفرق الحقيقى بين وزن جسمين ضئيلاً قل عدد الأفراد الذين يستطيعون التمييز الصحيح بين وزنهما، وحاول ثيرستون بطريقة مائلة لبناء موازين اتجاهات بأن يطلب من الأفراد المقارنة بين فقرات أو عبارات تتعلق باتجاه معين مثلى مشنى، والحكم على كل من الأزواج من حيث شدة الاتجاه، أى أيهما أكثر إيجابية وأيهما أكثر سلبية، وأطلق على ذلك أسلوب المقارنات الثنائية Paired-Comparisons.

العبارات كبيرةً كما هو الحال في معظم موازين الاتجاهات، حيث إن ١٠ فقرات تتطلب إجراء $\frac{9 \times 1}{2} = 45$ مقارنة ثنائية؛ لذلك اقترح ثيرستون أسلوبًا آخر يتطلب إجراؤه جهدًا أقل، وأطلق عليه أسلوب الفترات المتساوية ظاهريًا Equal-Appearing Intervals.

خطوات بناء موازين الاتجاهات باستخدام أسلوب ثيرستون:

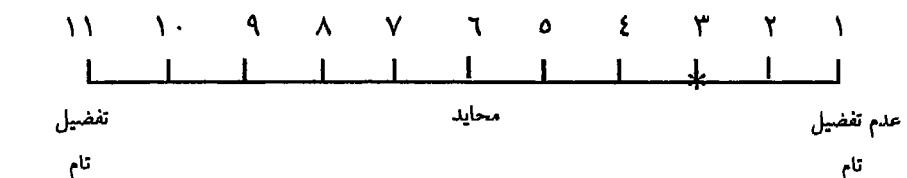
يمكن بناء موازين الاتجاهات باستخدام أسلوب ثيرستون باتباع الخطوات التالية:

أولاً: تجميع عدد كبير من الفقرات التي تتعلق بموضوع الاتجاه المراد قياسه، وليكن ١٠٠ فقرة مثلاً تعبر عن درجات تفضيل متباعدة للموضوع، أي تراوح بين تفضيل أو موافقة تامة، وعدم تفضيل أو عدم موافقة تامة. ويمكن الاسترشاد في ذلك بمراجعة الأدبيات المتعلقة بموضوع الاتجاه ومن المقابلات الشخصية الاستطلاعية. وينبغي أن تكون صياغة الفقرات واضحة وتعبر بدقة عن فكرة واحدة خالية من الغموض، وتسمح بالقبول أو الرفض بدرجات مختلفة فيما يتعلق بموضوع الاتجاه، ولا يجب أن يعبر محتواها عن حقائق، وأن تبدو جميع الفقرات مميزة، أي تحتمل استجابات مختلفة من الأفراد الذين تختلف شدة اتجاهاتهم نحو الموضوع، وتكتب كل فقرة على بطاقة صغيرة. فمثلاً إذا كان الاتجاه المراد قياسه نحو التعليم فإن بعض الفقرات ربما تكون كالتالي:

- (أ) لا يمكن أن نكون مواطنين صالحين بدون تعليم.
- (ب) أذهب إلى المدرسة فقط لأنني مُجبر على ذلك.
- (ج) ينبغي أن يحظى التعليم بمزيد من التمويل.
- (د) لا يجب أن أبدل جهداً كبيراً إلا للنجاح في الامتحانات.
- (هـ) التعليم مفيد في حياتنا المعاصرة.
- (زـ) أشك فيما إذا كان التعليم قد أسهم في تطوير الحياة أم لا.

ثانيًا: تُعرض بطاقات الفقرات على عدد من المحكمين الأكفاء، الذين يتراوح عددهم بين ٤٠، ٦٠ محكماً وربما أكثر، ويطلب من كل منهم على حدة تصنيفها في مجموعات أو أقسام منفصلة وفقاً لدرجة آرائهم. واقتراح ثيرستون أن يكون عددها ٩ أو ١١ مجموعة أو قسماً، فالمجموعة الأولى يُعين لها الرقم (١) وتدل على درجة مرتفعة من التفضيل، والمجموعة الأخيرة يُعين لها الرقم (١١) مثلاً وتدل على درجة منخفضة من التفضيل، بينما الأرقام من (٢) إلى (١٠) تدل على درجات متفاوتة من الأفضلية،

والرقم ٦ يدل على النقطة المحايدة Neutral . وهنا يفترض أن المسافات بين المجموعات المتتالية تبدو كما لو كانت متساوية من وجهة نظر المحكم . وينبغي مراعاة ألا يقوم المحكمون بتقدير الفقرات من حيث موافقتهم أو عدم موافقتهم على ما تشير إليه كل فقرة ، وإنما يضع الفقرة على ميزان أفضلية Favorability Scale من حيث إدراك كل منهم لمدى إيجابية أو سلبية الفقرة فيما تعكسه من الاتجاه الذي يتم تقديره . فمثلا الفقرة الأولى المبنية أعلاه يفحصها المحكم ويقرر ما إذا كانت تعكس عدم أفضلية وليس عدم أفضلية تامة من حيث الاتجاه نحو التعليم بغض النظر عما إذا كان يوافق أم لا يوافق على ما تشير إليه الفقرة ، أي أن عليه أن يفكر فقط في الدرجة التي تعكس بها الفقرة «الأفضليّة» لموضوع الاتجاه أي التعليم مستقلا عن اتجاهه نحو الموضوع ، ويضع علامة (*) مقابل الرقم المناسب على ميزان الأفضليّة كالتالي :



وتكرر العملية لكل فقرة من الفقرات التي تم تجميعها مبدئياً. وهذا يتطلب أن يكون المحكمون على دراية جيدة بالمفهوم أو الموضوع أو القضية المتعلقة بالاتجاه، وكذلك بردود فعل وعمليات تفكير المجموعة العمرية من الأفراد الذين يُسنى ميزان الاتجاه من أجلهم. وبالطبع لا تتوقع أن يتفق المحكمون اتفاقاً تاماً على المجموعة أو القسم الذي يرون أن كل بطاقة تنتمي إليه، حيث ستجد تشتيتاً أو توزيعاً لأرقام الأقسام المختلفة لكل منها. وبين هذا التوزيع عدد المحكمين الذين عينوا رقماً معيناً مناظراً لأحد الأقسام ١١، ١٠، ٩، ...، ١ لكل فقرة أو بطاقة وليس من الضروري أن يكون عدد البطاقات متساوياً في كل قسم.

ثالثاً، نجد قيمة وسيط Median التوزيع لكل فقرة، وهذه تعد بمثابة قيمة ميزان الفقرة Item Scale Value لأنها تدل على موقع الفقرة على متصل الأفضلية، وبذلك تشير إلى شدة الاتجاه Intensity Scale، وقد سبق أن أوضحتنا في الفصل الثالث كيفية حساب قيمة الوسيط بيانياً وحسابياً.

رابعاً، نوجد تشتت التوزيع باستخدام مقاييس التشتت التي أوضحتها أيضاً في الفصل الثالث، ولكن يُفضل استخدام مقياس نصف المدى الريعي لأنّه يتعلق

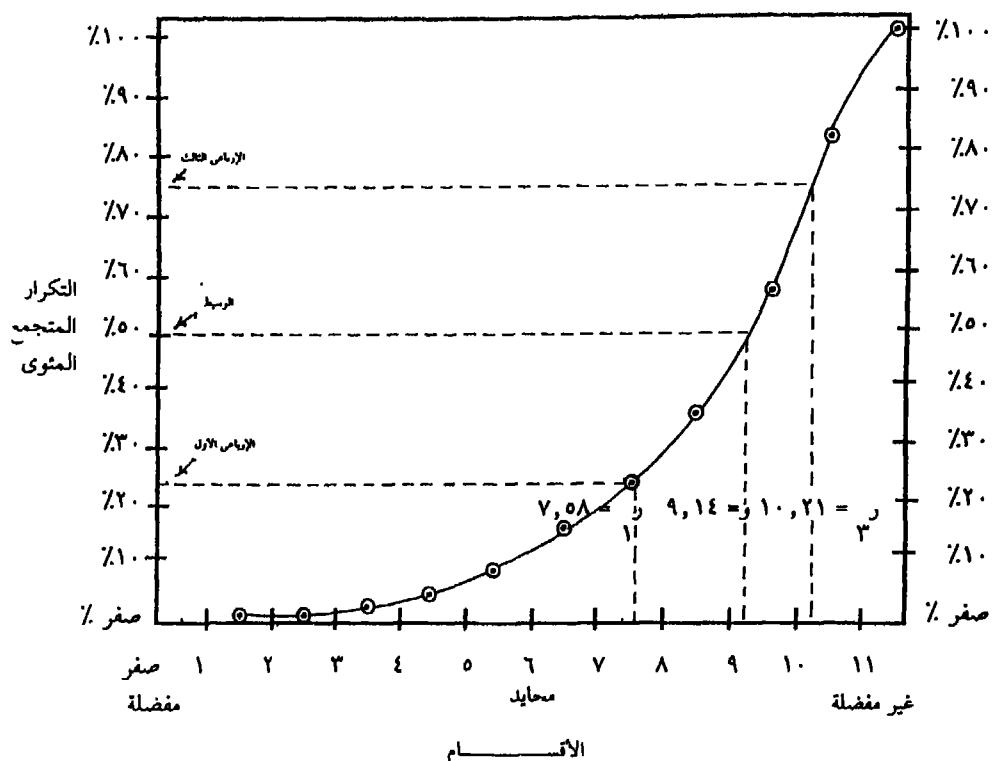
بالوسيط ، ويحدد مدى ٥٪ الوسطى من الحالات . فإذا كانت قيمة نصف المدى الريبي لفقرة ما كبيرة ، فإن هذا يعني أن هناك تشتتاً كبيراً في الأحكام المتعلقة بفقرة معينة مما يتطلب حذفها نظراً لأنها ربما تبدو غامضة ، وبالتالي لا ينبغي تضمينها في ميزان الاتجاه في صورته النهائية . وبعد حذف مثل هذه الفقرات تُستبعن الفقرات التي وسيطها أو قيم ميزانها موزعة توزيعاً متعدلاً على المتصل الافتراضي للميزان . وبذلك ربما نحصل على قيم للوسيط مثل ١,٧ ، ١,٨ ، ٢,٢ ، ٢,٩ ، ٣,٥ ، ٣,٤ ، ٣,٢ ، ٣,١ . فمن المرغوب فيه أن يكون لدينا مجموعة من الفقرات قيم وسيطها تمتد من أحد طرفي المتصل إلى قرب الطرف الآخر بحيث تكون جميع تداريج الاتجاه من الموجب إلى السالب مُمثلة ، غير أنه يُحتمل أن نجد عدداً قليلاً من الفقرات متجمعة عند نقطتين متعددة على طول الميزان ، وعندئذ يمكننا اختيار واحدة من الفقرات المتجمعة عند نقطة معينة يكون اتفاق المحكمين عليها متقارباً . فمثلاً إذا كانت قيمة وسيط كل من فقرتين ٣,١٩ ، ٣,١٥ ووجدنا أن توزيع الأحكام للفقرة الأولى أقل تشتتاً من توزيعها للفقرة الثانية ، فإنه ربما نختار الفقرة الأولى ونستبعد الثانية ؛ فالفارقتان اللتان وسيط كل منهما ٤ ، ٣ على الترتيب مثلاً ، يقعان على المتصل ويتبعان عن بعضهما البعض وحدة واحدة لفرق الذي يمكن ملاحظته Just Noticeable Difference (JND) . لذلك ينبغي اتخاذ قرار فيما يتعلق بمدى المتصل ، وعدد الفقرات التي تحتاج إلى اختيارها . وعموماً فإنه يكفي عادة اختيار ٢٠ أو ٢٥ فقرة ، وإن كان من الممكن اختيار عدد أقل . وربما تقرر بناء صيغتين متكافتين من ميزان الاتجاه إذا كان لدينا عدد كاف من الفقرات يسمح بذلك ، ولكن ينبغي أن تُتوزع أزواج الفقرات عشوائياً على كل من الصيغتين بحيث يكون لكل زوج منها نفس قيمة الوسيط ونصف المدى الريبي ، وتُتوزع الفقرات عشوائياً أيضاً داخل كل صيغة . ويمكن الإفاده من الصيغتين المتكافتين في دراسة تغير الاتجاهات .

ويمكن توضيح كيفية إيجاد قيم الوسيط ونصف المدى الريبي لإحدى فقرات ميزان الاتجاه نحو التعليم الذي أشرنا إليه ، ولتكن الفقرة الأولى . ونفترض أن توزيع الأحكام كما هو مبين بجدول (١-١٢) التالي :

النسبة المئوية للمجتمعة	النسبة المئوية	عدد المحكمين	رقم المجموعة أو القسم
صفر	صفر	صفر	١
صفر	صفر	صفر	٢
٢	٢	١	٣
٤	٢	١	٤
٨	٤	٢	٥
١٦	٨	٤	٦
٢٤	٨	٤	٧
٣٦	١٢	٦	٨
٥٨	٢٢	١١	٩
٨٢	٢٤	١٢	١٠
١٠٠	<u>١٨</u>	<u>٩</u>	<u>١١</u>
	<u>% ١٠٠</u>	<u>N = ٥٠</u>	

جدول (١-١٢) يوضح توزيع افتراضي لبيانات التصنيف لفقرة ميزان الاتجاه نحو التعليم (٣٠٠ محكم)

ويلاحظ من جدول (١-١٢) أن المحكمين الذين بلغ عددهم ٥٠ محكماً، قد صنفوا الفقرة في المجموعات أو الأقسام التي عددها ١١ قسماً، ولم يصنف أحد من المحكمين الفقرة في القسمين (١)، (٢) في العمود الأول، أى أن عدد المحكمين (صفر) لكل من هذين القسمين، وحصلنا على النسبة المئوية لعدد المحكمين في العمود الثالث بقسمة عدد المحكمين في كل قسم على العدد الكلى للمحكمين وضرب الناتج في ١٠٠ . والنسبة المئوية المجتمعة في العمود الرابع حصلنا عليها بالجمع المتتالي للنسب المئوية المناظرة لكل قسم في العمود الثالث . ويمكن رسم المنهجي المجتمع الصاعد النسبي بتمثيل الأقسام على المحور الأفقي، والنسبة المئوية المجتمعة على المحور الرأسى، ووضع نقط أعلى متصرفات الأقسام تنتظر النسب المئوية المجتمعة كما هو موضح بشكل (٢-١٢) التالي :



شكل (٢-١٢) يوضح المعنى التكرارى المجتمع الصاعد لفقرة ميزان الاتجاه نحو التعليم

ويتضح من شكل (٢-١٢) أن قيمة وسيط هذه الفقرة ٩,١٤ ، ونصف المدى

$$\text{الرباعي} = \frac{R_3 - R_1}{2} = \frac{2,63 - 7,58}{2} = \frac{10,21}{2} = 1,32$$

ويمكن بالطبع الحصول على هاتين القيمتين جبرياً (سبق توضيح ذلك في الفصل الثالث) أي أن قيمة ميزان الفقرة = ٩,١٤ ، والشتت = ١,٣٢

خامساً: نختار الفقرات بحيث تكون الفترات التي تفصل بينها متساوية، وقيمة نصف المدى الرباعي لها منخفضة، ويمكن اختيار سعة الفقرات وحدة واحدة أو نصف وحدة أو غير ذلك. فإذا قررنا اختيار ٢٠ فقرة مثلاً، فإنه ربما نختار الفقرات التي تقترب قيمة وسيط كل منها من ٢,٥ ، ٣,٥ ، ٤ ، ٤,٥ مثلاً، وهكذا، وبذلك تكون فترات ميزان الاتجاه متساوية ظاهرياً. وربما نود في بعض الحالات أن يكون هناك تمييز دقيق في جزء معين من الميزان عن غيره من الأجزاء، وعندئذ يمكن أن نختار الفقرات التي قيمتها متقاربة في هذا الجزء.

سادساً: تُرتب الفقرات التي تم اختيارها بطريقة عشوائية، وتُدون في قائمة دون ذكر قيمة ميزان كل منها Scale Values. وعندما يستخدم ميزان الاتجاهات يُطلب من سُيُّطِق عليهم إبداء موافقتهم أو عدم موافقتهم فقط على كل فقرة، أو اختيار عدد محدد من الفقرات التي تعبّر عن موافقتهم بدرجة أفضل. وننظر فقط إلى الفقرات التي أبدى كل منهم موافقته عليها، ونحسب متوسط قيم ميزان هذه الفقرات (الوسيط)، وهذه القيمة المتوسطة تحديد موقع الفرد على متصل الاتجاه (التفضيل - عدم التفضيل). ومن الناحية النظرية ينبغي أن يُبدي الفرد موافقته فقط على عدد قليل من الفقرات المتجاورة في قيمة ميزان كل منها التي تقترب من موقعه الحقيقي على متصل الاتجاه. فإذا ما أبدى عدد كبير من الأفراد موافقته على فقرات غير متجاورة في قيمة ميزان كل منها، فإن هذا يمكن أن يكون دليلاً على أن الميزان متعدد الأبعاد، وهذا ربما يتطلب تقسيم الميزان إلى موازين فرعية أحادية البُعد. وعلى الرغم من أن هذا الأسلوب في بناء موازين الاتجاهات يتعلق بدرجة أساسية بتحديد موقع كل فقرة على متصل الميزان الافتراضي، إلا أنه ينبغي الحفطة عند معالجة قيم الميزان كأعداد فعلية أو كوحدات يمكن جمعها أو تبادلها على المتصل الخطي. فنحن هنا لا نتعامل مع فترات متساوية وإنما مع فترات متساوية ظاهرياً Equal-Appearing Intervals، أي أنها وحدات سيكولوجية وليس وحدات عددية، حيث يمكن أن تكون متساوية أو غير متساوية عددياً. كما أن تقسيم المتصل إلى (11) قسم يُعد تقسيماً اعتبارياً. لذلك لا يجب أن نعتقد أن الدرجة (٨) تساوى ضعف الدرجة (٤) في قيمتها العددية، كما أنها لا نستطيع المقارنة بين الدرجات من ميزان إلى آخر في بطارية تشتمل على عدد من موازين الاتجاهات.

بعض ميزاته وكيفية أسلوب الفترات المتساوية ظاهرياً :

يتميز هذا الأسلوب الذي اقترحه ثيرستون Thurstone بأنه يؤدي إلى موازين للاتجاهات تُمكّن من التمييز بين عدد كبير من الأفراد في موقع كل منهم على ميزان فتراته المتساوية ظاهرياً كما سبق أن أوضحنا. فعندما يوجد متوسط قيم ميزان الفقرات التي وافق عليها كل فرد نستطيع أن نُحدّد موقع مختلفة لاتجاهاتهم، وبذلك يمكن التمييز الدقيق بينهم فيما يتعلق بالاتجاه نحو موضوع معين. ويرى بعض مؤيدي هذا الأسلوب أن اعتماد قيم ميزان كل فقرة Scale Value على اتفاق مجموعة أكفاء من المحكمين الذين يقومون بفحص كل فقرة واستبعاد الفقرات التي لم تحظ بقدر كبير من الاتفاق يُمكّنا من استخدام الميزان بمزيد من الثقة في ثبات درجاته.



غير أنه قد وجّه كثير من النقد لهذا الأسلوب، وبخاصة فيما يتعلق بالوقت والجهد الذي يتطلبه في بناء موازين الاتجاهات من حيث التجميع المبدئي لعدد كبير من الفقرات وتقسيمها بواسطة عدد كبير أيضًا من المحكمين كل على حدة، وحساب قيم ميزان كل فقرة، ومقدار تشتت هذه القيم استناداً إلى هذا التقييم، واختيار الفقرات التي سوف يشتمل عليها الميزان في صيغته النهائية، وفقاً لمحكمات معينة ، وإيجاد متوسط درجات كل فرد يطبق عليه الميزان لإيجاد موقعه النسبي على متصل الاتجاه موضع القياس. كذلك من المحتمل أن يتحدد موقع واحد على المتصل لفردين متوسط قيم ميزان الفقرات التي أبدوا موافقتهم عليها متساوياً على الرغم من اختلاف قيم وسيط الفقرات في الحالتين، إذ يمكن أن يوافق أحدهما على فقرات قيمة وسيط كل منها ٢ ، ٤ ، ٦ على الترتيب ومتوسطها ٤ ، ويوافق الآخر على فقرات مختلفة وسيط كل منها ٣ ، ٤ ، ٥ ومتوسطها ٤ أيضاً، مما يجعل تفسير درجة اتجاه كل منها موضع تساؤل، نظراً لاختلاف نمط اتجاه كل منها.

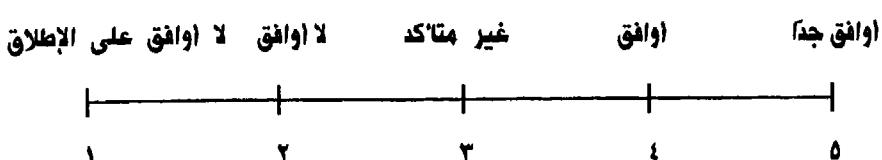
كما أن تصنيف المحكمين للفقرات ربما يتأثر باتجاهاتهم الشخصية وخلفياتهم الثقافية مما يؤدى إلى اختيار فقرات لا تمثل موقع ثابتة على متصل الاتجاه. إذ إن أسلوب ثيرستون يستند إلى افتراض أساس يتعلق بعدم اختلاف قيم ميزان الفقرات باختلاف عينات المحكمين. غير أن بعض الدراسات أوضحت أن هذا الافتراض يتحقق إذا لم يكن هناك تحيز واضح من جانب المحكمين فيما يتعلق بمتصل الاتجاه المراد قياسه. أما إذا كان المحكمون متاثرين بشكل ما بموضع الاتجاه فإن هذا ربما يؤدى إلى تغيير في قيم ميزان الفقرات (Ager & Dawes, 1965).

وعومماً فإن أسلوب الفقرات المتساوية ظاهرياً الذي اقترحه ثيرستون لا يتباين بالسلوك بدرجة أفضل من غيره من الأساليب الأخرى التي سنوضحها.

أسلوب التقدير الجماعي ليكرت:

الأسلوب الثاني لبناء موازين الاتجاهات يسمى أسلوب التقدير الجماعي Summated Ratings، أو أسلوب ليكرت Likert Technique. وهذا الأسلوب يتطلب جهداً ووقتاً أقل مما يتطلبه أسلوب ثيرستون، ويؤدي إلى نتائج مماثلة، حيث إنه لا يعتمد على تقييم المحكمين، لذلك يُعد من الأساليب الشائعة الاستخدام في القياس والبحوث التربوية والنفسية والاجتماعية. ويعتمد هذا الأسلوب على القياس الرتبى للاتجاهات، حيث يقدم للفرد قائمة تشتمل على عبارات أو فقرات، ويطلب منه إبداء

موافقته أو عدم موافقته بدرجات متفاوتة تعكس شدة اتجاهه، فالأفراد الذين تكون موافقتهم أقل شدة، والأفراد الذين يتبعون موقعهم على هذا المتصل يتوقع أن يكون سلوكهم مختلفاً في المواقف الاجتماعية المتعلقة بموضوع الاتجاه. غير أن ارتباط السلوك الفعلى بالاتجاه لا يزال مثار جدل بين علماء النفس، وتحدد شدة الاتجاه جزئياً بإعطاء أوزان مختلفة لاستجابة الفرد لكل فقرة. فمثلاً إذا كانت الفقرة «أحب مهنتي» فإنه يمكن أن يستجيب الفرد على ميزان رتبى متدرج يشتمل على خمس نقاط كالتالى:



فالاستجابة «أوفق جدًا» يُعين لها الوزن الرقمي (٥)، والاستجابة «لا أوفق على الإطلاق» يُعين لها الوزن الرقمي (١) وهكذا. وكلما زاد الوزن الرقمي دل ذلك على حب الفرد لمهنته والعكس صحيح، فإذا استجاب ١٠٠ فرد على هذه الفقرة، فسوف نحصل على ١٠٠ استجابة تتراوح أوزانها الرقمية بين (١) ، (٥). فجميع الأفراد الذين يستجيبون «أوفق جدًا» يحصلون على الوزن الرقمي (٥)، أما الذين يستجيبون «لا أوفق على الإطلاق» يحصلون على الوزن الرقمي (١). ولكن ليس بالضرورة أن جميع من أبدوا موافقتهم التامة على الفقرة يكون لديهم نفس المشاعر نحو مهنتهم؛ لأن فقرة واحدة لا تكفي بالطبع لتقدير هذه المشاعر. لذلك يستند أسلوب التقدير الجمعي على مجموعة من الفقرات من أجل التوصل إلى تقييم أكثر دقة، والتمييز بين الأفراد في شدة اتجاههم نحو موضوع معين على ميزان متدرج مثل المبين أعلاه. فإذا كان الميزان يشتمل على خمس نقاط، وكان لدينا (١٠) فقرات، فإن أعلى درجة يحصل عليها فرد سوف تكون (٥٠) وأقل درجة (١٠)، أي أن المدى أصبح (٤١) للفقرات العشر بدلاً من (٥) للفقرة الواحدة، وكلما زاد عدد الفقرات زاد المدى، أي زاد تباين الدرجات مما يؤدي إلى زيادة دقة القياس. وبالطبع لا يقتصر هذا الأسلوب على نمط الاستجابات الموضع أعلاه، وإنما يمكن استخدام أي ميزان رتبى متدرج آخر، مثل:

(١) موافق، غير متأكد ، غير موافق.

(٢) موافق بشدة، موافق، غير متأكد ولكن موافق إلى حد ما، غير متأكد ولكن غير موافق إلى حد ما، غير موافق، غير موافق على الإطلاق.

(٣) دائمًا ، غالباً، أحياناً، نادراً، إلخاً.

ويُفضل ألا يزيد عدد هذه الأقسام عن خمسة لكي يتمكن الفرد من التمييز بينها واختيار درجة موافقته بدقة. كما يلاحظ أن النمطين ١، ٢ اشتتملا على القسم «غير متأكد» حيث يفترض أن الفرد الذي يختاره لا يكون له رأي، ومع هذا يستخدم أحياناً في مثل هذه الموازين. ولكن النمط الثاني اشتمل على ستة أقسام لكي يحدد الفرد توجهه سواء جهة اليمين إذا اختار «غير متأكد ولكن غير موافق إلى حد ما» أو جهة اليسار إذا اختار «غير متأكد ولكن غير موافق إلى حد ما»، وذلك لأن الاستجابة «غير متأكد» لا تفيد في تحليل البيانات. وبالطبع فإن استخدام القسم «غير متأكد» في موازين الاتجاهات يعتمد على خصائص عينة المستجيبين ومستوى تعليمهم ومدى افتراضهم بالمطلوب منهم.

خطوات بناء موازين الاتجاهات باستخدام أسلوب ليكرت:

يمكن بناء موازين الاتجاهات باستخدام أسلوب التقدير الجمعى (أسلوب ليكرت) باتباع الخطوات التالية:

أولاً : تجميع عدد كبير من الفقرات التي تتعلق بموضوع الاتجاه المراد قياسه مثل الاتجاه نحو المعلم، أو نحو مجال دراسي معين، أو نحو موضوعات عامة مثل الانفتاح الاقتصادي، أو تقنيات الاتصال وغير ذلك؛ ونظرًا لأن كل فقرة في الميزان التجمعي تهدف للحصول على تباين متسبق فيما يتعلق بالاتجاه، فإن معظم الفقرات ينبغي أن تتضمن مدى متسعاً من حيث شدة التفصيل *Favorableness* ، مع مراعاة أنه لا توجد فقرات محابية فعلاً في مثل هذا الميزان أو فقرات متطرفة . والفقرات المتطرفة في الشلة في كلا الاتجاهين تؤدي إلى تباين أقل من الفقرات الأقل تطرفاً، ويكون تميزها ضعيفاً، ويمكن أن يتراوح عدد الفقرات بين ٤٠ ، ٤ ، ٦٠ فقرة تكون مقسمة تقسيماً متعدلاً بين الفقرات الموجبة والفقرات السالبة.

ثانياً : تطبيق الفقرات على مجموعة من الأفراد مماثلة للمجموعة المستهدفة أى التي سيطبق عليها ميزان الاتجاه في صيغته النهائية فيما بعد، ويمكن أن تشتمل هذه المجموعة على عدد من الأفراد يساوى عشرة أمثال عدد الفقرات (Nunnally, 1978). ويُفضل زيادة عدد الأفراد عن ذلك إذا أمكن، ويراعى جعل شروط أو ظروف التطبيق

مماثلة بقدر الإمكان لظروف التطبيق على المجموعة المستهدفة، ويُطلب من كل فرد أن يُبدي رأيه فيما يتعلق بكل فقرة ويُحدّد شدة موافقته، أى ما إذا كان موافقاً جداً، أو موافقاً، أو غير متأكد، أو غير موافق، أو غير موافق على الإطلاق؛ لذلك ينبغي أن تكون مجموعة الأفراد مماثلة للمدى الكلّي للاتجاه المراد قياسه بقدر الإمكان.

ثالثاً : تعيين درجة أو وزن رقمي لاستجابة الفرد على كل فقرة وفقاً للأوزان التي أشرنا إليها، مع مراعاة أن تعكس هذه الأوزان الرقمية للفقرات السالبة .

فمثلاً: إذا كانت الفقرة كالتالي: «تعلّمني المدرسة ما يساعدني في الحصول على عمل» فإن: موافق جداً = ٥ ، موافق = ٤ ، وغير متأكد = ٣ ، وغير موافق = ٢ . وغير موافق على الإطلاق = ١ .

أما إذا كانت الفقرة كالتالي : «المدرسة مضيعة للوقت» فإن : موافق جداً = ١ ، موافق = ٢ ، وغير متأكد = ٣ ، وغير موافق = ٤ ، وغير موافق على الإطلاق = ٥ .

وتجمع درجات الفرد في جميع الفقرات التي استجاب لها، فمثلاً إذا استجاب «موافق جداً» للفقرة الأولى (الدرجة = ٥) ، «وغير متأكد» للفقرة الثانية (الدرجة = ٣)، «وموافق» للفقرة الثالثة (الدرجة = ٤) ، «وغير موافق» على الفقرة الرابعة (الدرجة=٢)، فإن درجة الفرد = ٥ + ٣ + ٤ + ٢ = ١٤ ، وبالمثل في الفقرة الثانية مع مراعاة الأوزان الموضحة .

رابعاً : إجراء تحليل فقرات Item Analysis لاختيار الفقرات التي تميز بين الأفراد بدرجة أفضل على متصل الاتجاه، وذلك بإيجاد الارتباط بين الدرجات الكلية التي يحصل عليها الأفراد واستجاباتهم لكل فقرة، والاستبقاء على الفقرات التي يكون ارتباطها مرتفعاً، وبذلك نحصل على ميزان متسق داخلياً. أو يمكن إيجاد معامل تميز كل فقرة كما سبق أن أوضحنا في الفصل السابع. ويفضل ترتيب قيم معامل الارتباط ترتيباً تناظرياً واختيار عدد متساو تقريباً من الفقرات السوجية والفقرات السالبة وضمهم معاً لتكون ميزان الاتجاه في صيغته الأولية.

خامساً : إيجاد معامل ألفا لكرونباك (α) Cronbach للفقرات التي تم اختيارها، وذلك بإيجاد تباين درجات كل منها وتبالين الدرجات الكلية وتطبيق الصيغة (٤-١٢) التي سبق أن أوضحناها في الفصل الرابع، أو تكوين مصفوفة الارتباطات بين درجات الفقرات والدرجات الكلية وإجراء تحليل المفردات باستخدام الحاسوب.

وسوف تفيد هذه المصفوفة أيضاً في الخطوة التالية. فإذا كانت قيمة هذا المعامل مرتفعة بدرجة كافية (٨٠، أو أكبر) فإنه يمكن أن يشتمل ميزان الاتجاه على هذه الفقرات في صيغته النهائية (حوالى ٢٠ فقرة). وينبغي ملاحظة ما ذكرناه سابقاً من أن قيمة معامل الثبات تعتمد على عدد أقسام ميزان التقدير المقابل لكل فقرة. وهذه الخطوة تعد بمثابة مؤشر لخاصة أحادية البُعد Unidimensionality، حيث تتحقق بها من أن جميع الفقرات تقيس شيئاً واحداً، وهو ما اهتم به ليكرت اهتماماً أساسياً.

سادساً : ترتيب الفقرات المُختارة بطريقة عشوائية في ميزان الاتجاه، أي الاستبيان، وبذلك يمكن استخدامه في قياس الاتجاه المطلوب حيث يُطبق على المجموعة المستهدفة، وإيجاد الدرجة الكلية لكل فرد بجمع الدرجات المناظرة لاستجاباته على الفقرات التي يشتمل عليها الاستبيان.

والدرجة الخام الكلية التي يحصل عليها الفرد يكون معناها محدوداً في ذاتها، ويُتضح معناها بدرجة أفضل إذا قورنت بدرجات الأفراد الآخرين الذين طبق عليهم ميزان الاتجاه، حيث يمكن أن يتضح موقعه على متصل الاتجاه بالنسبة لموقع هؤلاء الأفراد.

ميزايا وعيوب أسلوب التقدير الجمحي :

يتميز هذا الأسلوب بأنه يجمع في بناء موازين الاتجاهات بين خبرة من يقوم بإعدادها والتفكير المنطقي في اشتقاء الفقرات من نطاق نظري شامل للاتجاه المراد قياسه. ويمكنه صياغة الفقرات بما يراه مناسباً مادام ملتزماً بالربط المنطقي بين محتوى الفقرات وموضع الاتجاه. ويستند هذا الأسلوب إلى عدد من الفروض أقل مما تستند إليه الأساليب الأخرى، ويتميز بسهولة حساب درجات الفقرات والدرجات الكلية ومقارنتها بغيرها من الدرجات على متصل الاتجاه. ويعتمد تفسير الدرجات على منطق بسيط مؤده أنه كلما زادت الدرجة الكلية التي يحصل عليها الفرد دل ذلك على زيادة شدة اتجاهه.

لذلك يُستخدم هذا الأسلوب بكثرة في تصميمات البحوث التي تستكشف بعض الظواهر النفسية الاجتماعية، ويتميز كذلك بالمرونة حيث يستطيع الباحث أن يجعل ميزان الاتجاه مشتملاً على عدد من الفقرات يراه مناسباً، وزيادة عدد الفقرات يمكنه من الكشف عن الفروق الفردية في الاتجاه المراد قياسه، وبخاصة إذا كان عدد أفراد

المجموعة كثيراً. غير أن بعض الباحثين يطلقون على أي ميزان متدرج يشتمل على عدد من الأقسام «ميزان ليكرت» على الرغم من أنهم لا يستخدمون في بنائه أسلوب تحليل الفقرات أو التتحقق من أن الفقرات تقسيس بُعداً أحادياً، وبذلك يُساء استخدام هذا الأسلوب، ولا يُفاد من نتائجه الفائدة المرجوة.

ولكن يُعبَّر على أسلوب التقدير الجماعي أنه يقتصر على مستوى القياس الرتبى، أي يسمح بترتيب الأفراد على متصل الاتجاه من حيث درجة موافقة كل منهم دون معرفة مقدار زيادة درجة موافقة الفرد عن غيره من الأفراد، أو تحديد مقدار التغير الذى حدث لديه فى الاتجاه بعد مروره بخبرات معينة، حيث إن ذلك يتطلب مستوى قياس فترى. كما أن الدرجة الكلية لفرد ما لا يكون لها معنى واضح نظراً لأن أنماطاً متعددة من الاستجابات للفقرات يمكن أن تؤدى إلى الدرجة الكلية نفسها. غير أنه يمكن التغلب على ذلك بالعنابة بتحليل الفقرات بحيث نضمن أن يشتمل الميزان فقط على الفقرات التي تميز جيداً بين الدرجات الكلية المرتفعة والدرجات الكلية المنخفضة، وقد سبق أن رأينا أن هذا ينطبق أيضاً بدرجة كبيرة على أسلوب ثيرستون نظراً لكثره أنماط الاستجابات الممكنة في هذه الحالة. غير أن البعض يرى أن اختلاف أنماط الاستجابات التي تؤدي إلى الدرجة الكلية نفسها يمكن إعزاؤه إلى تباين الخطأ في الاستجابات وليس إلى تباين حقيقى في الاتجاه. ونظراً لتنوع الفقرات التي يشتمل عليها الميزان فإن هذه الاختلافات في أنماط الاستجابات لكل فقرة التي لا تتعلق بالاتجاه الذي يجري قياسه يلغى بعضها البعض الآخر، ومع هذا فإنه ينبغي الاهتمام ببعض أنماط الاستجابات لكل فقرة أكثر من الاهتمام بالدرجات الكلية نظراً لأن معناها يختلف باختلاف هذه الأنماط.

الأسلوب التراكمي (اجتماع):

إن موازين الاتجاهات التي تُبنى استناداً إلى الأسلوبين السابقين تشتمل على فقرات غير متجانسة فيما يتعلق بالأبعاد المختلفة لموضوع الاتجاه المراد قياسه. فمثلاً في بناء ميزان اتجاه متعلق بالتعليم ربما لا نفصل بين الفقرات أو العبارات التي تتعلق بالممارسات السائدة في التعليم الإلزامي، والفقرات أو العبارات التي تتعلق بفكرة

التعليم الذاتي غير الرسمي، أو غير ذلك من جوانب أو أبعاد الاتجاه نحو التعليم. فالجمع بين جوانب أو أبعاد مختلفة في ميزان واحد يؤدي إلى صعوبات في تفسير الدرجات التي يحصل عليها الأفراد تفسيراً واضحاً، وبخاصة أن أنماط هذه الدرجات تكون متعددة كما سبق أن أوضحنا.

لذلك اقترح جستمان Guttman أسلوباً للتنبّل على هذه المشكلة أطلق عليه الأسلوب التراكمي Cummulative Technique، ويمكن باستخدام هذا الأسلوب بناء موازين اتجاهات أحادية البُعد Unidimensional. ويعُد ميزان الاتجاه أحادي البُعد فقط إذا أدى إلى ميزان تراكمي، ونعني بذلك أن الفقرات ترتبط فيما بينها بحيث إذا وافق الفرد على الفقرة (٢) فإنه يوافق أيضاً على الفقرة (١) التي تسبقها، ومن يوافق على الفقرة (٣) فإنه يوافق أيضاً على الفقرتين (١)، (٢) اللتين تسبقانها، وهكذا.

فمثلاً إذا طلبنا من الأفراد إبداء موافقتهم أو عدم موافقتهم على العبارات التالية المتعلقة بالاتجاه نحو منظمة الأمم المتحدة:

(أ) منظمة الأمم المتحدة تُقدِّر جميع الشعوب.

(ب) منظمة الأمم المتحدة هي أملنا في التعايش السلمي.

(ج) منظمة الأمم المتحدة تُعد قوة بناء في العالم.

(د) ينبغي استمرار المشاركة في منظمة الأمم المتحدة.

إذا كانت هذه الفقرات تمثل ميزاناً تراكمياً، فإننا نتوقع أن الفرد الذي يوافق على الفقرة (أ) سوف يوافق على الفقرات (ب)، (ج)، (د)، والفرد الذي لا يوافق على الفقرة (أ)، ولكنه يوافق على الفقرة (ب) نتوقع أن يوافق على الفقرتين (ج)، (د)، وبذلك يمكن ترتيب جميع استجابات الأفراد في نمط يماثل الموضح بجدول (٢-١٢) التالي، حيث تُقدر درجة واحدة لكل استجابة «موافق» على أي من الفقرات الأربع:

الاستجابة «غير موافق» على الفقرة				الاستجابة «موافق» على الفقرة				الدرجة
أ	ب	ج	د	أ	ب	ج	د	
			x	x	x	x	x	٤
		x		x	x	x		٣
	x	x		x	x			٢
x	x	x		x				١
x	x	x	x					صفر

جدول (٢-١٢) يوضح نمط الاستجابات على أربع فقرات في ميزان تراكمي

خطوات بناء موازين الاتجاهات باستخدام أسلوب جتمان:

عند بناء ميزان اتجاه تراكمي ينبغي كتابة أو تجميع عدد من الفقرات تتطلب الاستجابة «موافق» أو «غير موافق» وترتيبها بحسب درجة الموافقة، وأحياناً يمكن ترتيبها عشوائياً دون محاولة تحديد ما إذا كانت الفقرات الفاصلة بينها متساوية، حيث إن الميزان التراكمي يُعد من المستوى الرتبى. ولكن ينبغي التتحقق من أن جميع الفقرات تمثل بُعداً أحادياً للاتجاه المراد قياسه، أي تَتَخَذْ نمطاً مثلياً مماثلاً لما هو موضح بجدول (٢-١٢). فإذا أمكن التوصل إلى هذا النمط التام يصبح لمجموعة الفقرات التي يتم اختيارها معنى واحد، أي إذا علمنا الدرجة التي يحصل عليها فرد يمكن دون الرجوع إلى استبيان الاتجاه معرفة الفقرات التي أبدى موافقته عليها.

ولكن يصعب الحصول في الواقع الفعلى على مثل هذا النمط المثلثي التام الذي يعكس بُعداً أحادياً لميزان الاتجاه، وذلك لأنّه يوجد عادة عدم اتساق في الاستجابات، أي توجد أخطاء . ويمكن توضيح ذلك بنمط الاستجابات الذي يوضحه جدول (٣-١٢) التالي :

الفرقات							
الدرجة	أقل شدة			أكبر شدة			الفرد
	هـ	د	جـ	بـ	أـ		
٤	+	+	(-	+	+		١
٤	+	+	+	(-	+		٢

الخطأ	صفر	١	١	صفر	صفر
-------	-----	---	---	-----	-----

جدول (٣-١٢) يوضح الخطأ في بعض أنماط الاستجابات

ويتضح من جدول (٣-١٢) أن الفرد الأول استجاب بالموافقة على الفقرتين (أ)، (ب)، ولم يوافق على الفقرة (جـ)، ووافق على الفقرتين (د)، (هـ) وحصل على الدرجة (٤).

فعدم موافقته على الفقرة (جـ) يُعد في هذه الحالة من الخطأ Error ، وكذلك لم يوافق الفرد الثاني على الفقرة (بـ) حيث يُعد ذلك أيضاً من الخطأ، لذلك وضعنا دائرة على الفقرتين (جـ)، (بـ) بالنسبة لـكل من الفردین على الترتیب، وربما يدل ذلك على أن میزان الاتجاه ليس أحادی الـبعد. أما نمط الاستجابات الموضح في جدول (٢-١٢) فيدل على میزان اتجاه تراکمی أحادی الـبعد. ويمكن التنبؤ بنمط الدرجات من معرفتنا للدرجة الكلية لـلفرد حيث إن نمط الاستجابات ليس به أخطاء (علامات سالبة)، أي يمكن إعادة تكوين هذا النمط Reproducible . وإذا كان عدد الأخطاء في نمط استجابات مجموعة من الأفراد كبيراً، فإن هذا يعني عدم إمكانية إعادة تكوين هذا النمط Nonreproducible .

وقد اقترح جتمان Guttman صيغة يمكن باستخدامها إيجاد معامل إعادة تكوين نمط الاستجابات Coefficient of Reproducibility كالتالي :

$$\text{معامل إعادة تكوين نمط الاستجابات} = 1 - \frac{\text{عدد الأخطاء}}{\text{عدد الاستجابات}} \dots \dots \dots (١-١٢)$$

والمقصود بالأخطاء كما أشرنا الابتعاد عن النمط المثلثي المثالي.

وعدد الاستجابات = عدد الأفراد × عدد الفقرات.

فإذا نظرنا إلى جدول (٣-١٢) نلاحظ أن عدد الأخطاء = ٢ ، وعدد الأفراد = ٢ ،
وعدد الفقرات = ٢ ، وبتطبيق الصيغة (١-١٢) نجد أن:
معامل إعادة تكوين نمط الاستجابات = $1 - \frac{2}{2 \times 2} = 1 - \frac{2}{4} = 1 - 0.5 = 0.5$

وقد اقترح جتمان أن يكون الحد الأدنى لقيمة هذا المعامل لمجموعة من الفقرات التي يشتمل عليها الميزان التراكمي للاتجاه ٩٠ ، لكن يُطلق على الميزان أحادي البعد. ونظراً لأننا اعتمدنا في المثال السابق على فقرتين فقط انخفضت قيمة هذا المعامل إلى ٥٠ . لذلك أوصى جتمان اختيار ١٠ أو ١٢ فقرة وتُطبق على عدد لا يقل عن ١١ فرد، ويتم اختيار هذه الفقرات من نطاق شامل للفقرات المتعلقة بموضوع الاتجاه.

ميزات وعيوب الأسلوب التراكمي :

يتميز هذا الأسلوب كما اتضح لنا بتركيزه على خاصية **البعد الأحادي** للفقرات في موازين الاتجاهات، وبذلك يمكن تحديد عدم اتساق استجابات الأفراد، وربما التعرف على الاستجابات غير الصادقة مما يزيد من الثقة في المعلومات التي يقدمها المستجيبون. ويسهل استخدام الأسلوب التراكمي إذا كان عدد الفقرات قليلاً، أما إذا زاد العدد عن ١٢ فقرة، فإن استخدامه يكون مجدهاً، بل غير عملي، كما يتميز بأنه يمكن إعادة تكوين أنماط الاستجابات إذا علمنا الدرجة الكلية التي حصل عليها الفرد في ميزان الاتجاه، وهذا بالطبع لا يتحقق في أسلوب ثيرستون أو ليكرت.

غير أنه ليس من الأساليب الأكثر فاعلية في قياس الاتجاهات نحو موضوعات معقدة أو في التنبؤ بالسلوك المتعلق بها. كما أن خاصية **البعد الأحادي** ربما تنطبق على مجموعة معينة من الأفراد دون الأخرى ، أو ربما تنطبق عليها في وقت ما ولا تنطبق في وقت آخر، وبخاصة أن التطورات التي حدثت في أساليب بناء موازين الاتجاهات تؤكد أهمية بناء موازين متعددة الأبعاد حيث أوضحت البحوث السيكولوجية أن الأفراد لا يستجيبون عادة استجابات أحادية **البعد**، مما يجعل استخدام أسلوب جتمان غير واقعى. كما أن هذا الأسلوب يُعد من الأساليب التي تستوي إلى النماذج الاحتمالية في حين أن التوجهات المعاصرة اهتمت ببناء نماذج احتمالية يُسند إليها في قياس السمات الإنسانية بما في ذلك الاتجاهات. ولعل النظرية المعاصرة في القياس التربوي والنفسى التي سنوضحها في الفصل الرابع عشر تُعد أحد هذه التوجهات.



بعض الأساليب الخالدة:

تُعدُّ أساليب بناء الموازين التي ناقشناها فيما سبق أكثر الأساليب استخداماً في قياس الاتجاهات، غير أن هناك أساليب خاصة أخرى أقل استخداماً ولكنها تفيد أيضاً في قياس كثير من الجوانب الوجدانية للأهداف التعليمية وكذلك في قياس الاتجاهات. وسوف نوضح بعض هذه الأساليب فيما يلى:

أسلوب تمييز مفاهيم Semantic Differential Technique :

قام أوزجود Osgood بجامعة إيلينوي الأمريكية عام ١٩٥٢ بنشر نتائج دراسته المتعلقة بمكونات معانى المفاهيم، حيث جمع عدداً كبيراً من الصفات ووضعها فى زواج من الصفات المتضادة ثنائية القطب Bipolar، وطلب من الأفراد تقدير سلسلة من المفاهيم أو الحكم عليها من حيث ما تعنيه لكل منهم باستخدام هذه الأزواج من الصفات. وأجرى تحليلاً عاملياً للبيانات حيث استخلص ثلاثة عوامل أو مكونات أساسية تنطوى عليها هذه التقديرات هي: التقييم Evaluation، والقوة Potency، والنشاط Activity. وفيما يلى بعض أمثلة لأزواج الصفات المتضادة المتعلقة بكل من هذه العوامل:

عامل التقييم: جيد - ردء ، عادل - غير عادل ، ، أمين - غير أمين

ناجح - غير ناجح ، موجب - سالب ، سار - غير سار

عامل القوة: قوى - ضعيف ، ثقيل - خفيف ، كبير - صغير

متشدد - متراهل ، سميك - رفيع ، مذكر - مؤنث

عامل النشاط: سريع - بطئ ، نشط - خامل ، ساخن - بارد

مثير - هادئ ، متوتر - مسترخي ، مندفع - متروى

وقد نشر أوزجود ومعاونوه (Osgood et al., 1957) نتائج هذه البحوث فى كتاب «قياس المعنى The Measurement of Meaning» وأشاروا إلى كثير من أزواج الصفات المتعلقة بكل من هذه العوامل أو المكونات الأساسية. واستند أوزجود فى ذلك إلى افتراض أن خصائص الأفكار والأشياء فى اللغة المكتوبة والمتحاطب بها يتم التواصل بها بدرجة أساسية بواسطة الصفات. واقتصر أسلوبياً عاماً يمكن باستخدامه الحصول على تقديرات للمفاهيم فى سلسلة من موازين للصفات المتضادة ثنائية القطب The Semantic Differential Technique. أطلق عليه «أسلوب تمييز معانى المفاهيم».

ويعُد أسلوباً لقياس معنى شيء ما بالنسبة للفرد، ويُستخدم عادة للإشارة إلى أي

مجموعة من موازين التقدير Rating Scales يشتمل كل منها على صفة ثنائية القطب. لذلك فإن هذا الأسلوب لا يُعد أداة قياس معينة أو اختبار، وإنما أسلوب قياس عام يتميز بالمرونة في قياس الاتجاهات وغيرها من الجوانب الوجودانية. ولتوسيع ذلك نفترض أن المثير أو المفهوم المراد تقييم اتجاه الفرد أو مشاعره نحوه هو مفهوم «التعلم»، فيمكن أن تكون موازين التقدير التي تشتمل على صفات متضادة ثنائية القطب كما هو موضح بشكل (١٢-٣) التالي:

التعلم	
بطيء	سريع
رديء	جيد
مفيد	غير مفيد
غير سار	سار
نشط	خامل
ضعيف	قوى

شكل (١٢-٣) يوضح صفحة تميز معانى المفاهيم

ويُطلب من كل فرد تقدير اتجاهه نحو المفهوم المدون أعلى الصفحة الخاصة بموازين ذات سبع نقاط يشتمل كل منها على صفة ثنائية القطب كما هو موضح بشكل (١٢-٣)، وأن يضع علامة في الموضع الذي يعبر عن توجه وشدة مشاعره نحو المفهوم. ويمكن استخدام أي عدد من المفاهيم أو المثيرات مثل: الأشخاص، والأشياء، والمفاهيم المجردة، والممارسات، والمؤسسات وغير ذلك. وبذلك يمكن تقييم اتجاه الأفراد نحو عدد كبير من المفاهيم أو العبارات في وقت قصير بدلًا من بناء ميزان اتجاه لكل مفهوم على حدة كما في الأساليب التي سبق أن ناقشتها.

وقد أوصت كثير من الدراسات بالاهتمام بعامل أو مكونة «التقييم» في بناء الموازين باستخدام أسلوب تميز معانى المفاهيم.

خطوات أسلوب تميز معانى المفاهيم:

يمكن اتباع الخطوات التالية عند بناء موازين تميز معانى المفاهيم:

(١) تحديد المفهوم أو المفاهيم المراد تقييرها: يعتمد اختيار عدد ونوع المفاهيم على الهدف من بناء ميزان الاتجاهات، ويفضل اختيار مجموعة من المفاهيم

للمترابطة التي يجمعها إطار مشترك، فكلما زاد تجانس المجموعة أمكن اختيار صفات ثنائية القطب واضحة المعنى. فمثلاً مجموعة من المفاهيم مثل: التعلم، والمعلم، والمدرسة، والدراسة، والكتب المدرسية – تعد مجموعة متجانسة نسبياً، ويمكن للمعلم تقييم مشاعر الطلاب نحو مفاهيم أساسية في وحدة أو مقرر دراسي معين.

(٢) **اختيار موازين ثنائية القطب مناسبة**: يعتمد اختيار مجموعة من الموازين على مدى مواءمتها وتمثيلها للمفاهيم المحددة، وينبغي عدم اختيار صفات طبيعية أو فيزيائية للأشياء، فمثلاً لا يجوز إذا كان المطلوب تقييم مفهوم «صخرة» مثلاً أن نختار صفة ثنائية القطب مثل (خشى - أملس). ويجب أن تكون الموازين ممثلة لمدى متسع من المشاعر التي يُحتمل أن تقتربن بالمفاهيم المحددة، وربما يتطلب ذلك تجريب هذه الموازين ميدانياً وإجراء تحليل عاملى للبيانات، غير أن ذلك ربما لا يكون مناسباً في كثير من الحالات. ويمكن الاسترشاد في اختيار الصفات بالقوائم التي أعدتها أو موجود Osgood في كتابه الذي أشرنا إليه، وكذلك اقتراح صفات أخرى مناسبة. وينبغي مراعاة أن يكون مستوى صعوبة الكلمات التي تدل على صفات مناسباً للمختبرين ومتلوقاً لهم.

(٣) **تصميم صفحات تدون فيها الاستجابات**: ينبغي أن يظهر كل مفهوم في صفحة مستقلة، حيث يكتب المفهوم أعلى الصفحة يليه الموازين المتعلقة به، بحيث يكون مقابلًا لمتصف كل ميزان، كما هو موضح بشكل (١٢ - ٣). كما ينبغي أن يحدث تبادل عشوائي بين أقطاب الموازين أي بين أزواج الصفات مثل (وجب - سالب، غير ناضج - ناضج)، وأن يكون عدد النقاط الفاصلة بين كل صفتين ثابتًا، ويفضل أن يكون سبع نقاط أو ربما تسع نقاط. غير أن هذا يعتمد على خصائص عينة الأفراد، فإذا كانوا أطفالاً صغاراً ربما يكتفى بخمس نقاط، ويمكن الاكتفاء بأزواج من الصفات عددها يتراوح بين ١٠ ، ١٥ زوجاً تتعلق بعدد من المفاهيم ول يكن عشرة تقريباً، ولكن يمكن أن يستجيب الأفراد لعدد أكبر من المفاهيم ول يكن ٢٠ مفهوماً بموازين عددها ٢ ميزاناً، ويفضل تقليل هذا العدد بالنسبة للأطفال.

(٤) **كتابة تعليمات الاستجابة**: ينبغي أن تشتمل صفحة الغلاف على تعليمات عامة تتعلق بالهدف من المقياس والمطلوب من المستجيب، وكذلك كيفية تدوين الاستجابات، وتوسيع ضرورة التعبير عن الانطباع الأول بسرعة، وإصدار أحكام مستقلة لكل زوج من الصفات المتعلق بالمفهوم المعين، والتعامل مع كل مفهوم بطريقة مستقلة، والتذويه بأن الفرد ينبغي عليه الحكم على المفهوم كما يراه وليس كما يراه أو يستجيب له الآخرون، مع مراعاة عدم تفسير العلاقة بين الصفة والمفهوم نظراً لأن هذا يؤدي إلى عدم صدق استجابة الفرد. فمثلاً يمكن أن تكون التعليمات كالتالى:

الهدف هو قياس معانى بعض المفاهيم، والمطلوب منك الحكم عليها استنادا إلى سلسلة من الموازين الوصفية. وسوف تجد في كل صفحة مفهوما مختلفا يليه مجموعة من الموازين، وعليك أن تقدر كل مفهوم وتضع علامة (X) فوق الخط الذى يناسب هذا التقدير كالتالى:

خامل : _____ : _____ : _____ : _____ : X : _____ نشط

قوى : _____ : X : _____ : _____ : _____ ضعيف

(٥) **تقدير درجات الموازين والمفاهيم** : لتقدير درجات الاستجابات \hat{Y} بين قيم تتراوح بين ١ ، ٧ إذا كان عدد نقاط كل ميزان سبع نقاط، أو بين ١ ، ٩ إذا كان عددها تسع نقاط وهكذا، بحيث إن النقطة القريبة من الصفة التي تمثل القطب السالب (تقييم منخفض) يُعين لها الدرجة (١) ، والنقطة القريبة من الصفة التي تمثل القطب الموجب (تقييم مرتفع) يُعين لها الدرجة (٧)، وتُحدد درجات الفرد وفقاً لموقع العلامات التي وضعها. فإذا كان لدينا ١٠ موازين لمفهوم معين، فإن أقصى درجة تكون ٧٠ ، وأقل درجة ١٠ .

ويمكن استخدام هذه الدرجات في المقارنة بين اتجاهات الفرد نحو مفاهيم مختلفة مثل مفهوم محاضرة رياضيات، ومفهوم فيلم علمي عن الكيمياء. كما يمكن مقارنة تقديرات فردين مختلفين لمفهوم معين، وكذلك يمكن ترتيب درجات مجموعة من الطلاب لتقييم الفروق بينهم في الاتجاه، ويمكن أيضاً عن طريق إعادة تطبيق الموازين على المجموعة بعد مرور مدة زمنية معينة تقييم التغيرات التي حدثت في اتجاههم.

مزايا وعيوب أسلوب تمایز معانی المفاهيم :

يُعد أسلوب تمایز معانی المفاهيم من الأساليب المفيدة في تقييم ما يراه الفرد أكثر جاذبية من المفاهيم والمواصفات الاجتماعية والشخصية. والاهتمام بعامل «التقييم» في استجابات الأفراد على الموازين المختلفة يفيد في تعرُّف رد فعل الفرد نحو كثير من الأشياء، إذ يمكن أن يطلب منه تقدير أفراد عائلته، وزملائه في العمل، وأصدقائه وعائلته، وأصحاب المهن الأخرى، والأنشطة، والأفكار المجردة، مثل الانتماء، والسلام، والديمقراطية، وكذلك البرامج المرئية، والمسموعة، والمقروءة، وغير ذلك مما يصعب تقييمه باستخدام الأساليب الأخرى التي تتطلب إجراءاتها جهداً كبيراً. غير أن هذا يتطلب العناية باختيار الموازين التقييمية لكي لا يؤدي ذلك إلى خطأ في التقدير. فمثلاً ينبغي التتحقق من أن الصفتين المتناقضتين يمثلان قطبى الميزان ويقعان

بالفعل على متصل متدرج، فأحياناً توجد صفات لها أكثر من نقىض، وصفات يصعب اختيار نقىضها أو ربما ليس لها نقىض. وقد أوضحت كثير من الدراسات أن صدق موازين تمايز معانى المفاهيم ثبات درجاتها مُرضياً بوجه عام، والارتباطات بين التقديرات المتعلقة بهذه الموازين ودرجات موازين ثيرستون وليكرت تراوحت قيمها بين ٧٤، ٨٢، ٠، ٠، كما أن قيم معامل الثبات باستخدام طريقة التجزئة النصفية تراوحت بين ٨٣، ٩١، ٠.

وعلى الرغم من ذلك فإن هناك بعض التساؤلات حول هذا الأسلوب يتعلق بعضها بما إذا كان يفيد بالفعل في قياس المعنى، فقد وجد أن بعض المفاهيم تناول تقديرات متماثلة من معظم الأفراد على الرغم من أنها ربما لا تكون متماثلة تماماً في المعنى وذلك لأن التقديرات تعتمد على معنى المضمنون Connotative Meaning وليس على المعنى بجميع جوانبه. كما أن بعض الصفات الثانية ربما لا تعبّر تعبيراً معقولاً عن مفهوم معين، فمثلاً إذا كان المفهوم «الديمقراطية» وكان الميزان ثانوي القطب «عادل - غير عادل» أو «أمين - غير أمين» فإن الأساس الذي يستند إليه تقييم الأفراد يكون غير واضح.

أسلوب الترتيب التصنيفي The Q Sort :

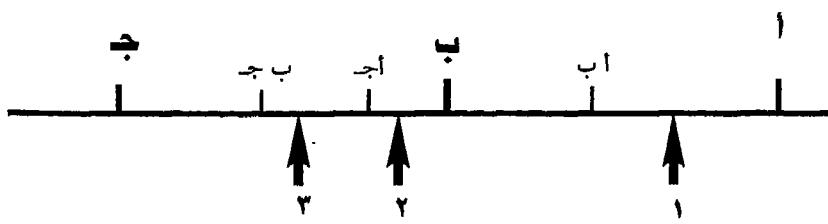
يُستخدم هذا الأسلوب أيضاً في قياس الاتجاهات والميول وغيرها من الجوانب الوجدانية، وقد اقترحه ستيفنسون (Stephenson, 1953, 1967). ويفيد بدرجة أساسية في أغراض الإرشاد والعلاج النفسي، والتغيرات السلوكية، والمقارنة بين تفضيلات أو آراء مجموعات مختلفة من الأفراد فيما يتعلق بموضوع ما، إلا أنه يختلف في منهجه عن أسلوب تمايز معانى المفاهيم، فهذا الأسلوب يهتم بالمقارنة ودراسة التماثل بين التقديرات المتعلقة بالاتجاهات والمشاعر بين الأفراد المختلفين. ويعتمد هذا الأسلوب على ترتيب مجموعة من الفقرات أو العبارات، أو الأشياء، أو الأشخاص أو غير ذلك وفقاً لمحك معين بواسطة الأفراد. وفي حالة قياس الاتجاهات يكون المحك هو شدة الموافقة، أو الأهمية، أو الأفضلية. وما يميّز هذا الأسلوب هو أن التراتيب التي نحصل عليها بواسطة الأفراد ينبغي إيجاد الارتباط فيما بينها والتوصيل إلى تجمعات من الأفراد الذين تكون تراتيبيهم متماثلة للكشف عن العامل الذي يمكن أن يفسر أنماط هذه التراتيب، وبذلك يختلف عن غيره من الأساليب في أنه يعتمد على الارتباطات بين الأفراد وليس على الارتباطات بين المقاييس.

ويتطلب هذا الأسلوب إعطاء الفرد مجموعة من البطاقات مدون على كل منها فقرة أو عبارة، أو سمة أو صورة أو غير ذلك، ويُطلب منه تصنيف هذه البطاقات وترتيبها في مجموعات وفقاً لموقعها النسبي على متصل أو بُعد واحد يتراوح بين «أكثر أهمية» و« أقل أهمية» ، أو بين «أكثر مشابهة» و« أقل مشابهة» ، أو «أكثر تميزاً» و« أقل تميزاً» أو «أفضل» ، و«أسوء» وهكذا. ويتم تحديد عدد البطاقات التي يسمح بها في كل مجموعة مسبقاً بحيث يقترب توزيعها من التوزيع الاعتدالى . ولعل هذا يماثل الخطوة الأولى في الإجراءات المتعلقة بأسلوب ثيرستون حيث يقوم المحكمون بتصنيف الفقرات فى 9 أو 11 مجموعة بهدف تحديد قيم ميزان كل فقرة استناداً إلى هذا التصنيف، أما في هذا الأسلوب فإن الهدف من التصنيف معرفة وجهة نظر الفرد أو اتجاهه نحو الشيء أو الموضوع المعين..

ونظراً لأن هذا الأسلوب يركز فقط على بُعد أحادى ، والصعوبات المتعلقة ببناء المقاييس وتطبيقها وتحليل بياناتها استناداً إليه يقلل من استخدامه في أغراض التقييم إذا ما قورن بغیره من الأساليب الشائعة الاستخدام .

أسلوب البسط : Unfolding Technique

اقتراح كومبس (Coombs, 1953 , 1964) أسلوباً جديداً في إطار نظرية التي أطلق عليه «نظرية البيانات Theory of Data» يُعرف باسم أسلوب البسط Unfolding Technique . ويُستخدم في استخدام سمة كامنة مثل الاتجاهات تطوى عليها تفضيلات مجموعة من الأفراد. ويعتمد هذا الأسلوب على تجميع عدد من الفقرات أو العبارات التي يفترض أنها تتعلق بالاتجاه المراد قياسه. ويُطلب من الأفراد اختيار الفقرات التي يرون أنها أكثر تفضيلاً بالنسبة لهم. ويفترض أن كل فرد يختار الفقرات التي تقترب من موقعه على متصل الاتجاه . ونظراً لأننا نتغاضى عن توجُّه الاتجاه، فكان الفرد طوى Fold ميزان الاتجاه Attitude Scale عند النقطة التي تحدد موقعه على المتصل لكل من الأفراد والفرقات. ويوضح شكل (٤-١٢) التالي مثالاً لهذا الميزان :



شكل (٤-١٢) يوضح مثلاً افتراضياً للميزان (ج) لموقع ثلاث فقرات وثلاثة أفراد على متصل اتجاه معين

ويوضح هذا الشكل توزيعاً افتراضياً لثلاث فقرات (١) ، (ب) ، (ج) استجابة لها ثلاثة أفراد (١) ، (٢) ، (٣) على متصل اتجاه معين، وتمثل الأسماء نقط المتتصف بين أزواج الفقرات. فمثلاً إذا طلب من الفرد (١) أن يختار الفقرتين الأفضل بالنسبة له، فإنه سوف يختار الفقرتين (١) ، (ب) على افتراض أن هذا الاختيار يعتمد على المسافة النسبية بين موقعه على المتصل وموقع الفقرتين اللتين اختارهما، كذلك الفرد (٣) سوف يختار الفقرتين (ب) ، (ج) . أما الفرد (٢) فعلى الرغم من أن موقعه على المتصل إلى يسار كل من الفقرتين (١) ، (ب)، إلا أنه سوف يختار أيضاً الفقرتين (١) ، (ب)، وهذا يصبح واضحاً إذا تصورنا أننا قمنا بطيّ المتصل عند موقع الفرد (٢) عندئذ تكون الفقرتان (١) ، (ب) أقرب إلى موقعه من الفقرة (ج) بسبب المسافات النسبية بين الفقرات الثلاث.

وإذا نظرنا إلى الشكل سوف نرى أن جميع الأفراد إلى يمين نقطة المتتصف أجد سوف يختارون الفقرتين (١) ، (ب)، بينما الأفراد إلى يمين نقطة المتتصف بـ ج ولكن إلى يسار نقطة المتتصف أجد سوف يختارون الفقرتين (ب) ، (ج) ، وبذلك يمكن استخدام توزيع اختيارات الأفراد لتحديد المسافات النسبية بين الفقرات على متصل الاتجاه.

وعلى الرغم من أن هذا الأسلوب يهدف لكشف السمات الكامنة، فإنه يمكن استخدامه في بناء موازين رتيبة للاتجاهات استرشاداً بالخطوات السابقة، ومن الجدير بالذكر أن استخدام هذا الأسلوب والإفادة منه يتطلب الرجوع إلى كتاب نظرية البيانات الذي أشرنا إليه.

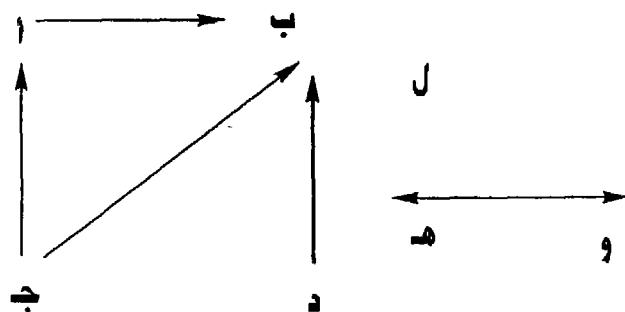
أساليب قياس العلاقات الاجتماعية : Sociometric Techniques

تهدف هذه الأساليب لدراسة التفاعل الاجتماعي بين أفراد مجموعة معينة، وقياس أنماط هذا التفاعل والتواصل. ويمكن عن طريق هذه الأساليب جمع معلومات عن توجهات التفاعل بين أعضاء مجموعة واحدة أو مجموعات فرعية، أو بينها وبين كل فرد على حدة، ويمكن أن يكون التفاعل سلوكياً فعلياً أو ما هو مرغوب أو متوفع أو متصور.

ونوع هذا التفاعل أو محتواه يتحدد أشكالاً مختلفة من السلوك الاجتماعي، مثل الجلوس بجوار، تناول الطعام مع، الشراء من، الإقراض، الزيارة، اللعب مع، الصداقة مع، التحدث مع، السكن بجوار، وغير ذلك. ويمكن الإفادة من ذلك في إطار العمل التربوي، حيث إن البنية الاجتماعية للصف الدراسي تتكون من أنماط

مختلفة من العلاقات التي تتطلب الدراسة، فالطلاب الذين يشعرون بالارتياح مع أقرانهم يستطيعون أن يستمروا قدراتهم بدرجة أفضل من يشعرون بالضيق منهم. كما أن سلوك المعلم وعلاقاته مع طلابه يؤثر في إدراك الطلاب لمعلمهم وفي تعلمهم. واتجاهات الطالب نحو مدرستهم يكون في كثير من الأحيان انعكاساً لمدى تقبل أقرانهم ومعلميهم لهم. لذلك فإن البيانات التي نحصل عليها من مقاييس العلاقات الاجتماعية تفيد في دراسة الأنماط التنظيمية المثلث لتكوين مجموعات فاعلة، والتكيف الاجتماعي للطلاب، والبنية الاجتماعية للمجموعات، وتأثير المدرسة والممارسات الصيفية والتتجددات التربوية في العلاقات الاجتماعية داخل المدرسة، والعلاقات القائمة بين مكانة الطلاب داخل جماعتهم، والقدرات العقلية والتحصيل الدراسي، وتفضيلات المعلمين لطلابهم. وتتعدد أساليب قياس العلاقات الاجتماعية حيث يمكن استخدام أساليب الملاحظة الفعلية للأفراد أثناء قيامهم بنشاطتهم المعتمدة لتعرف طبيعة هذه الأنشطة وال العلاقات بين الأفراد المشاركون فيها من حيث المبادأة والتواصل. كما يمكن استخدام استبيانات التفاعل الاجتماعي Sociometric Questionnaires ، والمقابلات الشخصية Interviews ، حيث يُطلب من كل فرد في المجموعة أن يحدد الأفراد الذين يفضل مشاركتهم له في بعض الأنشطة، والأفراد الذين لا يفضل مشاركتهم، وأحياناً يطلب من الفرد تحديد أي عدد من الأفراد، ولكن عادة يكون عدد الأفراد محدداً مسبقاً، ويفترض في مثل هذه الحالات أن يكون الفرد أميناً في تفضيلاته؛ لأنه سوف يتربى على ذلك إجراء تقييمات معينة تتعلق بالجامعة التي يتمتع بها الفرد.

ولعل الأسلوب الذي اقترحه مورينو ومعاونوه Moreno et al. يعد من الأساليب المبكرة التي تتعلق بالتمثيل الشكلي للعلاقات الاجتماعية Sociograms ، التي تُعرف في وقتنا الحاضر باسم الأشكال الموجهة Directed Graphs ، ويستخدم في تحليل بنية العلاقات بين أفراد جماعة معينة . ويمكن توضيح ذلك بالشكل (٥-١٢) التالي :



شكل (١٢-٥) يوضح العلاقات بين أفراد جماعة معينة

ويتضح من شكل (٥-١٢) أن الجماعة تتضمن على سبعة أفراد تربط بينهم علاقات مماثلة بالأسهم، وبعضهم لا تربطه علاقات حيث لا توجد أسهم متوجهة منه أو إليه. فالفرد (أ) تربطه علاقة بالفرد (ب)، والفرد (ج) تربطه علاقة بكل من (أ)، (ب)، والفرد (هـ) تربطه علاقة متبادلة مع الفرد (و)، والفرد (ل) لا تربطه علاقة بأحد، وهكذا.

وتستخدم مثل هذه الأشكال إذا كانت الجماعة قليلة العدد، أما إذا كانت تتضمن على عدد كبير نسبياً (١٠ أفراد أو أكثر) وبنية العلاقات الاجتماعية أكثر تعقيداً، فإنه يفضل في هذه الحالة استخدام مصفوفات العلاقات الاجتماعية Sociometric Matrices. ويحدد على كل من بعد المصفوفة الأفقي والرأسي الحروف المنشورة للأفراد، ويبدون في خلاياها الرقم (١) إذا كانت هناك علاقة بين الفرد وغيره من الأفراد، والرقم (صفر) إذا لم تكن هناك علاقة بينه وبين أي منهم، ويجرى تحليل بيانات المصفوفة من حيث العلاقات الأحادية والثنائية ، وعدم وجود علاقات.

واقتصر مورينو مؤسراً لجاذبية الفرد في جماعته وذلك بقسمة العدد الكلى للاختيارات والرفض، وقسمة الناتج على العدد الكلى للاختيارات، كما اقترح مؤسراً لتماسك الجماعة Group Cohesiveness Index وذلك بقسمة عدد الاختيارات المتبادلة مثني على عدد الأزواج الممكنة [$\frac{n(n-1)}{2}$].

وفيما يتعلق بصدق مقاييس العلاقات الاجتماعية ، فإنه ربما يصعب التتحقق منها إلى حد ما. أما الثبات فقد أوضحت بعض الدراسات أن اتساق الاختيارات بمدروز الزمن يبدو مرتفعاً إلى حد ما، غير أن هذا يتأثر بالمستوى العمري لأفراد الجماعة، واستقرارها ومدى شدة التفضيل.

أساليب بناء موازين الاتجاهات متعددة الأبعاد :

تُعدُّ جميع أساليب بناء موازين الاتجاهات التي ناقشناها فيما سبق أساليب أحادية البعد، أي تحاول تحديد موقع الفرد على متصل ذو بعد تقييمي واحد. وعلى الرغم من أن هذا يتفق مع تعريف الاتجاه الذي قدمناه في مستهل هذا الفصل، إلا أن بعض علماء النفس عرَّفوا الاتجاه بأنه مفهوم مركب متعدد الأبعاد، ويطلب قياس الاتجاه بهذا المفهوم أساليب متعددة الأبعاد لبناء موازين الاتجاهات Multidimensional Scaling Procedures .

وتهدف هذه الأساليب لتحديد الأبعاد المناسبة التي تنطوي عليها استجابات الأفراد فيما يتعلق بموضع الاتجاه أو مجموعة من الموضوعات. وعلى الرغم من أنها تتفق أن هذه الأبعاد المتعددة تختلف باختلاف الأفراد والموضوعات، إلا أنها لا يجب أن تتغير بتغيير مجموعة معينة من الاستجابات.

ويُعدُّ أسلوب التحليل العاملی Factor Analysis من الأساليب المتعارف عليها في استخلاص الأبعاد المتعددة من مجموعة من البيانات، وقد استخدمنا أو زجود Osgood كما سبق أن أشرنا في استخلاص أبعاد تميز معانٍ المفاهيم. وقد أوضحنا في كثير من الفصول السابقة استخدامات هذا الأسلوب في مجال القدرات العقلية والاستعدادات الخاصة والميول. كما يُعدُّ أسلوب الموارين متعددة الأبعاد Multidimensional Scaling من الأساليب الأخرى المماثلة للتحليل العاملی، ولكنه يعتمد على المسافات بين مجموعة من النقاط التي تمثل الدرجات، فهو أسلوب إحصائي لا ينطوي على معايير مترابطة في التحليل العاملی الذي يعد من الأساليب الإحصائية البارامترية، ويهدف أيضًا للتوصيل إلى مجموعة من الأبعاد. وموقع كل متغير يمثل الاتجاه على هذه الأبعاد، وبذلك يمكن تحديد عدد من الأبعاد التي تنطوي عليها معتقدات الفرد التي يستند إليها في الحكم على موضوعات اتجاهات مختلفة. ومن الجدير بالذكر أن هناك برامج حاسوب يمكن استخدامها في إجراء التحليل متعدد الأبعاد يتعلّق بهذا الأسلوب ، مثل برنامج INDSCAL الذي أعدده كارول ، وشانج (Carroll & Chang , 1970) . وقد أشرنا في الفصل الخامس إلى البرنامج الذي يجري التحليل العاملی .

تحقيق عالم على أساليب بناء موازين الاتجاهات :

لعلنا لاحظنا مما سبق تنوعُ أساليب بناء موازين الاتجاهات وتبين منهاجياتها وإجراءاتها حيث إن كلًا منها يستند إلى إطار نظرى معين، غير أن معظمها يؤدي إلى درجات للاتجاه تمثل موقع الأفراد على متصل تقييمي ثالثى القطب يتصل بموضوع الاتجاه. ومع هذا فإنها تختلف فيما بينها من حيث مستويات القياس، وخصائص الفقرات التي يتم اختيارها، وكذلك الخصائص السيكومترية .

فقد اعتمد أسلوب ثيرستون على أحکام مجموعة من الخبراء في تقدير قيم ميزان الفقرات التي تمثل موقع كل مفردة على متصل الاتجاه، وافتراض أن هذه الأحكام تكون على ميزان متساوي الفقرات مما يتربّع عليه افتراض أن قيم ميزان الفقرات ودرجات الاتجاه تكون فتراتها متساوية، في حين أن الأسلوب التراكمي الذي اقترحه جتمان يؤدي إلى ترتيب الفقرات على متصل الاتجاه دون افتراض تساوى الفقرات،

لذلك فإن ميزان الاتجاه يكون من المستوى الرتبى، أما أسلوب التقدير الجماعي الذى اقترحه ليكرت فإنه لا يعكس درجات متباعدة من الأفضلية أو الموافقة، وإنما يصنف الاستجابات إلى استجابات موجبة أو سالبة، ومع هذا يؤدى إلى ميزان رتبى أيضاً. فيما يتعلق بأسلوب تمایز معانى المفاهيم الذى اقترحه أوزجود، فإنه يعتمد على فقرات (صفات) ثنائية القطب، ولذلك لا نستطيع تصنيفها إلى فقرات موجبة وأخرى سالبة، وإنما يفترض فقط أنها ترتبط بالبعد التقىيمى للفضاء السيمانتى Semantic Space الذى يشتمل على ثلاثة أبعاد كما سبق أن أشرنا. أما النقاط أو المواقع السبعة التى تقع على الميزان ثنائى القطب، فإنه يفترض أنها تمثل فترات متزايدة وبالتالي يفترض أن درجات الاتجاه تفصلها فترات متزايدة كما في أسلوب ثيرستون.

وهذا الاختلاف فى مستويات القياس يؤدى إلى اختلاف فى الإفادة من الفقرات المحايدة Neutral Items فى تقىيم الاتجاهات. ففى أسلوب ثيرستون تُعد الفقرات المحايدة مؤشرات صادقة للاتجاه لأنها تُشكل جزءاً من ميزان الاتجاه، كذلك يمكن أن يشتمل ميزان الاتجاه وفقاً لأسلوب جثمان على فقرات محايدة، ولكننا لا نستطيع تمييزها عن غيرها من الفقرات نظراً لأن أسلوب جثمان يهتم بالتحقق من إمكانية ترتيب الفقرات على بعد أحادى، لذلك يمكن أن تكون هناك موافقة أو عدم موافقة على جميع الفقرات أو تتواءز بين الموافقة والمحايدة وعدم الموافقة، أما الفقرات التي لا نستطيع تصنيفها إلى فقرات موجبة أو سالبة وفقاً لأسلوب ليكرت فإنه يتم حذفها، حيث يفترض أن الفقرات المحايدة لا تشير بصدق إلى الاتجاه. كذلك فى أسلوب تمایز معانى المفاهيم تُستبعد الصفات ثنائية القطب (الفقرات) التي لا تتمى بوضوح إلى بعد «التقييم» ومع هذا يسمح بالاستجابة المحايدة على هذا البعد. وجميع الأساليب تهدف لتحديد الفقرات الواضحة المتعلقة بالاتجاه المراد قياسه، ولكن اختلاف محكّات اختيار الفقرات يؤدى إلى خصائص مختلفة للمحنّيات المميزة Trace lines لهذه الفقرات. ففى كل من أسلوب جثمان وليكرت ، وأسلوب تمایز معانى المفاهيم تكون هذه المحنّيات المميزة اطرادية Monotonic ، أي أنه كلما زادت درجة الاتجاه يزداد احتمال موافقة الفرد على الفقرة. بينما فى أسلوب ثيرستون تكون هذه المحنّيات غير إطرادية Nonmonotonic ،أى أن احتمال الموافقة على فقرة معينة يفترض أن يزداد كلما اقترن درجة الاتجاه للفرد من قيمة ميزان الفقرة Item's Scale Value من أيّ من طرفى المتصل ثنائى القطب، مما يجعل المحنّى المميز للفقرة يتخد شكل حرف U.

كذلك فى أسلوب جثمان تكون الفقرات متراكمة، بينما فى الأساليب الأخرى لا تكون كذلك، ولهذا فإن كل درجة من درجات الاتجاه فى الميزان التراكmic لجثمان

تقترن بنمط واحد فقط من الاستجابات للفقرات على ميزان الاتجاه، بينما يمكن الحصول على أنماط متعددة للاستجابات للدرجة نفسها في الأساليب الأخرى، لذلك لا نستطيع التنبؤ بالاستجابات للفقرات من درجات الاتجاه إلا في حالة استخدام أسلوب جتمان.

ولعل طبيعة موازين الاتجاهات التي تستند إلى جميع هذه الأساليب يمكن أن تضمن درجة عالية من الموضوعية، فجميع الاستجابات يمكن تقدير درجاتها باستخدام مفاتيح معينة أو باستخدام برماج حاسوب.

وسوف نلخص الاختلافات التي ذكرناها بين الأساليب الأربع الأساسية في بناء موازين الاتجاهات في جدول (٤-١٢) التالي:

الأساليب				
تمازز معاني المفاهيم	ليكرت	جتمان	ثيرستون	
فقرات	فقرات	فقرات	فقرات	الصيغة
-	طيفي المتصل	نقاط على مدى المتصل	نقاط على مدى المتصل	الموقع على المتصل
-	نوعي	رتبي	فترى (اعتباري)	مستوى تكميم الفقرات
فترى	رتبي	رتبي	فترى (اعتباري)	مستوى تكميم الدرجات
غير تجمعي	غير تجمعي	تجمعي	غير تجمعي	طبيعة المتصل
لا	لا	نعم	نعم	وجود فقرات محابية
اطرادي	اطرادي	اطرادي	غير اطرادي	المنحنى المميز
خطي	خطي	درجى	على شكل حرف U	

جدول (٤-١٢) يوضح جوانب الاختلاف بين الأساليب الأساسية في بناء موازين الاتجاهات

ثبات وصدق موازين الاتجاهات :

فيما يتعلق بثبات درجات موازين الاتجاهات، يُعد معامل الاتساق الداخلي، ومعامل الاستقرار من المعاملات المناسبة لتقدير ثبات درجات موازين الاتجاهات. فمعامل الاتساق الداخلي لدرجات ميزان اتجاه يشتمل على ١٠ فقرات واعتُنى بإعداده يمكن أن تقترب قيمته من ٨٠، . (Anderson, 1981)، وإذا زاد عدد الفقرات إلى ٢٠ فقرة يمكن أن تصل قيمة هذا المعامل إلى ٩٠، . ، لذلك يمكن بناء موازين اتجاهات تميّز درجاتها باتساق داخلي معقول. غير أن القيم التقديرية لمعامل استقرار الدرجات يستخدم بدرجة أقل، وربما يرجع سبب ذلك إلى مشكلة تفسير هذه القيم. فمن الضروري قبل تفسيرها معرفة درجة الاستقرار المتوقعة للاتجاه كتكوين فرضي عبر المدة الزمنية التي تفصل بين مرتب تطبيق الميزان. فالأدلة البحثية القليلة المتوفرة توضح أنه يمكن الحصول على قيم معقولة عبر مدة زمنية تتراوح بين أربعة وخمسة أسابيع (Anderson, 1981). أما أسلوب جثمان فإنه يتضمن داخله تقديرات الثبات، حيث يتطلب حساب معاملين أحدهما معامل إمكانية إعادة التكوين Reproducibility Coefficient ، والآخر معامل إمكانية بناء الميزان تراكميًّا Scalability Coefficient ولتقدير هذين المعاملين مما يتم التتحقق من توافر خاصية التراكم التي سبق أن أشرنا إليها، وهذا المعاملان يتعلقان بالاتساق الداخلي لميزان الاتجاه.

غير أن المشكلات الرئيسية لموازين الاتجاهات تتعلق بصدق هذه الموازين، فقد سبق أن أوضحتنا في الفصل الحادى عشر أن مشكلات الأنماط العقلية للاستجابات مثل مراعاة القبول الاجتماعى تشير إلى الاستجابات التى لا تنسق مع المشاعر الحقيقية للفرد، ولكن الفرد يلجأ إلى تزييف استجاباته بما يراه متوقعاً أو مقبولاً، وكذلك الإذعان أو الموافقة على معظم فقرات ميزان الاتجاه يلجأ إليها الفرد إذا لم يكن متاكداً أو لديه مشاعر متناقضة.

وعلى الرغم من أن هذه الأنماط العقلية للاستجابات تُقلل من صدق موازين الاتجاهات، إلا أنه يمكن باستخدام فروض مناسبة تتعلق بهذه الأنماط التتحقق الإيميريقى من صدق التكوين الفرضى لهذه الموازين. ويمكن كذلك التتحقق بسهولة من صدق محتوى الموازين التى تستند فى بنائها إلى كل من أسلوبى ثيرستون وليكرت، إذ يمكن إعادة كتابة الفقرات ومراجعةها عدة مرات للتوصل إلى اتفاق المحكمين على وضوحاً وملاءمتها . كما يمكن باستخدام تحليل المفردات فى أسلوب ليكرت، وإيجاد قيمة نصف المدى الريبيعى فى أسلوب ثيرستون الكشف عن الفقرات الغامضة.

كذلك يمكن التحقق من الصدق التلازمي لموازين الاتجاهات عن طريق تحديد مدى تميز ميزان الاتجاه بين مجموعتين من الأفراد يكون معلوماً اختلافهم في الاتجاه المراد قياسه. وكذلك يمكن التتحقق من ذلك بإيجاد الارتباط بين درجات ميزان اتجاه معين وموازين أخرى تتعلق بموضوعات تمثل موضوع الاتجاه المراد قياسه، مثل ميزان اتجاه نحو التدين، وميزان اتجاه نحو أماكن العبادة.

إرشادات عامة لبناء موازين الاتجاهات:

اتضح لنا مما سبق أن موازين الاتجاهات تشتمل على مجموعات من الفقرات أو العبارات يستجيب لها الفرد على ميزان معين، وهذه الفقرات أو العبارات تتعلق بموضوع أو شخص أو خبرة أو قضية معينة، وتعتمد نوعية الميزان على نوعية فقراته. لذلك ينبغي العناية ببناء وصياغة هذه الفقرات استرشاداً بالمقترحات أو المحركات التالية التي ينطبق بعضها على جميع موازين الاتجاهات، بينما ينطبق البعض الآخر على نوع معين من الموازين (Babbie, 1973 , Edwards, 1957) :

- (١) تجنب الفقرات أو العبارات التي تشير إلى حقائق، حيث إن الفقرات المتعلقة بالاتجاهات تتطلب أن يعكس الفرد مشاعره ومعتقداته حولها وليس تقرير صحة أو خطأ الفقرات ، فتقرير ذلك يتصل بمقاييس العوامل المعرفية بينما يجب تجنب ذلك في المقاييس الوجودانية .
- (٢) تجنب الفقرات التي يكون لها تفسيرات متعددة، حيث إن الفقرة ينبغي أن تحتمل تفسير واحد فقط لكي يعرف الفرد كيف يستجيب لها .
- (٣) تجنب الفقرات غير المتعلقة بموضوع الاتجاه المراد قياسه .
- (٤) ينبغي أن تكون جميع الفقرات واضحة ومصاغة بلغة سلسة ويسهلة تناسب أعمار المختبرين بحيث يمكنهم فهمها، حيث إن هذه الفقرات لا تهدف لقياس الذكاء أو الفهم .
- (٥) ينبغي أن تكون كل فقرة أحادية الُّبُُعد، أي تتعلق بمفهوم واحد فقط ، كما ينبغي أن لا تكون مزدوجة مثل «لا أحب المعلم الذي يعامل الطلاب معاملة غير عادلة ولا يسمح لهم بالمناقشة»، فالفرد ربما يوافق على أحد جزئي الفقرة، ولا يوافق على الجزء الآخر .
- (٦) تجنب اقتراح استجابة معينة، مثل «هل تواافق على أن التعليم المجاني يشجع الطلاب على التكاسل؟».

- (٧) تجنب الفقرات التي يحتمل أن يوافق عليها أو لا يوافق عليها جميع الأفراد، فمثل هذه الفقرات لا تميز بين درجات الموافقة أو الأفضلية، لذلك ينبغي اختيار فقرات تغطي المدى الكلى لمتصل الأفضلية، فلا يجوز مثلاً أن تكون الفقرة «احترم الأشخاص الذين يؤدون عملهم بأمانة».
- (٨) ينبغي اختيار عدد متساوٍ من الفقرات الموجبة والسلبية بقدر الإمكان لتقليل تأثير تحيز الاستجابات.
- (٩) تجنب استخدام كلمات تشير إلى عموميات مثل «دائماً» ، «جميعاً» ، «أبداً» «مطلقاً» حيث إن اشتمال الفقرات عليها يحدد استجابات الموافقة أو غير الموافقة آلياً.
- (١٠) ينبغي الحيطة عند استخدام كلمات مثل «فقط» ، «بقدر الإمكان» ، «مجرد» ، أو التقليل من استخدامها.
- (١١) ينبغي أن تكون الفقرات قصيرة، ويُفضل ألا يزيد عدد كلماتها عن ٢٠ كلمة.
- (١٢) تجنب استخدام فقرات تشتمل على نفي مضاعف، أى نفي النفي، حيث إن مثل هذه الفقرات يصعب فهمها، والأفضل جعلها موجبة.
- (١٣) ينبغي أن تشير الفقرات إلى الحاضر أو المستقبل وليس الماضي، وذلك لأن قياس الاتجاهات يتعلق بالوضع الراهن لاتجاهات وأثارها المتوقعة.
- (١٤) ينبغي توزيع الفقرات في ميزان الاتجاه عشوائياً، ولكن ينبغي التتحقق من عدم وجود أربع أو خمس فقرات موجبة أو سالبة متالية.
- (١٥) ينبغي مراعاة أن تكون الفقرات الأكثر حساسية تقع بالقرب من منتصف ميزان الاتجاه.
- (١٦) يُفضل ألا يزيد عدد نقاط متصل الاتجاه المراد قياسه عن سبع نقاط.

أنواع أخرى من مقاييس الاتجاهات:

تعتمد موازين الاتجاهات التي نقاشناها فيما سبق على تصميم موافق افتراضية يستجيب لها الفرد، ونستدل من هذه الاستجابات على اتجاه الفرد نحو موضوع الاتجاه، غير أن استجابات الفرد ربما لا تتفق مع سلوكه الفعلى المتعلق بالاتجاه، كما أن موازين الاتجاهات تُقصِّر ردود فعل الفرد أو استجاباته على مدى ضيق نسبياً من المحتوى، ولا تمنحه الفرصة للتغيير عن اتجاهاته بأسلوبه الشخصي. لذلك فإنه يمكن

قياس الاتجاهات بأساليب الملاحظة المباشرة Observational Techniques، وأساليب المقابلة الشخصية وجهًا لوجه . وتُعد هذه الأساليب وغيرها من الأساليب السابقة من النوع غير التكُنْرى Undisguised Techniques، حيث يكون الفرد عادة مدركًا أنه يجري تقييم اتجاهه، أى أن هدف القياس يكون واضحًا باستثناء أسلوب تمييز معانى المفاهيم الذى يقدم أحياناً للمستجيبين على أنه مقياس للمعنى .

ولكن يمكن قياس الاتجاهات بأساليب أخرى من النوع غير المباشر أى التكُنْرى Disguised Techniques، حيث يكون هدف القياس خافياً على المستجيب، وتبير ذلك أنه إذا كان هدف أداة القياس خافياً أو غير واضح، فإن احتمال تزيف الفرد لاستجاباته يكون أقل مما لو كان الهدف واضحًا، وبخاصة إذا كان سيترب على التائج قرارات تؤثر على الفرد، وبذلك يمكن الحصول على معلومات صادقة . وفي مثل هذه الحالات ينقاد الفرد إلى الظن بأن استجاباته سوف يتم تفسيرها كمؤشرات لمتغير آخر يختلف عن المتغير الذي يجرى قياسه أو تقييمه .

وتوجد أنواع متعددة من هذه الأساليب غير المباشرة سوف نشير إلى بعض منها فيما يلى :

أسلوب اختيار الخطأ ، وقد اقترحه هاموند Hammond's Error-Choice Technique . ويعتمد هذا الأسلوب على جعل الفرد يظن أنه يتخد حكمًا موضوعيًا، أى أن هناك إجابة صحيحة . ومثال ذلك اختبارات المعلومات Information Tests التي تبدو كما لو كانت تقييس معلومات الفرد ولكنها في الواقع الأمر تقييم الاتجاهات . وقُستخدم في هذه الاختبارات مفردات اختيار من متعدد تتناول مشكلات معينة تبدو كما لو كانت مشكلات موضوعية ، ولكنها تهدف لتقييم معتقدات الفرد . ويستند هذا الأسلوب إلى افتراض أن الاتجاهات تؤدي إلى تحيز الاستجابات بطريقة متسقة، ويمكن استخدام هذه الاستجابات المتحيزة في الاستدلال على اتجاه الفرد .

أسلوب القبول الظاهري : وقد اقترحه كوك Cook's Plausibility Technique . ولا يعتمد هذا الأسلوب على استجابات الموافقة أو عدم الموافقة لمجموعة من الفقرات، وإنما يطلب من الفرد الحكم على كل فقرة من حيث صدقها وفاعليتها، ومتضمناتها التقديمية، ومعقوليتها الظاهرية . فمثلاً مهمة المحكمين في أسلوب ثيرستون تقييم استحسان كل فقرة وتصنيفها في إحدى المجموعات التي عددها 11 مجموعة . ويمكن الحصول على قياس غير مباشر للاتجاه، إذا استخدمنا تقييم المحكمين في الاستدلال على اتجاههم، ويستند هذا الأسلوب إلى افتراض أن اتجاه

الفرد سوف يؤدي إلى تحيز حكمه على التقبل الظاهري للفقرة، أي أن الفقرة سوف تكون معقوليتها الظاهرية بالنسبة للأفراد الذين يوافقون على القضية التي تؤيدتها أكثر من معقوليتها بالنسبة للأفراد الذين لا يوافقون عليها.

أسلوب تقدير استجابات الآخرين : ففي هذا الأسلوب Estimation of Others Responses يطلب من الفرد الاستجابة لأى سؤال كما لو كان فرد أو مجموعة من الأفراد الآخرين يجيبون عن هذا السؤال. ويستند هذا الأسلوب إلى افتراض أن الأفراد يُسقطون معتقداتهم واتجاهاتهم ومقاصدهم على الأفراد الآخرين عند استجابتهم «كما يعتقد الآخرون»، ولكنهم في الواقع الأمر يستجيبون بالطريقة التي يعتقدونها أنفسهم.

ولا يتطلب هذا الأسلوب بناء أدوات قياس معينة، وإنما يمكن أن يطلب من الفرد أن يوضح الطريقة التي يعتقد أن معظم الأفراد الآخرين سوف يستجيبون لأى فقرة من فقرات ميزان اتجاه مقتن. ونظرًا لأن تقديرات اعتقاده عن استجابات الآخرين يفترض أنها مكافحة لاستجاباته عن نفسه، لذلك يمكن استخدام الطرق المعتادة لحساب درجة اتجاهه في ميزان الاتجاه.

أساليب تعتمد على اختبارات الإدراك والذاكرة؛ إذا كان الإدراك Perception انتقائيًا وكذلك الذاكرة، فإن الأفراد الذين لديهم اتجاهات مختلفة يجب أن يستجيبوا استجابات مختلفة لمثيرات متعددة. وأحد هذه الأساليب يتطلب عرضًا سريعاً لصورة تفصيلية تشتمل على عناصر للاتجاه المراد قياسه، فمثلاً إذا أردنا دراسة الاتجاه نحو تحرر المرأة، فإنه يمكن أن نعرض على مجموعة من الأفراد صورة رجال ونساء يعملون في أحد المكاتب الحكومية لمدة ثانيةين أو ثلاثة ثانية، ثم يطلب منهم إجابة بعض الأسئلة المفتوحة أو المقيدة المتعلقة بالأنشطة والأدوار التي يقوم بها كل من الرجال والنساء، كما يمكن أن تتطلب الأسئلة إجابات لا تتعلق مباشرة بالصورة نفسها (مثل: من الجالس بجوار المكتب؟ حيث لا يوجد مكتب بالصورة). غير أن تفسير الاستجابات التي نحصل عليها باستخدام هذا الأسلوب يعد أكثر صعوبة من الأساليب المباشرة أو غير التذكرية.

ومن الجدير بالذكر أن هناك أساليب إسقاطية متعددة يمكن استخدامها في قياس الاتجاهات وغيرها من السمات ودراستها، سوف نوضحها في الفصل الثالث عشر عند مناقشتنا لأساليب قياس الشخصية الإنسانية.

تحقيق على الأساليب غير المباشرة في قياس الاتجاهات:

ينبغى الإشارة إلى أن هناك العديد من أوجه النقد التي توجّه للأساليب غير المباشرة في قياس الاتجاهات وغيرها من السمات الإنسانية لعل أهمها هو الجانب الأخلاقي المتعلق بإخفاء هدف القياس عن المستجيبين (يمكن الرجوع إلى الفصل الثاني لمراجعة أخلاقيات القياس). كما أنه ليس هناك ما يبرر استخدام هذه المقاييس المتخفية أو غير المباشرة في مجال الإرشاد أو العمل التربوى حيث تتطلب أن تكون العلاقة بين المرشد وعملاته أو المعلم وطلابه متميزة بالأمانة والانفتاح، واللجوء إلى إخفاء هدف القياس لا يكون ضروريًا أو مطلوبًا من أجل الحفاظ على هذه العلاقة، علاوة على أن ابتكار أساليب أو مواقف غير مباشرة أو متخفية يتطلب جهدًا ووقتاً كبيرين. كذلك أوضحت الدراسات صعوبة التتحقق من صدق هذا النوع من المقاييس وفاعليتها في التمييز بين المجموعات المختلفة، والتباين بالسلوك.

بعض مشكلات قياس الاتجاهات:

إن الاستناد إلى المنهجيات والأساليب والإرشادات التي أوضحتها فيما سبق يُعد من الأمور الأساسية في بناء موازين الاتجاهات، غير أنه مهمًا بلغت العناية بإعداد هذه الموازين تظل هناك عوامل دخيلة تؤثر في استجابات الأفراد وتقلل من صدق الميزان المتعلق بالاتجاه المراد قياسه. فالدرجة التي يحصل عليها الفرد في ميزان الاتجاهات Attitude Scale لا يمكن أن تكون خالية تماماً من الخطأ، أو دقة الدقة الكاملة نظرًا لأنها تكون نتاج تفاعل معقد بين عوامل متعددة أحدها هو السمة أو الاتجاه المعين، ويتعلق بعض هذه العوامل بما سبق أن أشرنا إليه في الفصل الحادى عشر عند مناقشتنا لمشكلات قياس الميل، ومعنى بذلك الاتجاهات العقلية في الاستجابات Response Sets. وهذه الاتجاهات العقلية تؤثر بطريقة منتظمة في استجابات الأفراد وتؤدي إلى تحيز هذه الاستجابات، أما العوامل الأخرى التي تؤثر في الاستجابات فقد سبق أن أوضحتها عند مناقشتنا لمفهوم الثبات في الفصل الرابع. لذلك سوف نقتصر فيما يلى على توضيح بعض الاتجاهات العقلية التي تؤثر في درجات الأساليب المباشرة في قياس الاتجاهات وبخاصة الموازين.

(١) النزعة إلى اختيار استجابة محاباة: فكثير من موازين الاتجاهات تشتمل على قسم «غير متأكد» أو «لا أدرى» وينزع بعض الأفراد إلى اختيار هذا القسم ليعبر عن استجابتهم لجميع الفقرات. وقد وجد كرونباك Cronbach أن قيمة ثبات الدرجات التي تعتمد فقط على عدد الاستجابات المحاباة بلغت ٧٣٪ ، لذلك فإن النزعة إلى اختيار هذا القسم المحابي Neutral Category تعد نزعة متسقة تؤدي إلى قدر كبير من التباين

ال حقيقي في هذه الدرجات . وربما يفسّر ذلك بأن بعض الأفراد يتزخون الخذر في استجاباتهم ، أو أنهم لا يستطيعون اتخاذ قرار ، أو يفتقرن إلى الملاحظة ، أو ربما يعني أن الفرد ينزع إلى الاستجابة يميناً أو يساراً على متصل الاتجاه . ويُقدّر أحياناً وزن معين لثل هذا القسم نظراً لأن التحيز الناشئ عن الاستجابة المحايدة ربما يعزى بدرجة كبيرة إلى عوامل مشتركة نهتم بقياسها ، ولا يعزى جميعه إلى الخذر أو أسباب أخرى لا تستند إلى أدلة . ويمكن التغلب على ذلك إلى حد ما بعد استخدام هذا القسم المحايد في ميزان الاتجاه لكي يضطر الفرد إلى الاستجابة إما إيجاباً أو سلباً .

(٢) **مراجعة القبول الاجتماعي** : تتطلب جمع الأساليب المباشرة في قياس الاتجاهات أن يقدم الفرد تقريراً ذاتياً عن اعتقاده أو اتجاهه نحو موضوع معين ، وتوجد مواقف ربما يختلف اتجاهه أو اعتقاده نحوها عما هو سائد في المجتمع ، ففي مثل هذه المواقف ربما لا يلحد الفرد إلى التعبير عن مشاعره الحقيقة ، بل يستجيب بطريقة تتفق مع ما هو مقبول اجتماعياً Socially Desirable . واختلاف إدراكات الأفراد فيما هو مقبول أو غير مقبول اجتماعياً يمكن أن يؤدي إلى التقليل من صدق نتائج موازين الاتجاهات . وقد سبق أن أوضحنا عند مناقشتنا لمشكلات قياس الميل كيفية التغلب على هذا الاتجاه العقلي في الاستجابة وغيره من الأنماط ، ولعل من المناسب أيضاً أن نطلب من الأفراد الاستجابة دون ذكر أسمائهم لكي يعبروا بحرية عن مشاعرهم الحقيقة أو اتجاهاتهم إذا كان الهدف بحثياً .

(٣) **الاستجابات المتطرفة** : تشير بعض الأدلة البحثية إلى وجود فروق متسلقة بين الأفراد في نزعتهم إلى استخدام الأقسام المتطرفة Extreme Categories في موازين الاتجاهات ، مثل «موافق جداً» أو «غير موافق على الإطلاق» . في بينما أشرنا أعلاه إلى نزعه بعض الأفراد إلى اختيار القسم الأوسط «محايد» ، توجد نزعه لدى أفراد آخرين إلى اختيار الأقسام المتطرفة . وقد أجريت دراسات متعددة لإيجاد العلاقة بين هذه الاستجابات وبين متغيرات مثل الذكاء ، والتحفظ ، والوثوقية وغيرها من سمات الشخصية . و يبدو أنه من الصعب بناء ميزان اتجاهات لا يتأثر بهذه التزعنة المحتمل حدوثها . غير أنه توجد بعض الأساليب الإحصائية التي يمكن استخدامها في تحديد مدى تأثير استجابات الأفراد بها ، غير أن التزعنة للإجابات المتطرفة لا تؤثر بعامة في صدق الميزان إلا تأثيراً طفيفاً .

(٤) **الإذعان أو الموافقة** : يشير هذا الاتجاه العقلي في الاستجابة إلى نزعه الفرد إلى الموافقة على معظم الفقرات ، وقد حظى هذا النمط بكثير من البحوث والدراسات ، وبخاصة بعد أن اقترح أدورنو (Adorno, 1950) نظرية المتعلقة بتكوين بنية الشخصية السلطوية Authoritarian Personality التي أشرنا إليها في مستهل هذا

الفصل، حيث تبيّن أن الصياغة الموجبة لجمع فقرات مقياس التوهم الذي استُخدم بدرجة أساسية في قياس هذه الشخصية، جعل الأفراد يستجيبون بنعم أو موافق بغض النظر عن محتوى الفقرات. وللحقيقة من حدوث هذا التحيز أعدت فقرات موجبة ولكن ذات معنى معاكس للمعنى الأصلي، فإذا كان محتوى الفقرات محدداً رئيساً للدرجات الأفراد سوف يكون هناك ارتباط سالب بين الدرجات الأصلية والدرجات عندما عُكس المعنى الأصلي. غير أنه وجد أن هناك ارتباطاً موجباً مما يؤكّد تأثير الاستجابات بالإذعان أو الموافقة. ويمكن تقليل تأثير هذا الاتجاه العقلي في الاستجابات بالعنابة بصياغة الفقرات بحيث لا يشوبها الغموض، وكذلك استخدام عدد متساوٍ من الفقرات الموجبة والسلبية وتوزيعها عشوائياً في ميزان الاتجاه المراد قياسه كأسلوب لتقييم تأثير الاستجابات بهذا الاتجاه العقلي.

(٥) التركيب اللغوي للفقرات: يتعلّق هذا التحيز بميزان الاتجاه وصياغة فقراته، ولكنه يؤثّر في استجابات الأفراد، فاستخدام أقسام مثل «موافق»، «غير موافق» على الإطلاق»، «غالباً»، «عادة»، «نادرًا» يؤدّي إلى تفسيرات متباعدة من جانب الأفراد، مما يسميه البعض «أحياناً» ربما يسميه البعض الآخر « غالباً» وتجمّع هذه الأخطاء المنتظمة يؤدّي إلى تحيز الدرجات. غير أنه لا توجد أدلة بحثية كافية حول مدى تأثير الفروق الكمية في معانٍ مثل هذه المفردات اللغوية في درجات موازين الاتجاهات.

بعض استخراجات مقاييس الاتجاهات:

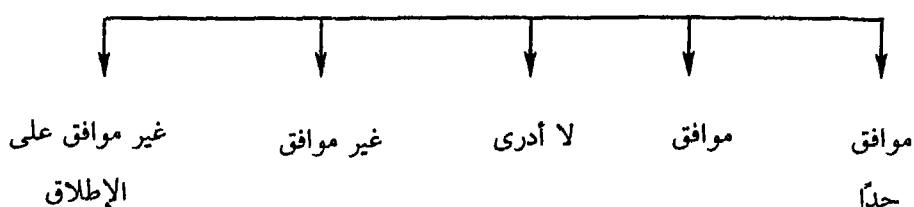
تقويم اتجاهات الطلاب:

ذكرنا في الفصل السابق أن تقويم الجوانب الوجدانية لدى الطالب لا يقل أهمية عن تقويم الجوانب المعرفية أو النفسية الحركية، فالأهداف التعليمية للمجالات الدراسية يتصل بعضها بالجوانب الوجدانية، ويجب أن تؤدي إلى تغييرات في سلوك الطالب وتشري تعلمه. فالتعلم والمشاعر الإنسانية مرتبطة ارتباطاً وثيقاً ولا يمكن اعتبارهما ثنائية، فكل طالب يتعلم شيئاً يكون بالطبع لديه اتجاه معين أو شعور ما نحو ما يتعلمه، فإذا كان اتجاه الطالب إيجابياً فإنه يشري ما يتعلم ويصبح سلوكه بناءً، وهذا يتطلب مراعاة العلاقات القائمة بين المعلومات المستقاة من المجالات الدراسية التي تتجمع لدى الطالب وشخصيته كفرد متميز. فمثلاً يجب على المعلم أن يجذب الطالب إلى المجال الدراسي و يجعله واعياً بأهميته في البيئة التي يعيش فيها لكي ينمي لديه اتجاهًا إيجابياً نحو هذا المجال.

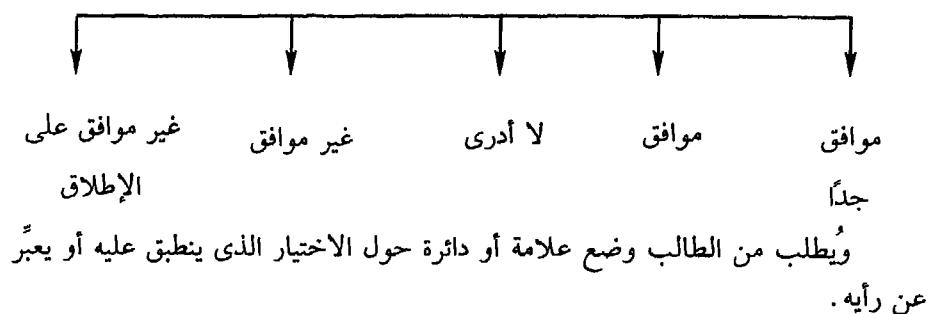
غير أن المشكلة التي تواجه كثيراً من المريين هي كيفية تحديد الأهداف المتعلقة بالاتجاهات وصعوبة قياس هذه الاتجاهات.

فالطريقة المباشرة لإجراء ذلك هي ملاحظة سلوك الطالب في مواقف محددة، كما يمكن استخدام المعاوزين المتدرجة واختبارات المواقف، ففي بناء المعاوزين المتدرجة يعطي الطالب عبارات جدلية أو خلافية معينة يلى كل منها عدداً من الاستجابات التي يمكن أن تتوضح درجة أو شدة موافقته على كل منها، ومثال ذلك:

(١) المدرسة مكان شيق بالنسبة لي



(٢) يجب أن تكون أبواب المدرسة مفتوحة طوال اليوم للطلاب



ويمكن وضع أوزان رقمية لكل اختيار، وجمع أوزان جميع الفقرات لكل طالب كما في حالة ميزان ليكرت، وذلك للاستدلال على مدى إيجابية أو سلبية اتجاه الطالب نحو القضية المطلوبة. ويمكن الاقتصر على «موافق» أو «غير موافق» إذا كان ميزان الاتجاه مصمماً لتلاميذ المرحلة الأولى. وهنا يُعطى المعلم الوزن (+ ١) إذا كان التلميذ موافقاً ، (- ١) إذا كان غير موافق، والعكس في حالة العبارات المنسفية، ثم يجمع أيضاً الأوزان ليستدل منها على مدى إيجابية أو سلبية اتجاه كل تلميذ.

ذلك يمكن قياس الاتجاهات بدرجة أفضل عن طريق اختبارات المواقف، وفيها يُقدم للطالب عبارة تتصل بموقف من المواقف الواقعية المرتبطة بموضوع أو

قضية موضع الاهتمام يليها عدد من الإجابات أو البدائل المقترحة، وعلى الطالب أن يختار العبارة التي تنطبق عليه، أى تعبّر عن رأيه تجاه القضية، ومثال ذلك:

رأيت زميلك يُحرّض على تعاطي المخدرات، ماذا تفعل؟

- (أ) أحاول أن أبيّن له عوائق ذلك
- (ب) أطلب منه أن يكُف عن هذا التحرير
- (ج) أحاول معرفة سبب تحريره على ذلك
- (د) أبلغ الأمر لمسئولى المدرسة
- (هـ) لا أعطى للأمر أى أهمية

ويفيد تقويم اتجاهات الطلاب في معاونتهم على تبصر ذاتهم مما ييسر نضجهم الشخصي والانفعالي، كما يفيد المعلم في تنظيم خبرات تعليمية تتعلق بالاتجاهات وتضمين المناهج الدراسية أنشطة لتقديم تغذية مرتبة تساعدهم في تعرف مشاعرهم المتعلقة بموضوعات أو أحداث معينة. وتعرف نتائج ذلك للصف الدراسي بأكمله يفيد المعلم في اتخاذ قرارات تربوية تتعلق بأساليب تنمية اتجاهات إيجابية لدى الطلاب بمعاونة المرشد النفسي والتربوى بالمدرسة من أجل مزيد من الدفع العاطفى في بيئة التعليم، وتحمّل مشاعر الآخرين، والمشاركة الجماعية في اتخاذ القرارات المتعلقة بالأنشطة المدرسية، مما يسهم في إشراط تعلم الطلاب وإيقابهم على المدرسة، وبذلك يمكن تحقيق الأهداف التربوية المرجوة.

تقويم البرامج التربوية:

تُستخدم مقاييس الاتجاهات في كثير من الأحيان في تقويم فاعلية البرامج، والمقررات الدراسية والتجديدات التربوية في المناهج والتقنيات التعليمية. إذ ربما يقوم معلم الدراسات الاجتماعية بتقويم فاعلية البرنامج التعليمي المتعلق بهذا المجال في تغيير اتجاهات الطلاب نحو كثير من الموضوعات المتعلقة بقضاياها ومشكلات مجتمعهم والمجتمعات الأخرى، وبالتالي في بقية المجالات الدراسية.

كما يمكن استخدام مقاييس الاتجاهات في دراسة معنييات الطلاب عند التحاقهم بالمدرسة، وأثناء الدراسة وفي نهاية العام. وكذلك يمكن قياس اتجاهاتهم نحو البرنامج التعليمي، وأساليب التعليم وطرائق التدريس، وملاءمة الكتب المدرسية، والممارسات داخل المدرسة، وقضاياها ومشكلات مجتمعاتهم المحلية. ويفيد ذلك في

تُعرف جوانب القسوة وأوجه القصور في البرامج التربوية والعمل على تطوير هذه البرامج.

التشخيص العلاجي :

يمكن الإفادة من مقاييس الاتجاهات في تعرّف أسباب نفور بعض الطلاب من متابعة الدراسة مما قد يؤدي إلى ضعف تحصيلهم أو رسوبهم المتكرر، فالطالب الذي يعتقد أن المدرسة مضيعة للوقت، أو أنها مكان لحسو الأذهان بمعلومات غير نافعة، أو أن المعلمين يستغلون سلطتهم في جعل الطلاب يمثّلون لآرائهم وتوجهاتهم، من غير المحتمل أن يبذل الجهد والوقت اللازم لتحقيق النجاح والتميز في دراسته، وهذا يتطلّب تحديداً لاتجاهات مثل هؤلاء الطلاب، وتصميم برامج وأنشطة تعليمية ثرية ومتعددة تسهم في تعديل اتجاهاتهم نحو المدرسة والتعلم.

الإرشاد التربوي والانتقاء :

تُستخدم أحياناً مقاييس الاتجاهات في الإرشاد التربوي للطلاب، فمثلاً بعض موازين الاتجاهات المقننة المتوفّرة تهدف لمسح اتجاهات الطلاب نحو معلميهم، ونحو طرائق التدريس والأهداف التربوية ومتطلباتها، ونحو الدراسة بعامة وتوجهاتها، لتحديد السلوك المميّز لكل منهم، وتعريفه باتجاهاته ومشاعره نحو مختلف مكونات العملية التعليمية، والتميّز بين الطلاب الذين لديهم اتجاهات إيجابية أو سلبية نحوها.

غير أن موازين الاتجاهات ربما تكون محدودة الفائدة في مجال انتقاء الأفراد إذا لم يكن موضوع الاتجاه مرتبطاً أو متعلقاً بالأداء في مهنة أو عمل معين، فالفرد الذي لديه اتجاه اجتماعي أو سياسي معين ربما لا يؤثر في أدائه لعمل معين، وبالتالي لا يصبح هذا الاتجاه مؤثراً يُستند إليه في انتقاء الفرد للالتحاق بهذا العمل. كما أن بعض الأفراد يلجهون كما سبق أن ذكرنا إلى تزييف استجاباتهم في موازين الاتجاهات وفقاً لمتطلبات الانتقاء، فإذا تقدم فرد للالتحاق بمهنة التدريس مثلاً، ربما لا يُظهر عدم رضاه عن العمل مع الأطفال، أو سلبية اتجاهه نحوهم، أو نحو المهنة ذاتها نظراً لحاجته الماسة إلى الالتحاق بأى عمل، وبالتالي تصبح قائمة استخدام مقاييس الاتجاه في عملية الانتقاء موضع تساؤل.

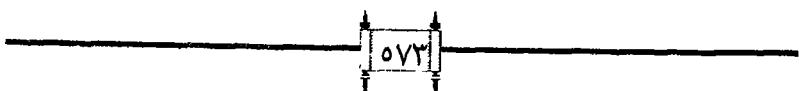
وعموماً، فإنه ينبغي عدم الاقتصار على نوع واحد من مقاييس الاتجاهات، وإنما يمكن استخدام مقاييس متعددة إذا تطلب الأمر ذلك. إذ يمكن استخدام موازين الاتجاهات إضافة إلى مقاييس أخرى غير مباشرة، وفي جميع الأحوال ينبغي مراعاة أوجه قصور كل من هذه المقاييس وارتباطه بال المجال المراد قياسه ويستلزم عمليات التقويم والتشخيص والإرشاد والانتقاء.

خلاصة

تُعدُّ الاتجاهات من الجوانب الوجданية الأساسية، حيث ترتبط بمشاعر الفرد ومعتقداته وحاجاته ودراوافه ورغباته، ويُستدل عليها من بعض الأنماط السلوكية التي تعكس في سلوك الفرد، أو من استجاباته اللغظية أو غير اللغظية لمقاييس الاتجاهات بأنواعها المختلفة. وهناك العديد من المفاهيم المتعلقة بمفهوم الاتجاه وتختلف عنه بدرجات متفاوتة مثل الميل والأراء والمعتقدات والقيم والدافع والعادات والروح المعنوية.

ولا تقتصر الاتجاهات على المكونة الوجданية وإنما تتضمن تفاعل هذه المكونة مع المكونتين المعرفية والسلوكية بما يجعل الفرد أكثر تهيئاً واستعداداً لكي يسلك بطريقة معينة نحو موضوع الاتجاه. وتميز الاتجاهات بخصائص أساسية من أهمها التوجُّه، والشدة، والمركيزية، والمرونة، والشمول، والاتساق. وتتعدد طرق وأساليب بناء مقاييس الاتجاهات، فبعضها يتعلق ببناء موازين مقننة تُعد بمثابة أساليب تقرير ذاتي، والبعض الآخر يعتمد على أساليب الملاحظة، والمقابلة الشخصية، أو يعتمد على أساليب غير مباشرة أو تنكرية، مثل الأساليب الإسقاطية، والأساليب الفسيولوجية. ومن أهم أساليب بناة موازين الاتجاهات أسلوب الفترات المتزاوية ظاهرياً (ثيرستون)، وأسلوب التقدير الجماعي (ليكرت)، وأسلوب التراكمي (جثمان)، وأسلوب تمایز معانى المفاهيم (أوزجود)، وأسلوب التصنيف التربى (ستيفنسون)، وأسلوب البسط (كومبس)، وأساليب قياس العلاقات الاجتماعية، والموازين متعددة الأبعاد، بأنواعها المختلفة. وتختلف هذه الأساليب في مستويات قياسها وخصائصها المختلفة. وتُستخدم موازين الاتجاهات استخدامات متعددة في الأغراض البحثية والتطبيقية أكثر من غيرها من المقاييس غير المباشرة، غير أن هناك بعض المشكلات المتعلقة بالاتجاهات العقلية للإجابات Response Sets ينبغي محاولة التغلب عليها سواء في تصميم موازين الاتجاهات وصياغة فقراتها، أو في تقيين عمليات تطبيق هذه الموازين لكي تكون نتائجها صادقة ومتسقة. ولعل تطوير مفهوم الاتجاه وأساليب قياسه في المجال التربوى يتطلب مزيداً من تأكيد علاقة الاتجاهات بالتعلم المدرسى واعتباره مدخلاً من مدخلات العملية التعليمية، وناتجاً أساسياً من

نواتجها، وما يترتب على ذلك من قرارات تربوية. وتوجيهه مزيد من البحوث نحو تطوير أساليب قياس الاتجاهات لبناء موازين تميز بنوعية جيدة وتوافر فيها الخصائص السيكومترية، ومحاولة التغلب على المشكلات الفنية المتعلقة بهذه الموازين التي تتخلل من صدقها وثبات درجاتها، واستخدام هذه الموازين في عمليات التقويم التربوى والنفسى.



الفصل الثالث عشر

قياس الشخصية

* مقدمة

* قياس الشخصية من منظور تاريخي

* مفهوم الشخصية وبعض نظرياتها

* تصنیف مقاييس الشخصية

* طرق وأساليب قياس الشخصية

(أولاً) مقاييس محددة البنية وواضحة الهدف (استبيانات الشخصية)

* استراتيギات بناء استبيانات الشخصية

* أمثلة لبعض استبيانات الشخصية

* استبيانات تستند إلى أساس نظري منطقى

* استبيانات تستند إلى التحليل العاملى

* استبيانات تستند إلى أساس إميريقى

* استبيانات جبرية الاختبار

* تقييم استبيانات الشخصية

(ثانياً) مقاييس غير محددة البنية وغامضة الهدف (الأساليب الإسقاطية)

* تصنیف الأساليب الإسقاطية

* أمثلة لبعض الاختبارات الإسقاطية

* تقييم الأساليب الإسقاطية

(ثالثاً) مقاييس غير محددة البنية وواضحة الهدف (الملاحظة، وال مقابلة،

ومقاييس مفهوم الذات)

* أساليب ملاحظة السلوك

- * الملاحظة في مواقف واقعية، ومواقف مضبوطة
 - * مصادر الخطأ في أساليب الملاحظة
 - * بعض الاقتراحات للتغلب على أخطاء الملاحظة
 - * موازين تقدير السلوك
 - * مصادر الخطأ في 'موازين تقدير السلوك'، وكيفية التغلب عليها
 - * أساليب المقابلة
 - * أنواع أساليب المقابلة
 - * ميزات وعيوب أساليب المقابلة
 - * تقييم أساليب الملاحظة والم مقابلة
- (رابعاً) مقاييس محددة البنية و GAMMA الهدف (مقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك)
- * مقاييس الأنشطة الفسيولوجية
 - * مقاييس الإدراك
 - * تقييم الأساليب الموضوعية
 - * تعقیب عام على طرق وأساليب قياس الشخصية
 - * استخدامات مقاييس الشخصية
 - * خلاصة

مقدمة :

عرضنا في الفصول السابقة أساليب وأدوات قياس التحصيل الدراسي والكتفيات والذكاء والاستعدادات الخاصة، أي الجوانب المعرفية، وتعلق جميعها بقياس الأداء الأقصى للفرد، كما عرضنا أساليب وأدوات قياس الميول والاتجاهات، أي الجوانب الوجدانية، وتعلق بقياس الأداء المميز للفرد. ونظرًا لأن شخصية الإنسان Personality بمفهومها العام تجمع هذه الجوانب وغيرها من المتغيرات المتعلقة بالتفكير، والدافعية، والسمات المزاجية، والمشاعر، والسلوكيات، في منظومة متكاملة تصف الأداء المميز للفرد أو نمط سلوكه المميز المتسق نسبياً، لذلك فإن أساليب تقييم الشخصية تتناول مجموعة متشعة من الخصائص المعرفية والوجدانية وغيرها من الخصائص التي يُطلق عليها جمِيعاً متغيرات الشخصية Variables. وهذه المتغيرات تمكِّن الفرد من فهم سلوكه وسلوك غيره من الأفراد والتبؤ به. فحياة الفرد تعد سلسلة متصلة من التفاعلات بينه وبين غيره من الأفراد، ويُصدر من خلال هذه التفاعلات العديد من الأحكام عن شخصيات من يتفاعل معهم في محاولة التعامل المشمر مع كل منهم. فالجار يحاول فهم شخصية جاره، والوالد يحاول تقييم سلوك وشخصية أبنائه، والمعلم يحاول معرفة الفروق الفردية بين شخصية طلابه، والطبيب يحاول التعرُّف على شخصية مريضه حتى يتَّسنى له معاونتهم في العلاج، وهكذا.

فمحاولات الأفراد فهم شخصية وسلوك الآخرين تُعد مشاركة من جانب كل منهم في عملية معقدة تتعلق بتقييم الشخصية الإنسانية. ولكننا بلا شك لا نستطيع أن نعتبر هذه المحاولات أسلوبًا للفيَّاس. ولقد حاول علماء النفس التمييز العلمي الدقيق بين شخصيات الأفراد باتباعهم الطرق والأساليب الكمية في القياس. فعالِم النفس يود معرفة مقدار سمة شخصية معينة لدى فرد ما، أو السمات التي تُميِّزه عن غيره من الأفراد، وأحياناً يود مقارنة فرد ما بفرد آخر، أو مقارنته بمتوسط مجموعة معينة، أو مقارنة سمات فرد الآن بسماته في وقت سابق. فتحديد مقدار أو إجراء مقارنة يتطلب عمليات قياس.

ويبينما يبدأ كثير من الأفراد بإصدار أحكام معينة على غيرهم ممن يتعاملون معهم ثم يُعدّلون من أحکامهم كلما زاد اندماجهم معهم؛ نجد أن عالم النفس يحاول عادة أن تكون أحکامه على الأفراد مبنية على أساس دقة باتباعه أساليب علمية في القياس.

لذلك سوف نتناول في هذا الفصل مفهوم الشخصية وبعض نظرياتها، وتصنيف المقاييس المتعلقة بهذه النظريات، والأساليب المتنوعة لقياس الشخصية، واستراتيجيات وطرق بناء الأنواع المختلفة من مقاييس الشخصية، وأمثلة متعددة لهذه المقاييس، وما يتعلق بكل نوع منها من مصادر خطأ وكيفية التغلب عليها، ونقدم تقييماً لكل أسلوب من أساليب القياس، وتقييماً عاماً لهذه الأساليب، ونوضح الاستخدامات المتعددة لمقاييس الشخصية.

قياس الشخصية من منظور تاريخي :

على الرغم من أن الدراسات العلمية للشخصية بعامة لم تبدأ قبل منتصف العقد الثالث من القرن العشرين، إلا أن محاولات تصنیف الأشخاص في أنماط أو أقسام استناداً إلى الخصائص الجسمية أو الفسيولوجية كانت سائدة قبل ذلك بزمن بعيد. ونظراً لأن هذه المحاولات المبكرة تبيّن عدم صحتها، فقد أصبحت غير مقبولة في وقتنا الحاضر. وقد أعد جالتون Galton أول استبيان للشخصية عام ١٨٨٠ لدراسة التصور، واستخدم هول Hall استبياناً للشخصية في دراسته لنمو المراهق، وكذلك استخدم كريبلين Kraeplin، وسومر Sommer اختبارات التداعي الحر، وأعد يونج Jong اختباراً للتداعي الحر عام ١٩٠٥ لقياس الانفعالات غير السوية.

ومما شجع على تطوير مقاييس مقتنة للشخصية، الحاجة إلى فحص الخصائص الشخصية للأفراد الملتحقين بالجيش في الحرب العالمية الأولى. ولعل قائمة وودورث لبيانات الشخصية Woodworth Personal Data Sheet تعد أول مقياس للشخصية محدد البنية تم بناؤه أثناء هذه الحرب، ونشر في صيغته النهائية عقب انتهاءها. وهذا يدل على أن نشأة مقاييس الشخصية ارتبطت بضرورات عملية نتيجة تحديات معينة، كما هو الحال في نشأة مقاييس الذكاء والجوانب المعرفية الأخرى. وتعد هذه القائمة بمثابة مقابلة شخصية مقتنة يجب فيها الفرد إما بـ «نعم» أو بـ «لا» عن ١١٦ فقرة أو عبارة تتعلق بالسلوك، وكان الهدف منها تحديد الأفراد الذين ربما يصعب عليهم التكيف للمواقف الضاغطة في الحياة العسكرية. وقد تبينفائدة هذه القائمة كأدلة فحص في ذلك الوقت، غير أنها لم تكن أدلة تشخيصية، كما لم تستند إلى أساس نظري لبنية السلوك سواء في بنائها أو في تفسير درجاتها؛ لذلك اتجه علماء النفس الكлиничى إلى بناء مقاييس أخرى تستند إلى أساليب أكثر تطوراً. ويُعد استبيان برنسرويتير Bernreuter Personality Inventory من المقاييس التي نالت قبولاً واسعاً خلال الأعوام التي تلت الحرب العالمية الأولى، حيث اشتمل على أربع خصائص

وصفية للشخصية هي: التزعات العُصبية، والاكتفاء الذاتي، والانطواء، والسيطرة. وقد اتسع نطاق استخدام هذا الاستبيان لمدة طويلة، ولكن قلًّا استخدامه في وقتنا الحاضر، حيث تلاه العديد من استبيانات الشخصية نتيجة النقد الذي وجّه للاستبيانات المبكرة للشخصية.

وبينما كان كثير من علماء النفس في الولايات المتحدة الأمريكية في العشرينات يقومون بإعداد استبيانات محددة البنية لقياس الشخصية وتطلب ورقة وقلماً، كان هناك اتجاه آخر يهدف لبناء مقاييس غير محددة البنية. فقد اهتم هرمان رورشاخ H. Rorschach Projective Techniques فى سويسرا بوضع أساس المقاييس الإسقاطية Techniques التى تستند إلى فكرة أن الفروق فى شخصية الأفراد تتعلق بالفروق فى إدراكمهم للعالم المحيط بهم.

لذلك بدأ التجريب على مثيرات غير محددة البنية، مثل التصميمات المجردة التى يمكن الحصول عليها من بقع الحبر لكي يستجيب لها الفرد بما يدركه. وقد نشر اختبار بقع الحبر The Rorschach Ink Blot Test عام ١٩٢١، ويعود من أكثر الاختبارات المعروفة فى قياس الشخصية حتى وقتنا الحاضر، ويُستخدم فى التشخيص الكlinيكي. غير أن الاهتمام بهذا النوع من القياس ازداد ببطء، فقد مررت عدة أعوام قبل نقل اختبار رورشاخ إلى الولايات المتحدة الأمريكية بواسطة ليڤي Levy ، وأجرت حوله أول دراسة للدكتوراه عام ١٩٣٢ بواسطة بيك Beck أحد تلاميذ ليڤي، حيث اهتم بالدراسة العلمية لخصائص هذا الاختبار. وقد استارت هذه الدراسة علماء النفس هناك على الرغم مما قوبلت به الأساليب الإسقاطية من ارتياح وشك ونقد من جانبهما. وعلى الرغم من ذلك قام موراي Murray ومورجان Morgan بإعداد اختبار تفهم الموضوع (TAT) The Thematic Apperception Test عام ١٩٣٥ ، وهو يختلف عن اختبار رورشاخ فى أن بنية مثيراته أكثر تحديداً من بنية بقع الحبر، إذ يشتمل على صور غامضة تمثل مناظر ومواقف متنوعة، ويُطلب من المفحوص أن يكون قصة تتعلق بهذه المثيرات الغامضة.

وقد نال هذان الاختباران اهتماماً متزايداً فى أواخر الثلاثينيات وأوائل الأربعينيات، ولعل هذا كان مرجعه إلى وقفة التأمل من جانب علماء النفس فيما يتعلق بالاستبيانات محددة البنية من أجل التغلب على المشكلات المتعلقة بينها وتفسير درجاتها. غير أن المقاييس الإسقاطية للشخصية، وبخاصة اختبار رورشاخ لم تحظ بعد ذلك بقبول عام.

ولعل إسهامات مورفي Murphy وجنسن Jensen عام ١٩٣٧ بمؤلفهما بعنوان *Approaches to Personality* الذي تناولا فيه بالتفصيل مدارس علم النفس، وطب الأمراض النفسية، وتوجيه الأطفال، وكذلك إسهامات ألبورت Allport المتعلقة بمفهوم الذات قد أثرت تأثيراً واضحاً في التنظير السيكولوجي في مجال الشخصية.

وقد حدث أثناء الحرب العالمية الثانية تطورات ملحوظة في أساليب بناء استبيانات الشخصية، فقد أعد عام ١٩٤٣ استبيان الشخصية متعدد الأوجه (MMPI)، واستند إلى الأساليب الإمبريقية في بنائه. وبعد هذا الاستبيان من أكثر مقاييس الشخصية استخداماً في وقتنا الحاضر ونال قدرًا كبيراً من الدراسات والبحوث. وتلى ذلك بناء استبيان كاليفورنيا السيكولوجي (CPI) عام ١٩٥٧ استناداً إلى الأساليب الإمبريقية أيضاً.

وقد استرشد علماء النفس بالتطورات السيكومترية، وبخاصة أساليب التحليل العاملى Factor Analysis ، وأساليب الاختيار الجبرى أو القهرى Forced-Choice في بناء استبيانات الشخصية. وبعد كاتل Cattell، وثيرستون Thurstone، وأيزنك Eysenck، وجيلفورد Guilford من الرواد الذين استخدمو أساليب التحليل العاملى في مجال قياس الشخصية. وتعد المحاولات الجادة التي قام بها جيلفورد في أوائل عقد الأربعينيات، وكاتل في نهاية العقد من الجهود البارزة في هذا المجال. فقد أعد كاتل Cattell استبيان العوامل الستة عشر للشخصية (16 PF)، حيث اعنى عناية كبيرة ببنائه استناداً إلى أسلوب التحليل العاملى.

واهتمت بحوث قياس الشخصية بعد ذلك بتطوير نظرية في الشخصية أكثر من اهتمامها بأساليب القياس من أجل صياغة تكوينات فرضية استناداً إلى هذه النظرية. لذلك نلاحظ أن مقاييس الشخصية تُستخدم حتى وقتنا الحاضر بكثير من الحرص والحيطة من جانب اختصاصي علم النفس والتربية نظراً لخصوصية هذا النوع من المقاييس كما سيوضح في هذا الفصل.

مفهوم الشخصية وبعده نظرياتها :

إن قياس وتقدير الشخصية يُعد مجالاً رئيساً من مجالات القياس التربوي والنفسي، ولكل يمكنا فهم مقاييس الشخصية فهماً واضحاً ينبغي أن نعرف المقصود بمفهوم الشخصية، ولماذا نحاول قياسها. فعندما تكون بقصد مفهوم أساسى في مجال السلوك الإنسانى ، فإننا نواجه مشكلة تعدد تعريفات هذا المفهوم . ولقد سبق أن

لاحظنا ذلك فيما يتعلق بمفهوم الذكاء. وينطبق ذلك أيضاً على مفهوم الشخصية، فتعريف الشخصية يختلف باختلاف وجهة نظر علماء النفس ومدارسهم الفكرية والسيكلولوجية. وقد استنتج ألبرت Allport من دراسته المسجية المبكرة التي أجرتها عام ١٩٣٧ أن هناك ما يقرب من ٥٠ تعرضاً على الأقل لمفهوم الشخصية. وربما يرجع ذلك إلى صعوبة تحديد ما يشتمل عليه هذا المفهوم المتسع الذي يعد من أكثر المفاهيم السيكلولوجية غموضاً (Smith & Adams, 1972).

إذ يرى جيلفورد (Guilford, 1959) أن الشخصية تنطوي على جميع خصائص الفرد المستقرة نسبياً، وتتضمن المظهر الجسماني، والذكاء والاستعدادات، والميول، والاتجاهات، والقيم، والأمراض النفسية، ويرى غيره أن مفهوم الشخصية يقتصر على السمات السلوكية غير المعرفية التي تكشف عن مفهوم الذات، والخلق، والنسق الفردي، وبخاصة المكونات الانفعالية والداعية لسلوك الفرد.

ويذلك يُستبعد الذكاء والاستعدادات والمظهر الجسماني، غير أن التوجهات الحديثة اقترحت ضرورةأخذ الجوانب المعرفية بعين الاعتبار من أجل التوصل إلى نظرية متكاملة للشخصية (Mischel, 1973). وتستند هذه التوجهات إلى أن جميع مظاهر السلوك الذي يكشف عن الشخصية من شأنها العقل، ولكن نستطيع متابعة دراسة الشخصية يجب عدم إغفال العقل، كما أن المواقف الموضوعية المحيطة بسلوك الفرد ربما تختلف اختلافاً جوهرياً عن المعنى الذاتي الذي يضفيه الفرد على هذه المواقف، حيث إن هذه المواقف ليست منعزلة عن الفرد.

^٤ لذلك لا يوجد تعريف صحيح للشخصية وتعريف خطأ، كما لا يوجد تعريف وحيد صواب، وإنما يعتمد التعريف على استخداماته المناسبة المتعلقة به، فمثلاً يذكر ألبرت (Allport, 1961) أن الشخصية هي التنظيم الديناميكي داخل الفرد لأنظمة النفسية الجسمانية التي تحدد سلوكه المميز وتفكيره. ويذكر سندربرج (Sundberg, 1977) أن الشخصية هي مجموعة العمليات التي يستخدمها الفرد أو الأفراد في تنمية انطباعات وتصورات، واتخاذ قرارات وتحقيق من فرضيات تتعلق بأنماط خصائص فرد آخر التي تحدد سلوكه في تفاعله مع البيئة. ويرى كراون (Crowne, 1969) أن الشخصية تعد نظاماً من الإمكانيات للسلوك، ويقصد بذلك أننا نلاحظ التنظيم والوحدة في السلوك، حيث إن سلوك الفرد ليس جزئياً أو منفصلاً، وإنما يكون متسبباً عبر الزمن وباختلاف المواقف، فالشخصية تتعلق بالفرد ككل وليس فقط بذاته أو اتجاهاته أو بعض دوافعه. فهي تتعلق بهذه جميعاً وبغيرها من المتغيرات وبأثر تفاعلاتها في

السلوك، لذلك يشير كراون إلى أهمية التمييز بين التحليل الكلينيكي لشخصية فرد معين من أجل فهم الشخصية الوظيفية للفرد ككل، وبين المدخل الذي يتبعه الباحثون في دراسة الشخصية حيث يعتمد على عزل بعض الأحداث السلوكية ودراسة أثر كل منها وكذلك أثرها المشترك في التجارب متعددة المتغيرات. كما أن هذا التعريف يؤكّد الإمكانيات للسلوك *Potentialities*، وليس السلوك في ذاته، فالاستدلال من السلوك عن محرّكاته وتوجّهاته أي إمكاناته هو الذي يُمكّنا من الفهم العام المتسع للسلوك وبالتالي للشخصية.

وعموماً فإن علماء النفس يمكن أن يتفقوا على عدد من الخصائص العامة لتعريف مقبول متسع للشخصية، على الرغم من أن استفادتهم إلى تعريفات توجّه دراساتهم يمكن أن تدلّ أو تتغاضى عن إحدى هذه الخصائص أو بعضها. ولعل التعريف التالي يقدم إطاراً يفيد في توجيه الحوار حول مفهوم الشخصية. فالشخصية تشير إلى تكامل جميع خصائص الفرد المستقرة نسبياً في تنظيم فريد يحدد محاولاته للتكيف مع بيئته دائمة التغيير، ويمكن تعديله في ضوء هذه المحاولات. ويتضمن هذا التعريف العناصر التالية:

- (١) تعدد السمات والإمكانات والقدرات التي تشتمل عليها الشخصية.
- (٢) انتظام هذه السمات والإمكانات والقدرات في نسق متكامل.
- (٣) تفرّد الشخصية، حيث تُميّز بين الفرد وغيره من الأفراد.
- (٤) تأثير الشخصية في العلاقات التفاعلية بين الفرد وبيئته وبيئه من الأفراد.
- (٥) اعتبار الشخصية من خصائص الفرد المستقرة نسبياً عبر مدة زمنية طويلة.

وقد حاولت نظريات الشخصية وضع تصورات نسقية لهذا المجال الذي يتميز بالتعقد بطرق مختلفة، لعل أكثرها شيوعاً اقتراح أبعاد أو سمات تتبادر حولها الشخصية، أو اقتراح تفسيرات توضح نمو الشخصية وдинامياتها، فالمدخل الأول يركز على طبيعة بنية الشخصية، أما المدخل الثاني فيركز على عمليات نمو الشخصية.

ويختلف توجّه الدراسات الإمبريالية في كل من المدخلين، فالنظريات البنائية تستند إلى نتائج اختبارات الشخصية في تحديد السمات الأساسية للشخصية، فقد توصلت بعض الدراسات إلى أبعاد ثنائية القطب، مثل الانطواء - الانبساط ، الهوس - الاكتئاب ، الذكورة - الأنوثة، وتوصلت دراسات أخرى إلى أبعاد متعددة للشخصية السوية أو العادية، فالقرير الموضوعي لشخصية الفرد في إطار هذا المدخل الذي يتعلق

بالسمات Trait Approach يتطلب تقدير موقعه أو مكانته في أبعاد متعددة. والشكل أو التكوين العام لهذه التقديرات أو الموضع المتعلقة بسمات الفرد يسمى الصفحة النفسية Psychogram أو البروفيل النفسي Profile . وقد سبق أن أوضحنا ذلك في الفصل السادس.

أما المدخل الثاني الذي يستند إلى نظريات التعلم ونظريات التحليل النفسي؛ فيركز على تحليل عمليات نمو الشخصية، ويؤكد العوامل الوراثية والبيئية تأكيداً كبيراً، وكذلك يؤكّد العمليات المهمة في التكيف الإنساني، مثل التعلم والدافعية والصراعات، وربط ميكانيزمات التعميم والتمييز ، وتكوين العادات السلوكية بمشكلات الشخصية .

وبالطبع توجد مداخل أخرى تستند إلى نظريات تجهيز المعلومات Information Processing وغيرها من النظريات المعرفية.

وسوف نوضح هذه المداخل المتعددة في سياق مناقشتنا لطرق وأساليب قياس الشخصية .

(تصنيف مقاييس الشخصية) :

اتضح لنا مما سبق أنه لا يوجد اتفاق بين علماء النفس على تعريف مفهوم الشخصية، مما أدى إلى تباين وتعدد النظريات التي حاول كل منها إلقاء الضوء على هذا المفهوم من منظور معين. وقد ترتب على ذلك تباين وتعدد طرق وأساليب قياس الشخصية وما يتعلّق بها من متغيرات. والحقيقة أن مجال قياس الشخصية يُعد من أكثر مجالات القياس النفسي والتربوي تنوعاً في طرقه وأساليبه، وربما يرجع ذلك كما سبق أن أوضحنا إلى ما يتضمنه مفهوم الشخصية من متغيرات ومفاهيم أخرى غير متجانسة. وعلى الرغم من تعدد طرق تصنيف مقاييس الشخصية، إلا أن هناك عناصر مشتركة يمكن أن تجمع بينها، فقياس أو تقييم الشخصية يُعد أسلوباً لجمع المعلومات عن فرد ما، وعلى وجه التحديد يتضمن الملاحظة الموضوعية المنظمة للسلوك تحت شروط محددة في علاقتها بمثيرات معينة، وللحصول على بيانات أو معلومات عن فرد ما يجب محاولة فهم أثر مختلف مكونات الموقف المتعلق بالثير على سلوكه، ويتحدد هذا الموقف الذي يواجهه الفرد ويستجيب له بخصائصه الفيزيائية، وبالتعليمات التي تقدم له ، وبالمطلب المحدد.

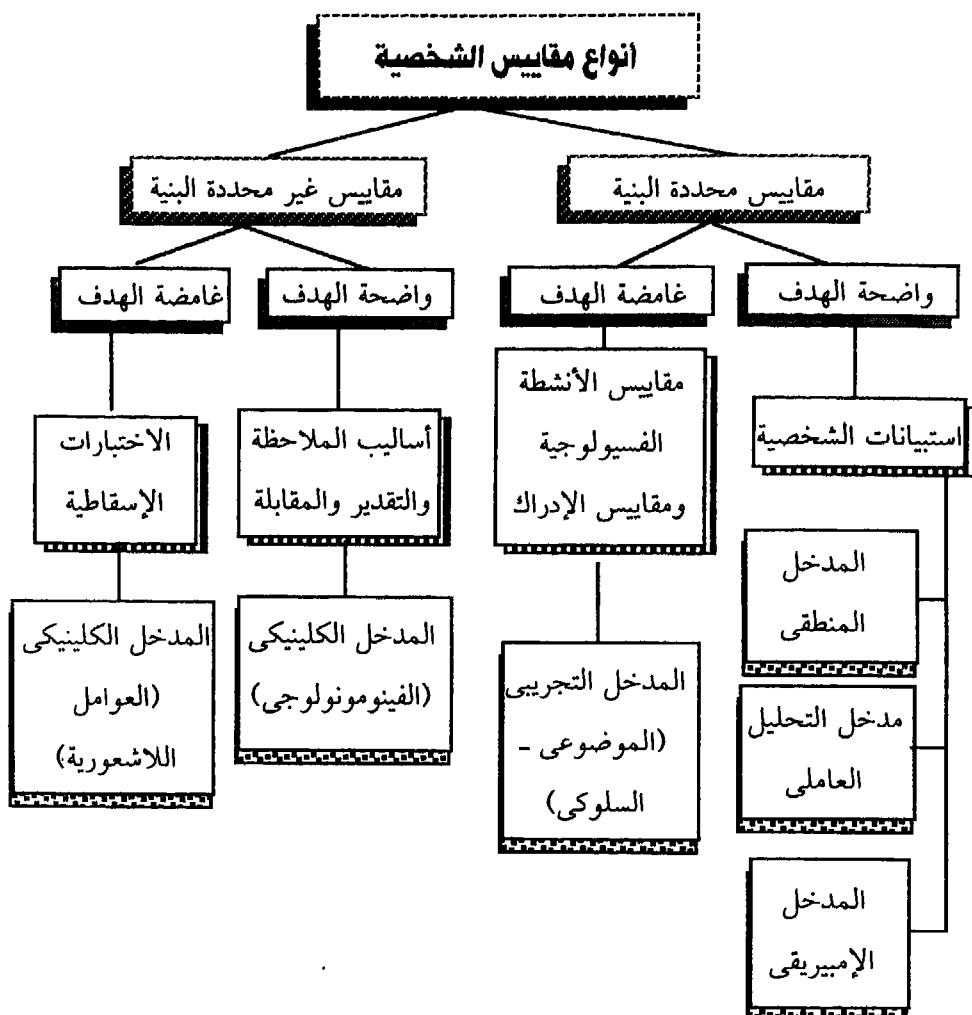
ونظراً لاختلاف المتغيرات المرتبطة بعملية جمع البيانات الشخصية عن الفرد، فإن هناك تبايناً واضحاً في طرق وأساليب جمع هذه البيانات. فطبيعة كل من الموقف، والمثير، والتعليمات المعطاة، والاستجابة المطلوبة، وكذلك كيفية تقديم الدرجات وتحليلها وتفسيرها، تُعد خصائص لعملية قياس أو تقييم الشخصية. إذ يستند إليها في عملية جمع البيانات وتسجيلها وتحليلها وتفسيرها. لذلك ينبغي اتباع أسلوب منظم في الحصول على هذه البيانات في إطار نظرية معينة أو فروض نظرية واضحة ومحددة المعالج، فاختلاف نظريات الشخصية يؤدي إلى طرق مختلفة لملاحظة سلوك الأفراد. أي أن هناك علاقات تفاعلية بين نظريات الشخصية وطرق وأساليب القياس أو التقييم.

ويرى أرندت (Arndt, 1974) أن نظرية الشخصية التي لا تسهم في تحديد وتسير وسائل جمع البيانات تُعد نظرية غير ذات جدوى من الناحية العملية والتطبيقية، كما أن البيانات غير المرتبطة بنظرية معينة تكون عديمة المعنى وقليلة الفائدة؛ لذلك ينبغي أن يستند تصنيف مقاييس الشخصية إلى إطار نظري أو فكري واضح، وأن يُراعى الفروق بين مختلف نظريات الشخصية، وتوجهات البحوث السيكولوجية في هذا المجال.

وتباين طرق وأساليب قياس أو تقييم الشخصية في مكونات عملية جمع البيانات التي أشرنا إليها. فالموقف الاختباري يمكن أن يكون شبيهاً بموافقات الحياة الواقعية، أو يكون محاكاً له، أو يكون موقفاً مختبراً محدد المثيرات، وكذلك يمكن أن يتضمن الموقف مثيرةً واحداً أو عدداً كبيراً من المثيرات محددة البنية أو غير محددة البنية (غامضة). وبعض المواقف الاختبارية يتطلب تسجيل استجابة واحدة وبعضاها الآخر يتطلب تسجيل عدد من الاستجابات، وهذه الاستجابات إما أن تكون تقريراً لفظياً ذاتياً، أو استجابات فسيولوجية ، أو أداءً معيناً، أما تقدير الدرجات فيمكن أن يكون موضوعياً ويمكن إجراؤه بواسطة الحاسوب أو يدوياً، كما يمكن أن يكون ذاتياً استناداً إلى أحکام الأخصائيين النفسيين. وفي الحالة الأولى يتم تفسير الدرجات من خلال الصفحات النفسية أو البروفيل Profile الذي تحصل عليه باستخدام برنامج الحاسوب، وفي الحالة الثانية يتم تفسير البيانات تفسيراً إكلينيكياً، بواسطة هؤلاء الأخصائيين.

لذلك يمكن تصنيف مقاييس الشخصية في بعدين أساسين كل منهما ثنائياً القطب (1969; Campbell, 1957; Mehrens & Lehman, 1969) من حيث بنية المثيرات، والهدف من الاختبار. فبنية المثيرات يمكن أن تكون محددة Structured أو غير محددة Nondisguised، والهدف من الاختبار إما أن يكون واضحاً Nonstructured أو غامضاً.

، وبذلك يمكن أن تقسم مقاييس الشخصية إلى أربعة أنواع يوضحها الشكل التخطيطي (١-١٣) التالي :



شكل (١-١٣) يوضح أنواع مقاييس الشخصية
(من تصميم المؤلف)

ويتبين من شكل (١-١٣) أن المقاييس محددة البنية يمكن أن تكون واضحة الهدف كما في استبيانات الشخصية **Personality Inventories** ، أو غامضة الهدف كما في اختبارات الأداء المقتنة ، ومقاييس الأنشطة الفسيولوجية . ويعتمد تحديد بنية المقاييس على ما يُسمح به للفرد من حرية الاستجابة . ففي المقاييس محدد البنية

يستجيب الفرد لعدد محدود من البذائل ويُقرر ما يراه منها منطبقاً عليه. أما في المقياس غير محدد البنية فإن الفرد يكون حرّاً في استجابته كما يراها ويعبر عنها. والمقياس واضح الهدف يعني أن الفرد يكون مدركاً للغرض الفعلى من تطبيقه عليه، أما المقياس غامض الهدف أو التنكرى، فإن تفسير الفاحص لنتائجها يختلف عما يفترضه المفحوص حول الهدف من المقياس، فالمقياس الذى يُطبق على أنه يقيس القدرة الإدراكية ويستخدم الفاحص نتائجه في الاستدلال على الاضطراب النفسي مثلاً يكون مقياساً محدد البنية، ولكنه غامض الهدف.

وقد سبق أن أوضحنا ذلك في الفصل السابق عند مناقشتنا لأساليب قياس الاتجاهات. ويستند بناء استبيانات الشخصية إلى المدخل السيكومترى أو المدخل الإيميريقي الذى ناقشناه في الفصل الحادى عشر فيما يتعلق بأساليب بناء مقاييس الميلول. أما اختبارات الأداء ومقاييس الأنشطة الفسيولوجية، فإنها تستند إلى المدخل الموضوعى السلوكي الذى يعتمد على الأساليب التجريبية في جمع بيانات عن سلوك الفرد في موقف يتم ضبط متغيراته.

وكذلك المقاييس غير محددة البنية يمكن أن تكون واضحة الهدف، مثل المقابلة الشخصية حيث تهتم بالجانب الكلينيكي، وتؤكّد المدخل الفينومونولوجي Phenomenological Approach، ويكون الفرد مدركاً لذاته وللبيئة المحيطة به، كما يمكن أن تكون غامضة الهدف مثل الاختبارات الإسقاطية Projective Tests التي تهتم بالجانب الكلينيكي أيضاً ولكن تؤكّد العوامل اللاشعورية. لذلك تعتمد هذه الاختبارات على مشيرات غامضة مثل التصميمات المبهمة ، والصور ، والكلمات ، والجمل ، والأفعال .

وعلى الرغم من أنه يمكن تصنيف مقاييس الشخصية في أقسام أخرى إضافة إلى هذه الأقسام الأربع، أو في أقسام تستند إلى أساس مختلف، إلا أن التصنيف الذي يوضحه شكل (١-١٣) يُعد مناسباً وكافياً لتوضيح المقاييس التي ستتناولها في الجزء التالى .

طرق وأساليب قياس الشخصية :

نظراً لتنوع طرق وأساليب قياس الشخصية التي يمكن تصنيفها في الأقسام الأربع الموضحة بشكل (١-١٣)، فسوف نتناول كل قسم منها على حدة، ونوضح الأساس النظري الذي يستند إليه بناء المقاييس المتعلقة به، ونقدم بعض الأمثلة للمقاييس شائعة الاستخدام في كل قسم منها، وتقديماً لهذه الأقسام. ونظراً لأن استبيانات الشخصية

والاختبارات الإسقاطية تُعد أكثر المقاييس شيوعاً واستخداماً، فسوف نبدأ مباحثتنا بالقسمين الأول والرابع، ثم ننتقل إلى القسمين الثاني والثالث.

أولاً : مقاييس مجردة البنية وواضحة الهدف (الاستبيانات الشخصية) :

تُعد استبيانات الشخصية Personality Inventories من المقاييس محددة البنية وواضحة الهدف بالنسبة للفرد المستجيب، كما تُعد من المصادر الأساسية للحصول على بيانات ومعلومات تتعلق بالعديد من سمات الشخصية، وتستند هذه الاستبيانات إلى نظريات السمات Trait Theories. فقد أوضح ألبورت Allport أن شخصية الفرد تكون من تنظيم ديناميكي من السمات التي تحدد أسلوبه المميز في التكيف مع بيئته. وقد اتفق ألبورت Allport وكائل Cattell في ضرورة قياس السمات الأساسية التي تُميز بين الأفراد، والتي يعتقدان أنها عوامل محددة داخل الفرد.

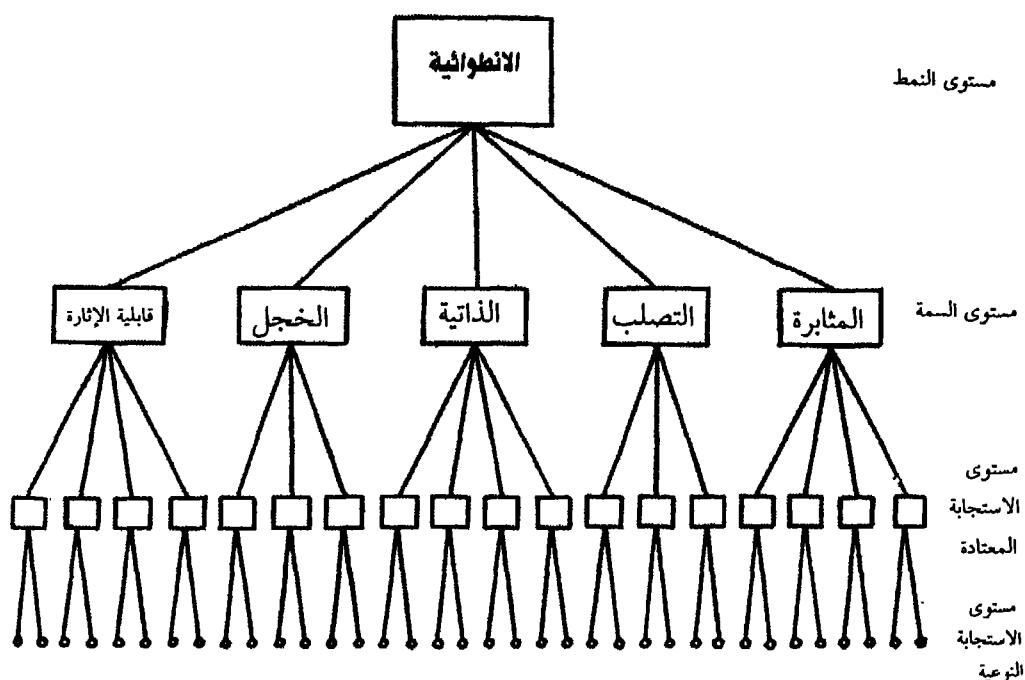
غير أن ألبورت يرى أن المدخل المتعلق بصياغة القوانين Nomothetic Approach ليس كافياً بذاته، ويجب أن يُعزز أو يُستكمَل بدراسة ووصف التنظيم الفريد المميز لكل فرد Idiographic Approach. بينما يرى كائل أن هذا التنظيم أو هذه البنية Structure يمكن قياسها بذاتها باستخدام الأساليب الإحصائية المتقدمة. وفي حين اهتم كائل Cattell، وجيلفورد Guilford في الولايات المتحدة الأمريكية بمدخل السمات Trait Approach، نجد أيزنك Eysenck في إنجلترا اهتم بمدخل الأنماط Types Approach، فمدخل السمات يتضمن تحديد عدد من السمات أو خصائص السلوك المستقرة نسبياً التي تُميز بين الأفراد ولكنها تمثل نزعة مشتركة بينهم، فالكلسل مثلاً خاصية مشتركة بين الأفراد على الرغم من تباين درجاتها لكل منهم.

لذلك تهدف نظريات السمات تحديد أهم سمات الشخصية الإنسانية وأكثرها دلالة ، وابتكر أساليب لقياس هذه السمات بدقة لدى كل فرد، أما مدخل الأنماط فيهدف لتصنيف الأفراد في أنماط وفقاً لتجمُعات معينة من السمات، أى أن الأنماط تُعد وحدات وصفية ذات رتب أعلى من السمات.

وقد أجرى أيزنك Eysenck دراسة رائدة عام ١٩٤٧ طبق فيها عدداً كبيراً من المقاييس على عينة تشتمل على ١٠٠٠ فرد من الجنود العاديين والعصابيين، إضافة إلى ملاحظات كلينيكية وجسمانية، وتوصل باستخدام أسلوب التحليل العائلي إلى

عواملين رئيسيين كل منهما ثناى القطب هما: الانطوائية - الانساطية ، والعصبية الاستقرار .

وقد أجري دراسة أخرى عام ١٩٥٢ اشتملت على عينات من الأفراد العاديين وأفراد من مرضى المستشفيات العقلية، وتوصل إلى عامل ثالث أطلق عليه عامل الذهان Psychoticism ويتعلق بجماعة خاصة. وتمثل هذه العوامل الثلاثة المستوى الأعلى في تصنيف أيزنك حيث تدل على أنماط الشخصية. وقد قام بتقسيم كل منها إلى سمات تدرج تحتها استجابات معتادة، وهذه تنقسم إلى استجابات نوعية. أي أن مدخل الأنماط أدى إلى تصنیف هرمي أو بنوي لسمات الشخصية ، فمثلًا نمط الشخصية المتعلق بالانطوائية Introversion ، وهو المستوى الأعلى ، وما يندرج تحته من مستويات أو ضيقه أيزنک بالشكل التخطيطي (٢-١٣) التالي (Eysenck, 1953) :



شكل (٢-١٣) يوضح النموذج الهرمي لبناء الشخصية فيما يتعلق ببعد الانطوائية

أما كاتل فقد ميز بين السمات المصدرية Source Traits والسمات السطحية Surface Traits، فالسمات المصدرية تكون من مجموعات من المتغيرات يبدو أنها ترتبط فيما بينها ارتباطاً متسلقاً إلى حد ما، بينما تمثل السمات السطحية المتغيرات التي تحدد الأحداث على المستوى السطحي للشخصية ، لذلك تُعد السمات المصدرية أكثر أهمية من السمات السطحية في تفسير السلوك والتبؤ به؛ لأنها أكثر استقراراً واتساعاً من السمات السطحية. فالسمات السطحية يمكن ملاحظتها بسهولة في سلوك الفرد، ولكن السمات المصدرية يمكن الكشف عنها فقط باستخدام الأساليب الإحصائية . ويمكن تحديد أو تعرف السمات المصدرية باستخدام أسلوب التحليل العاملی للعلاقات القائمة بين السلوك السطحي ، كما ميز كاتل بين السمات من حيث كونها عامة أو مُتفرّدة ، موروثة أو متاثرة بالبيئة ، ودينامية أو تتعلق بالقدرات ، أو المزاج . فالسمات العامة يشتراك فيها جميع الأفراد ، بينما السمات المتميزة تميز فرداً بعينه ، والسمات الدينامية توجه الفرد نحو هدف معين ، والسمات المتعلقة بقدراته تحدد قدراته على تحقيق الهدف ، وسماته المزاجية تتعلق بالجوانب الانفعالية لنشاطه الموجه نحو الهدف .

وقد استند كاتل على ثلاثة مصادر رئيسة مختلفة في الحصول على البيانات في دراساته هي :

- (١) سجل حياة الفرد (L Data) : وتعمل بملاحظات الأفراد عن شخصية فرد يعرفونه جيداً.
- (٢) استبيانات التقرير الذاتي (Q Data) : Self-Rating Questionnaire وتعمل بملاحظات الفرد عن نفسه.
- (٣) اختبارات موضوعية (T Data) Objective Tests : وتعمل بأداء الفرد في اختبارات مواقف محددة.

وقد استخدم البيانات المستمدة من هذه المصادر في تكوين مصفوفات معاملات الارتباط بين متغيرات الشخصية لكل نوع من أنواع البيانات الثلاثة، وأجري تحليلًا عاملياً على هذه المصفوفات، وتوصل إلى ١٦ عاملًا أولى أو سمات مصدرية للشخصية، ويتعلق معظم هذه العوامل بسمات مزاجية، فيما عدا عامل يتعلق بالقدرة، وعامل له طبيعة دينامية رأى كاتل أنه يمثل سمة سطحية.

وعلى الرغم من أن كاتل Cattell اعتمد في قياسه لسمات الشخصية على عوامل يتراوح عددها بين ١٦ ، ٢١ عاملًا ، إلا أن أيزنك Eysenck فضل الاعتماد على ثلاثة

عوامل عريضة نسبياً وغير مترابطة. وقد أشار أيزنك (Eysenck, 1965) في مؤلفه «بنية الشخصية وقياسها Personality Structure and Measurement» أن العوامل التي توصل إليها كاتل تحتاج لمزيد من إمعان النظر، وأن نموذجه الهرمي الذي يشتمل على عاملين رئيسين يتميز على نموذج كاتل. غير أن هناك بعض أوجه القصور الفنية والإحصائية التي تشوب دراسة أيزنك أشارت إليها بعض الدراسات الأخرى، مما يجعل ملاحظته السابقة تتطلب مزيداً من البحث. ومع هذا فإنه يبدو أن هناك اتفاقاً فيما بينهما في عامل الانبساطية والقلق اللذين يُعدان عاملين من الرتبة الثانية عند كاتل، لذلك نلاحظ أن نظرية كاتل المتعلقة بسمات الشخصية تُعد إطاراً أفاد في بناء عدد من استبيانات الشخصية.

استراتيجيات بناء استبيانات الشخصية :

يستند بناء استبيانات الشخصية إلى استراتيجيات مماثلة لاستراتيجيات بناء استبيانات الميلول التي سبق أن أوضحناها في الفصل الحادى عشر. ومع هذا سوف نلقى مزيداً من الضوء على هذه الاستراتيجيات في مجال قياس الشخصية. فقد أشرنا إلى ثلاث استراتيجيات رئيسية يستند إلى إحداها في بناء مجموعات من الفقرات المتماثلة استرشاداً بأحكام الخبراء، ويُستند إلى الثانية في بناء مجموعة من الفقرات التي تميز بالانساق الداخلى، ويُستند إلى الثالثة في انتقاء فقرات على أساس إمبريقي اعتماداً على المقارنة بين مجموعتين من الأفراد.

استراتيجية تستند إلى أحكام الخبراء : وتعتمد هذه الاستراتيجية في تجميع الفقرات - بحيث تشكل ميزان قياس - على آراء الخبراء، وقد أطلق عليها لانيون، وجودستاين (Lanyon & Goodstein, 1971) وكذلك ويجيتز (Wiggins, 1973)، الاستراتيجية المنطقية النظرية Rational-Theoretical Strategy، وأطلق عليها جولدبرج (Goldberg, 1974) استراتيجية الحدس Intuitive Strategy، بينما أطلق عليها مالونى (Maloney & Word, 1976) استراتيجية المحتوى Content Strategy.

ويمكن أن تستند أحكام الخبراء على الحدس أو الإحساس العام، كما يمكن أن تستند إلى الاستدلالات المنظمة من نظرية في الشخصية. فالفقرات التي يرون أنها تتعلق بوضوح بما يودون قياسه يمكن تجميعها في استبيان الشخصية. ولعل بعض المجالات العامة تستخدم هذه الاستراتيجية في تكوين بعض الأسئلة المتعلقة بنواحي

شخصية وتضع لها عنواناً مناسباً، مثل «كيف تنجح في حياتك العملية؟» أو «كيف تكون محبوباً من أصدقائك؟» وهكذا. غير أن الاستبيانات المبكرة قد حاولت إسناد الفقرات إلى أساس نظري، مثل قائمة وودورث لبيانات الشخصية، واستبيان برنروتر للشخصية اللذين أشرنا إليهما في مستهل هذا الفصل.

ويمكن إجراء تحليل للفقرات التي اختارها المحكمون للتوصل إلى مجموعة من الفقرات المتعلقة بما يقيسه الاستبيان؛ لذلك فإن الاستبيان الذي يُبنى على هذا الأساس يفتقر إلى الصدق التجريبي، إذ ربما يتميز فقط بصدق ظاهري نظراً لأنه اعتمد على الأحكام الذاتية للقائم ببناء الاستبيان، ومع هذا فإن ذلك يُعد خطوة أولية لالقاء الضوء على تكوينات فرضية معينة في مجال الشخصية.

استراتيجية تستند إلى الاتساق الداخلي للفقرات : وتعتمد هذه الاستراتيجية اعتماداً أساسياً على أسلوب التحليل العاملى Factor Analysis فى انتقاء فقرات تميز بالاتساق الداخلى. وقد اتضح لنا ذلك فى دراسات إيزننك Eysenck، وكاتل Cattell، وجيلفورد Guilford في هذا المجال ، كما سبق أن اتضح لنا ذلك في بناء بعض مقاييس الذكاء والاستعدادات الخاصة والميول. وتهدف هذه الاستراتيجية لانتقاء مجموعة من الفقرات ترتبط فيما بينها ارتباطاً مرتفعاً وترتبط بغيرها منمجموعات الفقرات ارتباطاً منخفضاً. ولعل جيلفورد يُعد أحد الرواد الذين استخدموه أسلوب التحليل العاملى فى بناء استبيانات الشخصية، وأعد استبيان العوامل بالاشتراك مع مارتن (GMI) Guilford-Martin Inventory، كما أعد استبيان مسح السمات المزاجية بالاشتراك مع زمرمان Guilford-Zimmerman Temperament Survey . ويقيس هذا الاستبيان عشر سمات باستخدام ٣٠ فقرة لكل سمة، ووجدوا أن كل مجموعة من هذه الفقرات متسقة داخلياً.

غير أنه يُعاب على أسلوب التحليل العاملى اختلاف العوامل التى يمكن أن تُستخلص من مصفوفات الارتباطات بين المتغيرات بتغيير الفقرات أو عينات الأفراد المستخدمة؛ لذلك نلاحظ أن ثيرستون (Thurstone, 1951) عندما أعاد تحليل بيانات هذا الاستبيان باستخدام عينات أخرى من الأفراد استخلص سبعة عوامل رئيسة بدلاً من عشرة. كما نلاحظ مما سبق أن إيزننك Eysenck استخلص من بياناته عاملين أو بعدين رئيسين، واستخلص كاتل Cattell ١٦ عالماً. كما أن الاستبيانات التى تستند إلى هذه الاستراتيجية تميز فقط بصدق العاملى Factorial Validity الذى يشير إلى الاتساق

الداخلى للفقرات ، ولكنها تفتقر - مثل الاستراتيجية السابقة - إلى الصدق التجربى . وكذلك فائدة مثل هذه الاستبيانات تكون محدودة فى المجال الكlinيکى الذى يتطلب استبيانات تُميز بين مجموعات معينة (Goldberg, 1972).

استراتيجية تستند إلى أساس إمبيريقي : يُطلق على هذه الاستراتيجية أحياناً استراتيجية المجموعات المحكمة Criterion-Keyed Strategy ، أو استراتيجية المجموعات المتناقضة Contrasting Groups Strategy . واستبيانات الشخصية التى تُبنى استناداً إلى هذه الاستراتيجية يعتمد فى انتقاء فقراتها على تميزها بين مجموعتين أو أكثر يختلفان فى خاصية أساسية معينة ، مما يجعل هذه الاستبيانات تتميز بصدق تجربى . وقد سبق أن أوضحتنا ذلك بالتفصيل عند مناقشتنا لاستبيانات الميلو فى الفصل الحادى عشر ، غير أن الفقرات التى يتم انتقاوها استناداً إلى هذه الاستراتيجية لا تكون متجانسة ، بل ربما تتضمن فقرات لا تتعلق بالسمة المراد قياسها مادامت هذه الفقرات تُميز بين مجموعات معينة من الأفراد؛ لذلك تفتقر هذه الاستبيانات إلى أساس نظرى يفسّر درجاتها .

ومع هذا فإنها تتميز بأن الفرد المستجيب يصعب عليه تزوير استجاباته للفقرات ، ولعل استبيان مينيسوتا المتعدد الأوجه (MMPI) ، واستبيان كاليفورنيا للشخصية (CPI) يُعدان من أهم الاستبيانات التى استندت فى بنائها على هذه الاستراتيجية .

ومن هذا يتضح اختلاف الاستراتيجيات الثلاث فى الأساس الذى تُبنى استبيانات الشخصية استناداً إليها ، فلكل منها ميزات وعيوب . ويمكن بناء استبيانات باستخدام أكثر من استراتيجية للإفاده من ميزات كل منها ، إذ يمكن مثلاً انتقاء فقرات تُميز بين أفراد مجموعتين متناقضتين وفي الوقت نفسه تكون متسقة داخلياً . أو يمكن تجميع فقرات على أساس نظري منطقى أو لا وإجراء تحليل عاملى على درجاتها لانتقاء مجموعات الفقرات المتجانسة ، وهكذا .

أمثلة لبعض استبيانات الشخصية :

نظراً لكثره وتنوع استبيانات الشخصية المتوفّرة ، حيث يبلغ عددها عدة مئات ، فسوف نوضح فيما يلى بعض أهم هذه الاستبيانات شائعة الاستخدام ، وسوف نراعى أن تكون مُستندة فى بنائها على كل من الاستراتيجيات الثلاث التي أوضحتناها .

استبيانات تستند إلى أساس نظري منطقي :

اعتمدت كثير من الاستبيانات المبكرة على قياس سمة واحدة من سمات الشخصية سواء كانت محدودة أو متعددة، لذلك تسمى الاستبيانات أحادية البُعد Uni-dimensional Inventories، وتؤدى إلى درجة كلية واحدة للسمة التي تقيسها. ومن أوائل هذه الاستبيانات قائمة وودورث للبيانات الشخصية عام ١٩١٧ ، واستبيان برنرويت للاكتفاء الذاتي عام ١٩٣٣ . واستند بناء كل منها إلى أساس نظري منطقي يتعلّق بمحفوٍ السمة المقاسة. غير أن الاستبيانات الأكثر حداًثة اعتمدت على قياس سمات متعددة للشخصية، واستندت في بناها إلى التحليل العائلي أو إلى أساس إيميريقي أو إلى كليهما معاً. وتؤدى إلى عدد من الدرجات الكلية لمجموعة الفقرات التي تشتمل عليها، لذلك تسمى الاستبيانات متعددة الأبعاد Multidimensional Inventories . وعلى الرغم من اختلاف أساليب بناء نوع الاستبيانات أحادية البُعد، إلا أن بعض الاستبيانات متعددة الأبعاد اعتمدت على الاستبيانات أحادية البُعد، فمثلاً اعتمد استبيان الشخصية متعدد الأبعاد لبرنرويت على كل من الاستبيانين المشار إليهما أعلاه وعلى استبيانات أخرى أحادية البُعد؛ لذلك سوف نوضح فيما يلى بياجاز قائمة وودورث ، ثم نقدم استبيان الشخصية لبرنرويت.

قائمة وودورث للبيانات الشخصية

:Woodworth Personal Data Sheet (WPDS)

تُعد هذه القائمة أول استبيان تقرير ذاتي للشخصية تم بناؤها أثناء الحرب العالمية الأولى، ونشرت في صيغتها النهائية عقب انتهاء هذه الحرب مباشرة. واستندت إليها العديد من الاستبيانات التي أعدّت حديثاً. وكان الهدف منها تحديد الأفراد المجندين الذين يعانون من مشكلات انفعالية تعيق أدائهم لمهامهم العسكرية، وتُعد بدليلاً مناسباً لل مقابلة الشخصية الفردية التي كان يجريها الأطباء النفسيون في ذلك الوقت حيث كان عددهم محدوداً. واعتبرت بمثابة مقابلة شخصية جماعية، ولكن باستخدام الورقة والقلم، وبذلك مهدت السبيل لبناء فقرات أو عبارات تتعلق بالأعراض المرضية من أدبيات الطب النفسي ومن دراسات الحالات التي تعاني من اضطرابات نفسية، ووُضعت في قائمة تشمل على ١١٦ سؤالاً يستجيب الفرد لها إما بـ «نعم» أو بـ «لا» ومن أمثلة هذه الفقرات ما يلى:

هل تتلعم أو يُعقل لسانك أثناء المحادثة؟
هل تُعاني من أحلام اليقظة؟
هل منظر بقع الدم يُشعرك بالدوار؟
هل تشعر عادة بأن صحتك جيدة؟
هل تشعر بالراحة أثناء النوم؟
هل تشعر بالسعادة معظم الوقت؟
هل تشعر بأن الناس يفهمونك ويتعاطفون معك؟

والدرجة التي يحصل عليها الفرد تشير إلى عدد الأعراض التي يشكو منها ، فالفرد الذي يقرر ذاتياً كثيراً من الأعراض يستيقى لمزيد من الفحص الكلينيكي ، أى أن هذه القائمة استُخدمت في التنبؤ بسوء التكيف ، وذلك بتطبيقها على مجموعات كبيرة من الأفراد المجندين .

وعلى الرغم من أن فقرات القائمة اختيرت منطقياً استناداً إلى محتواها ، إلا أن الفقرات التي استجاب لها ٢٥٪ أو أكثر من عينة الأفراد العاديين استجابة تشير إلى أعراض نفسية عصبية تم استبعادها استناداً إلى افتراض أنها لا تقيس أعراضًا لسوء التكيف مادام هؤلاء العاديون استجابوا مثل هذه الاستجابات التي ربما لا تكون كذلك إذا أجرى لهم مقابلة شخصية فعلية ، وبهذا الأسلوب يقل احتمال الأخطاء الموجبة . وعموماً فإن الاستجابات لفقرات هذه القائمة وغيرها من استبيانات الشخصية التي تضمنت فقرات مأخوذة منها تميز فقط بالصدق الظاهري أو صدق المحتوى الذي تميز به كثير من اختبارات القدرات والكتابات ، على الرغم من صعوبة التحقق من ذلك في استبيانات الشخصية كما سنوضح فيما بعد .

استبيان الشخصية لبرنرويتز

:The Bernreuter Personality Inventory (BPI)

يُعد هذا الاستبيان امتداداً لاختبار الاكتفاء الذاتي Self - Sufficiency Test الذي سبق أن أعده برنرويتز عام ١٩٣٣ بعنوان : قائمة التفضيل الشخصي Personal Preference Blank التي تضمنت ٦٠ سؤالاً يجيب عنها الفرد إما بـ «نعم» أو بـ «لا» أو بـ «؟» وتقدم مؤشرات حول مدى اعتماد الفرد على غيره من الأفراد ، فالفرد الذي لا يعتمد على غيره يتسم بالاكتفاء الذاتي . واعتمدت هذه القائمة على تجميع الفقرات

وتحrirها ، وإعداد محك تقدير الدرجات المتعلق بالاستجابات الثلاث ، وتجريبيها على عينة من الأفراد ، وإجراء تحليل الفقرات ، ومراجعة القائمة في ضوء نتائج التحليل ثم إعادة تجسيدها ، وإعداد المعايير المئوية . وبعد اختبار الاكتفاء الذاتي من النوع أحادي البعد ، أما استبيان الشخصية لبرنرويتر فيُعد من المحاولات الأولى لبناء الاستبيانات متعددة الأبعاد . فقد حاول برنرويتر التتحقق من صحة افتراض أن سلوك الفرد في موقف معين ربما يكشف عن سمات شخصية مختلفة ، وعندئذ يمكن تقدير أوزان متباعدة لهذه السمات ، وبذلك يمكن بناء استبيانات تستخدم في تحليل سمات متعددة في آن واحد .

وقد جمع برنرويتر فقرات الاستبيان من عدة استبيانات سابقة ، مثل استبيان السيطرة - الشخص الذي أعده ألبورت Allport ، واستبيان الشخصية لشيرستون Thurstone ، واستبيان الانطواء - الانبساط الذي أعده ليرد Laird ، واختبار الاكتفاء الذاتي الذي أشرنا إليه ، وغيرها من الاستبيانات التي كانت متوازنة . وحاول التتحقق من أن مجموعة الفقرات تقيس كل متغير من المتغيرات التي حددتها بالدقة نفسها التي تقيس بها الاستبيانات الأصلية هذه المتغيرات .

ويشتمل الاستبيان على ١٢٥ سؤالاً يجيب عنها الفرد بـ «نعم» أو بـ «لا» أو بـ «؟» ، وتقيس سمات العصبية ، والاكتفاء الذاتي ، والانطواء - الانبساط ، والسيطرة - الشخص ، والثقة بالنفس ، والميل الاجتماعي . ومن أمثلة هذه الفقرات ما يلى :

هل يتحسن عملك إذا امتدحك أحد؟

هل تميل إلى ممارسة الرياضة أكثر مما تميل إلى الأنشطة العقلية؟

هل تتجنب عادة استشارة غيرك؟

هل تخجل في معظم الأحيان؟

هل تشعر أن الناس من حولك يراقبونك؟

وقد بيَّنت معاملات الارتباط بين الدرجات أنها لا تقيس ست سمات مختلفة للشخصية ، إذ وُجد أن درجات العصبية ترتبط ارتباطاً مرتفعاً بدرجات الانطواء ، ولكن برنرويتر استطاع الحصول على ست درجات وذلك بتعيين أوزان عدديَّة مختلفة للفقرات . وأوضحت نتائج الدراسات التي أجرتها أن قيم معامل ثبات درجات كل مقياس فرعى معقولة بالنسبة للاستبيان ، أما صدقه في قياس السمات الست يعتمد على صدق الاستبيانات الأصلية التي جُمعت منها فقرات الاستبيان ، ونظرًا لأن صدق بعضها

في قياس السمة المتعلقة بها كان منخفضاً، فإن صدق استبيان برنرويتر تطلب المزيد من الدراسات، وقد أعد مفتاح تصحيح خاصاً بكل مقياس، وكرامة تعليمات توضح كيفية التصحيح، وكذلك صحيفة التقرير السيكولوجي.

ومما هو جدير بالذكر أن هناك بعض الاستبيانات الأخرى متعددة الأبعاد استندت في بنائها إلى أسس نظرية منطقية، ولعل أكثرها شيوعاً قائمة التفضيل الشخصي التي أعدتها إدواردرز (EPPS) Edwards Personal Preference Schedule في عام ١٩٥٣ ، ١٩٥٩ . ويُستخدم هذا الاستبيان بكثرة في مراكز الإرشاد النفسي، ونال قدرًا كبيراً من الدراسات والبحوث، وسوف نتناوله بمزيد من التفصيل فيما بعد.

استبيانات تستند إلى التحليل العامل

أشرنا فيما سبق إلى الإطار النظري لمدخل التحليل العامل في بناء الاستبيانات محددة البنية الذي اهتم به كل من أيزنك Eysenck ، وجيلفورد Guilford ، وكاتل Cattell . فاستبيانات الشخصية المتوفرة في الوقت الحاضر تفترض أن الأفراد يتميزون بخصائص أو سمات مستقرة نسبياً ويمكن الكشف عنها في مواقف متعددة، وتختلف هذه الخصائص من فرد إلى آخر، ويمكن قياسها. ولعل هذا الافتراض يتمثل جيداً في الاستبيانات التي يستند بناؤها إلى مدخل أو أسلوب التحليل العامل Factor Analysis . ومما ساعد في إجراء العمليات الحسابية والتحليلات العددية التي يتطلبها استخدام هذا الأسلوب توافر الحاسوب الذي يمكن عن طريق برمجياته المتقدمة تحويل البيانات متعددة المتغيرات المستمدة من عدد كبير من الاستبيانات وغيرها من مقاييس الشخصية .

وقد اعتمد جيلفورد Guilford في بنائه لاستبيانات الشخصية على إيجاد العلاقات بين عدد كبير من الاستبيانات المختلفة، وأجرى تحليلاً عاملاً على هذه العلاقات بغرض التوصل إلى الأبعاد أو العوامل التي تشتهر فيها هذه الاستبيانات، ومن ثم استخدام هذه العوامل المستخلصة في بناء استبيان جديد يقيس السمات الأساسية للشخصية . وقد أعد جيلفورد عدداً من الاستبيانات استناداً إلى هذا المدخل، ثم قام بتحديد العوامل المشتركة بينها واستخدامها في بناء استبيان مسح السمات المزاجية الذي سنوضحه فيما يلى.

كذلك استخدم كاتل Cattell بمعهد قياس الشخصية والقدرات بولاية إلينوي الأمريكية هذا المدخل في بناء استبيان العوامل الستة عشر للشخصية، ولكنه اتخذ توجهاً مختلفاً. فبدلاً من إيجاد العلاقات بين استبيانات متعددة استخدم قائمة ألبورت،

وأودبرت (Allport & Odber, 1936) التي اشتغلت على ١٨٠٠ صفة تتعلق بالسلوك الإنساني، ثم حاول اختزالها إلى عدد أقل من السمات التي تُعبر عن معانى المجموعات المختلفة من هذه الصفات، وتوصل إلى ١٧١ سمة (Cattell, 1957). وطلب من عينة من طلاب الجامعات تقييم زملائهم وأصدقائهم في هذه السمات، وأوجد مصفوفة معاملات الارتباط بينها، ثم أجرى تحليلًا عاملياً على هذه المصفوفة وتوصل إلى ٣٦ عاملًا أطلق عليها كما سبق أن ذكرنا السمات السطحية Surface Traits . وأجرى تحليلًا عاملياً مرة أخرى على هذه العوامل وتوصل إلى ١٦ عاملًا Traits ، Source Traits أساسياً متميزاً يفسر جميع المتغيرات، أطلق عليها السمات المصدرية واستند إلى هذه العوامل الستة عشر في بناء استبيانه الذي سنوضحه بعد قليل. ويؤكّد كاتل أن هذه العوامل هي الأبعاد الرئيسية التي تُعد ضرورية وكافية لوصف وتفسير الفروق الفردية في الشخصية التي أشارت إليها الأدباء السيكولوجية.

استبيان جيلفورد وزمرمان لمسمح السمات المزاجية

: Guilford - Zimmerman Temperament Survey (GZTS)

يشتمل هذا الاستبيان على ٣٠٠ فقرة استمدّت من استبيانات سابقة أعدّها جيلفورد استناداً إلى أسلوب التحليل العاملی. وتنطوي هذه الفقرات على عشرة عوامل يمثل كل منها ٣٠ فقرة مصياغة صياغة موجبة بدلاً من الأسئلة، ويستجيب الفرد لها إما بـ «نعم» أو بـ «لا» أو بـ «؟» وذلك لأن استخدام الأسئلة يزيد من حساسية الفرد في استجاباته للفقرات، أما الفقرات الموجبة أو الإثباتية يجعل الفرد يستجيب مباشرة. ومن أمثلة هذه الفقرات ما يلى :

أنت شخص نشط في جميع الأوقات.

يمكنك أن تفكّر في عذر معقول إذا تطلب الأمر ذلك.

نادرًا ما تُعيّد التفكير في أخطائك الماضية.

تحب أن تبدأ العمل في مشروع جديد بحماس شديد.

والسمات التي يقيسها هذا الاستبيان هي :

(١) النشاط العام (الحيوية مقابل الخمول).

(٢) الضبط (الجدية مقابل عدم الاهتمام).

(٣) السيطرة (القيادة مقابل الخضوع).

(٤) المقدرة الاجتماعية (الصداقات المتعددة مقابل قلة الصداقات).

(٥) الازان الانفعالي (اعتدال المزاج والتفاؤل مقابل تذبذب المزاج والتشاؤم).

(٦) الموضوعية (الواقعية مقابل الحساسية الزائدة).

(٧) علاقات المودة (اللباقة والاحترام مقابل العداوة والرغبة في السيطرة).

(٨) الفكر المتأمل (الازان العقلى مقابل التشتت العقلى).

(٩) العلاقات الشخصية (الثقة بالناس والمؤسسات مقابل الشك والانتقاد).

(١٠) الذكورة (الاهتمام بالأنشطة والمهن المتعلقة بالذكور وكف الانفعالات مقابل الأنشطة والمهن المتعلقة بالإثاث والتعبير عن الانفعالات).

ويمكن باستخدام هذه الاستبيانات الحصول على درجات منفصلة لكل من هذه السمات العشر، وكذلك يمكن رسم صفحة نفسية للفرد. وقد اشتمل الاستبيان أيضاً على ثلاثة مفاتيح تكشف عن تزيف الاستجابات أو تهادن الفرد المستجيب، وأعدَّت معايير لهذا الاستبيان. وقد وُجد أن قيم معاملات الارتباط بين المقاييس الفرعية العشرة منخفضة إلى حد ما فيما عدا عدد قليل منها، فوسط هذه القيم بلغ حوالي ٢٢، ٠، ٠، ٦٩، ٠، وبين كل من مقاييس السيطرة والمقدرة الاجتماعية ٦١، ٠، وبين كل من مقاييس علاقات المودة والعلاقات الشخصية ٥٠، ٠، مما يدل على أن السمات التي يقيسها الاستبيان مترابطة إلى حد ما.

وقد تراوحت قيم معامل ثبات درجات المقاييس الفرعية باستخدام طريقة التجزئة التصفية بين ٧٥، ٠، ٨٧، ٠، وهي تعدَّ قيماً معقولاً بالنسبة لمقاييس السمات المزاجية. وقد أجريت سلسلة من الدراسات للتحقق من الصدق التنبؤي للاستبيان، حيث أوضحت نتائجها أن هناك ارتباطاً منخفضاً موجباً بين ثلاثة من المقاييس الفرعية والمعدل العام لطلاب الجامعات في الولايات المتحدة الأمريكية، كما أوضحت أن تقديرات العمل في المهن التجارية الترويجية ارتبطت ارتباطاً موجباً دالاً بدرجات كل من مقاييس الضبط والفكر المتأمل، بينما أوضحت إحدى الدراسات عدم وجود ارتباطات دالة بين أيٌّ من المقاييس العشرة وتقديرات العمل في المهن الإدارية التنفيذية. غير أن هذا النوع من الصدق كما سبق أن ذكرنا في الفصل الخامس يعتمد

على المحك الذى يتم اختياره ومدى ملائمة موضوعه، وتقديرات العمل تكون فى كثير من الأحيان مشوبة بعنصر الذاتية.

وعموماً تبيّن أن هذا الاستبيان أكثر صدقًا وفائدة إذا استُخدم مع الأفراد العاديين من الطلاب والعاملين في مجالات الصناعة مما لو استُخدم مع مرضى العيادات النفسية (Khan, 1962 ; Watley & Martin, 1962). فهذا الاستبيان لم يضم بفرض التمييز بين الأسواء وغير الأسواء، أو كأدلة لتشخيص الأعراض المرضية. ومع هذا توجد بعض المؤشرات على إمكانية الإفادة منه في هذا الشأن، فقد وجد أن الأفراد الذهانين بعامة Psychotics تكون درجاتهم في كل من مقاييس المقدرة الاجتماعية، والتفكير المتأمل منخفضة (Guilford, 1956).

وقد أجرى بنديج (Bendig, 1960) دراسة عاملية أخرى لدرجات المقاييس الفرعية العشر بعد تطبيقها على أربع عينات من مجموعات عمرية مختلفة، واستطاع استخلاص ثلاثة عوامل متعمدة ومتسمة عبر هذه المجموعات وهي: علاقات المودة، والنشاط الاجتماعي، والابساط - الانطواء.

استبيان العوامل الستة عشر لكتال

: Cattell Sixteen Personality Factor Questionnaire (16 - PF)

يتكون هذا الاستبيان من صيغتين متكافتين (أ) ، (ب) تشتمل كل منها على 187 فقرة، حيث يتراوح عدد الفقرات التي تقيس كلا من هذه العوامل بين ١٠ ، ١٣ ، ١٧ فقرة، ويستجيب الفرد لكل فقرة إما بـ «نعم» أو بـ «أحياناً» أو بـ «لا»، ويطلب منه ألا يفكر كثيراً في الإجابة، وإنما يعطي استجابة سريعة كما لو كان في مواقف حياتية فعلية مماثلة لما تصفها الفقرات، كما يطلب منه تجنب الاستجابة «أحياناً» إلا إذا تعذر عليه الاختيار بين البديلين الآخرين، وألا يترك أي فقرة دون إجابة حتى لو لم تتطبق عليه بدرجة جيدة، وأن يتحرى الصدق في استجاباته قدر الإمكان دون محاولة الاستجابة بما يرضي الفاحص. ومن أمثلة فقرات هذا الاستبيان ما يلى:

أحب مشاهدة الألعاب الرياضية الجماعية.

المال لا يخلق السعادة.

أفضل الناس الذين يستطيعون تكوين صداقات بسرعة.

والعوامل الستة عشر التي يقيسها الاستبيان هي:

- (١) الانعزالية مقابل الانطلاق.
- (٢) البلادة مقابل توقُّد الذكاء.
- (٣) الانفعالية مقابل النضج الانفعالي.
- (٤) الخضوع مقابل السيطرة.
- (٥) السكون مقابل التحمس.
- (٦) العَرَضِيَّة مقابل الشعور بالوجود.
- (٧) التهُبُّ مقابل الجسارة أو الإقدام.
- (٨) الصلابة مقابل الحساسية.
- (٩) الثقة بالآخرين مقابل الشك.
- (١٠) التمسك بالتقاليد مقابل التحرر.
- (١١) البساطة مقابل التعقد.
- (١٢) الثقة بالنفس مقابل عدم الأمان.
- (١٣) التحفظ مقابل التجريب.
- (١٤) الاعتمادية مقابل الاكتفاء الذاتي.
- (١٥) التراخي مقابل ضبط النفس.
- (١٦) الاستقرار الانفعالي مقابل التوتر.

ويحصل الفرد على درجة كلية لكل عامل من هذه العوامل، حيث يفترض أن هذه الدرجات تعطى صورة عن شخصيته بعامة. ويمكن أن يُجرى التصحيح يدوياً أو باستخدام برنامج حاسوب مُعد لهذا الغرض، وكذلك رسم صفحة نفسية لدرجات الفرد. ويمكن الرجوع إلى دليل الاستبيان لمعرفة المعنى السيكلولوجي لكل من هذه العوامل الستة عشر، حيث تُفسَّر العوامل على أنها سمات الشخصية، وهذا الاستبيان يناسب الأفراد من أعمار ١٦ عاماً أو أكثر. وقد أعدَّت معايير للاستبيان (التساعيات المعيارية Stanines) التي سبق أن أوضحتها في الفصل السادس، واستندت إلى عينات كبيرة من طلاب المدارس الثانوية والجامعات. ويُعد الاستبيان نوعاً من المقابلة المقترنة المنظمة دون أن تكون وجهاً لوجه مع الفاحص، وهدفه محدد وواضح بمعنى

أن الفرد المُختبر يعلم أن الاستبيان يقيس الشخصية على الرغم من أن بعض فقراته ربما لا تشير مباشرة إلى ذلك.

وعلى الرغم من الجهد الجاد الذى بذلها كاتل وتعاونوه فى بناء هذا الاستبيان، إلا أنه لا يخلو مثل غيره من الاستبيانات محددة البنية واضحة الهدف من بعض أوجه القصور، ولعل أهمها عدم استقلالية العوامل الستة عشر، وعدم التجانس العاملى لفقرات المقاييس الفرعية، وعدم كفاية المعلومات المتعلقة بالجماعات المرجعية التى استندت إليها المعايير. كما أن قيم معامل ثبات درجات هذه المقاييس باستخدام طريقة التجزئة التصفية كانت منخفضة حيث تراوحت بين .٥٤ ، .٩٣ ، .٠٠ . . وعند ضم الصورتين معاً فى مقياس واحد وتقدير الثبات بطريقة التكافؤ وجد أن قيمة معامل الثبات حوالى .٥٠ ، . . وقيمة الثبات الناتجة من إعادة تطبيق الاستبيان بعد حوالى أسبوع كانت أقل من .٨٠ ، . .

غير أن أدلة الصدق الأساسية اعتمدت على أسلوب بناء الاستبيان باستخدام مدخل التحليل العاملى، فكثير من العوامل الستة عشر تناولت العوامل التى تم التوصل إليها من التقديرات والبيانات التجريبية مما يعزز صدق هذه العوامل. وقد أفادت البيانات المتعلقة بصدق الاستبيان فى اقتراح أفضل طرق استخدام الاستبيان فى مجال الصناعة والمجالات التعليمية والكلينيكية، ولكن البيانات الإحصائية لم تكن كافية بحيث يمكن الإفاده منها فى معرفة ما إذا كانت درجات الاستبيان تنبأ بالفعل بالسلوك المراد قياسه مما يتطلب مزيداً من الدراسات والبحوث فى هذا الشأن.

ومن الجدير بالذكر أن كاتل أعد استبيانات أخرى للشخصية مماثلة لهذا الاستبيان بحيث تُناسب الأعمار الصغيرة، وهى: استبيان الشخصية للمرحلة الثانوية (١٢ - ١٧ عاماً)، واستبيان الشخصية للأطفال (٨ - ١٢ عاماً)، واستبيان الشخصية للمرحلة الدراسية المبكرة (٦ - ٨ أعوام).

كما أعد صيغة قصيرة للاستبيان الأصلى ، وصيغتين متكافئتين للراشدين من ذوى الذكاء أقل من المتوسط، وهذا يدل على أن هذه الاستبيانات تناسب تطبيقات عملية عديدة ومتعددة ، وكذلك تُعد مجالاً خصباً لدراسات والبحوث فى مجال الشخصية. فتتوفر مثل هذه الاستبيانات التى تقىس الشخصية بدءاً من عمر ستة أعوام وما فوق يُعد أمراً ضرورياً للدراسات الطولية المتعلقة بنمو الشخصية إذا افترضنا أن هذه الاستبيانات تميز جميعها بالصدق.

استبيانات تستند إلى أساس إمبيريقي :

أوضحنا فيما سبق أمثلة لاستبيانات تستند في بنائها إلى أساس نظري منطقى يتعلق بمحنوى الفقرات وصدقها الظاهرى فيما يقيسه استبيان الشخصية، كما أوضحنا أمثلة لاستبيانات تستند إلى أسلوب التحليل العاملى الذى يمكننا من انتقاء فقرات متسبة داخلياً أو متجانسة. غير أن هناك اتجاهًا آخر يؤكّد أهمية بناء استبيانات تتميز بصدق تجريبى وذلك بانتقاء الفقرات التي تميز بين مجموعتين من الأفراد أو أكثر. وقد سبق أن أوضحنا في الفصلين التاسع والحادي عشر أن كلاً من مقاييس بینیه للذکاء، واستبيان الميول المهنية لسترونج قد استندا إلى أساس إمبيريقي في انتقاء فقراتهما حيث اعتمدَا في ذلك على مجموعات محكمة Criterion Groups تم اختيارها بطريقة مستقلة عن أي من المقاييس. وسوف نوضح فيما يلى أمثلة لبعض استبيانات الشخصية المهمة التي استندت أيضًا في بنائِها وانتقاء فقراتها إلى هذا المدخل الإمبريقي Emperical Approach الذي لا يفترض إلا عدًّا قليلاً من الافتراضات عند انتقاء فقرات الاستبيان، حيث ينبغي فقط أن تكون هذه الفقرات مرتبطة بمحك أو محکات خارجية مناسبة.

استبيان مينيسوتا للشخصية متعدد الأوجه

: Minnesota Multiphasic Personality Inventory (MMPI)

تُعد مشكلة تشخيص سوء التكيف Maladjustment من المشكلات المهمة التي شجَّعت علماء النفس علىبذل مزيد من الجهد في بناء استبيانات الشخصية. فلقد اتضح لنا أن وودورث Woodworth أعد قائمة البيانات الشخصية لتحديد المشكلات الانفعالية للأفراد المجندين أثناء الحرب العالمية الأولى، وتبعها عدد آخر من الاستبيانات المماثلة. ولعل أهمها جميًعاً استبيان الشخصية متعدد الأوجه الذي يطلق عليه (MMPI) وأحياناً (Mult) فهو يُعد من استبيانات التقرير الذاتي الأكثر شيوعاً واستخداماً، والدليل على ذلك أنه أجرى عليه ما يزيد عن ٧٠٠ بحث ودراسة (Buros, 1988)، وصدرت حوله كتب مرجعية في عدة مجلدات مثل : An MMPI Handbook (Dahlstrom , Welsh & Dahlstrom, 1960 , 1972 , 1975) ، وتزداد الدراسات والبحوث حوله عاماً بعد عام .

ونال أيضًا نصيبياً وافرًا من النقد والتعليق سواء من الأوساط السينكلوجية أو على صفحات الجرائد والمجلات الأمريكية ، ومع هذا فإنه لا يزال يُمثل اتجاهًا أساسياً في بناء استبيانات الشخصية (Reynolds & Sundberg , 1976) . وقد أعدَّ هذا الاستبيان

هاثاواي Hathaway وهو عالم نفس كلينيكي، وماكنلى McKinley وهو عالم في الطب النفسي العصبي بجامعة مينيسوتا الأمريكية عام ١٩٤٢ ، حيث تبينت لهما الحاجة إلى بناء استبيان صادق يمكن استخدامه في التشخيص الكلينيكي لمرضى العيادات النفسية.

ولم يستندا في بنائهما إلى نظريات معينة تتعلق ببنية الشخصية ، وإنما استندتا إلى المدخل الإيميريقي ، فقد بدأ العالمان باختيار ما يقرب من ألف فقرة من بطاقات الفحوص السكلينيكية ، وكتب الطب النفسي ، وأساليب الفحص العصبي ، ومقاييس الاتجاهات الشخصية والاجتماعية التي سبق بناؤها ، ثم قاما بتطبيق هذه الفقرات على حوالي ٢٠٠ من المرضى المصابين بأمراض عصبية نفسية في مستشفيات جامعة مينيسوتا ، وكذلك على حوالي ١٠٠ من الأفراد العاديين من الممرضات وأقارب المرضى ، وعلى عدة مئات من الطلاب المتقدمين للالتحاق بالجامعة ، وعلى عينة من سكان إحدى المدن التابعة لهذه الولاية .

وقد تم انتقاء ٥٥٠ فقرة اشتمل عليها الاستبيان في صيغته النهائية على أساس تمييزها بين هؤلاء المرضى من مثل الفُصاميين والمكتثفين ، وبينهم وبين الأفراد العاديين . وقد زاد عدد هذه الفقرات إلى ٥٦٦ فقرة في الطبعة المعيارية الأكثر شيوعاً واستخداماً التي صدرت عام ١٩٤٧ (Dahlstrom et al., 1972) ، كما صدرت صيغة مختصرة للاستبيان عام ١٩٥٤ اشتملت على ٤٢٠ فقرة فقط (Olson, 1954) ، وتناولت فقرات الاستبيان مدى متسعًا من الموضوعات المتعلقة بجوانب مختلفة للشخصية ، مثل : النواحي الصحية ، والاتجاهات الجنسية والاجتماعية والدينية والسياسية والأسرية والمهنية ، والمخاوف المرضية ، والحالات الانفعالية المختلفة ، والروح المعنوية ، وما يتعلق بالذكورة والأنوثة ، واتجاه المفحوص نحو الاستبيان .

وُصنفت هذه الفقرات في عشرة مقاييس فرعية كلينيكية ، وأربعة مقاييس فرعية تُستخدم في التتحقق من مدى صدق استجابات المفحوص .

والمقاييس الكلينيكية هي :

(١) توهُّم المرض .

(٢) الاكتئاب .

(٣) الهستيريا التحولية .

(٤) الانحراف السيكوباتي .

(٥) الذكورة - الأنوثة .

(٦) البارانويا .

(٧) السيكاثينيا (الوهن النفسي) .

(٨) الفُصام .

(٩) الهوس الخفيف .

(١٠) الانطواء الاجتماعي .

أما مقاييس التحقق من صدق الاستجابات فهى :

(١) المقياس (؟) : وتقدير الدرجة بعدد الفقرات التي لا يستطيع المفحوص الإجابة عنها بـ «نعم» أو بـ «لا»، وينبغي أن تكون هذه الدرجة أقل ما يمكن .

(٢) مقياس الكذب (ل) : يتضمن الفقرات المتعلقة بأمور مقبولة اجتماعيا إلا أنها لا تتطبق على الناس في عالم الواقع، مثل «لا أقول الصدق دائمًا». فعلى الرغم من أن الإجابة المعتادة تكون بـ «نعم» إلا أن الإجابة المقبولة اجتماعيا تكون بـ «لا». وتُسْتَمِد درجة هذا المقياس من مثل هذه الفقرات، فالدرجات التي تزيد عن درجة معينة تشير إلى تعمُّد المفحوص تزيف استجاباته .

(٣) مقياس الشيوع أو عدم الشيوع (ف) : يتضمن الفقرات التي يجب عنها ١٠٪ أو أقل من الأفراد بالصورة التي تُصْحِّح بها، وتعتمد الدرجة على الفقرات التي يُجبِّ عنها المفحوص بحيث يُظهر نفسه بمظاهر غير سوي لأسباب معينة يراها لصالحه أو ربما لعدم فهمه تعليمات الاستبيان ، أو ربما لأن لديه اضطراباً فعلياً .

(٤) مقياس التصحيح (ك) : يعكس الموقف الدافعى للمفحوص فيما يتعلق بالكشف عن مشكلاته، ومحاولته إنكارها لإظهار نفسه بمظاهر سوى، ويبدو أن درجات هذا المقياس تُفيد في زيادة صدق المقاييس الclinician التخديمية السابقة .

ومن أمثلة فقرات هذا الاستبيان ما يلى :

تُلْقِنِي الصُّوْضَاء بِسَهْوَةٍ .

أَعْمَلُ تَحْتَ ضَغْطٍ شَدِيدَةٍ .

نُومِي مُتَقْطَّعٌ وَمُضْطَرِبٌ .

لَا أَقُولُ الصَّدْقَ دَائِمًا .

يَدُوْ أَنَّه لَا يَفْهَمُنِي أَحَدٌ .

أَحَبُّ وَالْدَى .

أَشْعُرُ عَادَةً بِأَنَّ الْحَيَاةَ ذَاتَ قِيمَةٍ .

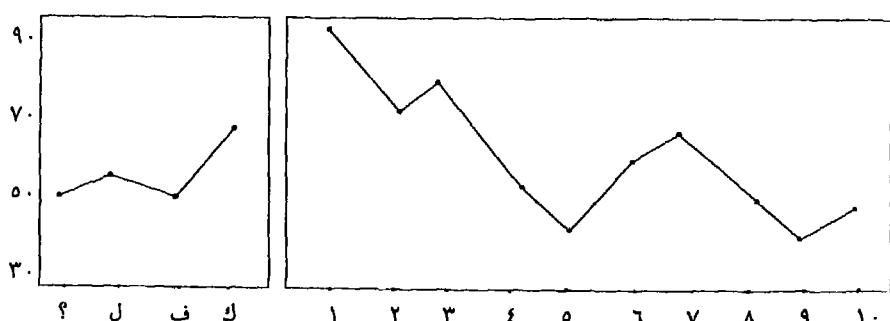
أَشْعُرُ بِالْخُوفِ عِنْدَمَا أَنْظَرَ إِلَى أَسْفَلِ مِنْ مَكَانٍ مُرْتَفَعٍ .

أَشْعُرُ بِالسَّعَادَةِ مُعَظَّمِ الْأَوْقَاتِ .

ويتطلب الاستبيان أن يجيب المفحوص إما بـ «صواب» أو بـ «خطأ» أو بـ «لأنني» ويدون إجاباته على ورقة إجابة خاصة. وبعد تصحيح الاستبيان وتقدير درجات كل مقياس فرعي تحول الدرجات الخام إلى درجات تائية T Scores أو ضمنها في الفصل السادس.

ويتم رسم صفيحة نفسية (بروفيل نفسي) Profile للمفحوص ، مع ملاحظة أنه أصبح لا يفضل استخدام أسماء أو عناوين المقاييس الكlinيكية الفرعية ، وإنما تُستَخدَم الأعداد من (١) إلى (١٠) ، والرموز (؟) ، (ل) ، (ف) ، (ك) لمقاييس الصدق ، وذلك لأن هذه المقاييس klinيكية أصبح لها معنى مختلف عن المعنى الأصلي الذي كان مقصوداً ، فالأرقام تركز على المعنى الواسع للمقياس الفرعي وليس على عنوانه.

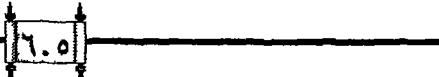
ويوضح شكل (٣-١٣) التالي مثلاً لصفحة نفسية لفرد عصبي :



شكل (٣-١٣) يوضح صفة نفسية لفرد عصبي

المقاييس klinيكية

المقاييس الصدق



ويتبين من شكل (٣-١٣) أن الصفحة النفسية اعتمدت على الدرجات التالية التي متوسطها ٥٠، وانحرافها المعياري ١٠. وكلما كانت إجابات المفحوص غير عادلة أو شاذة ارتفعت الدرجات التالية، فالدرجة التالية التي تزيد عن ٧٠ في أي مقاييس على حدة ربما لا تدل على اضطراب متعلق بما يشير إليه مقاييس فرعى معين. غير أن تفسير الصفحات النفسية لا يُعد أمراً بسيطاً ، حيث يعتمد ذلك على تحليل نمط الارتفاعات والانخفاضات في جميع المقاييس الفرعية Pattern Analysis وليس على الارتفاع المطلق لدرجة كل مقاييس على حدة. فالبروفيل أو النمط يتضح من عدد الدرجات المرتفعة في المقاييس الفرعية. فمثلاً أعلى الدرجات في شكل (٣-١٣) هي درجات المقاييس الفرعية الثلاثة الأولى، لذلك تسمى مقاييس العُصبية Neurotic درجات Triads. ويقارن هذا النمط بالمفحوصين الآخرين من ذوى نمط ١ - ٣ - ٢ ، وكذلك بالدلالات الكlinيكية المتماثلة لهؤلاء المفحوصين. وبالطبع ينبغي أولاً عند تفسير الصفحات النفسية الاهتمام بدرجات مقاييس الصدق الأربع لتعرف مدخل المفحوص في استجاباته لفقرات الاستبيان .

وقد اهتمت كثير من الدراسات ببحث الأنماط المختلفة للدرجات، ودلالتها الكlinيكية، فقد حاولت هذه الدراساتربط بين الأنماط الشائعة أو المشتركة وبين عبارات وصفية تشخيصية، وتصميم برامج حاسوب تفسّر الدلالات الكlinيكية لكل من هذه الأنماط آلّياً (Graham, 1977)، فالتفسير كما ذكرنا ينبغي أن يعتمد على تحليل عدة مقاييس فرعية في آن واحد، مما يُعد عملية معقدة تتطلب استخدام الحاسوب، حيث يُجرى هذا التحليل والتفسير بسرعة ودقة وكلفة منخفضة؛ لذلك تقوم بذلك بعض شركات الحاسوب في الولايات المتحدة الأمريكية، حيث يتم تحليل وتفسير ما يزيد عن عشرة آلاف صفحة نفسية شهرياً في أحدها للمفحوصين بهذا الاستبيان (Butcher & Dahlstrom, 1980) ، في برنامج الحاسوب يرسم الصفحة النفسية، ويدون أسفلها عبارات وصفية تفسّر أنماط الدرجات المتعلقة بالمقاييس الفرعية لكل مفحوص وفقاً لقواعد محددة لتقدير الدرجات وبيانات ترميزية مصممة للبرنامج. وعلى الرغم من أن هذا يُعد توجّهاً جديداً نحو التفسير الموضوعي لدرجات استبيانات الشخصية، إلا أن قبول هذا التوجّه يعتمد على الاقتناع بفكرة الموضوعية في تفسير هذه الاستبيانات، حيث يرى البعض أن هذا التفسير ينبغي أن يستند إلى الأحكام المستبورة للأخصائيين الكlinيكيين .

ويُستخدم الاستبيان في التشخيص الكlinيكي لمن هم في مستوى المرحلة الثانوية وما بعدها، وهو متوافر في صيغته الأجنبية على شكل بطاقات منفصلة مدون على كل

منها إحدى الفقرات لكي يسهل تطبيقه فردياً، كما أنه متوافر في كتيب أو على شرائط تسجيل بعرض التطبيق الجماعي، وكذلك توجد صيغة عربية لهذا الاستبيان، وترجم أيضاً إلى اللغة الأسبانية.

وقد وجَّه بعض النقد إلى هذا الاستبيان لأنخفاض ثبات درجات بعض المقاييس الفرعية التي اشتمل عليها، حيث بيَّنت بعض الدراسات أنَّ قيم معامل الثبات بطريقة إعادة التطبيق على الأفراد العاديين (معامل الاستقرار) تراوحت بين ٠٠٦٠ ، ٠٩٠ . بفضل زمن أسبوع واحد، وتقل هذه القيم إذا زادت المدة الزمنية الفاصلة. وقد برر مؤلفاً الاستبيان انخفاض الثبات بأنَّ الأساليب السيكومترية المتعارف عليها في تقدير الثبات لا تصلح لاستبيانات الشخصية، وحججتها في ذلك أنَّ كثيراً من سمات الشخصية متغيرة، وأنَّ البيانات المستملدة من إعادة التطبيق تُعد مُؤشراً لتبين السمات وليس تقديرًا لثبات الاستبيان، حيث يمكن توقع هذا التباين في السمات لدى من يعانون أمراضاً نفسية، وبخاصة من يتلقون علاجاً.

فقد أوضح بيرجن (Bergin, 1971) أنَّ المقياس الفرعى الشانى (مقاييس الاكتئاب) بخاصة تتغير درجاته عقب تلقى المرضى علاجاً نفسياً فاعلاً، وكذلك وجد أنَّ قيم معامل الثبات بطريقة التجزئة النصفية كانت منخفضة حيث تراوحت بين ٠٠٥ ، ٠٠٩٦ ، ٠٧٠ . ووسيطها ، وهذا يؤثر على تفسير الصفحة النفسية للمفحِّس لتأثيرها بالأخطاء العشوائية وعدم استقرار درجاتها؛ لذلك اقترح بعض الباحثين ضرورة انتقاء فقرات جديدة تحل محل بعض فقرات الاستبيان التي تُسهم في انخفاض الثبات، وذلك في الدراسات المستقبلية.

كذلك أوضحت الدراسات أنَّ صدق الاستبيان كان موضع تساؤل ، وذلك لقلة عدد أفراد عينة المرضى (أقل من ٥٠ فرداً في كل مجموعة كلينيكية) التي استُخدمت في انتقاء فقرات الاستبيان وعدم تمثيلها لهؤلاء المرضى، ولعل ذلك ربما أدى إلى تعين بعض الفقرات في المقاييس الفرعية استناداً إلى عوامل الصدفة (Sax, 1974) . كما أنَّ الصدق التلازمي أي مدى تميز الاستبيان بين المجموعات الكلينيكية والأفراد العاديين لم يكن مُرضياً، وأكَّد ذلك مؤلفاً الاستبيان. فمما لا شك فيه أنَّ بناء الاستبيان استناداً إلى الأساليب الإميريقية يعتمد اعتماداً كبيراً على طبيعة المجموعات المحكمة Criterion Groups ، والتشخيص الكلينيكي لا يكون ثابتاً ويختلف من مستشفى إلى آخر ، ومن عيادة نفسية إلى أخرى ، وذلك لاختلاف الأساس النظري الذي يُستند إليه في التشخيص (Kline, 1976) . غير أنَّ الدراسات والبحوث الكثيرة التي أجريت حول

هذا الاستبيان (ما يزيد عن ٧٠٠ دراسة) تُقدّم دليلاً عن صدق التكوين الفرضي للستبيان، وربما كان هذا أحد أسباب استمرار استخدامه حتى وقتنا الحاضر على الرغم من افتقاره إلى أدلة سيكومترية رصينة تؤيده.

ومن بين أوجه النقد الأخرى تداخل الفقرات في أكثر من مقياس فرعى، أي تكرار كثير من الفقرات في أكثر من مقياس، فمثلاً المقياس الفرعى المتعلق بالفضام (المقياس ٨) يشتمل على ٧٨ فقرة، ١٦ فقرة منها فقط غير متكررة، ولعل هذا قد أدى إلى قيم مرتفعة للارتباط بين المقاييس الclinique الفرعية حيث تراوحت بين ٦٤ ، ٦٤ ، ٨٧ ، بحسب عدد أفراد العينات المستخدمة (Dahlstrom & Welsh , 1960). لذلك اهتمت بعض الدراسات بإجراء تحليل عاملی لهذه الارتباطات، وتوصلت إلى عاملين فقط يفسران تباين درجات المقاييس العشرة، أحدهما عامل عام ربما يتصل بالقلق Anxiety أو الفُضامية Neuroticism، ويتأثر تأثيراً كبيراً بالاتجاه العقلی للمفحوصين من حيث مراعاة القبول الاجتماعي Social Desirability. ولعل الارتباط المرتفع بين درجات المقاييس الفرعية ، وما توصلت إليه دراسات التحليل العاملی يجعل صدق تحليل أنماط الاستجابات في الصفحات النفسية موضوع تساؤل .

كذلك أكدت كثیر من الدراسات (Dahlstrom et al. , 1972; Hathaway & McKinley, 1967) أهمية مراعاة الخصائص الديموغرافية للمفحوصين عند تفسير الصفحات النفسية، مثل العمر ، والنوع ، والموطن ، والتعليم ، والمستوى الاجتماعي الاقتصادي وغير ذلك، فما تدل عليه درجات الاستبيان ترتبط بهذه العوامل ارتباطاً واضحاً، حيث تبيّن أن نمط واحد لأكثر من صفحة نفسية ربما يختلف معناه باختلاف هذه الخصائص .

وكما أوضحنا في الفصل الثاني ، فإن المعايير التي أقرتها الجمعية الأمريكية لعلم النفس (APA) أكدت ضرورة إعادة تقييم الاختبارات والمقياس التربوية والنفسية كل عدة أعوام ، ونظرًا لأن هذا الاستبيان أُعدَّ منذ زمن بعيد ، فإن تغير نوعية الحياة وتعقدها أدى إلى تغير واضح في التعبيرات الانفعالية والسلوك ، وهذا يجعل من الصعب تفسير نتائج بعض المقاييس الفرعية التي يشتمل عليها الاستبيان ، مثل مقياس الذكرة والأئمة ، والفضام ، مما يتطلب تحديث معايير هذه المقاييس ، وكذلك استحداث معايير للمجموعات العصرية المختلفة. لذلك تُنظم الندوات الدولية المتخصصة التي تُعقد كل عام تقريبًا في الولايات المتحدة الأمريكية حول المقاييس النفسية حلقة دراسية تتناول المراجعة الفنية لهذا الاستبيان (Butcher, 1981) .

وعلى الرغم من تعدد أوجه النقد التي وُجّهت إلى هذا الاستبيان، إلا أن هذا لا يقلل من قيمة المعلومات التي يقدمها للأخصائى الكlinيكي، والتي يمكن اعتبارها فروضاً يمكن أن يتحقق منها استناداً إلى مصادر بيانات أخرى. فالاستبيان يتميز بشيوع استخدامه وسهولة تطبيقه، وتصحيحه وتفسير نتائجه في ضوء التقنيات المعاصرة، بل يُعد من أكثر استبيانات الشخصية استخداماً في التشخيص الكlinيكي، والعلاج النفسي، والبحوث السيكلوجية، ويفيد في أغراض الوصف واتخاذ القرارات. كما أثر تأثيراً واضحاً في تطوير إجراءات التقييم الموضوعي للشخصية استناداً إلى الأساليب الإمبريالية بعد ما كان التركيز في الماضي على الأساليب النظرية المنطقية. وتتضمن مقاييس فرعية للتحقق من صدق الاستجابات، وتُعد جزءاً متكاملاً من إجراءات التقييم والتفسير. كما أن نظم ترميز المقاييس الفرعية دلالاتها واستخدام الحاسوب في رسم الصفحات النفسية وتفسيرها يُعد من التطورات التي تسهم في موضوعية التصحيح والتفسير.

ولعل ما يميز هذا الاستبيان أن كثيراً من الاستبيانات الأخرى تم إعدادها باستخدام فقرات أو إجراءات مستمدة منه. ومن أمثلتها استبيان كاليفورنيا السيكلوجي الذي سنوضحه فيما يلى حيث يُطبق على الأفراد العاديين، ومقاييس الفلق الظاهر لتايلور الذي استُخدم أولاً في الأغراض البحثية ثم أصبح يُستخدم في قياس القلق، واستمدت فقراته مباشرة من استبيان الشخصية متعدد الأوجه.

استبيان كاليفورنيا السيكلوجي

:California Psychological Inventory (CPI)

يمكن اعتبار هذا الاستبيان صيغة مماثلة لاستبيان الشخصية متعدد الأوجه (MMPI) الذي ناقشناه ، ولكنه يُطبق على الأفراد العاديين وليس على الحالات المرضية كما في الاستبيان الأول. وقد أعدَّ هذا الاستبيان جوخ H. G. Gough ، ويشتمل على ٤٨٠ فقرة من نوع الصواب أو الخطأ استمد نصفها من الاستبيان الأول، أما النصف الثاني فقد أعد خصيصاً لهذا الاستبيان بحيث يناسب الخصائص الشخصية للمرأهقين والراشدين. ويشتمل الاستبيان على ١٨ مقاييساً فرعياً تقيس أبعاد الشخصية العاديه (السوية) Normal Personality ، وقد صنفت هذه المقاييس في أربعة أقسام هي : فاعلية العلاقات الاجتماعية التفاعلية، الضبط الذاتي، التوجُّه الأكاديمي، الاتجاه نحو الحياة (Gough, 1968) ، ويشتمل الاستبيان على ١٨ مقاييساً فرعياً تتعلق بهذه الأقسام الأربعـة ، وهذه المقاييس هي :

(أ) مقاييس العلاقات الاجتماعية التفاعلية (النفوذ الاجتماعي، والثقة بالذات):

- | | |
|------------------------|----------------------|
| (١) السيطرة | (٤) الحضور الاجتماعي |
| (٢) المكانة الاجتماعية | (٥) تقبل الذات |
| (٣) المقدرة الاجتماعية | (٦) الإحساس بالكيان |

(ب) مقاييس الضبط الذاتي (التنشئة الاجتماعية ، والتضجع ، وتحمل المسؤولية):

- | | |
|------------------------|-----------------------|
| (٧) تحمل المسئولية | (١٠) التحمل |
| (٨) التنشئة الاجتماعية | (١١) الانطباع الجيد |
| (٩) ضبط النفس | (١٢) الانتفاء للجماعة |

(ج) مقاييس التوجُّه الأكاديمي (إمكانات التحصيل ، والكفاءة العقلية):

- | |
|-------------------------------|
| (١٣) الإنجاز عن طريق الامتثال |
| (١٤) الإنجاز المستقل |
| (١٥) الكفاءة العقلية |

(د) الاتجاه نحو الحياة (الأسلوب العقلى ، ونمط الميل):

- | |
|------------------------------|
| (١٦) الاتجاه النفسي - العقلى |
| (١٧) المرونة |
| (١٨) الأنوثة |

ومن أمثلة فقرات هذا الاستبيان ما يلى:

- أحب الشعر .
يتوقع مني الناس الكثير .
حياتي الأسرية سعيدة دائمًا .
أشعر أحياناً بأنني على وشك التمزق .
لم أحارو مطلقاً تعمد الكذب .
أحياناً أكون فضولياً .

وتُستخدم مقاييس الإحساس بالكيان، والانطباع الجيد، والانتماء للجامعة في التحقق من الصدق. وقد أعد المقياسان الأول والثاني من استجابات الأفراد العاديين الذين يطلب منهم تزيف هذه الاستجابات لإظهار أنفسهم بمظهر غير سوي أو سوئ على الترتيب، بينما درجة المقياس الثالث عبارة عن عدد الاستجابات الشائعة.

وقد تم إعداد 11 مقاييساً من بين المقاييس الباقية عن طريق مقارنة استجابات مجموعات مختلفة من الأفراد، بينما المقياس الرابع الآخرى (الحضور الاجتماعي، وقبول الذات، وضبط النفس، والمرونة) أعدَّ استناداً إلى صدق محتواها. أى أن صدق معظم المقياس استند إلى محكّات اجتماعية تتعلق بالحياة الواقعية، وذلك بهدف قياس أبعاد مهمة للشخصية وموائمة لواقع الحياة الاجتماعية. فالمحكّات التي استُخدمت هي: المعدل التصحيلى العام في المقررات الدراسية لطلاب العينة، وتقديرات الأقران، والمشاركة في الأنشطة المصاحبة. ولعل هذا يُعد من أهم ميزات هذا الاستبيان، حيث يؤكد إمكانية استخدام الأسلوب الإمبريقي في قياس الشخصية العادية (السوية).

وقد استُخدمت أيضاً معايير الدرجات التائية لكل مقياس فرعى، وبيّنت نتائج الدراسات أن هناك ارتباطات مرتفعة نسبياً بين هذه المقياس، وكذلك أوضحت الدراسات العاملية أن هناك عاملين يفسران التباين في درجات كل من مقاييس القسم الأول المتعلقة بالعلاقات الاجتماعية، ومقاييس القسم الثاني المتعلقة بالضبط الذاتي. كما أن ثبات الدرجات لا يختلف كثيراً عن استبيان مينيسوتا (Megagee, 1972) حيث تراوحت قيمة باستخدام طريقة إعادة التطبيق بين 57 ، 77 ، 0 . . . ومع هذا فإن الاستبيان يفيد في كثير من المجالات التي تتطلب تقييم الشخصية، بل إن هذا الاستبيان ربما يعد من أفضل استبيانات الشخصية المتوفّرة حالياً، وقد أجريت حوله كثير من الدراسات للتحقق من صدقه، وتمت مراجعته وأعيد تقييّنه عدداً من المرات أكثر من أي استبيان آخر استند إلى أساس إمبريقي.

فمعاييره اعتمدت على استجابات 600 من الذكور، 700 من الإناث من أعمار ومستويات اجتماعية مختلفة، وعلى الرغم من أن صدق كل مقياس فرعى على حدة كان منخفضاً، إلا أنه يمكن استخدام درجات بعض المقياسات الفرعية في معادلة انحدار للتنبؤ ببعض المحكّات، مثل تقديرات الطلاب، والسترب ، وجناح الأحداث وغيرها من المحكّات المناسبة.

استبيانات جبرية الاختيار:

تهدف هذه الاستبيانات لضبط أحد الاتجاهات العقلية Response Set التي تؤثر في استجابات الأفراد في معظم استبيانات التقرير الذاتي وتقلل من صدقها، وهي مراعاة القبول الاجتماعي (SD) Social Desirability. وسوف نعرض فيما يلى مثلا لأهم هذه الاستبيانات التي حاول مؤلفها التغلب على هذه المشكلة باستخدام فقرات جبرية الاختيار Forced Choice Format، وقد سبق أن أوضحنا هذا النوع من الفقرات في استبيانات الميل في الفصل الحادي عشر.

قائمة إدواردرز للتفضيل الشخصي

:Edwards Personal Preference Schedule (EPPS)

يُعد أسلوب الاختيار الجبرى مدخلاً مختلفاً في بناء استبيانات الشخصية، ففى معظم هذه الاستبيانات يتزع الأفراد إلى إظهار أنفسهم بمظهر مرغوب فيه، فعندما تكون هناك استجابة واحدة مقبولة اجتماعياً، فإن الأفراد يختارونها دون الاهتمام بمحظى الفقرة. وقد اهتم إدواردرز بمشكلة مراعاة القبول الاجتماعي (SD) التي تؤثر في دقة درجات استبيانات الشخصية، وحاول التغلب عليها في قائمة التفضيل الشخصي التي اعنى عناية كبيرة ببناء فقراتها، وذلك بأن يكون الفرد مُجبراً على أن يختار إحدى فقرتين متكافئتين في درجة القبول الاجتماعي (Edwards, 1959). وتُعد هذه القائمة من القوائم المبكرة التي صُممَت لتقييم شدة الحاجات Needs التي افترضها موارىExplorations (Murray, 1938) بجامعة هارفارد في كتابه «استكشافات في الشخصية» in Personality حيث وضح فيه نظاماً للحاجات يشتمل على 15 حاجة شخصية Personality Needs مستقلة نسبياً. وقد قام إدواردرز ببناء مجموعات من الفقرات يتعلق محتواها بهذه الحاجات. وتشتمل القائمة على 225 زوجاً من الفقرات استناداً إلى نظرية موارى في الشخصية، وبذلك استطاع إدواردرز تجنب المداخل الذاتية التي لا تستند إلى إطار نظري في بناء الفقرات لكي تتميز بصدق المحتوى. وقد حاول تقدير كل فقرة من حيث «مراعاة القبول الاجتماعي» وكُون منها ثنايات متوازية في هذا الشأن، ويُطلب من الفرد أن يختار من بين كل زوج الفقرة التي تنطبق عليه، أي تُعبر عن حاجاته الشخصية بدرجة أفضل، وبذلك لا يستطيع تزييف استجاباته.

فالفرد يكون مُجبراً على الاختيار بين حاجتين Needs ، أي يختار واحدة على حساب الأخرى في ضوء محتوى كل منها ، وبذلك يمكن تمثيل انتقاء الفقرات على أحد المقاييس بالنسبة لانتلاقتها على مقياس آخر، مما يؤدي إلى درجات تُنسب إلى إطار مرجعى فردى Ipsative Scores التي سبق أن أوضحتها بالتفصيل في الفصل الحادى عشر. أي أن شدة كل من هذه الحاجات Needs لا يُعبر عنها بصورة مطلقة، وإنما في علاقتها بشدة حاجات الفرد الأخرى. فالتركيز هنا يكون على الفرد ذاته وليس على معايير الجماعة المرجعية؛ لذلك يمكن أن يحصل فردان على درجات متساوية في الاستبيان، ومع هذا ربما يختلفان اختلافاً واضحًا في شدة حاجاتهما. وتحويل هذه الدرجات الفردية إلى معايير مئنية ربما لا يكون مناسباً، بل وربما يؤدي إلى صعوبات في تفسير وتحليل الدرجات إحصائياً.

والحاجات الشخصية التي تتضمنها القائمة هي:

- (١) الإنجاز: تخطي الصعب ، بذل أقصى جهد، إنجاز شيء متميز.
- (٢) الانقياد: الامتثال بدلاً من القيادة، تقبل الآخرين ومدحهم، المسيرة.
- (٣) النظام: الحاجة للترتيب، التنظيم، التخطيط المسبق، المدخل المنظم.
- (٤) الاستعراض : الحاجة إلى تركيز انتباه الآخرين نحوه، وتحقيق ذلك عن طريق المظهر الخارجي أو استخدام العبارات اللفظية.
- (٥) الاستقلال الذاتي: تجنب الالتزامات المفروضة، عدم التقيد، الاستقلالية، التحرر.
- (٦) الانتقام: الولاء، الإخلاص، تقديم العون، الألفة.
- (٧) التدخل الاجتماعي: تحليل سلوك ودوافع الآخرين، الفهم من خلال تحليل الذات والآخرين.
- (٨) تقديم العون: تشجيع الآخرين وتعاونهم، تعاطف الآخرين في أوقات الشدة.
- (٩) السيطرة: إدارة الحوار، الإشراف والريادة، التأثير، اتخاذ القرارات للآخرين.

(١٠) الاستكناة: الشعور بالنقض، الحاجة إلى العقاب، الحاجة إلى الاعتراف بالخطأ.

(١١) الحاجة إلى التربية الأساسية: معاونة الأصدقاء والمحاجين، الصفح والتسامح، المودة والكرم والتعاطف.

(١٢) التغيير: أداء أنشطة جديدة و مختلفة ، التجريب ، القيام برحلات، تجنب الأعمال الروتينية، تكوين صداقات جديدة.

(١٣) التحمل: العمل بجد، إنهاء الأعمال المطلوبة، عدم التشتت.

(١٤) الجنسية الغيرية: المشاركة في جميع أنشطة الجنس الآخر.

(١٥) العدوانية: المخالفة، النقد المفتوح، الانتقام ، التوبيخ .

ومن أمثلة فقرات هذه القائمة ما يلى :

(١) (أ) أحب أن أعاون أصدقائي عندما تعرّضهم مشكلة.

(ب) أحب بذل أقصى جهدى في أي عمل يوكل إلىـ.

(٢) (أ) أحب أن أوضح ما يتراوّى لى عن الأشياء.

(ب) أحب أن أصفح عن أصدقائي الذين يغضبوننى أحياناً.

(٣) (أ) أشعر بالانقباض إذا فشلت في موقف معين.

(ب) أشعر بالتوتر إذا تحدثت أمام مجموعة من الأفراد.

وإلى جانب هذه المقاييس الفرعية الخمسة عشر قدم إدواردر مقياساً لاستقرار الصفحة النفسية، ومقاييساً لاتساق الاستبيان يشتمل على ١٥ زوجاً من الفقرات المتطابقة في شكلها وموزعة عشوائياً في الاستبيان، وبذلك يكون عدد مرات الاختيار المتطابق لكل فرد يمكن تحويله إلى مئينيات استناداً إلى بيانات معيارية ، كما يمكن تحليل اتساق استجابات الفرد بإيجاد الارتباط بين درجاته في الفقرات الزوجية والفقرات الفردية لكل مقياس فرعي .

وقد استمدت معايير القائمة من ١٥٠٠ طالب وطالبة بالجامعات ، ٩٠٠ من الراشدين من ٤٨ ولاية أمريكية، وأعدّت معايير مئينية منفصلة لكل من هاتين المجموعتين، وكذلك لطلاب المدارس الثانوية .

وبيّنت الدراسات أن قيم ثبات درجات المقاييس الفرعية وذلك بإعادة تطبيق الاستبيان بفواصل زمني قدره أسبوع واحد تراوحت بين ٧٤ ، ٨٨ ، ٠ ، كما هو موضح بدليل القائمة، وقيم الثبات بطريقة التجزئة النصفية تراوحت بين ٦٠ ، ٠ ، ٨٧ ، ٠ . أما الارتباطات بين المقاييس الفرعية فقد تراوحت قيمها بين ٣٤- ٤٦ ، ٠ ، وهى قيم أقل من نظيرتها فى استبيان كاتل أو استبيان مينيسوتا.

وعلى الرغم من اهتمام إدواردر بضبط متغير مراعاة القبول الاجتماعي في انتقاء لفقرات القائمة، واستناده إلى العلاقة المنطقية بين محتوى كل فقرة وبين تقسيم الحاجات الشخصية الذى افترضه موراي كمحك وحيد لهذا الانتقاء، فإن هذا يؤكّد تركيزه على ضبط هذا الاتجاه العقلى على حساب الجوانب الأخرى لصدق القائمة. والدليل على ذلك أن الدراسات المتفقة التى حاولت التحقق من الصدق يصعب تفسير نتائجها نظراً لأنها لم تأخذ بعين الاعتبار النظام المرجعى الفردى الذى تستند إليه درجات القائمة، فمتوسط قيم الارتباطات بين المقاييس الفرعية كان سالباً، ومتوسط ارتباط جميع هذه المقاييس بأى متغير خارجي تقترب قيمته من الصفر (Hicks, 1970) مما يجعل استخدام معامل الارتباط فى التتحقق من صدق القائمة لا يكون مناسباً.

ومن بين أوجه النقد الأخرى التى وجّهت لهذه القائمة ضرورة التمييز بين مراعاة القبول الاجتماعي ومراعاة القبول الشخصى. كما أنه ربما لا يكون هناك اتفاق بين قيم مراعاة القبول الاجتماعى التى تحصل عليها باستخدام كل فقرة على حدة بالقيم التى تحصل عليها من ثانيات الفقرات، فمراعاة القبول الاجتماعى للفقرة الواحدة ربما تتأثر بالمضمون الذى تعبّر عنه. وللحقيق من هذا الافتراض تم الحصول على قيم جديدة لمتغير مراعاة القبول الاجتماعى لثلاثين زوجاً من هذه الفقرات، وتبين أنها مختلفة اختلافاً ملحوظاً عن القيم الأصلية التى سبق تحديدها، والتى تم مزاوجة الفقرات استناداً إليها (Corah et al. , 1958) . كما أنه بترت تساؤلات تتعلق بما إذا كان استخدام فقرات جبرية الاختيار قد أدت بالفعل إلى ضبط متغير مراعاة القبول الاجتماعى (Feldman & Corah, 1960 ; Wiggins 1966) . وقد اقترحت أنسناري Anastasi, 1976) ضرورة مراجعة هذه القائمة للتغلب على أوجه قصورها الفنية وبخاصة نمط الفقرات وتفسير الدرجات، وكذلك إجراء دراسات صدق تعتمد على تحليل نمط الدرجات التى تستند إلى إطار مرجعى فردى.

ومن الجدير بالذكر أن إدواردر أعد مقاييساً آخر أكثر طموحاً من هذه القائمة أطلق عليه استبيان الشخصية لإدواردر (EPI) Edwards Personality Inventory ، ويشتمل على ٣٠٠ فقرة مدونة فى خمسة كتيبات ، واستخدم أسلوب التحليل العاملى

وتقديرات متغير مراعاة القبول الاجتماعي في بناء الاستبيان حيث يقيس ٥٣ بعدها من أبعاد الشخصية، بمعدل ١١ إلى ١٥ بعدها في كل كتيب، وتتطلب الاستجابات أن تكون إما بـ «نعم» أو بـ «؟» أو بـ «لا»، وتمت صياغة الفقرات بحيث تصف الآخرين ويدى الفرد رأيه حول ما تصفه الفقرة عن نفسه لزيادة صدق الاستجابات. ويستغرق تطبيق كل كتيب ٤٥ دقيقة تقريباً، وإجمالي زمن تطبيق الاستبيان حوالي ثلث ساعات؛ لذلك كانت عينة المعايير متغيرة وغير مماثلة حيث اقتصرت على ٢٠ طالب، ٣٢٩ طالبة بجامعة واشنطن وعدد أقل من بعض المدارس الثانوية، وُجُدَّ أن قيمة الثبات باستخدام طريقة كيودر - ريتشاردسون .٨٠، تقريراً، ولكن لا توجد دراسات كافية تتعلق بصدق الاستبيان.

تقييم استبيانات الشخصية:

بعد عرضنا الموجز للعديد من الأمثلة لاستبيانات الشخصية التي تستند في بنائها إلى أساليب مختلفة وتقيس سمات متعددة ومتباعدة، وستستخدم في أغراض متنوعة، سوف نوضح فيما يلى بعض الملاحظات العامة على هذه الاستبيانات وما يتعلق بها من مشكلات :

(١) **تعدد المعانى التى تشير إليها الفقرات:** تشمل استبيانات الشخصية على فقرات محددة البنية يصوغها المؤلف في عبارات معينة، وهنا تبرز المشكلة اللغوية في صياغة العبارات وما تحمله من معنى ربما يختلف من فرد مستجيب إلى آخر، أو يختلف عما يقصده المؤلف. فمثلاً إحدى عبارات استبيان الشخصية متعدد الأوجه كالتالى: «والدى كان رجلاً طيباً» ، وعندما يجيب الفرد عن هذه الفقرة سوف يتبين له أنها تناسب الأفراد الذين توفى والدهم، أو ربما يفسرها على أن والده ليس كذلك.

وعبرة مثل «هل عادة تُوجّهُ المخوار في المواقف الجماعية؟» فالفرد عليه أن يقرر ما تعنيه المواقف الجماعية ، وهل هي مواقف أسرية أم مواقف في العمل أم مع الأصدقاء؟ وهل تتعلق بمواقف رسمية أم غير رسمية؟ وهل المقصود بكلمة «تُوجّه» المبادأة بالحديث، أو التحدث معظم الوقت، أو التعليق على المخوار؟ وهل المقصود بكلمة «عادة» معظم الأحيان، أم في جميع الأحيان. أم دائمًا؟ فصياغة عبارات استبيانات الشخصية تعد أكثر صعوبة من صياغة فقرات مقاييس الجوانب المعرفية أو المهارية، وتحتمل تأويلات وتفسيرات متباعدة من جانب المختبرين.

(٢) أقسام الاستجابات : يستجيب الفرد لفقرات الاستبيان إما بـ «نعم» ، أو بـ «لا» ، أو بـ «لا» ، وهذا النمط المقيد لأنواع الاستجابة لا يسمح بوصف الإطار المرجعي الذي استند إليه الفرد في استجابته . وأحياناً يضطر الفرد لاختيار الاستجابة «نعم» على الرغم من أنها لا تصف على وجه التحديد ما يراه الفرد عن نفسه ، لذلك يلجأ بعض الأفراد إلى تبرير استجاباتهم في عبارات لفظية يكتبها على هوا من ورقة إيجابية استبيان الشخصية .

(٣) عدم معرفة الفرد لدراوئه وسلوكيه : تفترض استبيانات الشخصية أن الفرد يعرف جيداً دوافعه وسلوكيه الماضي والحاضر بحيث يمكنه أن يستجيب بصدق لفقرات الاستبيان . غير أن هذا ربما يعتمد جزئياً على مدى اضطرابات الشخصية ، فكلما كان الفرد يعاني من اضطرابات مرضية شديدة تكون استجاباته المتعلقة بسماته الشخصية غير دقيقة ، نظراً لأنه لا يكون واعياً بدرجة كافية بدوافعه وسلوكيه . كما أن الاستجابات لبعض فقرات الاستبيان ربما تتطلب معلومات لم يلاحظها الفرد من قبل ، أو ربما تتطلب منه إصدار أحكام أو تفسيرات تفوق قدراته ، وهنا يتطلب التمييز بين التفسيرات أو الاستدلالات وبين الحقائق . ففقرة مثل : «هل تفضل العمل الفردي على العمل الجماعي؟» يستطيع أي فرد عادي أن يُجيب عنها بيسر وسهولة ، بينما تُوجد فقرات تتطلب من الفرد اتخاذ قرار معين يستند إلى فحص ذاتي لم يتحقق منه سابقاً لتفسير سلوكه ، فالفقرات يكون لها مضمون ظاهر ومضمون كامن ، وهذا المضمون الأخير هو الذي يعد صادقاً . فالفقرة «هل كرهت والدك أو والدتك في وقت ما؟» يجيب عنها الفرد عادة بـ «لا» مع أن ذلك ربما لا يكون صحيحاً بالنسبة له ، ولكنها ربما تكشف عن شيء يتعلق بسمة أو أكثر من سمات شخصيته . غير أنه ليس ضرورياً أن تتطلب فقرات استبيانات الشخصية معرفة الفرد لنفسه ، ففقرات مثل «ذهب إلى المسجد بانتظام» و«أميل إلى قراءة المجالات العلمية» تتعلق بحقائق عن الفرد يسهل عليه ملاحظتها ، ومع هذا ربما لا يكون الفرد صادقاً في استجاباته لهذه الفقرات . فالاتجاهات العقلية تؤثر تأثيراً لا يمكن إغفاله في استجابات الأفراد لفقرات هذه الاستبيانات .

(٤) الاتجاهات العقلية في الاستجابات : يتضح تأثير الاتجاهات العقلية ؛ أي تُميز الاستجابات أو تزييفها Response Sets بدرجة أكثر حدة في استبيانات الشخصية عن غيرها من الاختبارات والمقياس ، فعلى الرغم من أن مصممي الاستبيانات يُعنون عناية كبيرة بصياغة فقراتها صياغة دقيقة وواضحة لا تحتمل سوء التفسير ، وب بحيث

تقس جوانب يمكن ملاحظتها وتعلق بحقائق عن حياة الفرد، إلا أن مشكلة تزيف المختبرين لاستجاباتهم تُعد من المشكلات الأساسية في استبيانات الشخصية. وقد اهتمت كثيرون من الدراسات بتعرف الدوافع التي تجعل الأفراد يلجأون لذلك، وعلى الرغم من الأبحاث المستفيضة التي اهتمت بضبط تحيز الاستجابات أو تزيفها، إلا أنها لم تستطع استبعاد هذه المتغيرات الداخلية من أي استبيان للشخصية.

فالفرد ربما يلجأ إلى تزيف استجاباته ليظهر بمظهر الشخص السوى لكنه لا يتم حجزه في مستشفى الأمراض النفسية مثلاً، أو ليقنع المسؤول عن العمل بأنه شخص مناسب لوظيفة معينة، وعلى العكس ربما يلجأ الفرد إلى تزيف استجاباته ليظهر بمظهر الشخص غير السوى للتهرب من الجندي، أو من جريمة ارتكبها، أو ليحظى بمزيد من التعاطف من جانب أفراد أسرته وأصدقائه، ومع هذا فإن تزيف الاستجابات ربما يكون غير متعمد أي لا شعوري. لذلك لعلنا لاحظنا فيما سبق أن بعض الاستبيانات حاولت مراعاة هذه المتغيرات الداخلية، وذلك بتصميم مقاييس فرعية تساعد في كشف تأثيرها، مثل مقاييس (ك) في استبيان مينيسوتا المتعدد الأوجه (MMPI) الذي يساعد في تقييم التزععات الشعورية وغير الشعورية للأفراد الذين يلتجأون إلى إظهار أنفسهم بصورة معينة، وبالتالي يمكن ضبط أو تحديد أثر هذه التزععات، وكذلك يساعد في زيادة فاعلية الاستبيان في التمييز بين الأفراد العاديين وغير العاديين من مرضى المستشفيات النفسية.

وقد سبق أن ناقشتنا بالتفصيل الأنماط المختلفة لهذه الاتجاهات العقلية، مثل مراعاة القبول الاجتماعي، والإذعان، وتطور الاستجابات، وغيرها في الفصلين السابقين. ولعل البحث عن أشكال من الاستبيانات تختلف عن تلك المتوفرة في وقتنا الحاضر ربما يسهم في التغلب بدرجة أفضل في استبعاد أثر هذه الاتجاهات العقلية التي تؤثر في صدق تفسير الدرجات.

فقد تبين لنا أن الاستبيانات التي تشتمل على فقرات جبرية الاختيار - Forced Choice لم تتغلب على هذه المشكلة، بالإضافة إلى صعوبات تفسير الدرجات التي تستند إلى إطار مرجعى فردى Ipsative Scores. وقد اقترح جولد برج Goldberg شكلاً جديداً من أشكال الفقرات التي تتطلب من المُختبر ما إذا كان قد قام بشاط معين في الماضي (مثل تسلق الجبال، أو التأثير على المعلمين) مثلاً؟ وما رأيه في هذا النشاط (هل يحبه أم لا يحبه)؟ وما احتمال قيامه بهذا النشاط مستقبلاً؟ ويرى أن هذا الشكل من الفقرات يزودنا ببيانات متنوعة باستخدام مثير أو عبارة واحدة ، كما يسمح بتحليل أنماط الاستجابات مثل، «قمت بهذا النشاط ، وأحبه ، ولكنني لن أقوم به في

المستقبل» أو «قمت بهذا النشاط، ولم أحبه ، ولكنني أود أن أقوم به في المستقبل»، وغير ذلك من أنماط الاستجابات.

(٥) **زمن تطبيق الاستبيانات**، تُعد الشخصية من الشخصيات المركبة من مكونات متعددة يتطلب قياسها عدة مقاييس فرعية يشتمل عليها الاستبيان، ولكن تميز الدرجات بالثبات ينبغي أن يشتمل كلاً من هذه المقاييس الفرعية على عدد مناسب من الفقرات المتعلقة به لكي تُمثل النطاق السلوكي للسمة المراد قياسها.

لذلك نلاحظ أن استبيانات الشخصية تشتمل على عدد من الفقرات أكبر بدرجة واضحة من عدد الفقرات التي تشتمل عليها معظم المقاييس الجماعية للذكاء أو الاستعدادات الخاصة. فاستبيان جيلفورد وزمرمان يشتمل على ٣٠٠ فقرة، واستبيان كاتل يشتمل على ١٨٧ فقرة، واستبيان مينيسوتا يشتمل على ٥٦٦ فقرة، واستبيان إدواردز يشتمل على ٢٢٥ فقرة، مما يتطلب وقتاً كبيراً في قراءتها، وبخاصة إذا كان الفرد المستجيب ضعيفاً في القراءة، بالإضافة إلى الزمن اللازم لتفسير ما تعنيه كل فقرة من فقرات الاستبيان كما ذكرنا في البند الأول، وبخاصة إذا استند بناء السفرات إلى أساس اميريقي حيث لا ترتبط الفقرات ارتباطاً مباشرًا بالسمة المراد قياسها. وتزداد حدة هذه المشكلة إذا تحرّى الفرد الصدق في استجاباته، ونظر إلى نفسه نظرة تحليلية لكي يختار الاستجابة التي تنطبق عليه أو تصف سلوكه.

(٦) **تعدد أسباب أنماط الاستجابات المتماثلة**، يجب بعض الأفراد عن فقرة معينة من فقرات الاستبيان بد «نعم» لأسباب ربما تكون مختلفة، فإذا استجاب فردان بد «نعم» لفقرة مثل: «أحب العمل الجماعي» فإن الأول ربما يشعر بالأمان بين أفراد الجماعة، بينما يود الآخر تكوين صداقات، وبذلك يمكن لفردين ببنية شخصيهما مختلفة أن يستجيباً استجابة واحدة لفقرة معينة، وهذا يجعل من الصعب التمييز بين أنماط الشخصية المختلفة.

(٧) **تجمّعات السمات**: يتطلب تفسير استجابات الفرد أن يكون هناك تجمع عام من السمات التي تقيسها فقرات الاستبيان. وعلى الرغم من محاولات جيلفورد، وكاتل، وأيزنك وغيرهم من المشغلين بالتحليل العامل تحديد أبعاد أو عوامل للشخصية، إلا أن هذه المحاولات لم تتوصل إلى اتفاق حول هذه العوامل، على عكس التائج المشرمة التي أمكن التوصل إليها في مجال القدرات العقلية، والميول، والاتجاهات. وربما يرجع ذلك إلى أن أبعاد أو عوامل الشخصية تتفاعل مع المواقف المختلفة بطرق متباينة، فالفرد المسيطر في العمل ربما يكون منقاداً في المترزل، أو مع أصدقائه

المقربين، فأنماط السمات الاجتماعية لدى الأفراد تعتمد على خصائص الموقف الذي يواجهه الفرد.

(٨) **تعدد أساليب بناء استبيانات الشخصية:** اتضح لنا مما سبق تعدد الأساليب المختلفة التي يستند إليها بناء استبيانات الشخصية. ولعل هذا يدعو إلى التساؤل حول أكثر هذه الأساليب فاعلية في هذا الشأن. فهل الأساليب النظرية المنطقية ذات جدوى؟ وهل الأساليب الإحصائية المتعلقة بالتحليل العاملی أكثر فاعلية من الأساليب الإمبريالية؟

والحقيقة أن إجابة هذا التساؤل تعتمد على معرفة الغرض الذي يُعنى من أجله استبيان الشخصية لكي يكون أكثر فاعلية. فمما لا يدعو إلى الشك أن الاستبيانات التي تستند في بنائها على الأساليب الإمبريالية، مثل استبيان الشخصية متعدد الأوجه (MMPI) تكون أكثر استخداماً في المجالات التطبيقية، وتهتم كثيراً من الدراسات والبحوث السيكولوجية بإجراء دراسات حولها، وهذه النتيجة تظل قائمة على الرغم من أن الاستبيانات التي تستند في بنائها إلى أسلوب التحليل العاملی ينبغي أن تتميز نظرياً على الأقل بمعزى الاستبيانات التي تستند إلى الأساليب الإمبريالية، بالإضافة إلى أنها يمكن أن تزودنا بدرجات يسهل تفسيرها (Pervin, 1975).

وربما يرجع عدم تحقق ذلك إلى أن الاستبيانات المتوفّرة التي تستند إلى التحليل العاملی لا تقيس جميع أبعاد أو عوامل الشخصية. وبالطبع فإن أداة القياس المثالیة يجب أن تشتمل على فقرات يتم انتقاءها استناداً إلى الأساليب الثلاثة.

(٩) **ثبات درجات استبيانات الشخصية:** هناك العديد من الأدلة التي تُبيّن أن ثبات درجات استبيانات الشخصية ، سواء معامل الاستقرار أو التكافؤ أو الاتساق الداخلي ، أقل بعامة من ثبات درجات غيرها من المقاييس ، وبخاصة مقاييس الجوانب المعرفية . ولكن هل هذا الانخفاض يرجع إلى طبيعة السمات المقاسة أم إلى عملية القياس ذاتها؟ فالشخصية بعامة لا تُعد خاصية مستقرة بدرجة مماثلة لبعض الخصائص الأخرى . فمزاج الفرد يختلف باختلاف المواقف ومن يوم إلى آخر ، وكذلك يميل السلوك إلى التغير من ظرف إلى آخر ، كما أن خصائص استبيانات الشخصية تُسهم في خفض ثبات درجاتها .

بعض المشكلات التي أشرنا إليها في البنود السابقة ، مثل عمومية صياغة فقرات الاستبيان ، واختلاف تفسيراتها من فرد إلى آخر يجعل استجابات الأفراد غير متسقة مما يؤثر في ثبات الدرجات .

كما أن ثبات أو اتساق الصفحات النفسية (البروفيل النفسي) الذي تزودنا به بعض هذه الاستبيانات يعتمد على عدد فقرات كل مقياس فرعي، حيث إننا نهتم باتساق درجات كل من هذه المقياسات على حدة. فالمقياس الفرعية القصيرة تؤدي إلى قيم منخفضة لمعامل الثبات.

ونظراً لأننا نتوقع تغيرات معينة في الشخصية نتيجة الخبرة، ونقلب الحالة المزاجية، فإننا يجب أن نتوقع وبالتالي انخفاض قيم معامل استقرار الدرجات عبر الزمن. ويمكن عندئذ استخدام طريقة التجزئة النصفية أو طريقة كيودر - ريتشاردسون التي تتطلب كل منها تطبيق الاستبيان مرة واحدة لتقدير قيمة معامل اتساق الدرجات، ولكن ينبغي أن تكون السمات أحادية البعد. وقد أوضحت الأدلة أن استبيان الشخصية الذي يعنى بنائه يمكن أن يحقق درجة مرتفعة من الاتساق الداخلي.

(١٠) **صدق استبيانات الشخصية:** على الرغم من أن الثبات يُعد من مشكلات استبيانات الشخصية، إلا أن الصدق يعد مشكلة أكثر تعقيداً، فكثير من هذه الاستبيانات تفتقر إلى دراسات تتعلق بجوانب الصدق المختلفة، فيما عدا مؤشرات عن صدقها الظاهري، أي تعلق الفقرات بالسمة المراد قياسها. غير أن هذه المؤشرات ليست فقط غير كافية للتحقق من صدق الاستبيان، بل إنها ربما تؤثر في الصدق تأثيراً سلبياً، فكلما بدت الفقرات مماثلة للسمة المراد قياسها أدى ذلك إلى تزييف المختبرين لاستجاباتهم (Nunnally, 1970). وقد سبق أن أوضحنا أنه يمكن التتحقق من الصدق التنبؤ لمقياس الجوانب المعرفية، وذلك باختيار محکات خارجية مناسبة، غير أنه يصعب التوصل إلى مثل هذه المحکات التي يمكن التنبؤ بها باستخدام استبيانات الشخصية.

فمثلاً ما الذي نستطيع التنبؤ به من سمة أو عامل «البساطة» مقابل «التعقد» في استبيان كاتل، أو عامل «الموضوعية» أو «العلاقات المودة» في استبيانات جيلفورد وزمرمان ، فدرجات مثل هذه الاستبيانات يمكن أن ترتبط بمتغيرات أو محکات أياً كانت. غير أنه إذا استُخدمت الأساليب الإمبريقية في بناء الاستبيانات، فإن تفسير الدرجات لا يُصبح مشكلة أساسية. فنحن لا نهتم في هذه الحالة بأسباب اختلاف استجابات مرضى العيادات النفسية عن استجابات الأفراد العاديين مثلاً، وذلك لأننا

نلاحظ فقط أن هناك فروقاً ذات دلالة إحصائية بين استجابات كل من المجموعتين، أما أسباب اختلاف الاستجابات فهذا من مهام الأخصائي الكlinيكي، ولكنه لا يهم أخصائي القياس النفسي الذي يحاول التتحقق من صدق الاستبيان.

وقد استندت بعض استبيانات الشخصية في التتحقق من صدقها إلى الأساليب الارتباطية، وذلك بإيجاد الارتباط بين درجات الاستبيان ودرجات استبيان آخر مماثل له يتميز بالصدق، كما أن بعضاً آخر استند في ذلك إلى أسلوب التحليل العاملی. فقد رأينا فيما سبق أن استبيان العوامل الستة عشر لكتال استند إلى أحد أساليب التحليل العاملی، بينما استند استبيان جيلفورد وزمرمان إلى نتائج التحليل العاملی.

غير أنه في معظم الأحيان يكون صدق التكوين الفرضي أكثر ملائمة لتقدير صدق استبيانات الشخصية، فسمات الشخصية تُعد بمثابة تكوينات فرضية، والتحقق من صدق هذه التكوينات يتطلب جمع بيانات من مصادر متعددة للتوصّل إلى أدلة كافية في هذا الشأن كما سبق أن أوضحنا في الفصل الخامس، فصدق التكوين الفرضي يصعب التتحقق منه في وقت قصير أو التعبير عنه بدليل واحد. لذلك ينبغي على مؤلف الاستبيان أن يقدم بيانات متعددة تُلقى الضوء على صدق التكوين الفرضي للاستبيان مسترشداً في ذلك بمدخل السمات المتعددة والطرق المتعددة - Multitrait Campbell & Fiske Multimethod Approach وفيسك اقترحه كامبل ، وأوضحته في الفصل الخامس.

فكما تعددت أدلة الصدق وتتنوعت الدراسات المتعلقة به، زادت الثقة في استبيان الشخصية وفي استخداماته المناسبة.

ثانياً مقاييس غير مجربة البنية وغامضة المدّفف (الأساليب الإسقاطية) :

تُعدّ الأساليب الإسقاطية Projective Techniques اتجاهًا آخر في تقييم الشخصية يختلف عن الاتجاه السيكومترى الذي تناولناه في الجزء السابق المتعلق باستبيانات أو قوائم الشخصية . وقبل أن نناقش هذه الأساليب ينبغي أن نوضح مفهوم «الإسقاط Projection» الذي تستند إليه . فقد قدّم فرويد Freud هذا المفهوم في الأدب السينمائي، حيث عرّفه بأنه العملية التي ينسّب بها الفرد دوافعه ومشاعره وعواطفه إلى أفراد آخرين أو إلى أشياء في العالم الخارجي كدفاع يحول بينه وبين

شعوره أو إدراكه لخصائصه التي تهدده، فاستخدام الفرد لعملية الإسقاط تسمح له بإنكار ما بداخله من دوافع غير مرغوبة، وبإسقاط هذه الدوافع لا تُصبح مُدركة كجزء منه وإنما تُصبح جزءاً من الآخرين (Freud, 1911). أما مفهوم «الأسلوب الإسقاطي Projective Technique» فقد أصبح معروفاً عقب مقالة فرانك (Frank, 1939) الذي أوضح فيها أهمية أدوات قياس الشخصية التي تميز بالغموض أو التأثير Disguised Devices، وعارض استخدام استبيانات الشخصية بحجج أنها تستخدم في تصنيف الأفراد، ولا تعطينا معلومات مفيدة تتعلق بالفرد كإنسان، وحث على استخدام اختبارات تساعده في فهم واضح للمشاعر والأفكار الخاصة لدى الفرد.

وعندما يُطبق مفهوم الإسقاط على الأساليب الإسقاطية فإنه يستخدم بالمعنى العام الذي يشير إلى العملية التي تؤثر بها خصائص الفرد، أي بنية شخصيته، في أساليب إدراكه وتنظيمه وتفسيره للبيئة وخبراته، وهذه التأثيرات يمكن أن تتضمن وتقاس بدرجة أفضل عندما يواجه الفرد مواقف أو مثيرات غامضة. وقد استخلص ليندزي (Lindzey, 1961) من مراجعته الشاملة لمفاهيم أساليب الإسقاط أن ما يُميز هذه الأساليب بدرجة أساسية هو حساسيتها للجوانب اللاشعورية أو الكامنة في الشخصية، وتتنوع الاستجابات التي يقدمها الفرد، وتعدد أبعاد الشخصية التي يتم قياسها في آن واحد، وعدم إدراك الفرد للغرض من الاختبار، وثراء البيانات والمعلومات المتعلقة بالاستجابات.

ولعل صعوبة التمييز بين مفهوم الإسقاط كما استخدمه فرويد وأطلق عليه الإسقاط الداعي أو غير السوى، والمفهوم كما استُخدم في المواقف الاختبارية يرجع إلى التشابه اللغوي بين المفهومين وتداخل المعنى في استخدام كل منهما. فالإسقاط من وجهة نظر فرويد يعتقد فيه الفرد بالفعل أن الصفات أو الخصائص التي أسقطها ليست صفات أو خصائصه، ولا يكون مُدركاً أو واعياً بذلك، وهدف التخلص من رغباته وأفكاره ودوافعه غير المرغوبة أو المقبولة، أما الإسقاط في إطار المواقف الاختبارية فهو العملية التي يستجيب بها الفرد عندما يقدم له عدد من المثيرات الغامضة استجابة نشطة وواعية. ولعل أفضل تعريف للأسلوب الإسقاطي هو ما اقترحه ليندزي (Lindzey, 1961) بأنه أداة تميز بالحساسية لمظاهر السلوك اللاشعوري أو غير الظاهر، وتسمح بـ^وتشجع الفرد على أن يقدم استجابات كثيرة متنوعة، أي متعددة الأبعاد، وتستدعي بيانات ثرية تتعلق بالاستجابات، ولا تتطلب إلا قدرًا ضئيلاً من إدراك الفرد للغرض من الاختبار. ويهتم الأخصائي клиينيكي الذي يستخدم هذا الأسلوب بالتوصل إلى تفسيرات كلينيكية شاملة من هذه البيانات.

فالأساليب الإسقاطية وما يتعلّق بها من اختبارات تسمح للفرد بأن يفرض تكويننا وتنظيمًا خاصًا على المثيرات، وبهذا تُعبّر عن تصوّر ديناميكي للشخصية.

ويمكن أن تتضح العلاقة بين هذه الأساليب وبين النظريات السيكوديناميكية في الأسس التي بُنيت عليها الاختبارات الإسقاطية التي سنوضحها بعد قليل. فهذه النظريات تؤكد الجوانب المزاجية الذاتية للفرد، وتعتبر أن السلوك يحدث نتيجة تفاعل قوي متعدد مثل: الدوافع، وال حاجات، والحوافر، والصراعات، وأن تكوين الشخصية يتضمن تنظيمات مختلفة مثل الشعور واللاشعور، وأن الشخصية هي العملية التي يفرض فيها الفرد تنظيمًا وتكونًا على مثيرات خارجية في البيئة المحيطة به. كما تهم هذه النظرية بالفهم الشامل المتكامل للسلوك من حيث العلاقة بين أجزاءه بدلاً من تفسير السلوك بطريقة جزئية أو على أساس سمات شخصية منفصلة كما في استبيانات الشخصية.

تعريف الأساليب الإسقاطية :

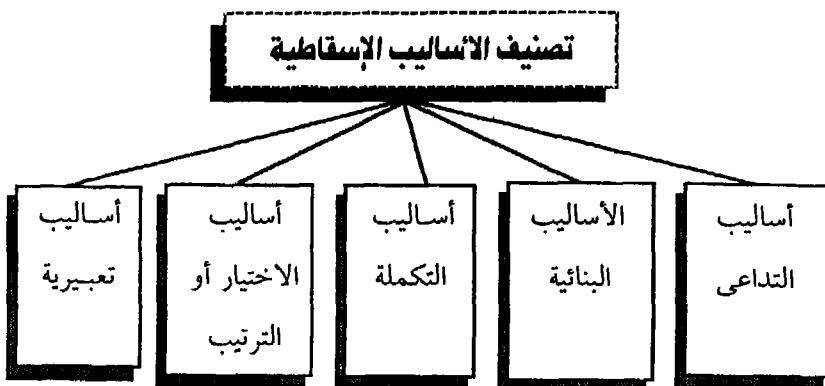
تتعدد طرق تصنيف الأساليب الإسقاطية، ولعل أبسطها هو التصنيف الثنائي إلى قسمين عاميين هما:

(١) **أسلوب يُركز على المحتوى Content - Centered :** يهتم هذا الأسلوب بتحليل استجابات الفرد في ضوء الموضوع أو المشاعر التي يعيّر عنها، حيث يفترض أن هذا المحتوى يكون تابعًا لصورات الفرد ويمثل تخيلاته. فاختبار تفهُّم الموضوع Thematic Apperception Test (TAT) الذي أعدَّه مواري Murray يعد مثلاً لهذا الأسلوب الإسقاطي، حيث يتم تحليل الاستجابات في ضوء الموضوعات Themes التي يعيّر عنها الفرد في قصص يكتتبها كاستجابة لسلسلة من الصور التي تميّز بالغموض النسبي.

(٢) **أسلوب يُركز على الشكل Form - Centered :** يستمد هذا الأسلوب أهميَّة التشخيصية بدرجة أساسية من خصائصه الشكلية، فما يراه الفرد في المثير الغامض يتعلّق بعدد من الأقسام الشكلية ويعُد اختبار بقع الحبر لورشاخ Rorschach Inkblot Test مثلاً لهذا الأسلوب، فهذا الاختبار يصنف استجابات الفرد من حيث بعض الأقسام الشكلية، مثل الحركة، والتزعّة إلى استخدام الكل أو الأجزاء في بقعة الحبر، وتنظيم الشكل، والحساسية للألوان والظلال. وتركز بعض المقاييس الإسقاطية

الأخرى على المحتوى والشكل معاً، مثل اختبار الرسومات، فتحليل الرسومات يستند بدرجة أساسية إلى الأقسام الشكلية، وكذلك على محتوى هذه الرسومات.

غير أن ليندزى (Lindzey 1959 , 1961) اقترح تصنيفاً أكثر شمولاً وتكاملاً من التصنيف السابق للأساليب الإسقاطية المتنوعة في خمسة أقسام عامة استناداً إلى نمط الاستجابة المطلوبة يوضحه الشكل التخطيطي (٤-١٣) التالي :



شكل (٤-١٣) يوضح تصنيف ليندزى للأساليب الإسقاطية

(١) **أساليب التداعي** Association Techniques : وفيها يُطلب من الفرد أن يستجيب بأقصى سرعة لمثيرات مثل الكلمات أو الصور أو المدركات ، ومن أمثلة ذلك اختبارات تداعى الكلمات، وإكمال الجمل، واختبار الرورشاخ، وكذلك جهاز كشف الكذب Polygraph الذى يعتمد على التأثيرات المعلومة للانفعالات الشديدة على ضغط الدم، والنفس ، وعرق راحة اليد.

(٢) **الأساليب البنائية** Construction Techniques : وفيها يُطلب من الفرد تكوين قصة أو رسم صورة دون أي قيد على استجابته، وتختلف هذه الأساليب عن أساليب التداعى فى أن الفرد فى تكوينه لصيغة فنية يقوم بسلوك معرفى معقد يوسع فيه ويعدّ ويتطور تداعياته الأصلية أثناء كتابة القصة أو رسم الصورة. ومن أمثلة ذلك اختبار تفهم الموضوع (TAT)، واختبار رسم صورة وكتابة قصة، واختبار بلاكى للصور Blacky Pictures Test .

(٣) **أساليب التكميلة** Completion Techniques : وفيها يُطلب من الفرد أن يكمل مهمة ناقصة بأى طريقة يراها مناسبة، والاستجابة فى هذه الحالة تكون أكثر بطئاً منها فى أساليب التداعى ، والمثيرات الأصلية تكون أكثر تعقيداً. ومن ناحية أخرى

تكون نواتج الاستجابات أقل تعقداً وأكثر تقيداً منها في الأساليب البنائية، ومن أمثلة ذلك اختبارات تكملة القصص Story Completion Tests.

وأكثر هذه الاختبارات شيوعاً قائمة الجمل الناقصة التي أعدها روتter Rotter، حيث تشتمل على . ٤ جملة ناقصة مثل:

«أتمنى لو أنني ...».

«معظم الناس ...».

«والذى ...».

ويُطلب من الفرد أن يكمل الجمل بما يختاره دون تحديد لزمن الإجابة، وهنا يفترض أن الطريقة التي يكمل بها الفرد الجمل تعكس ما ينطوي عليه من دوافع وصراعات وحاجات أو غير ذلك من ديناميات الشخصية. وتُستخدم هذه القائمة وغيرها من الأساليب البنائية في المجالات الكlinيكية، وكذلك في اتخاذ القرارات المتعلقة بانتقاء الأفراد.

(٤) **أساليب الاختيار أو الترتيب Choice or Ordering Techniques** : وفيها يُطلب من الفرد أن يختار من بين عدد من الصور أو البديل بديلاً أو أكثر وفقاً لمحك معين، مثل المواءمة أو الفرق أو الاستحسان، وغير ذلك، وأحياناً يُطلب منه ترتيب البديل استناداً إلى أحد هذه المحكمات، ومن أمثلة ذلك اختبار تومكينس وهورن Tomkins - Horn Picture Arrangement Test عن خطوط، وتمثل أنشطة يقوم الفرد بترتيبها لكي تكون قصة ، واختبار زوندوى Szondi Test.

(٥) **أساليب تعبيرية Expressive Techniques** : وفيها يُطلب من الفرد القيام بدور نشط في الرسم أو الطلاء أو اللعب كوسيلة للتعبير عن شخصيته، وهنا يكون التركيز على التأثير، وأسلوب أو نمط التعبير. وكما أوضح ألبورت فإن هذه الأساليب يفترض أن لها قيمة علاجية وتشخيصية، حيث إن الفرد هنا لا يكشف عن نفسه فقط، بل يعبر عن نفسه أيضاً بطريقة تؤثر في تكييفه الشخصي. ومن أمثلة ذلك مواقف اللعب مع الأطفال، ورسم شخص Draw - A - Person Test ، وأساليب الطلاء.

وقد قدم ألبورت Allport, 1933 مفهوم الأسلوب التعبيري Expressive Technique ليشير إلى استجابات الفرد التي يتطلبه اختبار الرورشاخ، وإجراءات اللعب Play Procedures التي تُستخدم عادة في الأساليب العلاجية، وكذلك الدراما

النفسية Psychodrama، والدراما الاجتماعية Sociodrama التي تُستخدم في المدارس وفي الإرشاد لتحسين العلاقات بين مجموعات من الأفراد.

أمثلة لبعض الاختبارات الإسقاطية:

نظرًا لأن أكثر الاختبارات الإسقاطية شيوعاً واستخداماً في المجالات الكلينيكية لتقييم الشخصية بما اختبار بقع الحبر لرورشاخ، واختبار تفهم الموضوع لمواري، فسوف نوضحهما بالتفصيل فيما يلي:

اختبار بقع الحبر لرورشاخ:

يُعد هذا الاختبار من الاختبارات الإسقاطية الأكثر شيوعاً، وقد أعده الطبيب النفسي السويسري هيرمان رورشاخ H.Rorschach أثناء الحرب العالمية الأولى وبعدة، ونشر حوله ما يزيد عن أربعة آلاف مقالة وكتاباً وبحثاً (Bernstein & Neitzel, 1980). وعلى الرغم من أن بقع الحبر قد استُخدمت في الماضي لاستدعاء استجابات من الأفراد في الدراسات المتعلقة بالتصور العقلي، إلا أن رورشاخ يُعد أول من أدرك بوضوح أهمية استخدام هذه الاستجابات في تقييم الشخصية. ومع هذا فإن هذا الاختبار أدى إلى انقسام شديد في الآراء بين علماء النفس، كما سبق أن أشرنا في عرضنا للإطار التاريخي لقياس الشخصية.

وعلى الرغم من أن رورشاخ كان متأثراً بنظرية فرويد وما تضمنه من مفاهيم، مثل اللاشعور وдинاميات الشخصية، إلا أنه لم يُصمم الاختبار استناداً إلى هذه النظرية، غير أن طبيعة بناء الاختبار يبدو أنها قد تأثرت بها بلا شك.

في بينما أكد رورشاخ في بعض الأحيان الجوانب الإمبريقية للاختبار، إلا أنه أكد أيضًا دور نظريات علم النفس الدينامي Psychodynamics عند التفكير في بنائه للاختبار.

ويتكون الاختبار من 10 بطاقات متماثلة، بعدها كل منها 5، 5 بوصة، خمسة منها ظلالها بيضاء وسوداء، وبطاقتين حمراء ورمادي، وثلاثة بطاقات متعددة الألوان: ويوضح شكل (١٣-٥) التالي (Brown, 1970) إحدى بقع الحبر المماثلة لبطاقات الاختبار.



شكل (١٣-٥) يوضح إحدى بقع الحبر المماثلة لبطاقات اختبار رورشاخ

وقد صمم رورشاخ المثيرات عن طريق وضع بقع الحبر على ورقة، وقام بطيء الورقة بحيث ينبع منها أشكال متماثلة ولكن غير محددة البنية ، ثم قام بتطبيق هذه البقع (المثيرات) على مجموعات مختلفة من مرضى المستشفيات النفسية ، واحتفظ بالبقع التي استدعت استجابات مختلفة من هذه المجموعات ، واستبعد البقع الأخرى .

واستخدم رورشاخ في محاولاته آلافاً من بقع الحبر حتى استقر في النهاية على عشر بقع ، وصمم إجراءات تقييم الدرجات استناداً إلى الملاحظات הקלينيكية ، وتبليورت هذه الإجراءات بعدما قام بتجريب البقع على مجموعات أخرى من الأفراد العاديين ، والمتخلفين عقلياً ، والفنانين ، وغيرهم من الأفراد الذين يتميزون بخصائص معلومة .

وعند تطبيق الاختبار تُعرض كل من هذه البطاقات على المفحوص ، ويُطلب منه أن يُوضّح ما يراه فيها أو ما يُمكّن أن تمثله بقعة الحبر ، ونظرًا لأن بقعة الحبر تُعد مثيراً غامضًا ، فإن استجابة الفرد تمثل إسقاطه لمعنى معين على المثير ، وهذه المعانى التي تم إسقاطها تُستخدم بدورها كأساس للاستدلال عن بنية شخصيته وдинاميتها .

ولا تزال بقع الحبر العشر تُستخدم حتى وقتنا الحاضر ، وأُعيد النظر في نظام تقييم الدرجات ، ولكن الجهود التي بذلها رورشاخ في البداية ظلت أساسية في معظم

نظم تقدير الدرجات المستخدمة حالياً. وعند تطبيق البطاقات ينبغي أن يحاول الفاحص جعل المفحوص في حالة عدم توتر، وأن يعمل على تكوين صلة وثام معه، وأن تكون التعليمات كافية فقط بالقدر الذي يمكنه من استكمال استجاباته. ويتم التطبيق عادة في مرحلتين، في المرحلة الأولى يقدم الفاحص كل بطاقة على حدة للمفحوص، ويطلب منه أن يوضح ما تعنيه البقعة بالنسبة له، ويسجل الفاحص الاستجابات والمواقع المختلفة في البطاقة التي ركز عليها المفحوص في استجاباته، وما الذي جعل البقعة تأخذ هذا الشكل. وكذلك ملاحظاته لبعض جوانب أداء المفحوص، مثل بطاقة، وكيفية قلب استجابته الأولى لكل بطاقة ، والזמן الكلى الذى استغرقه فى كل بطاقة، وكيفية قلب البطاقة على الوجه الآخر. وفي المرحلة الثانية يقوم المفحوص بفحص البطاقات للمرة الثانية للتعليق على مظاهر بقع الحبر التى جعلته يستجيب استجابة معينة.

ويمكن أن يطلب منه الفاحص أثناء ذلك توضيح بعض الأمور المتعلقة باستجاباته، وبذلك يمكن للفاحص أن يحدد الأساس الذى بنى عليه المفحوص استجاباته بدقة أكبر. وأهم ما يميز إجراءات تطبيق هذا الاختبار حرية المفحوص في اختيار ما يراه في بقعة الحبر، وأين يراه، ومحدّدات إدراكته.

وتتعدد نظم تقدير الدرجات ، إذ يوجد خمسة نظم متمايزة على الأقل، على الرغم من أن النظام الذى اقترحه كلوifer Klopfer ، وبك Beck (Exner, 1976) يُعد من النظم شائعة الاستخدام. غير أن هناك اتفاقاً عاماً بين جميع النظم على أن تفسير الدرجات ينبغي أن يستند إلى أبعاد معينة هى : الموضع Location، والمحدّدات Determinants، والمحتوى Content . فالموقع يشير إلى الجزء من بقعة الحبر الذي استند إليه المفحوص استناداً أساسياً في استجابته، إذ يمكن أن يستخدم المفحوص البقعة ككل ، أو تفاصيل عامة، أو تفاصيل بسيطة، أو الظلال البيضاء، ويمكن للفاحص التتحقق من ذلك أثناء المرحلة الثانية. وموضع الاستجابات يعتقد بأنه مؤشر للمدخل العقلى لحياة الفرد من حيث اعتماده على تعميمات متaramية، أو يراعى التفاصيل بانتظام ، أو تضييع لديه المعالم في تركيزه الشديد على قضايا جانبية غير ذات أهمية .

والمحدّدات تشير إلى مظاهر البقعة التي تحدّد الاستجابة، مثل الشكل أم الصيغة، والظلال أم التركيب، واللون ، والحركة. وتقدير درجات المحدّدات يعد عملية معقدة وأكثر ذاتية من تقدير الموضع أو المحتوى، فلكل من هذه المظاهر تقدیرات معينة ومؤشرات مختلفة.

والمحتوى يشير إلى القسم الذى يتسمى إليه المُدرك. ويتم تقدير محتوى الاستجابات فى ثلاثة أقسام تتعلق بالإنسان أو الحيوان أو المُجردات كالعدوان أو العاطفة، بحسب مادة المثير الذى يراه المفحوص. وكذلك يمكن تقدير ذلك بحسب شيوخ محتوى إدراك الفرد أو أصالته، فتكرار عدد معين من الاستجابات لكل بقعة حبر هو الذى يحدد شيوخها أو أصالتها. ويختلف تفسير شخصية الفرد تبعاً للمحتوى الذى يراه فى معظم البطاقات، فمثلاً يمكن المقارنة بين تفسيرين ربما يتم التوصل إليهما من مجموعتين من الاستجابات؛ إحداهما رأى فيها المفحوص حيوانات تصارع، ورأى فى الأخرى أناساً يشاركون مشاركة كاملة فى عمل ما.

وفي كثير من الحالات يكون محتوى الاستجابات على شكل قصص ويفسرها الفاحص على أنها تعبير عن خصائص شخصية المفحوص وأسلوب حياته. كما يمكن تفسير المحتوى بطريقة رمزية، فمثلاً الانفجار ربما يرمز إلى عدون شديد، والشعل يرمز إلى الميل إلى العدوان، والخنزير يرمز إلى الشره أو النهم، والعنكبوت، والأخطبوط، والساحر، والغول، ربما يرمز كل منها إلى اتجاهات سلبية نحو الأب المسيطر، والنعامة ربما ترمز إلى محاولة الابتعاد عن الصراعات، والأقسام العامة للمحتوى، مثل الطعام والحيوانات التى يؤكل لحمها، والكائنات العدوانية والحيوانات المفترسة والشخصيات التى تمثل القوة والسلطة يمكن تفسيرها رمزياً تحليلياً (Klopfer & Davidson, 1962). وينبغي الإشارة إلى أنه ينبغي تقييم كل استجابة استناداً إلى الأبعاد الثلاثة السابقة، أي الموضع والمُحدّدات والمحتوى. فالتفسير لا يتم على أساس استجابة واحدة، ولكن تفسّر الاستجابة في علاقتها بالتنظيم الكلى للاستجابات، فعدد الاستجابات التى تنتهي إلى كل من هذه الأبعاد، وكذلك النسب المختلفة التى يتم حسابها منها يحدد تفسير بروتوكول الاختبار ككل. فمثلاً الاستجابات المتعددة الجيدة التى تشير إلى الكل ربما تشير إلى الانفعالية، والحركة ربما تشير إلى التخيل، ونسبة عدد استجابات الحركات الإنسانية إلى عدد استجابات اللون ربما ترتبط بدرجة التفكير العقلى أكثر من التوجه العملى، وبطء الاستجابة ربما يفسّر على أنه قلق، والعدد القليل من استجابات اللون والحركة ربما يفسّر على أنه اكتئاب، والاستجابات الكثيرة التي تتعلق بالظلال ربما تشير إلى ضبط الذات (Schafer, 1954).

لذلك فإن تفسير الاستجابات لهذا الاختبار يتطلب فاحصاً مدرباً تدرّباً عالياً ولديه خبرات كلينيكية واسعة. فالتفسيرات إذن ينبغي أن تعتمد على النمط العام للإجابات، وكذلك على سلوك الفرد أثناء استجاباته للاختبار، فالفاحص يستخدم ملاحظاته عن

هذا السلوك كمصدر إضافي للبيانات عند تفسيره للاستجابات. فالمحظوظ الذي يطلب دائمًا العون والإرشاد من الفاحص ربما يفسر سلوكه هذا بأنه سلوك شخص غير استقلالي ، والمحظوظ الذي يبدو متوترًا أو يسأل أسئلة مراوغة ثم ينظر إلى الوجه الآخر للبطاقات ربما يفسر هذا السلوك بأنه سلوك شخص متشكك ، وهكذا.

ثبات وصدق اختبار رورشاخ :

إن الذي يطلع على الدراسات والمراجعات التي أشار إليها الكتاب السنوي للقياس العقلي Buros Mental Measurement Yearbook في طبعاته المتتالية سوف يلاحظ وجهات نظر ناقدة متعددة فيما يتعلق بالخصائص السيكومترية لاختبار رورشاخ. ثباتات الدرجات يتباين تبايناً كبيراً ربما بسبب تنوع وثراء الاستجابات التي تُستخدم في تقدير معظم خصائص شخصية الفرد، وتباين التفسيرات، مما يؤدي إلى مشكلات تتعلق بتقدير ثبات وصدق الاختبار.

فمعظم طرق تقدير ثبات درجات الاختبارات التي تبني استناداً إلى الأساليب الإحصائية مثل استبيانات الشخصية تناسب هذه الاختبارات لأنها تؤدي إلى بيانات كمية، أما في الاختبارات الإسقاطية وبخاصة اختبار الرورشاخ ، فإننا لا نستطيع الرجوع إلى جدول معايير لمعرفة دلالة أنواع الدرجات كما في الاختبارات المقننة.

لذلك فإن إحدى طرق تقدير ثبات الرورشاخ هو تحديد مدى اتفاق مجموعة من المحكمين في تقدير الدرجات. ولكن هذه الطريقة غير شائعة الاستخدام بسبب اختلاف نظم تقدير درجات هذا الاختبار، وهذا الاختلاف يؤدي إلى عدة مشكلات في تقدير الثبات بالأساليب المتعارف عليها. فعدم استقرار الدرجات ربما يرجع إلى عدم استقرار طريقة القياس، أو إلى التغيير المؤقت الذي يحدث في الجوانب المزاجية للفرد. كما أن بطاقات الرورشاخ تختلف عن فقرات استبيانات الشخصية أو مقاييس الجوانب المعرفية مما يجعل من الصعب التتحقق من الثبات بطريقة التكافؤ. وكذلك تختلف استجابات الفرد لبقع الحبر من بطاقة إلى أخرى، مما يجعل استخدام طريقة التجزئة التصفية غير مناسبة؛ لذلك توصلت دراسات فيرنون (Vernon, 1933)، وهيرتس (Hertz, 1934) ، وكرونباك (Cronbach, 1949) ، وسوبين (Zubin, 1965) التي استخدمت طرقاً مختلفة لتقدير الثبات إلى نتائج متعارضة.

ونظراً لافتقار الاختبار إلى الثبات، فإن صدقه سوف يكون كذلك، فتقدير صدق اختبار رورشاخ يُعد أمراً بالغ التعقيد، إذ يمكن تتبع عدد كبير من الدراسات في هذا الشأن وجميعها أدت إلى نتائج متناقضة، كما لا يوجد اتفاق فيما بينها حول أفضل طرق تقدير صدق هذا الاختبار. وربما تكون إحدى طرق إجراء ذلك هي التحقق من مدى مطابقة التفسيرات الناتجة عن الاختبار بالبيانات المستمدة من مصادر كلينيكية أخرى، ومع هذا فقد تبانت نتائج الدراسات في هذا الشأن.

ويرى بعض الباحثين أن الأساليب المتعارف عليها في تقدير صدق الاختبارات والمقاييس يصعب تطبيقها على اختبار رورشاخ. إذ من الصعب التوصل إلى محكّات مناسبة يمكن باستخدامها التتحقق من صدق تفسيرات الاختبار للمستويات العميقة أو اللاشعورية للسلوك، كما يصعب تحديد العلاقات بين أجزاء الاختبار والسلوك الفعلى.

لذلك حاول بعض علماء النفس بناء اختبارات بقع حبر أيضاً ولكن تمييز بمزيد من الموضوعية والدقة، ومثال ذلك أسلوب بقع الحبر لهولتزمان Holtzman Inkblot Technique (HIT) الذي يشتمل على صيغتين متكافئتين تتكون كل منهما من ٤٥ بقعة حبر، تتطلب كل منها استجابة واحدة فقط. وتم انتقاء بقع الحبر استناداً إلى قيم معامل الثبات بطريقة التجزئة النصفية، وتمييزها بين استجابات الأفراد العاديين والمرضى لكل بقعة منها. ويمكن تفسير درجات كل استجابة في ٢٢ بُعداً، كثير منها يماثل أبعاد تفسير درجات اختبار رورشاخ ، وتشمل الموضع والمُحدّدات والمحتوى.

ونظراً لأن هذا الأسلوب أقرب إلى أسلوب بناء استبيانات الشخصية، فإن الثبات في هذه الحالة أفضل من ثبات اختبار رورشاخ (Holtzman, 1968, 1975). غير أنه تبيّن من مراجعة الأدبيات أنه لا توجد بيانات كافية للمقارنة بين الاختبارين. كما حاول ساندفن (Sandven, 1975) التوصل إلى منهجة جديدة تضم منهجة بناء استبيانات الشخصية، والخصائص الأساسية للأساليب الإسقاطية أطلق عليها القياس الإسقاطي Projectometry. غير أن المجال لا يسمح بتوضيح هذه منهجة وتطبيقاتها المقترحة .

اختبار تفهّم الموضع :

أعد اختبار تفهّم الموضع The Thematic Apperception Test (TAT) كريستينا مورجان C. Morgan ، وهنرى موراي H. Murray عام ١٩٣٥ بالعيادة النفسية بجامعة هارفارد كأسلوب للكشف عن الأفكار اللاشعورية والخيالات Fantasies . ويختلف هذا الاختبار عن اختبار الرورشاخ في أن مثيراته أكثر انتظاماً وأقل غموضاً واستجاباته تكون أكثر تحديداً . فهذا الاختبار يتطلب من الفرد أن يحكى قصصاً تتعلق بصور معينة بحيث يمكن للأفراد تفسير الدوافع السكانمة لدى الفرد، وانفعالاته، وعواطفه، وصراعاته وحاجاته النفسية التي افترضها موراي في نظريته للشخصية . وهذا الاختبار يلي اختبار رورشاخ في شیوع استخدامه ، وفي عدد الدراسات والبحوث التي استثارها . وقد استمد الاختبار أهميته أولاً في العيادات النفسية ، ولكنه أصبح تدريجياً أداة بحث في علم النفس الارتقائي ، وعلم النفس الاجتماعي ، والشخصية ، والدراسات الأنثروبولوجية عبر الثقافات ، كما يستخدم في تقييم الشخصية في مجالات الإرشاد وعلم النفس الصناعي .

ويتكون الاختبار من ٣٠ صورة باللونين الأبيض والأسود مرسومة على بطاقات عبارة عن ٤ مجموعات متداخلة تتكون كل منها من ١٩ بطاقة إضافة إلى بطاقة بيضاء ، وتحتفل كل صورة من حيث محتواها ودرجة غموضها . وتبين الصور في درجة انتظامها وبنيتها ، إذ تراوح بين الانتظام التام أو البنية المحددة بدرجة كبيرة إلى البنية غير المحددة على الإطلاق كما في البطاقة البيضاء .

ويمكن تطبيق الاختبار على الراشدين والراهقين ، وتعرض البطاقات على الفرد واحدة تلو الأخرى مصحوبة بتعليمات بسيطة ، ويُطبق الاختبار عادة على مرتين أو أكثر ، وإجراءات التطبيق المعتمد هو أن يُطلب من المفحوص أن يحكى قصة كاملة تتعلق بكل من ١٩ بطاقة يتم اختيارها بحيث تناسب عمره ونوعه ، وكذلك البطاقة البيضاء ، ويُطلب منه أن يستغرق خمس دقائق في كل قصة ، ويحكى ما يحدث في الصورة الآن ، والأحداث التي أدت إليها . وما سوف تنتهي إليه ، ويجب أيضاً أن تتضمن القصة تفاصيل عن الشخصيات التي في الصورة ، وأنشطتهم وتفكيرهم ومشاعرهم .

ويوضح شكل (٦١٣) إحدى الصور المماثلة لصور الاختبار (Kleinmuntz, 1967).



شكل (٦١٣) يوضح إحدى الصور المماثلة لبطاقات اختبار نهم الموضوع

وهناك أنظمة أو أساليب متعددة أيضاً لتقدير الاستجابات بعضها كمي والآخر غير كمي، غير أنه نظراً لتعقد الأساليب الكمية وما تتطلبه من وقت طويل، فإن الفاحصين يميلون إلى استخدام الأساليب غير الكمية في التفسير. ومعظم هذه الأساليب تأخذ بعين الاعتبار بطل القصة، وال حاجات، والضغطوط، والموضوعات، والمواضيع، والنواتج. فالبطل Hero هو الشخصية الرئيسية في كل صورة والتي يتتطابق معها المفحوص، حيث تدور القصة حول بطل واحد يسهل تحديده من بين الشخصيات. كذلك يتم التركيز على خصائص هذا البطل وبخاصة دوافعه وحاجاته. ومعظم نظم التقييم بما في ذلك نظام موراي Murray الأصلي تهتم بشدة واستدامة أو استمرارية وتكرار أو شيع كل حاجة Need كمؤشرات عن أهمية ومواءمة تلك الحاجة.

أما الضغوط Press فتشير إلى القوى البيئية التي تتدخل مع أو تُيسّر تحقق الحاجات المختلفة، وهنا أيضاً يكون الاهتمام مركزاً على الشدة والاستمرارية والتكرار في تقييم الأهمية النسبية لهذه القوى. كذلك تكرار الموضوعات المختلفة Themes (مثل الكتاب)، والنواتج (مثل الفشل) تعد أيضاً مؤشرات أساسية لهذه الأهمية.

ويمكن أن يُراعي التفسير طول القصة وتنظيمها وأصالتها إلى جانب محتواها، وكذلك نط الموضوعات، ونوعية البيئات، وموافق العلاقات الشخصية التي يصفها المفحوص.

و عموماً فإن كثيراً من الاختصاصيين يستندون في تفسيراتهم لهذا الاختبار على فرضية أن دوافع، و نزعات، وأفعال، و مشاعر، و صراعات بطل القصة أو رمزها الرئيسي تعد جميعها أيضاً من خصائص المفحوص سواء كان واعياً أو غير واع.

ثبات و تطبيق اختبار تفهم الموضوع :

إن الخصائص السيكومترية لهذا الاختبار لا تختلف كثيراً عن اختبار رورشاخ، وذلك نظراً لأنه لا توجد إجراءات مقتنة لتطبيق الاختبار و تصحيفه و تفسير نتائجه. فإحدى طرق تقدير الثبات في هذا الاختبار تعتمد على اتفاق المحكمين في تحليل كل منهم لبروتوكول الاختبار. ولكن تكون هناك درجة عالية من الاتفاق ينبغي أن يكون المحكمون على درجة عالية من الخبرة والتمرس فيما يتعلق بنظام تقييم الاستجابات للحصول على قيم لمعامل الثبات تقترب من .٩٠ . على الأقل لكي تكون هناك ثقة في درجة الاتفاق (Murstein, 1963). أما الثبات بطريقة إعادة تطبيق الاختبار فإن قيمة تذبذب بحسب المدة الزمنية الفاصلة بين مرتب التطبيق.

و قد وجد أن قيمة معامل الاستقرار تقل كثيراً بمدورة مدة زمنية تبلغ عدة أشهر؛ لذلك يرى كثير من مستخدمي الاختبار أن معامل الاستقرار لا يناسب هذا الاختبار كما هو الحال في اختبار رورشاخ، وذلك لأننا نتوقع اختلافاً في موضوع القصة و محرارها الذي يقدمه الفرد من وقت إلى آخر، وكذلك الحالة النفسية للفرد ربما تتغير بمدورة الزمن، و اختلف صيغ القصص ينبغي أن يعكس هذا التغيير دون أن يكون هذا دليلاً على عدم الثبات. غير أنه ينبغي التركيز على اتساق أقسام لاستجابات أو عدم اتساقها بدلاً من التركيز على المقارنة بين القصص في مرتب التطبيق. أما طريقة التجزئة النصفية في تقدير الثبات فإنها لا تناسب هذا الاختبار كما هو الحال في اختبار رورشاخ. فقد بينت الدراسات أن قيمة معامل الثبات باستخدام هذه الطريقة كانت منخفضة.

ويُعد تقدير صدق هذا الاختبار من الأمور المعقّدة أيضاً، فالسؤال الذي يتadar إلى الذهن هو: هل القصص تعكس حقيقة شخصية الفرد المختبر أم أنها مجرد ردود فعل نمطية للموقف أى للصور؟ وحتى إذا افترضنا أن القصص تعكس شخصية الفرد، فهل افتراض أن الفرد يتطابق مع البطل، وأن مشكلات البطل والضغوط المؤثرة عليه تعكس حقيقة جوانب حياة المختبر يُعد افتراضياً صحيحاً؟

ومما يزيد من حدة المشكلة هو عدم الاتفاق على طريقة واحدة لتقدير الاستجابات ونظام التفسير. ونظرًا لأن الاختبار يستخدم في الوصف العام للشخصية، فإن مشكلة تحديد المحك عند تقدير الصدق المرتبط بمحك تعد من المشكلات المعقّدة. غير أن هناك اتفاقاً بين معظم الاختصاصيين في هذا المجال على أن هذا الاختبار يتميز بصدق محتواه في تقييم الشخصية.

ومن الجدير بالذكر أن هناك اختبارات أخرى تُعد تعديلات لاختبار تفهُّم الموضوع صُمِّمت بحيث تناسب صغار السن، مثل اختبار القصة المصور لسيموندر Symonds Picture Story Test ، واختبار تفهُّم الموضوع للأطفال (CAT) ، واختبار الصور لبلاكي Blacky Pictures ، والاختبار السمعي لتفهُّم الموضوع Auditory Iowa Picture Apperception Test (AAT) ، واختبار أيوا لتفسير الصور Interpretation Test (IPTT) . كما أن هناك تعديلات أخرى للاختبار لقياس الاتجاهات نحو المدرسة ، والسلطة ، والعمل ، وغير ذلك تُستخدم أيضًا في مجالات الإرشاد المهني والمشروعات البحثية المختلفة.

وكذلك أعد ريتزل ومعاونوه (Ritzel et al. , 1980) اختباراً بديلاً لاختبار تفهُّم الموضوع، وذلك بانتقاء صور تسمح بتواءن القصص الإيجابية والسلبية، وكذلك بأنشطة متعددة ومستويات مختلفة من الطاقة لبطل القصة، وذلك بهدف تطوير منهجة اختبار تفهُّم الموضوع بحيث يمكن تقييم الاستجابات تقييماً كميًّا.

مقارنة بين اختبار تفهُّم الموضوع واختبار رورشاخ :

يُعد كل من اختبار تفهُّم الموضوع واختبار رورشاخ أداة كلينيكية تتطلب أحصائيات متعرّساً لاستخدامها ولتقييم استجابات المفحوصين.

ويُفيد اختبار رورشاخ في الكشف عن الأضطرابات في عمليات التفكير وبخاصة عند المرضى الذين يتعرضون لأضطرابات ذهانية، بينما يهتم اختبار تفهُّم الموضوع بالتوافق الاجتماعي. وعلى الرغم من اختلاف الاختبارين فيما يقيسه كل منهما، إلا أن لكل منهما ميزاته وعيوبه. فاختبار تفهُّم الموضوع يهتم اهتماماً رئيساً بمحفوظي القصص وخصائص البطل الذي يتطابق معه المفحوص، ويختلف تفسير الاستجابات باختلاف هذه الخصائص. أما محتوى الاستجابات فليس له نفس القدر من الأهمية في اختبار رورشاخ، فالجزء من بقعة الحبر الذي تشير إليه استجابات المفحوص لا يكون مهمًا بقدر أهمية محددات المدرك (صيغة أو لون منطقة معينة من البقعة).

ونظراً لأن هناك العديد من الأدلة التي أوضحت افتقار كل من الاختبارين إلى الخصائص السيكومترية مثل الثبات والصدق، وبخاصة اختبار رورشاخ، فإن كثيراً من علماء النفس الكلينيكي يفضلون استخدام الاختبارين معاً في التشخيص والتوصيل إلى فهم متكملاً للشخصية.

تقييم الأساليب الإسقاطية:

اتضح من عرضنا السابق نوع الأساليب الإسقاطية واختلاف أدواتها وما تشمل عليه من مثيرات، وما تتطلبه من استجابات، وما تتوصل إليه من تفسيرات. فالثيرات التي تستدعي الاستجابات الإسقاطية يمكن أن تكون كلمات أو جمل أو قصصاً ناقصة، أو بقع حبر، أو صوراً محددة أو غير محددة البنية، أو عرائس ولعب، أو رسوماً أو لوحات ملونة، أو غير ذلك. وقد صنفها ليندزى Lindzey كما رأينا في خمسة أقسام عامة. ونظراً لأن الأساليب المختلفة تكشف عن أوجه متابينة إلى حد ما للشخصية على الرغم من تداخل هذه الأوجه، فإنه يفضل استخدام أكثر من أسلوب في عملية التقييم.

ميزات الأساليب الإسقاطية:

تميز الأساليب الإسقاطية بأن مثيراتها وتعليماتها غير محددة البنية Unstructured أو غامضة مما يسمح بحرية الاستجابة وتنوعها، وتباين الأساليب في درجة غموض مثيراتها. بعضها مثل اختبار رورشاخ تكون مثيراته أكثر غموضاً من اختبار تفهم الموضوع أو إكمال الجمل مما يؤدي إلى تباين ملحوظ في مدى الاستجابات. ومعظم هذه الأساليب تكون غامضة الهدف أيضاً، مما يجعل من الصعب على الفرد المختبر تزيف استجاباته، وذلك لأن طرق تقييم الاستجابات أو تقدير الدرجات وتفسير الاستجابات يكون غير مألف له. وكذلك تسمع هذه الأساليب بحرية الأخصائي الكلينيكي في سير وتوسيع ما يقصده المختبر. كما أن معظمها لا يحتاج إلى مهارة في القراءة، ويمكن أن تُطبق على الأميين والأطفال الصغار، وتتميز أيضاً بعدم وجود استجابة صحيحة وأخرى خطأ، أو استجابات محددة مسبقاً، حيث يتم تفسير الاستجابات كيفياً أو تصنيفياً بطرق مختلفة. بعضها مثل اختبار رورشاخ يركز التفسير على الخصائص الشكلية، وفي غيره مثل اختبار تفهم الموضوع أو إكمال الجمل يكون التركيز على المحتوى. وتستند معظم الأساليب على فرضية أن إضفاء

الفرد بنية على المثيرات الغامضة أو ردود فعله لهذه المثيرات يعكس نزعاته وتصوراته الشخصية وبخاصة نزعاته اللاشعورية وميكانيزماته غير اللفظية.

عيوب الأساليب الإسقاطية :

من أهم عيوب هذه الأساليب افتقارها إلى التقنيات Standardization، فالاختبارات الإسقاطية ليست اختبارات بالمعنى السيكومترى الدقيق، وإنما أدوات أو أساليب يستخدمها أخصائى مدرس تدرّباً عالياً في تقدير الشخصية والكشف عن جوانبها. وعلى الرغم من أن هناك جهوداً لتقنين عرض المثيرات، إلا أنه تظل هناك فروق في المداخل التي يستخدمها الفاحصون المختلفون. كما أن نتائج التقييم ووصف شخصية المختبر تعتمد على الأحكام المستبورة للفاحص.

فالفاصل لا تقتصر مهمته على تطبيق الاختبار، وإنما يُعد جزءاً من الأسلوب الإسقاطي ذاته والمواد المستخدمة؛ لذلك يتباين الفاحصون في مهارتهم وفاعليتهم في التوصل إلى مواصفات شخصية المُختبرين، وينعكس ذلك على صدق الأسلوب المستخدم. فتقدير الشخصية باستخدام الأساليب الإسقاطية يُعد فناً أكثر من كونه علمًا. ومع هذا فإن جهوداً قد بُذلت لتقنين هذه الأساليب، وبخاصة اختبار رورشاخ من حيث بناء صيغ من الاختبار تعتمد على الاختيار من متعدد بدلاً من الاستجابات غير المقيدة. ففي هذه الحالة تُقدم مع كل بقعة حبر قائمة باستجابات يختار الفرد من بينها الاستجابة القريبة مما يراه في البقعة. وبدائل الاستجابات تم انتقاءها من سجلات الأفراد العاديين وتقارير غير العاديين. كما أن أسلوب بقع الحبر لهولتزمان Holtzman الذي سبق أن أشرنا إليه يُعد من الأساليب الإسقاطية التي تمثل استبيانات الشخصية في الأساس الفكري الذي استند إليه وفي بنائه ، وبذلك يُعد أكثر تقنيتنا من اختبار رورشاخ.

ومن بين أوجه قصور الأساليب الإسقاطية افتقارها أيضاً إلى المعايير Norms ، أو استنادها إلى معايير مقتضبة أو مستندة إلى جماعات مرجعية غير محددة تحديداً دقيقاً. وعدم توافر هذه المعايير يؤدي إلى تفسير الاستجابات في ضوء الخبرة الكlinيكية، مما يسمح بتدخل كثير من الجوانب الذاتية والتحيز في عملية التفسير. وقد أسلحت هذه العوامل مجتمعة في القيم المنخفضة لمعامل الثبات وعدم وجود أدلة

كافية تتعلق بالجوانب المختلفة لصدق هذه الأساليب وجدوى استخدامها كما سبق أن أوضحنا.

وعلاوة على ذلك فإن جميع الأساليب الإسقاطية تتطلب التطبيق الفردي مما يجعل عملية التطبيق وتقييم الاستجابات وتفسيرها مكلفاً. كما يجب أن يقوم بتطبيق هذه الأساليب، وبخاصة اختبار رورشاخ، واختبار تفهم الموضوع، والدراما النفسية، والاجتماعية، واللعب، واختبارات الرسم، أخصائيون تلقوا تدريباً خاصاً على استخدامها وتفسير نتائجها.

وعلى الرغم مما يشوب هذه الأساليب من عيوب، إلا أنها تزود الأخصائيين ببيانات تتعلق بديناميات الشخصية يصعب الحصول عليها باستخدام الاستبيانات أو أي أساليب أخرى. غير أنه ينبغي عدم الاستناد إلى نتائجها في اتخاذ قرارات مهمة تتعلق بالأفراد في ضوء الطرق التفسيرية الحالية. كما لا ينبغي استخدامها في الإرشاد المهني أو انتقاء الأفراد، ولكن يمكن الإفاده من هذه الأساليب في الأغراض البحثية من أجل التوصل إلى إجراءات كمية لتقييم الاستجابات لكي تصبح أكثر موضوعية، وكذلك لبحث صدق متغيراتها في علاقتها باختبارات أخرى تم التحقق من صدقها، والتحقق من صدقها التجربى في ضوء مجموعات مختلفة مثل مرضى البارانويا، والفصام. فإذا أجريت مثل هذه البحوث، فإن الأساليب الإسقاطية يمكن أن تصبح واحدة من أفضل الاختبارات السيكلوجية (Kline, 1976).

ثالثاً : مقاييس غير محددة البنية ومتضمنة الهدف

(الملاحظة، والمقابلة، ومقاييس مفهوم الثالثاً :

اعتمدت مقاييس الشخصية التي أوضحناها في الجزء السابق على الاستبيانات أو الاختبارات الإسقاطية التي تتطلب ورقة وقلمًا، وتستدعي استجابات في موافق اختبارية اصطناعية. وعلى الرغم من أن هذه المواقف تكون مقتنة وتسمح بضبط المتغيرات الدخيلة، إلا أنها تختلف عن مواقف الحياة الفعلية. إذ ربما يستجيب الفرد لهذه المقاييس استجابات لا تتفق ونمط سلوكه المميز في حياته اليومية؛ لذلك لا نستطيع أن نعتمد على هذه المقاييس وحدتها اعتماداً تاماً في تقييم متغيرات الشخصية، وإنما ينبغي استخدام أساليب أخرى تتعلق بمشاهدة السلوك ملاحظة مباشرة في موقف طبيعية، أو إجراء مقابلة وجهاً لوجه. وتوارد هذه الأساليب وضوح الهدف من القياس، حيث إن إخفاء الهدف ربما يعوق الحصول على المعلومات المتعلقة بخاصائص شخصية

الفرد، وتُعدُّ بعض أساليب الملاحظة Observations ، والمقابلات المباشرة Direct Interviews ، والسير الذاتية للفرد كما يقرّها بنفسه Autobiography ، أمثلة المقاييس غير محددة البنية وواضحة الهدف، حيث تستند إلى المدخل الفينومولوجي Phenomenological Approach . ويرى أصحاب هذا المدخل أن تحليل الشخصية إلى مجموعة من المكونات أو السمات يتنافى مع التنظيم المتكامل والдинامي للشخصية، كما يرون أن اهتمام نظريات التحليل النفسي بالتنظيم اللاشعوري في وظائف الشخصية يغفل الجوانب المهمة التي ينبغي التركيز عليها، مثل الإدراك، والمعانٍ، والمشاعر ، والذات. ويؤكدون أنه يمكن النظر إلى الفرد على أنه يستجيب للعالم الخارجي في ضوء إدراكاته المميزة له، فمجاله الفينومولوجي يُعد جزءاً من بيئته الفيزيائية التي يدركها والتي يكون لها معنى خاص بالنسبة له، بينما تُعد «الذات Self» جزءاً من هذا المجال المتعلق به شخصياً؛ لذلك فإن الفرد يعمل جاهداً لتحقيق ذاته، أي إحداث تطابق أو تناcq بين ذاته الواقعية وذاته المثالية (Rogers, 1961, 1964).

وسوف نقدم فيما يلى أمثلة لبعض أساليب تقييم الشخصية التي تستند إلى هذا المدخل ، مثل أساليب ملاحظة السلوك ، وال مقابلة . وتحتفل هذه الأساليب في مدى تحديد بنية المواقف التي تتضمنها .

(أ) أساليب ملاحظة السلوك Observation Techniques

تلعب أساليب الملاحظة المباشرة لسلوك الأفراد دوراً أساسياً في تقييم الشخصية سواء في الفصل المدرسي ، أو موقف العمل ، أو في مراكز الإرشاد ، أو المواقف الكلينيكية . وقد سبق أن أوضحنا في الفصل العاشر دور الملاحظة وال مقابلة في تقييم الأفراد في مواقف العمل المختلفة . وسوف نوضح بایيجاز في هذا الجزء بعض أنماط أساليب الملاحظة سواء في مواقف واقعية أو محاكاة لهذه المواقف ، أي مواقف اصطناعية ، وكيفية استخدامها في جمع معلومات تتعلق بشخصية الفرد ، والتنبؤ بسلوكه المستقبلي . وتميز هذه الأساليب بأنها يمكن من خلالها ملاحظة سلوك الفرد كما يحدث تلقائياً ، وتكون الملاحظة مستقلة عن قرارة الفرد أو رغبته في تقرير سلوكه .

الملاحظة في مواقف واقعية ، ومواقف مضبوطة :

يمكن الحصول على بيانات تتعلق بسلوك الأفراد عن طريق ملاحظتهم في مواقف واقعية Natural Observation لا يتحكم فيها القائم بالملاحظة ، حيث يُسجل

سلوك الفرد أو مجموعة من الأفراد أثناء نشاطهم المعتاد دون أن يحاول التدخل أو وضع قيود أو شروط معينة على موقف معين. فمثلاً يمكن أن يقوم المعلم بـ الملاحظة سلوك الأطفال أثناء اللعب في فناء المدرسة، أو يقوم الأخصائي الـ كلينيكي بـ زيارة منزل العميل لملاحظة الأنماط السلوكية لأفراد أسرته. وعلى الرغم من أن تواجد القائم بالـ الملاحظة ربما يؤدي إلى إحداث تغييرات في سلوك من يلاحظهم ، إلا أن عملية الملاحظة تقييد في الحصول على معلومات أولية أو انطباعات تتعلق بالفرد.

أما الملاحظة في مواقف تجريبية مضبوطة Controled Observation فإنها تتم داخل المختبر أو في مواقف تجريبية يتـ حكم فيها القائم بالـ الملاحظة، حيث إن إجراء ذلك ميدانياً يكون مكلفاً وغير مناسب ويفـ ظد إلى نتائج معقولة ومرتبكة؛ إذ يمكنه تنظيم الموقف وتعديلـه وإدخـال متغيرات تستـ دعـيـ استـ جـابـاتـ معـيـنةـ وـفقـاًـ لـ اـحـتـيـاجـاتـ الـ بـحـثـيـةـ لـكـيـ يـحدـدـ كـيـفـيـةـ اـسـتـ جـابـةـ الـأـفـرـادـ لـهـذـاـ المـوـقـفـ،ـ مـثـلـ الـمـلـاـحـظـةـ غـيرـ الـمـبـاـشـرـ لـلـتـفـاعـلـ الـاجـتمـاعـيـ بـيـنـ مـجـمـوعـةـ الـأـفـرـادـ أـثـنـاءـ أـدـائـهـمـ مـهـمـةـ مـعـيـنةـ يـتمـ تـنـظـيمـهـاـ تـنـظـيمـاـ منـاسـبـاـ.ـ فـيـ هـذـهـ الـحـالـةـ رـبـماـ يـرـاقـبـ الـقـائـمـ بـالـمـلـاـحـظـةـ هـذـاـ التـفـاعـلـ مـنـ خـلـالـ مـرـآـةـ ذاتـ اـتـجـاهـ واحدـ بـحـيثـ لاـ يـسـطـيعـ الـأـفـرـادـ رـؤـيـتـهـ،ـ وـبـذـلـكـ لـاـ يـؤـثـرـ تـواـجـدـهـ فـيـ سـلـوكـ الـمـعـتـادـ لـلـأـفـرـادـ.ـ أـوـ رـبـماـ يـوـدـ مـشـارـكـتـهـمـ فـيـ الـعـلـمـ دـوـنـ عـلـمـهـ بـأـنـهـ يـقـومـ بـمـلـاـحـظـهـمـ،ـ فـعـنـدـئـذـ تـكـونـ الـمـلـاـحـظـةـ بـالـمـشـارـكـةـ الـفـعـلـيـةـ Participant Observationـ.ـ وـفـيـ هـذـهـ الـحـالـةـ يـسـطـيعـ الـقـائـمـ بـالـمـلـاـحـظـةـ تـسـجـيلـ الـعـدـيدـ مـنـ مـظـاـهـرـ سـلـوكـ الـانـفـعـالـيـ والـحـرـكـيـ وـالـتـفـاعـلـيـ.ـ وـجـمـيعـهـاـ تـُـعـدـ مـنـ الـمـظـاـهـرـ غـيرـ الـلـفـظـيـ لـسـلـوكـ الـفـرـدـ،ـ وـتـُـقـدـمـ صـورـةـ وـاقـعـيـةـ عنـ شـخـصـيـتـهـ بـدـرـجـةـ أـفـضـلـ مـنـ التـقـرـيرـ الذـاتـيـ لـلـفـرـدـ عـنـ نـفـسـهـ.ـ كـذـلـكـ رـبـماـ يـهـتـمـ باـحـثـ نفسـيـ بـسـلـوكـ مـحـدـدـ،ـ مـثـلـ اـسـتـ جـابـاتـ الـأـفـرـادـ لـمـوـقـفـ إـحـاطـةـ مـعـيـنةـ رـبـماـ يـنـدـرـ حـدوـثـهـ فـيـ الـحـيـاةـ الـوـاقـعـيـةـ،ـ فـعـنـدـئـذـ يـمـكـنـهـ تـصـمـيمـ مـوـقـفـ تـجـريـبـيـةـ مـحـكـمـةـ تـؤـدـيـ إـلـىـ هـذـهـ الـاستـ جـابـاتــ.

وقد سبق أن أوضحنا في الفصل العاشر أنواعاً أخرى من المواقف الاصطناعية ، مثل عينات العمل Work Samples ، والمحاكاة Simulation ، والاختبارات situational Tests .

مـعـاـدـرـ الـخـطاـءـ فـيـ أـسـالـيـبـ الـمـلـاـحـظـةـ :

إن أساليب الملاحظة شأنها شأن غيرها من أساليب قياس الشخصية أو تقييمها تؤثر فيها مصادر خطأ متعددة. ومن بين هذه المصادر ما يلى:

(1) التحيز؛ تعتمد أساليب الملاحظة اعتماداً أساسياً على القائم بالـ المـلـاـحـظـةـ ،ـ والإـجـراءـاتـ الـمـنـظـمةـ الـتـيـ يـلـتـزمـ بـهـاـ أـثـنـاءـ الـمـلـاـحـظـةـ ،ـ وـمـاـ تـنـطـلـهـ مـنـ تـسـجـيلـ دـقـيقـ

للسلوك المستهدف. غير أنه من المتعارف عليه أن فردين يُمكّنهما إدراك الحدث نفسه إدراكاً مختلفاً، مما يؤدى إلى أخطاء الملاحظة، فالعامل الإنساني يلعب بلا شك دوراً في ذلك. فمشاعر القائم بالملاحظة تجاه فرد معين أو جماعة معينة أو نشاط ما ربما تؤدي إلى تحيزه إيجاباً أو سلباً، مما يؤثر في مصداقية الملاحظة. غير أنه يمكن التغلب على ذلك جزئياً بتنمية عملية الملاحظة وإجراءاتها، واقتصار القائم بالملاحظة على تسجيل السلوك وفق أقسام محددة مسبقاً.

(٢) **القيم الشخصية**: يمكن أن تؤثر أيضاً القيم الشخصية للقائم بالملاحظة فيما يراه أو يدونه. إذ ربما يُركز في ملاحظته على ما يراه متسقاً مع قيمه ويغفل ما عدا ذلك أو العكس. كذلك ربما تلعب الاتجاهات العقلية دوراً في توجيه الملاحظة وجهة معينة مما يؤدى إلى اختلاف النتائج إذا ما قام فرد آخر بالملاحظة.

(٣) **توقعات جماعة معينة**: سلوك جماعة منظمة من الأفراد، مثل أعضاء أحد النوادي، أو إحدى الجمعيات، أو طلاب فصل معين يكون عادة متسقاً ومتواافقاً مع معاييرها وأخلاقياتها مما يجعل القائم بالملاحظة مدركاً لما هو مقبول في الجماعة، وما هو غير مقبول. وربما تتأثر ملاحظته بهذا الإدراك المتعلق بتوقعاتها لمسايرة الجماعة أو الامتثال لها.

بعض الاقتراحات للتغلب على أخطاء الملاحظة:

على الرغم من الصعوبات المتعلقة بالحصول على بيانات دقيقة و المناسبة للغرض من الملاحظة، إلا أنه يمكن التغلب على هذه الصعوبات وتقليل أخطاء الملاحظة لكي تكون أكثر صدقاً بالاسترشاد بالاقتراحات التالية:

(١) **تحديد جوانب السلوك المراد ملاحظته**: فقد سبق أن أوضحتنا في كثير من الموارد أن تحديد محتوى المجال المراد قياسه في ضوء الأهداف المرجوة يُعد خطوة أساسية في كثير من الاختبارات، وبخاصة الاختبارات التحصيلية. كذلك فإن تحديد جوانب أو أقسام معينة من السلوك الذي نود ملاحظته في ضوء الغرض من الملاحظة والسمات المرجوة يُعد من الأمور الضرورية.

فإحدى المشكلات التي يواجهها القائم بالملاحظة هو عدم معرفته ما ينبغي ملاحظته، كما أنه لا يستطيع ملاحظة وتسجيل كل ما يحدث في موقف معين. فلكل تكون الملاحظة منظمة ينبغي تحديد طبيعة البيانات والمعلومات المناسبة التي ينبغي الحصول عليها، وكذلك الأسلوب الذي سوف يستخدم في تسجيل عينات ممثلة من

السلوك. ويمكن الاسترشاد في ذلك بأسلوب معاينة الأحداث Incident Sampling Technique، الذي يتعلق بمشاهدة وتسجيل أحداث محددة، أو أسلوب المعاينة الزمنية Time Sampling، الذي يزيد من فاعلية الملاحظة عن طريق إجراء سلسلة متتابعة من الملاحظات يستغرق كل منها دقائق محدودة لمدة يوم واحد أو أكثر. وهذه الأساليب لا يكون لها تأثير في الاستجابة كرد فعل Unobtrusive Measures المشكلة الرئيسية في قياس الشخصية هي أن عملية القياس ربما تؤثر في سلوك الفرد المستجيب، وبالتالي تؤثر في القياسات التي تحصل عليها، أي أن القياسات تتفاعل مع الفرد وتؤدي إلى تغيير نظرته إلى نفسه، أو أن يختار دوراً معيناً لمحاولة تحقيق ما يرى أنه يتفق ومطلب القائم بالمشاهدة.

(٢) تدريب القائمين بالملاحظة: على الرغم من التحديد الإجرائي لجوانب السلوك المراد ملاحظته، إلا أنه يظل هناك احتمال عدم اتفاق بين القائمين بالملاحظة لأسباب متعددة. لذلك ينبغي تنظيم جلسات لتدريبهم على كيفية الملاحظة الموضوعية لعينة من السلوك، والتمييز بين الحقائق والتفسيرات في تسجيل ملاحظاتهم وكتابة التقارير ، والمقارنة بين النتائج التي توصل إليها كل منهم ومناقشة الاختلافات فيما بينها من أجل تحقيق الأنساق في التفسيرات، وتفصيل أقسام السلوك المرجو ملاحظته.

وتتطلب أساليب الملاحظة استخدام بعض الأجهزة التي يمكن أن تراوح بين أجهزة بسيطة وقليلة الكلفة نسبياً، مثل أجهزة التسجيل، وساعات التوقيت، وأجهزة مرتفعة الكلفة، مثل الأجهزة الإلكترونية المعقدة التي يمكن استخدامها في عرض المثيرات، وتسجيل السلوك، وبخاصة إذا كان الهدف قياس الاستجابات أو ردود الفعل الإدراكية والحركية، مما يتطلب تدريب القائمين بالملاحظة أيضاً على كيفية استخدام هذه الأجهزة استخداماً مناسباً.

(٣) تكميم الملاحظات: فمشاهدة بعض مظاهر السلوك، مثل التفاعل الاجتماعي والاستجابات الانفعالية، وسلوك المعلم وإدارته للفصل المدرسي تتطلب التعبير عن هذه الملاحظات بمقاييس كمية. وهذه المقاييس تمثل عادة تكرار حدوث سلوك معين في مدة زمنية محددة. غير أنه ينبغي في هذه الحالات تصميم خطة زمنية دقيقة يحدد في ضوئها عدد مرات حدوث السلوك الذي نهتم بملاحظته.

(٤) تحديد إجراءات تسجيل السلوك: فالبيانات التي يتم جمعها عن طريق الملاحظة ينبغي أن تتميز بالدقة، وهذا يتطلب تسجيل السلوك فور حدوثه حتى لا يتغير ذلك بالأخطاء التي تنتهي عن العمليات الانتقائية في التذكر .

ولعل التقنيات المتطرورة للملاحظة وتنظيم البيانات سوف تسهم في التغلب على مشكلات التسجيل الفوري الفاعل للسلوك أو الأحداث التي تجري ملاحظتها. فمثلاً يمكن تطوير نظام ترميزى لأقسام السلوك أو الأحداث المعينة ، كأن يُعطى الرمز (أ) لحدث أو سلوك معين ، والرمز (ب) لحدث أو سلوك آخر مرتبط بالظاهرة موضع الملاحظة وهكذا. كما يمكن تحديد مرات تكرار الرمز في مدة زمنية معينة أثناء عملية الملاحظة .

ونظراً لأن إجراء ذلك ينبغي أن يستند إلى أساليب منهجية أكثر اتساقاً ودقة ، فسوف نوضح فيما يلى بعض هذه الأساليب التي تُستخدم في تسجيل الملاحظات ، وهي موازين تقييم السلوك Behavioral Rating Scales، وقوائم Checklists . المراجعة .

موازين تقييم السلوك :

تُعد موازين تقييم السلوك من أكثر الأساليب استخداماً في تقييم الشخصية ، حيث تشمل على السلوك المرجو ملاحظته في موقف واقعية . فالتقدير Rating يتضمن الحكم على خصائص أو سمات الفرد في موقف مضبوطة ، ويختلف الحكم على السلوك Behavioral Rating عن الحكم على السمات Trait Rating في أن الأول يتعلق بالتعتميم والاستدلال . وميزان تقييم السلوك يُعد أداة تُستخدم في تسجيل قيم تقديرية لما نود قياسه ، وبذلك يفترض أنه يزود القائم بالملاحظة بوسيلة لتقدير السلوك بمزيد من الدقة والموضوعية .

وقد سبق أن قدمنا في الفصل الثالث عشر أنواعاً متعددة من موازين تقييم الاتجاهات ، وأوضحنا أساليب بنائها وتحليل فقراتها ، وكذلك كيفية إعدادها في الفصل المدرسي: غير أن هذه الموازن تُستخدم في تقييم الفرد لاتجاهاته بنفسه ، أي أنها تُستخدم في التقدير الذاتي Self-Rating ، بينما تُستخدم موازين تقييم السلوك بواسطة القائم بالملاحظة لتقدير سلوك فرد يفترض أن دوره يكون سليماً في عملية التقدير أو الحكم ، ويعين القائم بالملاحظة قيمة رقمية للسلوك الملاحظ . ومع هذا فإن موازين التقدير لا تختلف في كل من الحالتين . فميزان تقييم السلوك يُعد أسلوبياً منظماً للحكم على سلوك ملاحظ وتسجيله وتقريره ، وبذلك يمكن التمييز بدرجة أكثر دقة بين الأفراد في سمة معينة . كما يمكن استخدامها في تقييم الإجراءات والعمليات والنتائج ، والنمو الشخصي والاجتماعي للطلاب في خصائص متعددة ، مثل الالتزام ، والتحمّس ،

ومراقبة الآخرين ، والانبساط ، وغير ذلك . وكذلك تُستخدم في تقدير أداء العاملين وسماتهم الشخصية كما سبق أن أوضحنا في الفصل العاشر ، وفي تقدير طلاب الجامعات لفاعلية أداء أساتذتهم .

أنواع موازين تقدير السلوك :

تعدد أنواع موازين تقدير السلوك Behavioral Rating Scales كما سبق أن أوضحنا في فصول سابقة ، غير أنها سوف نتناول مرة أخرى بإيجاز بعض هذه الأنواع الرئيسية كما هو موضح بالشكل التخطيطي (٧-١٣) التالي :



شكل (٧-١٣) يوضح بعض أنواع موازين تقدير السلوك

موازين تقدير رقمية : Numerical Ratings

تُعد أبسط أنواع الموازين ، حيث يُعين القائم بالملحوظة قيمة عدديه أو رقمياً يدل على درجة توافر سلوك أو خاصية معينة لدى الفرد ، مثل النشاط العام ، أو الجدية ، أو الكفاءة الشخصية ، وغير ذلك . ومعظم الخصائص الوجданية والشخصية يمكن تقديرها بموازين تقدير رقمية . ويعد أسلوب تمييز معانى المفاهيم Semantic Differential Technique الذى سبق أن أوضحناه بالتفصيل في الفصل الثاني عشر نوعاً من موازين التقدير الرقمية ، ويمكن أيضاً باستخدامه تقدير كثير من هذه الخصائص . وفي موازين التقدير الرقمية تُقدم الخاصية أو السمة في عبارة معينة يليها قيم تتراوح بين ١ ، ٥ مثلاً ، ومثال ذلك ما يلى :

« يتسم بالاكتفاء الذاتي في علاقته بالآخرين »

٥ ٤ ٣ ٢ ١

حيث ١ = بدرجة مرتفعة ، ٢ = بدرجة فوق المتوسط

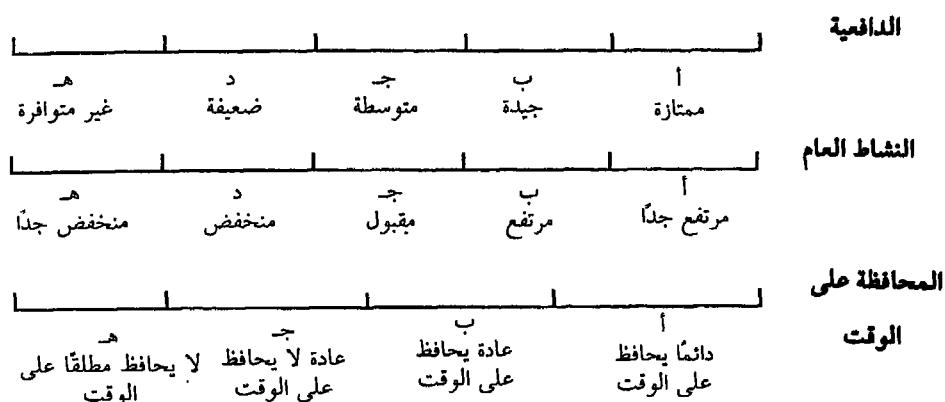
٣ = بدرجة متوسطة ، ٤ = بدرجة أقل من المتوسط

٥ = بدرجة منخفضة

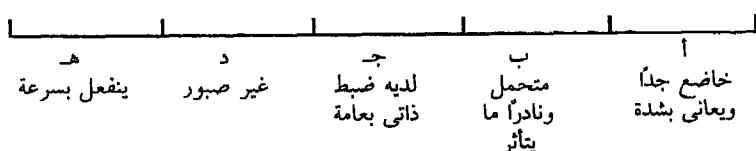
وعادة تكون هذه القيم العددية وما تشير إليه مشتركة بين جميع فقرات ميزان التقدير.

موازين تقييم بيانية : Graphic Rating Scales

يُعد هذا النوع من الموازين أكثر استخداماً من موازين التقدير الأخرى، ويطلب أيضاً أن يعين القائم بالملاحظة قيمة عددية للخاصة المعينة، غير أنه في هذه الحالة يكون هناك خط أفقى أو رأسى يمثل خصائص أو أوصاف للسلوك على مُتصل، ويُقسم الخط إلى عدد من المسافات تمثل عدد التقديرات لكل قسم مُحدد بعنوان معين أسفله. ويستطيع القائم بالملاحظة وضع علامة في أي مكان يراه مناسباً على الخط دون تقدير بنقط معينة . ويوضح شكل (٨١٣) بعض الأمثلة لأنواع موازين التقدير البيانية:



كيف يستجيب للإحباط أو المواقف غير السارة؟



شكل (٨١٣) يوضح أمثلة لبعض موازين التقدير البيانية

ومن ميزات هذا النوع من موازين التقدير سهولة تسجيل التقديرات ، وإمكانية تقييم عدد من الخصائص أو السمات في صفحة واحدة مع جعل الميزان مشتركاً بين الخصائص جميعها، وبذلك يمكن رسم صفحة نفسية (بروفيل) لهذه الشخصيات لكل فرد.

- وقد أوضح جيلفورد Guilford بعض القواعد التي يمكن الاسترشاد بها في إعداد موازين التقدير البيانية نوجزها فيما يلى :
- (١) ينبغي أن تشغل كل سمة أو خاصية صفحة مستقلة ، حيث إن هذا يسمح بتقدير كل سمة على حدة قبل الانتقال إلى السمة التالية .
 - (٢) يجب ألا يقل طول الخط عن ١٢ سم ، ولكن لا ينبغي أن يزيد عن ذلك كثيراً لكي يتسع استيعابه .
 - (٣) يجب أن يكون الخط متصل دون فجوات ، حيث إن الفجوات تعطى انطباعاً بأن المتغير منفصل .
 - (٤) ينبغي تبادل الكلمتين المماثلين لطرف الخط بطريقة عشوائية لتجنب التزعع الآلية لوضع علامة على أحد الطرفين دون الآخر .
 - (٥) يُفضل استخدام ٣ أو ٥ كلمات وصفية على الخط؛ اثنان منها طرفيان ، والباقي في الوسط .
 - (٦) ينبغي أن تكون الكلمات الوصفية موجزة وحروفها صغيرة ، بحيث يكون هناك فراغ بين كل كلمتين .
 - (٧) ينبغي استخدام كلمات وصفية متعارف عليها وأن تكون مصاغة صياغة لغوية صحيحة .
 - (٨) لا يجب أن تكون الكلمتان طرفيان متطرفيان جداً في معناهما لكي لا يت Jenningsهما القائم باللحظة .
 - (٩) ينبغي أن تكون الكلمات الوصفية المحايدة مقابلة لمتصف الخط .

موازين تقدير جبرية : Forced - Choice Rating Scales

سبق أن أوضحنا في الفصلين الحادى عشر والثانى عشر أن موازين التقدير الجبرية اقترحت لتقليل تزييف الاستجابات في مقاييس الميل والاتجاهات .
وتُستخدم هذه الموازين لتقليل تحيز القائم باللحظة في تقديره لخصائص وسمات شخصية الأفراد . ومثال ذلك نزعته إلى إعطاء تقييمات مرتفعة أو منخفضة لفرد معين؛ لذلك تشتمل كل فقرة من فقرات الميزان على أربع عبارات اثنين منها متكافئان في القبول الاجتماعي ، والاثنان الآخران متكافئتان في عدم القبول الاجتماعي ، غير أن كل عبارتين يختلفان في درجة تمييزهما بين مجموعات محددة معينة Criterion Groups . ونظراً لأن القائم باللحظة لا يكون على دراية بالعبارات

المميزة، فإنه يصعب تحيز تقديراته، حيث يتطلب منه وضع علامة على العبارة التي تصف الفرد بدرجة أفضل، والعبارة التي تصفه بدرجة أقل، وبذلك يتناول أكثر من سمة في آن واحد. ومثال ذلك في تقدير أداء المعلم ما يلى:

- يتحلى بالصبر مع الطلاب بطبيئي التعلم.
- يثق في نفسه أثناء الشرح.
- يهتم ويتابع طلاب الفصل.
- يُخبر الطلاب مقدماً بأهداف الدرس.

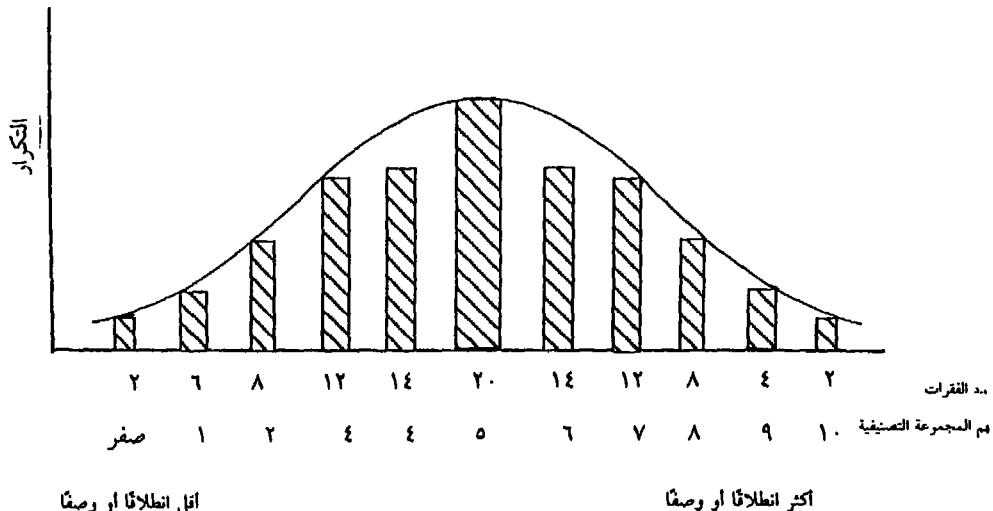
ومثال آخر:

- لا يعرف متى وأين يتدخل في الحوار.
- يتحمل المسئولية بسهولة.
- لا يستمع لاقتراحات الآخرين.
- يقدم مقترنات بناءً.

وقد سبق أن أوضحنا بالتفصيل مزايا وعيوب هذا النوع من موازين التقدير.

وتُقدر الدرجات بتعيين درجة واحدة موجبة (+1) لكل عبارة مقبولة اجتماعياً وصادقة، ودرجة سالبة واحدة (-1) لكل عبارة غير مقبولة اجتماعياً وصادقة.

ويوجد أسلوب آخر بديل لموازين التقدير الجبرية سبق أن ناقشناه في الفصل الثاني عشر وهو أسلوب التصنيف الترتيبى Sort الذي اقترحه ستيفنسون Stephenson, 1953، ويُستخدم في أغلب الأحيان في تقدير مفهوم الذات Self Concept، حيث يُعطى القائم بالتقدير عدداً من العبارات الوصفية المتعلقة بالشخصية، ويُطلب منه تصنيفها في مجموعات مرتبة ترتيباً تنازلياً وفقاً لمدى انطباقها على الفرد أو وصفها له أو شيوعها. وينبغى أن يضع عدداً معيناً من الفقرات أو العبارات في كل قسم بحيث يكون توزيع التقديرات اعتدالياً كما هو موضح بشكل (٩-١٣) التالي:



شكل (٩-١٣) يوضح التصنيف الترتيبى Sort Q لعدد قدره ١٠٠ فقرة أو عبارة

ويتبين من شكل (٩-١٣) أن معظم الفقرات أو العبارات تقع وسط التوزيع ، ويقع عدد قليل منها في كل من طرفيه .

موازين تقدير قوائم مراجعة : Checklist Rating Scales

يُعد هذا النوع من موازين التقدير أكثر فاعلية في تقييم الخصائص والسمات الشخصية . فالقائم باللحظة يضع علامات على جميع العبارات التي تشتمل عليها قائمة مراجعة والتي يتصرف بها فرد معين . وقد سبق أن أوضحتنا هذه القوائم في الفصل الحادى عشر .

وعلى الرغم من أن العبارات تكون مرتبة عشوائياً في القائمة ، إلا أنها تتراوح بين عبارات مفضلة وعبارات غير مفضلة وتعلق بسمات متعددة . وللحصول على الدرجة الكلية للتقدير ، يوجد متوسط درجات تقديرات جميع العبارات . ويمكن استخدام أسلوب الفقرات المتساوية ظاهرياً الذي اقترحه ثيرستون في تحديد القيم المناظرة لكل عبارة . وقد ناقشنا ذلك بالتفصيل في مجال قياس الاتجاهات في الفصل الثاني عشر .

كما يمكن استخدام هذه القوائم في تقيير المشكلات الشخصية وذلك في مجال الإرشاد المدرسي من أجل التعرف على الطلاب الذين لديهم مشكلات شخصية معينة . ومن أمثلة ذلك قوائم مونى للمشكلات Mooney Problem Checklists ، وتتكون هذه القوائم من أربع صيغ للمستويات العمرية المختلفة ، ويشتمل كل منها على عدد من

المشكلات يتراوح بين ٢١٠ ، ٣٣٠ مشكلة في المجالات الصحية، والجسمية، والنمو، والمنزل، والأسرة، والعلاقات بين الجنسين، والأخلاقيات، والتدين، والزواج، والأمن الاقتصادي، والمدرسة أو العمل، والمجالات الاجتماعية والترويحية. ويطلب من الفرد وضع خط تحت المشكلات الأقل أهمية، ويوضع دائرة حول المشكلات الأكثر أهمية، ويكتب ملخصاً عن مشكلاته. وتقدر درجات تناظر عدد المشكلات في كل من هذه المجالات، وتفسر الاستجابات بطريقة ذاتية، حيث إنه لا توجد معايير لهذه القوائم.

ومن الأمثلة الأخرى قائمة الصفات (ACL) التي أعدّها جوخ ، وهيلبرن (1965 , Gough & Heilburn)، وتشتمل على ٣٠٠ صفة مرتبة أبجدياً، وتعلق بمشكلات شخصية، مثل شارد الفكر، ومُضحك، ويستغرق تطبيقها ٢٠ دقيقة. وتحتلت عن قوائم مونى بأن تفسير درجاتها يستند إلى معايير الدرجات التائية، وتُعد هذه القوائم من نوع التقرير الذاتي .

وكذلك قائمة سلوك الطفل (CBCL) Child Behavior Checklist التي أعدّها أشنباك ، وإدلبوك (Achenbach & Edelbrok, 1991) لكي يستخدمها الآباء في تقدير الكفايات والمشكلات السلوكية لأطفالهم، حيث قاما بتحديدها باستخدام التحليل العاملى .

وفيما يلى أمثلة أخرى لبعض موازين تقدير السلوك المتوافرة في مجال تقييم الشخصية في الصدف المدرسي :

(١) بروفيل تقدير السلوك (BRP-2) : Behavior Rating Profile (BRP-2)

أعدّ براون ، وهاميل (1990 , Brown & Hammill) واستناداً في بنائه على المدخل البيئي Ecological Approach، حيث حصل على معلومات من مصادر بيئية متعددة . ويشتمل على موازين تقدير للطلاب والمعلمين والآباء والعلاقات الاجتماعية، وبذلك يمكن تقويم الطالب بطريقة فردية في موافق متعددة بواسطة عدد من المحكمين، ويناسب الأعمار التي تتراوح بين ٦،٥ ، ٨،٥ عاماً .

(٢) نظام تقدير المهارات الاجتماعية

: Social Skills Rating System (SSRS)

أعد جريشام ، والليوت (Gresham & Elliot , 1990) وبعد بمثابة بطارية تسمح بتقديرات متعددة للسلوك الاجتماعي الذي يمكن أن يؤثر في علاقات المعلم بالطلاب ، وعلاقة الطلاب فيما بينهم ، والتحصيل الدراسي . وت تكون البطارية من صيغتين :

إحداهما تتعلق بالمعلمين والآباء، والأخرى تتعلق بالطلاب في مختلف المراحل التعليم العام.

(٣) ميزان تقييم الكفاية الاجتماعية والتكييف المدرسي لوالكر، ومكون

Walker - McConnell Scale for Social Competence and School Adjustment (W-M Scale):

وقد أعدَّ عام ١٩٨٨ لتعريف الأطفال الذين يفتقرُون إلى المهارات الاجتماعية، ويستغرق تطبيقه خمس دقائق فقط. ويشتمل على ثلاثة موازين تقييم فرعية تتعلق بالسلوك الاجتماعي الذي يُفضله الأطفال لدى معلميهم، ولدى أقرانهم، والتكييف المدرسي. ويمكن أن يطبقه المعلم الذي أتيحت له فرصة ملاحظة أطفال صف معين مدة شهرين على الأقل، وتوجد معايير لهذه الموازين من الصف الأول إلى السادس.

(٤) موازين تقييم السلوك لبيركس

: Burks Behavior Rating Scales (BBRS)

أعدَّها بيركس (Burks, 1977) لتعريف بعض أنماط السلوك غير المقبول للأطفال في المرحلتين الابتدائية والإعدادية، وذلك بهدف الحصول على معلومات تقييد المعلم وأولئك الأمور في تعرف السلوك المضطرب لدى بعض الأطفال، والتنبؤ بأدائهم في فصول الثنائيات الخاصة

وتُستخدم موازين تقييم قوائم المراجعة إذا كنا نهتم بمعرفة ما إذا كانت خاصة أو سمة معينة متوفرة أم لا. لذلك ينبغي تحديد الخصائص أو السمات المرجو ملاحظتها، وأن يكون لكل فرد قائمة مراجعة خاصة به، ويتم ملاحظته فردياً في ضوء العبارات أو الفقرات التي تشمل عليها القائمة.

وقد سبق أن أوضحنا في الفصل السابق بعض الإرشادات العامة لبناء موازين التقدير.

مصادر الخطأ في موازين التقدير، وكيفية التخلُّب عنها:

هناك العديد من مصادر الخطأ الشائعة في أساليب التقدير، بعضها يرجع إلى موازين التقدير، وبعضها الآخر يرجع إلى القائم بعملية التقدير وإلى السمات المراد تقاديرها. وفيما يلى توضيح لأهم مصادر هذه الأخطاء:

(١) **تأثير الظاهرة Halo Effect** : ويحدث هذا الخطأ عندما يؤثر الانطباع العام للقائم بالتقدير عن الفرد في كيفية تقديره له في سمات مختلفة. فإذا كان لديه انطباع جيد بعامة عن فرد معين، فإنه يميل إلى إعطائه تقدير مرتفع في معظم السمات،

وعكس ذلك صحيح. ويؤدي هذا الخطأ إلى قيم مرتفعة زائفة لمعامل الارتباط بين سمات تبدو غير مرتبطة. كما تؤدي إلى إخفاء الفروق الفردية، وذلك لأن تأثير الهالة يجعل التقديرات متراكمة عند طرف ميزان التقدير، مما يقلل من صدق التقديرات.

ويمكن التغلب إلى حد ما على هذا المصدر من الخطأ بتعريف السمات تعريفاً إجرائياً دقيقاً، وكذلك بتبادل الكلمتين أو الصفتين المماثلتين لطرف الميزان بطريقة عشوائية في موازين التقدير البيانية كما سبق أن أشرنا.

(٢) التساهل في التقديرات Leniency : يتزع بعض القائمين بعملية التقدير إلى التساهل في تقديراتهم للأفراد، وبذلك يتتجنبون استخدام الطرف الأيسر الذي يمثل القيمة المنخفضة على ميزان التقدير. فكثير منهم لا يفضل الحكم على سمة معينة لدى الفرد بأنها ضعيفة جداً لكي لا يبدو متشددًا. ويمكن أيضاً أن يحدث العكس حيث يتزع البعض إلى التشدد في التقديرات، وبذلك يتتجنب استخدام الطرف الأيمن لميزان التقدير.

فالخطأ في كلتا الحالتين يُعد خطأً متسقاً في اتجاه تحيز القائمين بعملية التقدير. وهذا المصدر من مصادر الخطأ لا يكون تأثيره ذات أهمية إذا قام بعملية تقدير السمات نفسها فرد واحد أو عدة أفراد، حيث إن اتجاه تحيزهم يكون معلوماً، غير أنه إذا قام عدة أفراد بعملية التقدير، وأجريت مقارنة بين تقديراتهم، فإنه ينبغي تعرف اتجاه تحيزهم لكي يؤخذ بعين الاعتبار.

وهذا المصدر من مصادر التحيز يقلل من مدى ميزان التقدير مما يقلل بالتالي من تميز التقديرات. ويمكن التغلب على ذلك باستخدام موازين التقدير جبرية الاختيار، أو جعل عدد الأقسام الأكثر تفضيلاً أقل من عدد الأقسام الأقل تفضيلاً على ميزان التقدير، وفي بعض الأحيان يمكن استخدام موازين التقدير البيانية بدلاً من الموازين جبرية الاختيار.

(٣) التزعة المركزية Central Tendency : ويحدث هذا الخطأ عندما يتتجنب القائم بالتقدير طرف الميزان ويركز تقديراته في الوسط، وهذا يقلل من الفروق الفردية في تقديرات الأفراد مما يقلل بدوره من ثبات هذه التقديرات. ويمكن التغلب على ذلك أيضاً باستخدام الموازين جبرية الاختيار، وكذلك بتوضيح الغرض من عملية التقدير للقائمين بها بحيث ي عملون على إبراز الفروق الفردية للتقديرات بدلاً من العمل على تلافيها.

(٤) غموض السمات المراد تقديرها Ambiguity Error : ويحدث هذا الخطأ عندما تكون السمات المراد تقديرها غير معرفة تعريفاً جيداً، فسمات مثل الدافعية، وسوء التكيف، والانقباض ربما تفسر تفسيرات مختلفة. كما أن الكلمات الوصفية التي

تحدد أقسام موازين التقدير الбинانية ربما لا تكون واضحة المعنى ، فالفرق بين «بدرجة كبيرة» ، و«مرتفع» غير واضح . وللتغلب على ذلك ينبغي العناية بتعريف السمات والخصائص تعریفًا دقیقًا بحيث يستند القائمون بالتقدير إلى إطار مرجعي محدد . كما يمكن وضع التعريف إلى جوار السمة على ميزان التقدير ، وانتقاء الكلمات الوصفية بحيث لا تحتمل أكثر من معنی واحد كما يلى :

الداعية (قدرة الفرد على الاستمرار في العمل حتى يستكمله)



(٥) **عدم قدرة القائم بالتقدير:** ترجع مصادر الخطأ السابقة في معظمها إلى عدم استخدام القائمين بالتقدير ممحكات منظمة . وكذلك إلى عدم تدريبهم على مهارات الملاحظة والتقدير وتسجيل البيانات ، وهذا يؤدي إلى انخفاض ثبات التقديرات . ويمكن التغلب على ذلك بتنظيم دورات تدريبية للقائمين بالملاحظة وذلك لتوضیح أهمية تحرّی الدقة والأمانة في التقدير ، والتعريف بمصادر الخطأ الشائعة وكيفية تجنبها أو التقليل منها ، وتطبيق ذلك في بعض المواقف المماثلة للواقع الفعلى .

كما ينبغي توضیح معنی السمات أو الخصائص المراد تقدیرها فی عبارات وصفیة سلوکیة يمكن ملاحظتها بحيث لا تتطلب إلا قدرًا ضئیلاً من الاستدلال . ولعل أسلوب الأحداث الحرجية Critical Incidents Technique الذي أوضحته في الفصل العاشر يساعد في تحديد السلوك الأكثر أهمية الذي ينبغي ملاحظته . وكذلك توضیح أقسام ميزان التدريج ودلالتها ، والتدريب على استخدامها في الملاحظة من خلال ورش عمل منتظمة .

ابا أساليب المقابلة : Interview Techniques

تُعد المقابلة من أكثر الأساليب المستخدمة في تقييم الشخصية ، وتُزودنا بمعلومات تختلف عما تزودنا به أساليب ملاحظة السلوك ، حيث إنه في المقابلة يتحدث الفرد عن نفسه ، بينما في الملاحظة يتم تقدیر سلوكه . فال مقابلة تُستخدم في أغراض متعددة في مجالات الإرشاد ، وال المجالات الكلينيكية ، والتربوية ، والاجتماعية ، والصناعية ، والقانونية ، واستطلاع الآراء ، وكذلك في اتخاذ قرارات تتعلق بانتقاء الأفراد وترقيتهم ، وفي تقويم البرامج التربوية لجمع بيانات تتعلق بخلفية الطلاب وتعرف وجهة نظرهم حول برنامج معین ، واستطلاع آراء المعلمین والإداريين حول محتوى وأسلوب تنفيذ البرامج التعليمية ، كما تُستخدم في الأغراض البحثية .

وال مقابلة تتضمن تفاعلاً لفظياً شفوياً وجهاً لوجه بين فردین أو أكثر، حيث يحاول أحدهم وهو القائم بال مقابلة Interviewer أن يتعرف مثلاً على مفهوم الذات لدى الفرد ، وعقائده، ومقاصده، وأسلوب تفكيره، وآرائه وغير ذلك من خصائصه الشخصية .

وعلى الرغم من أن المقابلة لا تُعد أداة قياس بالمفهوم السيكومترى ، لأنها لا تؤدى عادة إلى بيانات كمية ، إلا أنها تقدم بيانات ثرية إذا قام بها أخصائى نفسى أو تربوى مدرب تدريباً جيداً، بل وأحياناً يمكن أن تُزودنا ببيانات كمية كما هو الحال في أساليب الملاحظة ، إذ يمكن اعتبار المقابلة استبياناً شفوياً إذا روعى تقنين الأسئلة التي تُوجه للفرد .

فال مقابلة تمدنا بمعلومات كلينيكية وفيزيولوجية ، وتعتمد هذه المعلومات على الإطار النظري الذى يستند إليه القائم بال مقابلة ، فهناك أطراً نظرية مختلفة للمقابلة تنسب إلى فرويد Freud ، وروجرز Rogers ، وساليفان Sullivan ، وغيرهم . فال مقابلة عند فرويد Freud التي تعتمد على التحليل النفسي يكون الاهتمام فيها مركزاً على تحديد طبيعة صراعات الفرد والكشف عن اللاشعور . وهنا يهتم الأخصائى بالمعانى المتعددة الشعورية واللاشعورية المقصودة وغير المقصودة فى مفردات لغة الفرد الذى تُجرى معه المقابلة ، ويساعد الأخصائى الفرد على استرجاع ذكريات ماضيه ليساعده فى إدراك منشأ الصراعات التى ربما يعاني منها . وهنا يعتمد الأخصائى فى تفسيره للمعلومات التى جمعها عن الفرد على أكثر مما تسمع به هذه المعلومات ، أي يعتمد على فرضيه الخاصة عن طبيعة القوى النفسية التى تحكم السلوك الإنساني ، وهذه الفروض النظرية تُوجه عملية المقابلة ويتم التفسير فى ضوئها (Freud, 1953) . أما المقابلة عند روجرز Rogers التي تعتمد على المدخل الفينيـزـوـلـوـجـي فتتركز حول العميل ، ويسأل فيها الأخصائى عدداً قليلاً من الأسئلة المباشرة ، ويعيد صياغة إجابات الفرد عنها ، ويُحثه على الاستمرار فى الإجابة ، ويتجنب أي تفسيرات لا تقوم على أساس المعلومات التى حصل عليها ، أي أن الأخصائى هنا يهتم بالطريقة التى يشعر أو ينظر بها الفرد إلى نفسه .

ويؤكد هذا المدخل مفهوم الذات Self Concept ، وهو الجزء من المجال الظاهري الذى يتكون من تركيبة من الإدراكات والقيم المتعلقة «بالذات» أو بالفرد كمصدر للخبرة والسلوك ، أي أن «الذات» هى المحور الرئيسي للخبرة . لذلك فإن

المقابلة تُعد أحد الأساليب المهمة التي يمكن باستخدامها الحصول على معلومات تتعلق بعفوم الذات (Combs, 1959).

أنواع أساليب مقابلة:

تنوع أساليب مقابلة وفقاً للغرض من استخدامها، غير أنه يمكن تصنيفها في نوعين رئيسيين هما: المقابلة غير المقنتة Unstandardized Interview، والمقابلة المقنتة Standardized Interview، وتباين المقابلة في مدى تحديد بنيتها، فهي تتراوح بين البنية المحددة تحديداً جيداً، والبنية غير المحددة إطلاقاً. فالمقابلة محددة البنية Structured Interview تعتمد على قائمة من الأسئلة مُعدة مسبقاً، أما المقابلة غير محددة البنية Unstructured Interview فتعتمد على مجموعة من الأسئلة، ولكن يمكن تعديلها بالحذف أو بالإضافة أثناء المقابلة. وتوجد أنواع متعددة من المقابلة بغض النظر عن مدى تحديد بنيتها، وتختلف فيما بينها من حيث الغرض الذي تُستخدم من أجله، فهناك المقابلة الشخصية لانتقاء الأفراد Personnel Interview، أو تعرف مشاكلهم أثناء العمل Trouble Shooting Interview ، والمقابلة المتعلقة بالاقتراع Poll Interview، والمقابلة الكlinيكية Clinical Interview التي تُستخدم في المجالات الطبية والنفسية والاجتماعية، مثل المقابلة الطبية التشخيصية Medical Diagnostic Interview، ومقابلة الانصباب أو مقابلة الاستجابة الدفاعية- Defensive- Response (Stress) Inter- view، والمقابلة الطبية النفسية المدخلية Psychiatric Intake Interview، والمقابلة لتحديد الحالة العقلية Mental Status Interview، والمقابلة لدراسة تاريخ الحالة Research Interview، وهناك أيضاً المقابلة البحثية Case History Interview، مثل المقابلة المسحية لتحديد الحالة النفسية الراهنة لمجتمع من الأفراد، والمقابلة التي تتم بعد إجراء تجربى معين Postexperiment Interview وكذلك توجد أساليب محاكاة المقابلة Interview Simulation باستخدام برامج حاسوب يتم تصميمها لإجراء نوع معين من المقابلة الاصطناعية ، وغير ذلك.

ونظراً لضيق المجال، فإننا لن نستطيع مناقشة جميع هذه الأنواع، وإنما سوف نقتصر فيما يلى على توضيح النوعين الرئيسيين السابق ذكرهما.

المقابلة غير المقنتة: يُعد هذا النوع من المقابلة غير محدد البنية، حيث إن الأسئلة التي توجه للفرد لا تكون محددة سلفاً، وإنما تتحدد بناء على استجاباته، فهذه المقابلة تمثل الحديث العادي بين فردين، وتُستخدم عادة في المقابلات الشخصية

للمتقدمين للوظائف أو للالتحاق بالكليات. وعلى الرغم من أنه يصعب الحصول على بيانات كمية استناداً إلى المقابلة غير المقنتة، إلا أنها تتميز بتنوع الاستجابات نظراً لمرونة الأسئلة، وتسمح للقائم بالمقابلة أن يختار المجالات التي يهتم بها الفرد الذي تُجرى معه المقابلة والتي لم تكن متوقعة، أي أنها تهتم بفردية الشخص وتجعله في حالة ارتياح ولا يلجأ إلى تمثيل الأدوار، وبذلك تساعد في التحليل الفينومولوجي. غير أنه يُعاب عليها صعوبة المقارنة الكمية بين الأفراد أو بين الأسئلة التي تُقدم لهم نظراً لأنها تختلف باختلاف المواقف والقائمين بالمقابلة، ولكن يمكن معالجة البيانات معالجة كيفية أو تحليلها تحليلاً موضوعياً. كما يمكن أن يعتمد التفسير اعتماداً مباشراً على ما تم التوصل إليه من معلومات، أو يمكن أن يتضمن التفسير فروضاً أبعد مما تسمح به هذه المعلومات.

وهذه المعالجة الكيفية لها أهمية في عملية تحليل المحتوى Content Analysis التي تُستخدم في تصنيف هذه المعلومات بأسلوب منظم، وتختلف طبيعة هذا التصنيف باختلاف محتوى المقابلة والغرض منها.

المقابلة المقنتة: على الرغم من أن هذا النوع من المقابلة يستخدم فيه نمط الأسئلة نفسها كما في المقابلة غير المقنتة، ويهدف لجمع معلومات مماثلة من كل فرد؛ إلا أن الأسئلة التي تُوجه لكل منهم تكون محددة قبل بدء المقابلة، وتكون موحدة في صياغتها ، وتُقدم بالترتيب نفسه لجميع الأفراد، لذلك تُعرف باسم «المقابلة وفق جدول Schedule Interview». وبذلك تيسّر المقابلة المقنتة عملية القياس والتقدير الكمي للاستجابات، وتسمح بالمقارنة بين الأفراد، والبيانات المستمدة منها ربما تكون أكثر ثباتاً من النوع الأول.

وأحياناً تُستخدم موازين التقدير Rating Scales التي أوضناها منذ قليل إلى جانب المقابلة المقنتة كوسيلة لتلخيص المعلومات التي يتم الحصول عليها، وعادة يسجل ما دار من حوار أثناء المقابلة، ومن هنا يتضح أن نقاط قوة المقابلة المقنتة تُعد نقاط ضعف للمقابلة غير المقنتة، والعكس صحيح.

مزایا وعيوب أساليب المقابلة:

تتميز المقابلة بتنوعها بأنها تتيح الفرصة للفرد أن يُعبر عن نفسه ويتسائل، كما تتميز نسبياً بسهولة الاستفسار من الأطفال والأمينين، والحكم على أسلوب تعبيرهم إلى جانب محتوى المقابلة. غير أنه يُعاب عليها صعوبة تحليل وتصنيف الاستجابات غير محددة البنية، وكلفة تدريب القائمين بالمقابلة، وإجراء المقابلة مع كل فرد على حدة..

ومهما بلغت فاعلية التدريب فإن الخصائص الشخصية للقائم بالمقابلة مثل معتقداته وتوقعاته ونزعاته ربما تؤثر تأثيراً منظماً في استجابات الأفراد، وهذا يُعد مصدرًا من مصادر الخطأ في المقابلة، ويطلق عليه الخطأ المنتقل Contagious Bias ، كما أنه ربما يُسيء فهم ما يقوله الفرد يُسيء فهم ما يقوله القائم بالمقابلة. غير أنه يمكن تقليل أثر تحيز القائمين بالمقابلة بالعناية بتصميم إجراءات المقابلة ومحتها، وكذلك بتدريبهم لتنمية مهارات المقابلة لديهم ، وجعلهم على دراية بالمشكلات المحتملة، وأن يتم انتقاهم بحيث يكون هناك قدر من الاتفاق بين خصائصهم وخصائص الأفراد الذين سوف تجري المقابلة معهم. فكلما قلت مصادر الخطأ والتحيز في عملية المقابلة ، كانت البيانات المستمدة منها أكثر صدقاً.

تقييم أساليب الملاحظة والمقابلة :

يتضح مما سبق أن أساليب الملاحظة والم مقابلة تهدف للحصول على معلومات مهمة تتعلق بخصائص أو سمات شخصية الفرد، على الرغم من أن أساليب الملاحظة تعتمد على ما يفعله الفرد، بينما تعتمد أساليب المقابلة على ما يقوله الفرد عن نفسه؛ لذلك فإن كلاً من هذه الأساليب شائع الاستخدام في التقييم المنظم لخصائص الأفراد. ويمكن اعتبار عمليات الملاحظة والم مقابلة فناً وعلمًا، فبعض القائمين بها يجرون هذه العمليات بفاعلية أكثر من غيرهم، ويكون لديهم مهارة في تكوين علاقة وثام بينهم وبين الأفراد الذين يلاحظونهم أو يجرون مقابلة معهم. وعلى الرغم مما تتميز به هذه الأساليب من مرونة وواقعية وما تؤدي إليه من بيانات واستجابات ثرية ومتعددة ربما يصعب الحصول عليها باستخدام الاستبيانات أو الأساليب الإسقاطية، إلا أنها أيضًا لا تخلي من مشكلات متعددة ، وهذه الأخطاء إذا لم نكن مدركين لها ونعمل على تجنبها أو تقليل أثرها، فإنها تؤدي إلى خفض ثبات وصدق البيانات أو المعلومات التي نحصل عليها .

وهذا لا يأتي إلا إذا اعْتَدْتُ بتصميم أساليب الملاحظة وما يتعلق بها من موازين تقدير، وكذلك أساليب المقابلة ، مع مراعاة تقسيم هذه الأساليب ، والاستناد إلى متوسط تقديرات عدد من يقومون بالملاحظة لكي تتلاشى عوامل التحيز، وبذلك نحصل على صورة أكثر دقة لخصائص شخصية الفرد، وبالتالي يزداد ثبات التقديرات. فزيادة عدد القائمين بتقدير فرد أو عدة أفراد يماثل زيادة عدد مفردات الاختبار مما يؤدي إلى رفع قيمة الثبات، فالثبات يشير إلى استقرار أو اتساق التقديرات ، أو اتفاق القائمين بالمقابلة .

وتباين قيم الثبات بتباين القائمين باللاحظة أو المقابلة ومدى اتفاقهم، إذ وجد بعض الباحثين أن هذه القيم تتراوح بين ٢٣ ، ٩٧ ، ٠ ، ٥٧ . (Ulrich & Trumbo, 1965). وربما يرجع هذا التباين إلى أن القائمين باللاحظة أو المقابلة ينظرون إلى أشياء مختلفة، إذ ربما ينظر أحدهم إلى جوانب قوة الفرد، بينما ينظر آخر إلى جوانب ضعفه. فاختلاف تقديراتهما في هذه الحالة يرجع إلى أن كلاً منهما تستند أحکامه إلى جوانب مختلفة للفرد. كما أن التوجه النظري لكل منها يؤثر في نوع المعلومات التي يبحث عنها والتي يستند إليها في اتخاذ القرار.

أما مصادر الخطأ المتسبة مثل تأثير الاهالة، وتأثير هاوثورن Hawthorn Effect الذي يؤدي إلى تغيير استجابات الأفراد، وغير ذلك من الأخطاء فإنها تؤدي إلى خفض صدق البيانات المستمدّة من الملاحظة والم مقابلة، حيث إن الصدق يتعلق بمعنى الدرجة. فالأخطاء التي تقلل من موضوعية الملاحظة أو المقابلة تؤدي إلى أحکام غير دقيقة، وبالتالي فإنها تؤدي إلى تحيز في معنى التقديرات أو صدق التقييم، وربما يفسر ذلك تباين قيم الصدق التنبؤى لهذه الأساليب. فقد أوضح أرليك Ulrich وترامبو Trumbo أن قيم معامل الصدق التنبؤى في مجموعة من الدراسات تراوحت بين ٥٠ ، ٥٠ و ٧٢ . باستخدام محكمات مختلفة ، مثل النجاح في الكليات ، والأداء في العمل ، وأكّد ذلك دراسة كارلسون Carlson وزملائه (Carlson, et al., 1971) . وعموماً فإن صدق هذه الأساليب يختلف باختلاف الغرض من استخدامها، وطريقة تسجيل المعلومات المستمدّة منها. فكلما كانت هذه الأساليب غير مقتنة قل ثبات نتائجها وصدقها، ومع هذا فإنها تزودنا عادة بمعلومات ربما يصعب الحصول عليها من مصادر أخرى، ولكن الاعتماد على هذه المعلومات غير المتسبة أو منخفضة الصدق سوف يكون محدوداً؛ لذلك يفضل أن تكون أساليب الملاحظة والم مقابلة مقتنة، وتُستخدم في أيدي ماهرة إذا كان اتساق النتائج يُعد أمراً ضرورياً.

ومن الجدير بالذكر، أنه لا ينبغي استخدام أساليب المقابلة في التتحقق من صحة فروض بحثية معينة، وإنما يمكن استخدامها في الملاحظات الكشفية لاشتقاق فروض يمكن بحثها بأساليب أفضل (Guilford, 1971) . وعلى الرغم من هذا كله، فإن أساليب الملاحظة والم مقابلة سوف تظل من الأدوات الأساسية لتقدير الشخصية. وبالطبع تزداد فائدتها في المواقف الكlinيكية ، وموافق الإرشاد ، وفي انتقاء الأفراد ، وموافق العمل حيث يكون لها استخدامات مختلفة.

(ابحث) مقاييس محددة البنية ونماذج المنهج

(مقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك) :

اتضح لنا مما سبق أن استبيانات الشخصية، والأساليب الإسقاطية، وأساليب الملاحظة والتقدير، والمقابلة، تُعد أدوات أساسية لتقدير الشخصية، غير أن هناك أساليب أخرى تعرف بالأساليب الموضوعية Objective Techniques، وتستند إلى توجهات مختلفة لا تزال موضع دراسة وبحث. فأساليب التقرير الذاتي، والأساليب الإسقاطية تؤكد جوانب الاختلاف في السلوك الإنساني، وترتكز على الوظائف العامة في شخصية الفرد، غير أن الأدوات أو المقاييس التي تستند لها يليجاز فيما يلى ترتكز على الأنشطة التي يمكن ملاحظتها، وعلى الموقف المحيط بها، فهذه المقاييس لا تعتمد على تقديم سؤال للفرد بطريقة مباشرة مما يتطلب تقريراً ذاتياً كما في الاستبيانات، ولا تعتمد على الافتراض الذي تستند إليه الأساليب الإسقاطية ، أو المثيرات غير محددة البنية. وهذه المقاييس غير مباشرة، وواضحة الهدف وموضوعية. وأمثلة ذلك مقاييس الأنشطة النفسية الفسيولوجية، ومقاييس الموضوعية للأداء، ومقاييس القدرة الإدراكية التي تكون محددة البنية .

وتعُد كثير من هذه المقاييس اختبارات أداء، حيث يوضع الأفراد في موقف مفزن، أي يُعرض المثير نفسه على جميع الأفراد في الظروف نفسها في المختبر، وتُستخدم الملاحظة المباشرة في الحصول على معلومات دقيقة يمكن الاستناد إليها. ولكلّي يتم ضبط الموقف تكون هذه الاختبارات محدودة في نوع وعدد المثيرات التي تقدم للفرد، وفي خصائص الشخصية التي يفترض قياسها.

فالهدف من هذه الاختبارات الحصول على معلومات تتعلق بسلوك معين في ظروف مفنة. وتهتم النظريات التي تؤكد استخدام هذه الأنوع من الاختبارات الموضوعية والمحددة البنية بالعمليات المشتركة بين جميع الأفراد . ويرى المؤيدون لها أن علم الشخصية يمكن أن يتقدم بدرجة أفضل بالدراسة المنظمة لخصائص الأداء في مواقف محددة ومعرفة تعريفاً جيداً . وهذا لا يقلل بالضرورة من أهمية فهم العلاقات بين الأجزاء عند تقييم الشخصية، ولكن البحث المنظم لهذه الأجزاء يجب أن يسبق أي محاولة لفهم الشخصية (Goldfried & Kent, 1972).

وتعتمد هذه النظريات إلى حد كبير على البحوث النفسية العامة في مجالات علم النفس الاجتماعي، وعلم النفس المعرفي، وبخاصة الإدراك والتعلم، وأحياناً يصعب

التميّز بينها كنظريات في الشخصية وبين النظريات العامة للسلوك. فقد اهتمت هذه النظريات بالخصائص العامة للأفراد، كما اهتمت بأجزاء السلوك بدلاً من اهتمامها بالسلوك ككل.

مقاييس الأنشطة الفسيولوجية :

تُستخدم مقاييس الأنشطة أو الحالة الفسيولوجية Physiological Measures في تقييم التغيرات التي تحدث في الاستجابات نتيجة للإثارة أو لمواقف ضاغطة. فأنماط الموجات الكهربائية للمخ، ونبض القلب، وضغط الدم، والاستجابة السيكوجلغلانية للجلد، ونشاط الغدة الدرقية، وقدرة العضلات ، ودرجة حرارة الجسم ، وكيمياء الدم، ترتبط بالإثارة الانفعالية واليقظة أو تركيز الانتباه ، ويختلف الأفراد في النمط المميز لهذه الأنشطة الفسيولوجية. فجهاز كشف الكذب Polygraph يُستخدم في تسجيل بعض هذه الأنشطة الفسيولوجية في آن واحد في موقف مفتوح أثناء المقابلة أو أثناء قراءة كلمات أو عرض أحداث معينة مرتبطة بجريمة اتهم فيها الفرد، أو سؤاله، ويكشف فقط الاستجابة الانفعالية التي تنتجه عن ذلك. وإذا استخدم الجهاز أخصائى ماهر، فإنه يمكنه تعرُّف ما إذا كان الفرد بريئاً أو مذنبًا. غير أن القضاء الأمريكي يرى أن هذا الأسلوب ليس صادقاً بدرجة تسمح بتقديمه كأدلة للحكم.

وأجهار رسم المخ الكهربائي (EEG) Electroencephalogram الذي يُستخدم في تحديد موجات المخ الكهربائية يُعد أيضاً من أدوات التقييم النفسي الفسيولوجي في المجالات الكlinيكية للكشف عن إصابات المخ، وكذلك في طب الأعصاب.

ويرى كثير من الباحثين أن الفروق الفردية في مستوى الوظائف الفسيولوجية، أو في مقدار التغير الذي يحدث فيه يرجع إلى عوامل تشريحية وليس إلى عوامل نفسية. لذلك يُفضلون استخدام المؤشرات الفسيولوجية في المقارنة بين استجابات الأفراد لمتغيرات متنوعة لمعرفة ما يسيطر على انتباه الفرد ، وما يستثيره، وما إلى ذلك (Averill & Opton , 1968). غير أنه قد بيّنت الدراسات على الشخصيات الفصامية أن استجاباتهم الفسيولوجية للمواقف الضاغطة Stress أكبر من استجابات الأسوية.

ومع هذا، فإن نتائج البحوث بعامة أوضحت عدم وجود أنماط فسيولوجية متمايزة تقترب بانفعالات معينة، أو بخصائص مختلفة للشخصية.

مقاييس الإدراك :

تُعد هذه المقاييس حلقة وصل بين مقاييس القدرات ومقاييس الشخصية، وقد أمكن التوصل إليها من تجارب علم النفس المعرفي في مجال المعرفة Cognition والإدراك Perception. فقد لاحظ علماء النفس أن هناك فروقاً فردية اتضحت لدى المشاركين في التجارب المتعلقة بتجهيز المعلومات Information Processing، والانتباه Attention.

وأطلق هؤلاء العلماء تسميات مختلفة على هذه الفروق ، غير أنهم أشاروا جمِيعاً إلى الفروق الشخصية في استراتيجيات الضبط ، وأطلقوا عليها مصطلح «الأساليب المعرفية Cognitive Styles» الذي يشير إلى كل من القدرة والدافعية. وقد أوضحت دراسات تجريبية متعددة أن هناك علاقات دالة بين الخصائص المتعلقة باتجاهات الفرد ودوافعه وانفعالاته ، وبين أدائه في مهام إدراكية أو معرفية ، وبعد اختبار رورشاخ الذي سبق أن أوضحتناه واحداً من الاختبارات الإدراكية Perceptual Test.

وتعد دراسات هيرمان ويتكن ومعاونيه (Witkin, et al., 1954, 1967, 1975) من الجهود البارزة في هذا المجال. فقد استطاعوا تحديد أحد الأبعاد الرئيسية للأساليب المعرفية أطلقوا عليه مصطلح «الاعتماد - الاستقلال عن المجال» Field Dependence VS. Field Independence .Psychological Differentiation.

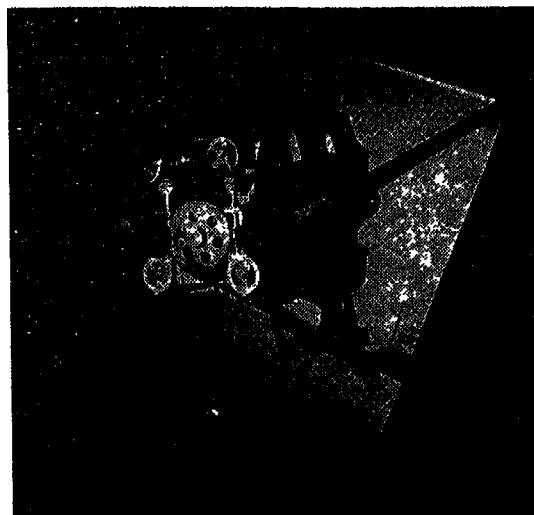
ويشير كلا المصطلحين إلى الفروق بين الأفراد في أساليب انتقاء وتنظيم موقف أو مجال معقد لإنتاج استجابة معرفية تُطلب منهم. فالفرد المستقل عن المجال أو المتمايز يمكنه فصل الأجزاء المطلوبة من المثير المربك الذي يُقدم له ، أما الفرد المعتمد على المجال أو غير المتمايز فإنه ستجيب للموقف ككل دون تحليله ، مما يدل على اهتمامه بالموقف ذاته . وقد توصل ويتكن Witkin إلى هذا البُعد المعرفي عن طريق مجموعة مبتكرة من الاختبارات التي يبدو أنها مترابطة وتُستخدم في تقييم هذا الأسلوب المعرفي . وهذه الاختبارات هي :

(١) اختبار الحجرة المائلة والكرسي المائل

: Tilting - Room - Tilting - Chair Test (TRTC)

ويُطلب من الفرد في هذا الاختبار أن يُعدّل من وضع الكرسي الذي يجلس عليه إلى الوضع الرأسي عندما يلمع أن الحجرة المائلة لا تحافظ على جسمه في وضع متوازن .

ويوضح شكل (١٣-١٠) التالي هذا الاختبار :



شكل (١٠-١٣) يوضح اختبار الحجرة المائلة والكرسي المائل

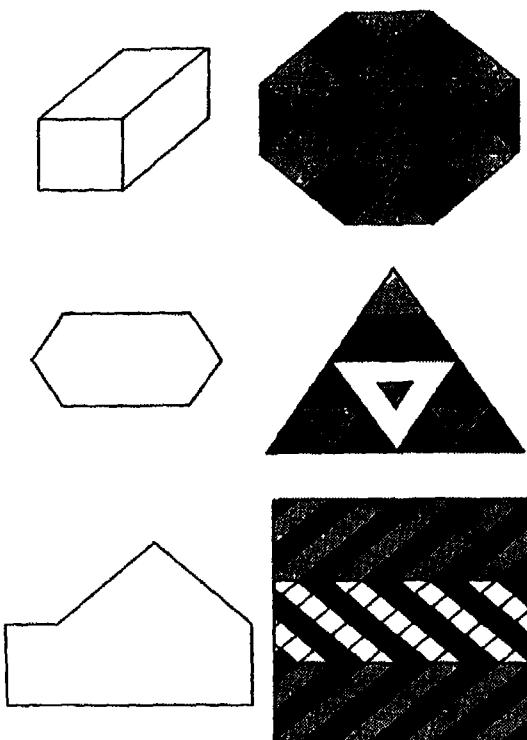
(٢) اختبار المؤشر والإطار (RFT) : Rod and Frame Test (RFT)

ويُطلب من الفرد تعديل مؤشر مضى في وضع رأسى عندما يكون المرجع البصري الوحيد هو الإطار الذى يمكن أن يميل بزوايا معينة. وترجع صعوبة هذا الاختبار إلى أن الحجرة تكون مظلمة إللاماً تاماً، والمربع المضى المحاط بالمؤشر يتم تدويره بزوايا مختلفة في الوقت الذى يعدل فيه الفرد أو يعيد تعديل المؤشر . ويُفترض أن الفرد لا يجب أن يُشتت انتباذه للأوضاع المختلفة للمؤشر ، وإنما يُحدد الوضع الرأسى للمؤشر بإحساسه الجسمانى الداخلى فقط ، فالأفراد المعتمدون أو الأقل تميزاً يستجيبون للتغيرات التى تحدث فى الوسط التجربى بدرجة أكبر من الأفراد المستقلين .

(٣) اختبار الأشكال المتضمنة (EFT) : Embedd Figures Test (EFT)

ويُطلب من الفرد تحديد شكل بسيط متضمن أو مطمور في شكل أكثر تعقيداً معطى له ، وتوجد صيغ مختلفة من هذا الاختبار تناسب الراشدين ، والأطفال ، وأطفال ما قبل المدرسة .

ويوضع شكل (١٢-١٣) التالي هذا الاختبار:



شكل (١٢-١٣) يوضح بعض مفردات اختبار الاشكال المتضمنة

وعلى الرغم من أن هذه الاختبارات إدراكيه، إلا أنها تُستخدم في تقييم خصائص الشخصية، فقد وجد أن الفرد المستقل عن المجال في الاختبارات الثلاثة يتسم بالاكتفاء الذاتي في علاقاته الاجتماعية، ولديه وسائل دفاعية وضوابط قوية توجه أفعاله، كما يتسم بالفکر المرتفع، ويلجأ إلى نفسه إذا ما صادف اضطراباً في شخصيته بدرجة أكبر من الفرد المعتمد على المجال.

وقد أوضحت كثير من الدراسات أن هذا الأسلوب المعرفي يعد سمة مستقرة نسبياً ومتسقة وتتميز بدرجة ما من العمومية، وترتبط كما ذكرنا بعدد من متغيرات الشخصية، مثل القيادة، والمسيرة الاجتماعية (Witkin et al., 1975).

ومن الجدير بالذكر أن هناك أنواعاً أخرى من الأساليب المعرفية، مثل أسلوب الجمود - المرونة، وأسلوب التحليلي - التركيبى، وأسلوب التأمل - الاندفاع ،

وأسلوب التمايز التصورى وغيرها؛ لذلك تعدد أدوات قياس هذه الأساليب. فاختبار الأسلوب التصورى The Conceptual Style Test هو الذى أعده كاجن وزملاؤه (Kagen et al. , 1963) يطلب فيه من الطفل انتقاء صورتين مستفتتين من بين مجموعة معطاة من الصور وتوضيح سبب اتفاقهما، ويميز هذا الاختبار بين الأطفال التحليليين والاطفال التركيبيين.

كما أن اختبار سجل مفهوم الدور Role Construct Repertory Test (RCRT) يعد من أساليب التقييم التى نالت اهتماماً ملحوظاً ، ويرتبط بالأساليب المعرفية ، وقد استند بناؤه إلى نظرية كيلي Kelly في الشخصية المتعلقة بالمفهوم الشخصى Personal Construct Theory التي تهتم بوصف الحالات الفردية ، حيث إن الفرد هو الذى ينظم ويؤسس عالمه بطريقته الخاصة . ويتكون الاختبار من عدة صيغ تشتهر جميعها فى أنها تتطلب من الفرد انتقاء أسلوبه الخاص فى المقارنة بين الناس بدلاً من استخدام النظام التصورى لمن قام ببناء الاختبار. فالاختبار عبارة عن قائمة تشتمل على ٢٤ عنواناً لأدوار متعددة مثل :

«المعلم الذى تُفضلة» ، «أكثر شخص ناجح تعرفه» ، «أقرب صديق إليك» ، وهكذا. ويكتب الفرد اسم الشخص المنشطر لكل من هذه الأدوار على بطاقة منفصلة، ثم يقوم الفاحص الكلينيكي باختيار ثلاثة بطاقات، ويطلب من الفرد أن يوضح كيفية اتفاقثنين منها واختلافهما عن الثالثة، إذ ربما يقول أن معلمه مخلص ، وصديقه عطوف ، وهكذا. ويستمر الفاحص إلى أن تنتهي جميع البطاقات، ثم يقوم بدراسة النتائج لتحديد النظام المفهومى Construct System للفرد، والأبعاد التي تحدد تفكيره فيما يتعلق بالآخرين ، وبذلك يمكن توجيهه فى جلسات العلاج النفسي إلى أبعد أخرى يحتاج إلى تعلمها لكي يغير من نظرته إلى الآخرين. وقد طور كيلي هذا الأسلوب بعد ذلك ليكون أكثر عمومية، حيث أوضح مفهوم التعقد المعرفي Cognitive Complexity الذى يشير إلى العدد الكلى للمفاهيم أو الأبعاد المختلفة التى يستخدمها الفرد فى تفكيره أو نظرته نحو الآخرين، ويوجد اتفاق عام على عدد المفاهيم الشخصية Personal Constructs التى يمكن أن تحدد موقع الآخرين بالنسبة لها، وتراوح بين البساطة المعرفية والتعقد المعرفي (Kelly , 1955 , 1958).

وكذلك توجد مقاييس تتعلق بالضبط المعرفي Cognitive Control، أو ما أطلق عليه روتter (Rotter, 1972) محل الضبط (الداخلى) مقابل محل الضبط (الخارجي) للتعزيز Locus of Control . ويشتمل مقياس روتter على أزواج من الفقرات ، تتعلق

إحدى فقرات كل زوج بتوقعات الفرد من حيث إنه يستطيع التحكم في أمور حياته أو ضبطها، والأخرى تتعلق بتوقعاته أن القوى الخارجية هي التي تحكم في أموره، ومثال ذلك :

« يعتمد نجاح الفرد على ما يبذله من جهد » ، « إن النجاح يعتمد على الحظ ». وقد نال هذا المتغير المعرفي اهتماماً ملحوظاً ، حيث أجريت حوله كثير من الدراسات ، وأوضح بعضها أنه يرتبط ارتباطاً دالاً بمتغيرات أخرى للشخصية . (Getter & Weiss , 1968; Stephens & Delys , 1973)

تقييم الأساليب الموضوعية :

تهتم الأساليب الموضوعية في تقييم الشخصية بجمع بيانات عن الفروق الفردية بطريقة منظمة ، وتصمم المقاييس المتعلقة بها بحيث يجعل الفرد يُركز على مهمة معينة ويكملاها بسرعة أو فاعلية ، وبملاحظة أو تسجيل بعض مظاهر أدائه الذي يرتبط ارتباطاً غير مباشر بالمهنة المعينة ، يحاول الفاحص الحصول على معلومات عن نمط الوظائف المميزة له ، وبذلك يمكن التغلب على الموقف الدفاعي للفرد عند تقييم شخصيته بطريقة مباشرة .

غير أن هذه المقاييس تتعلق بنمط معين من الأداء يتحدد بالإجراءات التي اتبعت في قياسه . فإذا استُخدم المقياس أو الاختبار كمتغير تابع في تجربة سيكلوجية ، فإن النتائج يكون لها أهمية نظرية على الرغم من عدم إمكانية التوصل إلى تفسيرات تسم بالعمومية ، وذلك لأن درجات الاختبار تفتقر إلى أساس منطقى ، كما هو الحال عندما يُستخدم مقياس لنشاط فسيولوجي معين كمؤشر للعصبية ، مما يتطلب التحقق من صدق هذا المؤشر بدرجة عالية من الثقة . غير أن كثيراً من الأساليب الموضوعية غير المباشرة تفتقر إلى دلالات الصدق المرجوة في تقييم الشخصية ، وتتطلب عملية جمع البيانات إجراءات تفصيلية مكلفة .

ومع هذا فإن هذه الأساليب لا تتأثر باستجابات الأفراد لها بالاتجاهات العقلية Response sets التي تلعب دوراً واضحاً في استبيانات الشخصية ، ويمكن استخدامها كجزء من عملية التقييم الشامل لشخصية الفرد ، على الرغم من أنها تتطلب مزيداً من البحوث لتعرف مدى فائدتها في التقييم الموضوعي للشخصية .

تحقيق عام على طرق وأساليب قياس الشخصية :

يتضح مما سبق تعدد أساليب قياس الشخصية، وتبين المداخل التي تستند إليها سواء في بناء الاختبارات والمقاييس، أو في تقدير الدرجات وتفسيرها، ومن بين هذه المداخل: المدخل المنطقي ، والمدخل السيكومترى ، والمدخل الإمبريقى ، والمدخل التجارىي ، والمدخل الكلينيكي التشخيصى ، والمدخل الكلينيكي الفيزنومونولوجي . وأحد الفروق المهمة بين هذه المداخل الرئيسية في تقييم الشخصية يتعلق بطريقة الحصول على البيانات وكيفية معالجتها.

فالأساليب الموضوعية السلوكية تعالج البيانات كعينة من سلوك الفرد، أما الأساليب الإسقاطية فتعتبر استجابات الفرد بمشابه مؤشر لدراوقة الداخلية وصراعاته ومشاعره، ومعظم الأخصائيين الذين يستخدمون هذه الأساليب تأثروا بدرجة كبيرة بنظرية فرويد والفكر المتعلق بالتحليل النفسي ، بينما نجد أن الأساليب السيكومترية تعتمد في تفسير درجات استبيانات الشخصية على الدرجات المعيارية بأنواعها المختلفة والصفحات النفسية (البروفيل) لل اختبار.

- غير أنه لا يوجد اتفاق بين علماء النفس على كيفية ضم نتائج الملاحظة أو المقابلة مع درجات هذه الاستبيانات للتوصل إلى قرارات تتعلق بالأفراد أو المرضى أو التنبؤ بسلوكهم المستقبلي ، وقد أدى ذلك إلى التساؤل حول ما إذا كان من الأفضل ضم البيانات بطريقة ذاتية أو كلينيكية أم بطريقة آلية أو إحصائية . فميزايا وعيوب الأساليب والبيانات الكلينيكية Clinical Methods في مقابل الأساليب والبيانات الإحصائية Actuarial Methods لا تزال مثار جدل حاد ونقاش مستمر . وقد برزت هذه المشكلة بصورة أوضح في كتاب ميهل (Meehl, 1954) «التنبؤ الكلينيكي في مقابل التنبؤ الإحصائي» ، حيث نقش مشكلة استخدام بيانات اختبارات ومقاييس الشخصية في التنبؤ وأوضح أن التنبؤ الكلينيكي يتضمن صياغة فروض خاصة بتكوين شخصية الفرد أو دينامياتها ، واستخدام هذه الفروض في التنبؤ . ويمكن أن يستند التنبؤ إلى الحدس أو الحكم الشخصي ، ولا يعتمد على العلاقات الإحصائية . فقد يعتمد الكلينيكي اعتماداً تاماً على الحدس Intuition ، وهنا ربما لا يستطيع تحديد أساس تنبؤه ، أو قد يستخدم المبادئ العامة المستمدة من إحدى نظريات الشخصية ويطبقها على الحالات الفردية .

أما في التنبؤ الإحصائي فإن الباحث يحدد العلاقة بين الدرجات المختلفة في الاستبيانات أو الاختبارات والسلوك المحك المتبناً به ، ويفسر الدرجة التي يحصل عليها الفرد بالرجوع إلى جداول المعايير ، ولا يعتمد هذا التنبؤ على نظرية معينة ، وإنما يقوم على أساس اعتبارات إحصائية فقط .

ومذ شر كتاب ميهل انقسم الأخصائيون الكلينيكيون إلى فريقين يتبنى كل منهما إحدى وجهتي النظر، وحاول كل منهما التوصل إلى أدلة نؤيد وجهة النظر التي يتبنها وقد قام ميهل بمراجعة عدد كبير من الدراسات التي استخدم بعضها المدخل الإحصائي، واستخدم البعض الآخر المدخل الكلينيكي، وتوصل إلى أن أكثر من ثلثي الدراسات أكدت تفوق المدخل الإحصائي على المدخل الكلينيكي في التنبؤ الكلينيكي، أما بقية الدراسات فقد أوضحت أن المدخل الإحصائي يتساوى معه في التنبؤ. كما أكد ميهل أن المدخل الإحصائي يستغرق استخدامه وقتاً وجهداً وكلفة أقل من المدخل الكلينيكي، وقصد بذلك أن الأدوات الإحصائية الكلينيكية يقضى وقتاً طويلاً في عمل يمكن أداؤه بدرجة أكثر فاعلية عن طريق شخص أقل كفاءة، أو باستخدام الآلة استناداً إلى الأساليب الإحصائية. ولكن ينبغي ملاحظة أن مناقشة ميهل اقتصرت على التنبؤ ولم ت تعرض لنوع البيانات المستخدمة في التنبؤ. فالتمييز بين نوعي التنبؤ اقتصر على كيفية ضم البيانات التي يتم جمعها لأغراض التنبؤ. وبالطبع لا يعكس الانقسام بين الفريقين الأساليب المستخدمة في التنبؤ فقط، وإنما يعكس أيضاً اختلافاً واضحاً في نظريات الشخصية وطرق التقدير التي يتبنها كل فريق (Meehl, 1959).

وعلى الرغم من ذلك فإنه يمكن استخدام كل من المدخلين في التفسير إذا علمنا جوانب القوة والضعف في كل من أساليب القياس أو التقدير، وفي كل من نوعي التنبؤ. فمعظم الكلينيكين يستخدمون عدداً من الأساليب الإحصائية التي تساعدهم في اتخاذ القرارات. إذ ربما يلجأ أحدهم إلى تحويل الدرجات الخام إلى درجات معيارية، أو الرجوع إلى جداول معايير، أو رسم صفحة نفسية (بروفيل)، أو استخدام معادلات الانحدار لاتخاذ قرار معين، هذا إلى جانب التأمل فيما وراء البيانات المتعلقة بالسلوك وجوانب الاختبار.

لذلك فقد أوضح كرونباك (Cronbach, 1960) أن لكل مدخل في الحصول على البيانات المتعلقة بالشخصية وتفسيرها أغراض مناسبة. فإذا واجه أخصائي القياس النفسي مشكلة كلينيكية تتطلب الفهم وليس التقويم البسيط فقط، مثل تحديد أسباب السلوك العدواني لدى طفل، فإنه لا يستطيع الإجابة عن ذلك، وإذا حاول الأخصائي الكلينيكي تعرف أسباب ذلك دون الاستناد إلى نظرية معينة مناسبة، فإنه ربما تكون تفسيراته غير صحيحة، فأيٌّ منها يمكن أن يتفوق على الآخر إذا عمل في مجاله الصحيح.

استخدامات مقاييس الشخصية :

تُستخدم مقاييس الشخصية في غرضين رئيين، أحدهما نظري بحثي ، والآخر عملي تطبيقي ، فالغرض النظري البحثي يُمكّنا من تحديد متغيرات الشخصية وتكويناتها الفرضية عن طريق قياس هذه المتغيرات ومحاولة الربط بينها وبين أنماط السلوك الملاحظ ، أما الغرض العملي التطبيقي فيُمكّنا من اتخاذ قرارات والتنبؤ بسلوك الأفراد في مواقف عملية متنوعة . فكثير من المشغلين بعلم النفس في مختلف المجالات التطبيقية ، أو مجالات البحث يهتمون اهتماماً خاصاً بقياس الشخصية ، ومن بين هذه المجالات: التشخيص ، والإرشاد والعلاج النفسي ، وانتقاء الأفراد ، وبحوث الشخصية وبناء النظريات ، وكذلك في تقويم البرامج . وسوف نوضح فيما يلى بایجاڑ کلا من هذه الاستخدامات :

التشخيص: تُستخدم مقاييس الشخصية في أغراض التشخيص الكlinيكي في المستشفيات والعيادات النفسية ، وذلك لتحديد مدى وطبيعة الاضطراب في الشخصية الذي يعاني منه المريض ، واقتراح العلاج المناسب . وهذا يتطلب معرفة بنية وتنظيم الشخصية الذي يؤدي إلى ما يعانيه المريض من اضطراب .

وقد عمل علماء النفس على مراجعة مفهوم التشخيص في مجال الشخصية في علم النفس وال التربية ، حيث إن النموذج الطبي Medical Model الذي يؤكد تحديد طبيعة المرض من تعرُّف أعراضه أصبح غير مقبول من جانب كثير من هؤلاء العلماء . فهذا النموذج لا يُمكّن الأخصائي من التوصل إلى تشخيص دون معرفة ظروف الفرد ، ولا يكون هناك اتفاق بين مجموعة من الأخصائيين حول مُسببات العرض المرضي ، وعدم اتساق تشخيصاتهم ، وعارضه بعض علماء النفس فكرة تصنيف الأنماط المرضية ، مثل النمط العصابي ، والنمط الذهانى ، واقتراهم محاولة فهم القوى والصراعات التي تدفع هؤلاء الأفراد إلى أن يسلكوا بطريقة معينة مميزة .

وقد اقترح النموذج السلوكي بدليلاً عن النموذج الطبي ، حيث يهتم بتعديل السلوك أكثر من اهتمامه بتشخيصه ، أي محاولة التخلص من السلوك غير المرغوب فيه وتعزيز السلوك المرغوب .

الإرشاد والعلاج النفسي: يهتم الإرشاد والعلاج النفسي بعلاج اضطرابات الشخصية بأساليب سيميولوجية ، فكلاهما يعتمد على التفاعل اللغوي بين المريض

والأشخاص النفسي بغرض مساعدة المريض على تغيير سلوكه لكي يتيسر له التكيف، لذلك تُستخدم مقاييس الشخصية قبل وأثناء وبعد العلاج، بحيث تُصاحب التشخيص الكlinيكي، وذلك بهدف تقدير التغير الذي يتيح عن أسلوب العلاج.

وقد نشأ علم الإرشاد النفسي كفرع متصل بهم بدرجة أساسية بالإرشاد التربوي والمهني. فأشخاص الإرشاد النفسي يقوم بتوجيه الفرد إلى مجالات الدراسة أو المهن أو الوظائف التي تتناسب مع استعداداته وقدراته وميوله. غير أنه سرعان ما أدرك هؤلاء الأشخاص أن لا يمكن تجزئه الفرد إلى مجموعة من الاستعدادات والميول المنفصلة دون النظر إلى شخصيته كنظام متكامل؛ لذلك اتجهوا إلى تطبيق كثير من اختبارات ومقاييس الشخصية للتوصل إلى هذه الصورة المتكاملة.

أما علم النفس الكlinيكي فقد نشأ مرتبطة ارتباطاً مباشرًا بموافق الصحة النفسية، وتضم الأسلوب التي يستخدمها الأشخاص الكlinيكي بحيث يمكن الإفاده منها في تشخيص الاضطرابات النفسية. ولكن نظراً لأنه لا يمكن فهم شخصية الأفراد في ضوء أعراض الأمراض النفسية ، بل يجب التعرف على الجوانب الإيجابية في شخصية هؤلاء الأفراد ، فقد استُخدم مفهومي الإرشاد النفسي والعلاج النفسي ليحل كل منهما محل الآخر (Missick, 1961).

انتقاء الأفراد: تُستخدم مقاييس الشخصية وبخاصة أساليب الملاحظة والمقابلة واختبارات المواقف والاستبيانات ، في انتقاء الأفراد وبخاصة في المجالات الصناعية والعسكرية . كما تُستخدم في ترقية الأفراد، أو قبولهم في برامج تدريبية معينة لزيادة فاعلية أدائهم. وتستخدم محكّات متعددة ، مثل تقديرات المشرفين على العمل ، والإنتاجية ، والنجاح في مؤسسات التدريب ، والمُعْدَل العام للطلاب . والمشكلة الرئيسة هنا هي ما إذا كان المتقدمون يتميزون بسمات شخصية معينة يتطلبها العمل المرشحون له ، وهذه السمات يمكن تحديدها باستخدام مقاييس الشخصية . كما أن هناك صعوبات تتعلق بعدم ثبات درجات أو تقديرات هذه المحكّات ، وكذلك صعوبات تتعلق بطبيعة سمات الشخصية المهمة في النجاح في المهن المختلفة .

وُتُستخدم أحياناً الاختبارات الإسقاطية في أغراض الانتقاء ، ولكنها تؤدي إلى نتائج غير مرضية .

بحوث الشخصية وبناء النظريات :

تُستخدم مقاييس الشخصية أيضاً في التتحقق من فرضيات تتعلق بطبيعة سمات الشخصية، وكيفية تنظيم هذه السمات ، وكيف تُستخدم في أغراض متعددة، وتلافي إساءة استخدامها. كما تُستخدم في التتحقق من كفاية طرق القياس ذاتها حيث تصمم البحوث للكشف عما إذا كانت أداة قياس معينة تقدير سمات شخصية بطريقة متسقة، وما إذا كانت السمات التي تقديرها هي ما يود الباحث قيادتها. فربما تدعوا حاجة مؤسسة صناعية معينة أو مستشفى للأمراض النفسية أو العقلية إلى بناء مقاييس جديد للشخصية لأغراض معينة، فهنا يقوم فريق البحث بدراسة مدى تحقيق المقاييس للهدف الذي أعدد من أجله.

ويمكن أيضاً تحديد العلاقات القائمة بين تكوينات فرضية، مثل الشخصية ، والذكاء ، والميل ، والاتجاهات ، والتحصيل باستخدام أساليب التحليل العاملى لتوضيح معانى هذه التكوينات الفرضية واستخدام مقاييسها استخداماً فاعلاً . وكذلك يمكن بناء نظام منطقى يمكن فيه قياس هذه التكوينات الفرضية التي يمكن أن تستدل منها على العلاقات القائمة بين الأنماط السلوكية .

ونظراً لأن النظريات المستحدثة في مجال الشخصية ربما تتطلب أساليب أو مداخل جديدة في القياس تعتمد على أساليب المذكورة Computer Models of Personality ، فإن استخدام الحاسوب في محاكاة الشخصية الإنسانية Computer Simulation يعد من الطرق المعاصرة في دراسة الشخصية ، والتوصل إلى فروض مهمة عن هذه الظاهرة . فبدلاً من تطبيق اختبارات ومقاييس الشخصية على الإنسان ربما يمكن اختبار الحاسوب بدلاً من الإنسان وفق برنامج معين ، لتحديد مجالات التشابه والاختلاف بينهما مما يسهم في فهم طبيعة الشخصية الإنسانية فهماً أفضل . ويمكن استخدام الحاسوب في إجراء مقابلات شخصية جماعية Mass Interview بغض النظر على الأفراد الذين ربما يحتاجون إلى علاج نفسي ، كما يمكن استخدامه - كما سبق أن أوضحنا - في تفسير بيانات الصفحات النفسية (البروفيل) المستمدة من استبيانات الشخصية كما في استبيان مينيسوتا متعدد الأوجه (MMPI) وغيره من الاستبيانات .

وتُستخدم مقاييس الشخصية أيضاً في المواقف العملية التي تتضمن قياس أثر برنامج تدريسي معين مثلاً على الشخصية . وهنا تكون عينة البحث مجموعة من الأفراد المتطوعين للاشتراك في التجربة . فمثلاً يمكن إجراء تجربة عن أثر الحرمان من الطعام لمدة طويلة على شخصية مجموعة من الأفراد .

تقويم البرامج:

على الرغم من أن مقاييس الشخصية ربما تكون محدودة الفائدة في تقويم البرامج التربوية والتدريبية، إلا أنها أحياناً تقييد في تعرف مدى فاعلية هذه البرامج في تعديل شخصية الأفراد. إذ يمكن أن تُطبق هذه المقاييس قبل بدء برنامج معين، وعقب الانتهاء منه لتعرف التغيرات التي حدثت نتيجة للبرنامج، وربما ترجع قلة استخدام مقاييس الشخصية في تقويم البرامج إلى ما سبق أن ذكرناه في هذا الفصل من تحيز الاستجابات، وصعوبة الحصول على محكّات مناسبة ومتسقة، ونوعية الفقرات أو المهام المتعلقة بموقف معين دون سواه، وغير ذلك من العوامل التي تعيق استخدام الفاعل لهذه المقاييس في هذا الغرض.

كما أن تطبيق مقاييس الشخصية في مواقف تجريبية ربما يتناهى وحقوق الإنسان وخصوصياته إذا ما طُبّقت عليه دون موافقته. وبعض هذه المقاييس يمكنه غامض الهدف، أو ربما يتعمّد المُجرب إخفاء الهدف عن المشاركون في التجربة مما يتناهى والشروط الأخلاقية لاستخدام الاختبارات.

لذلك ينبغي مراعاة شروط وضوابط تطبيق الاختبارات والمقاييس التي أوضحتها بمزيد من التفصيل في الفصل الثاني.

خلاصة

عرضنا في هذا الفصل الأساليب والمقاييس المختلفة الشائعة الاستخدام في قياس الشخصية، وهي: الاستبيانات، والاختبارات الإسقاطية، والملاحظات وموازين التقدير، والمقابلة، واختبارات الأداء، والأنشطة الفسيولوجية، واختبارات الإدراك. وقد اتضح لنا مدى تعقد مفهوم الشخصية واختلاف تعريفاتها، حيث إن هناك عدم اتفاق بين علماء النفس في تصورهم لبنية أو مكونات الشخصية، وذلك لاختلاف الأطر النظرية التي يتبناها كل فريق منهم. وقد أدى هذا إلى عديد من المشكلات في قياس الشخصية، وذلك لاختلاف المبادئ المنظمة المستخدمة في التوصل إلى معلومات عن شخصية الأفراد، فبعضهم يستند إلى نظريات التحليل النفسي أو إلى المدخل الفينومولوجي، وبعضهم يستند إلى مدخل السمات التي يتم التوصل إليها باستخدام أساليب التحليل العاملی، وبعضهم يستند إلى الأساليب الإمبريقية، في حين يستند فريق آخر إلى الأساليب التجريبية أو الموضوعية - السلوكية.

ونظرًا للتعدد هذه التوجهات والأساليب فقد اقتُرُح تصنيف اختبارات ومقاييس الشخصية من حيث مدى تحديد بنيتها، ووضوح هدفها، وطريقة إجابتها. وحاولناربط بين نظريات الشخصية وطرق القياس والتقييم المختلفة. فنظريات الشخصية تؤدي إلى أنواع مختلفة من البيانات وتتأثر بها، كما أن طرق وأساليب قياس أو تقسيم الشخصية تعتمد على فروض معينة تتعلق بطبيعة الشخصية وخطوط وإجراءات عملية القياس. فاستبيانات الشخصية تكون محددة البنية وواضحة الهدف، وتستند في بنائها إلى أساليب التحليل العاملی، مثل استبيان العوامل الستة عشر لكاتل (16 PF) أو استبيان مينيسوتا متعدد الأوجه (MMPI)، وغيرها من الاستبيانات التي تعتمد على التقرير الذاتي للفرد. أما الاختبارات الإسقاطية ف تكون غير محددة البنية، وغامضة الهدف، حيث تشتمل على مثيرات غامضة مثل بقع الحبر، أو الصور التي تدور حول موضوع، أو الجمل الناقصة، أو الرسوم، أو اللوحات، أو اللعب وغير ذلك، ومن أهم هذه الاختبارات اختبار رورشاخ، واختبار تفهم الموضوع حيث يعبر الفرد عما يراه ويُسقط مشاعره ودوافعه وعواطفه على هذه المثيرات الغامضة.

ويمكن تصنيفها في أساليب التداعى، والأساليب البنائية، وأساليب التكملة، وأساليب الاختيار أو الترتيب، والأساليب التعبيرية. وتُستخدم أساليب الملاحظة

والتقدير، والمقابلة بكثرة في تقسيم جوانب معينة من الشخصية، وبخاصة في المواقف الكlinينية، وفي أغراض انتقاء الأفراد، وتتفاوت في مدى تحديد بنيتها ووضوح هدفها وأساليب بنائها. وتتعدد أنواع موازين التقدير التي يمكن استخدامها في أساليب الملاحظة المقننة، كما تُستخدم أحياناً في المقابلة.

وتُوجد مجموعة أخرى من الأساليب التي تناسب المجالات البحثية، وتعلق بالأنشطة الفسيولوجية واختبارات الإدراك التي تُعد حلقة وصل بين مقاييس القدرات ومقاييس الشخصية. وتعلق اختبارات الإدراك بالأساليب المعرفية، ومن بينها اختبار الحجرة المائلة والكرسي المائل، واختبار المؤشر والإطار، واختبار الأشكال المتضمنة التي أعدّها ويتمكن وزملاؤه لقياس الاستقلال أو الاعتماد الإدراكي على المجال.

وتحتفل مقاييس الشخصية اختلافاً ملحوظاً في ثبات درجاتها وصدقها، ويرجع ذلك إلى اختلاف مصادر الأخطاء المنتظمة وغير المنتظمة التي تؤثر في الموقف الاختباري وفي درجات المقياس. فاستبيانات الشخصية تتأثر بتحيز الاستجابات والاتجاهات العقلية للفرد وغير ذلك من الأخطاء. والاختبارات الإسقاطية تتأثر بذاتية القائم بتطبيق الاختبار وتصحيحه وتفسير نتائجه، كما أنه لا توجد أدلة كافية تعزز استخدام هذه الاختبارات . وكذلك تفتقر أساليب الملاحظة والمقابلة والأساليب الموضوعية - السلوكية إلى مؤشرات مرتفعة للثبات والصدق . وعلى الرغم من محاولة القائمين ببناء مقاييس الشخصيةربط استجابات أو أداء الأفراد في هذه المقاييس بسلوكهم أو أدائهم في مواقف خارج نطاق الموقف الاختباري ، إلا أنهم يختلفون في تحديد أهمية السلوك الذي سترتبط به نتائج الاختبار، وفي مدى كفاية الأدلة المتعلقة بهذه العلاقات .

ونتيجة لهذه الاختلافات لاحظنا أن مؤيدى كل طريقة يؤكدون مزايا طريقتهم ويشرون إلى قصور الطرق الأخرى، وهذا ربما يبين لنا مدى تعقد قياس الشخصية؛ لذلك لا نستطيع القول بأن هناك أسلوباً أمثل لقياس الشخصية، وإنما ينبغي النظر إلى كل من هذه الأساليب نظرة فاحصة ناقدة، وانتقاء ما يناسب الغرض من الاختبار، وفيما تُستخدم نتائجه . فاختبارات ومقاييس الشخصية تُستخدم في العديد من المجالات klinينية، والإرشاد والعلاج النفسي، وانتقاء الأفراد، وتقدير البرامج، كما تُستخدم في أغراض البحثية والتحقق من صحة الفروض المتعلقة بشخصية الأفراد، والعلاقات بين متغيرات الشخصية، والمتغيرات الأخرى المتعلقة بالذكاء، والإدراك، والميول وغيرها .

وعلى الرغم من أن الأساليب المختلفة لقياس الشخصية تُقدم معلومات ومؤشرات مهمة عن طبيعة وتنظيم الشخصية، إلا أن التفسيرات المستمدّة من أي منها ينبغي أن تُضم إلى البيانات التي يتم جمعها باستخدام الأساليب الأخرى، وكذلك إلى البيانات المستمدّة من اختبارات التحصيل، والذكاء، والاستعدادات الخاصة ، والميول، والاتجاهات التي ناقشناها في الفصول السابقة من أجل فهم أفضل لشخصية الفرد، والتبنّي بسلوكه المستقبلي . فالشخصية الإنسانية تُعد نظاماً مفتوحاً Open System تتفاعل فيه المكونات المعرفية والوجدانية والمزاجية وغيرها ، كما تتفاعل مع المتغيرات البيئية وتتأثر بها بدرجة كبيرة كما تؤثر فيها، مما يؤدي إلى تباين ملحوظ في سلوك الأفراد ومدى تكيفهم مع أنفسهم ومع بيئتهم المحيطة بهم .



توجهات معاصرة في القياس
والتحكيم التربوي والنفسى



الفصل الرابع عشر

بحرين التطورات النظرية المعاصرة

* مقدمة

أولاً: نظرية الاستجابة للمفردة الاختبارية

* نظرية الاستجابة للمفردة الاختبارية من منظور تاريخي

* الأنواع الرئيسة لنماذج الاستجابة للمفردة الاختبارية

* الأساس المنطقى لنماذج الاستجابة للمفردة الاختبارية

* تصنیف نماذج الاستجابة للمفردة الاختبارية

* مفهوم المترافق المميز للمفردة (ICC)

* النماذج الاستاتيكية اللوغاريتمية أحادية البعد

* نموذج راش

* نموذج البارامترین

* نموذج البارامترات الثلاثة

* مزايا وعيوب نماذج الاستجابة للمفردة الاختبارية

ثانياً: نظرية الثبات متعدد الأوجه

* بعض أوجهه قصور المفهوم الكلاسيكي للثبات والصدق

* بعض المفاهيم الأساسية لنظرية إمكانية التعميم

* دراسات إمكانية التعميم

* دراسات القرارات

* تحديد النطاقات الشاملة في دراسات إمكانية التعميم

* تحليل مكونات التباين في دراسات إمكانية التعميم

* العلاقة بين نظرية إمكانية التعميم ونظرية الصدق

* صدق التكوين الفرضي من منظور نظرية إمكانية التعميم

ثالثاً: إسهام علم النفس المعرفي في تطوير أساليب القياس والتقويم

* علاقة النظرية المعرفية بالعملية الاختبارية

* تصميم الاختبارات في ضوء النظرية المعرفية

* خلاصة

مقدمة :

عرضنا في الأبواب الثلاثة السابقة الأسس النظرية وأهم الجوانب النطبيقية المعرفية والوجودانية والشخصية للقياس والتقويم التربوي والنفسى ، ولاحظنا مدى تعدد وتنوع أدوات قياس هذه الجوانب ، وثراء الأساليب التى يمكن استخدامها فى هذا المجال الحيوى الإنسانى ، وتطبيقاتها فى مختلف مجالات تقويم الأفراد والبرامج التربوية والنفسية والتدريبية والمهنية . وسوف نتناول بإيجاز فى هذا الباب بعض أهم التطورات المعاصرة المتعلقة بالأسس النظرية ، وأساليب بناء الاختبارات والمقاييس التربوية والنفسية وتحليل وتفسير البيانات المستمدة منها ، وماترتب على ذلك من توجهات جديدة فى استخدامات هذه الاختبارات والمقاييس وتطبيقاتها المتنوعة .

فالذى يطلع على أدبيات القياس المعاصر يلاحظ بوضوح إسهام التطورات التكنولوجية إسهاماً كبيراً فى تقدم علم القياس والتقويم التربوى والنفسى الذى أدى بدوره إلى إعادة النظر فى النظريات والنماذج السيكومترية المتعارف عليها ، وتطوير نماذج ونظريات معاصرة للتغلب على كثير من المشكلات المتعلقة ببناء الاختبارات والمقاييس فى الجوانب المختلفة للسلوك الإنساني ، بما يجعلها أكثر مرنة وملائمة لطبيعة الظواهر السلوكية .

وبالطبع يصعب تناول أو توضيح جميع جوانب القياس والتقويم التى شملتها التطور فى هذا الباب ، إذ إن كلا منها ينبغى أن يفرد له عدة مجلدات ، ولكننا سوف نلقي الضوء على بعض أهم هذه التطورات التى تجعل الدارس والباحث والممارس أكثر استبصاراً فى بنائه أو انتقاده أو استخدامه لأدوات القياس والتقويم التربوى والنفسى والاجتماعى ، بحيث يمكنه الرجوع إلى المصادر المتخصصة التى تناولت هذه التطورات . وسوف نتناول فى هذا الفصل بعض أهم التطورات النظرية فى هذا المجال ، ونوضح بعض القضايا والتوجهات المتعلقة بتطبيقاتها فى الفصل الخامس عشر .

ومن أهم التطورات المعاصرة التى سوف نلقي عليها الضوء فى هذا الفصل نظرية الاستجابة للمفردة الاختبارية (IRT) ، ونظرية الثبات Item Response Theory ، ونظرية متعدد الأوجه Generalizability Theory ، وبعض إسهامات علم النفس المعرفي فى تطوير أساليب القياس والتقويم . ويرجع اختيارنا لهذه الموضوعات إلى أهميتها واستخداماتها المتعددة فى مختلف المجالات التطبيقية فى وقتنا الحاضر .

أولاً: نظرية الاستجابة للمفردة الاختبارية

Item Response Theory (IRT)

اعتمدت الاختبارات والمقاييس التي أوضحتها في البابين الثاني والثالث في أساليب بنائتها وتحليل مفرداتها على أساس ومبادئ النظرية الكلاسيكية للاختبارات Classical Test Theory (CTT). وقد أوضحت بعض هذه الأسس والمبادئ في الباب الأول، وأوضح كل من جاليسن (Gulliksen, 1950) في كتابه نظرية الاختبارات العقلية Theory of Mental Measurement ، وجيزلي (Ghiselli, 1964) في كتابه نظرية القياس النفسي Theory of psychological Measurement ، وماجنسون (Nunnally, 1987) في كتابه Test Theory ، ونانلى (Magnusson, 1966) في كتابه Psychometric Theory - أساس ومبادئ هذه النظرية بالتفصيل في مجال الاختبارات مرجعية الجماعة أو المعيار (NRTs).

ويعتمد انتقاء مفردات هذه الاختبارات في إطار هذه النظرية على نتائج بعض الإحصاءات المتعلقة بمايلي:

(١) **تجانس مفردات الاختبار Item Homogeneity**: ونقصد بذلك أن الأفراد الذين يجيبون إجابة صحيحة عن إحدى مفردات اختبار يميلون إلى أن يجيبوا إجابة صحيحة عن بقية مفرداته، وبذلك يحصل كل منهم على درجة مرتفعة في الاختبار، ويحدث العكس بالنسبة للأفراد الذين يجيبون إجابة خطأ عن إحدى المفردات. ويُطلق على مقياس تجانس المفردات اسم «معامل التمييز Discrimination Index»، الذي يمكن الحصول على قيمته باستخدام معامل الارتباط الثنائي المتسلسل بين درجات الأفراد على المفردات ودرجاتهم الكلية في الاختبار.

(٢) **الاتساق الداخلي للاختبار Internal Consistency**: ونقصد بذلك أن يكون احتمال إجابة فرد إجابة صحيحة عن إحدى مفردات اختبار دالة مطردة لموقع الفرد على متصل السمة أو القدرة التي يقيسها الاختبار. فإذا كان نمط إجابات الأفراد الذين يحتللون نفس الموقع على هذا المتصل عن جميع مفردات الاختبار متتشابهاً، يمكن الحصول على معاملات الاتساق الداخلي باستخدام معامل كيودر - ريتشاردسون Kuder - Richardson الصورتين ٢٠، ٢١، أو معامل ألفا لكروباك.

(٣) **صعوبة مفردات الاختبار Item Difficulty**: وتُقدر بنسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن مفردة معينة من مفردات الاختبار.

(٤) التشتيت الأقصى للتوزيع درجات الاختيار Maximal Dispersion

ونقصد بذلك أن توزيع درجات الأفراد على متصل السمة أو القدرة التي يقيسها الاختبار يتخذ شكل المنحنى الاعتدالى. لذلك يُعنى بانتقاء المفردات التي تتراوح قيم معامل صعوبتها بين ٤٠، ٧٠، .. ، لكى يكون انتشار الدرجات أكبر ما يمكن. ويعد هذا الشرط قيداً تفرضه النظرية الكلاسيكية للاختبارات مرجعية الجماعة أو المعيار على توزيع الدرجات. وقد نقاشنا هذه الإحصاءات فى الفصول الأربع الأخيرة من الباب الأول.

ولكن ما المشكلات المتعلقة بالاختبارات التي يتم بناؤها في إطار هذه النظرية؟

إن المشكلة الرئيسية هي أن جميع الخصائص السيكومترية للاختبارات التي أشرنا إليها تعتمد على خصائص عينة الأفراد التي يُجرى عليها الاختبار، وعلى خصائص عينة المفردات التي يشتمل عليها الاختبار، مما يؤدي إلى نظام قياس نفسى وتربيوى غير مستقر.

فمعاملات صعوبة المفردات تتذبذب وتتغير بتغير سمات أو قدرات أفراد العينة المراد اختبارها، وقياس سمات أو قدرات الأفراد تتذبذب وتتغير بتغير صعوبة مفردات الاختبار، وثبات الدرجات المستمدة من الاختبار تتذبذب وتتغير بتغير مستوى وانتشار سمات أو قدرات عينة الأفراد، والدرجات المستمدة من هذه الاختبارات لا يكون لها معنى أو دلالة في ذاتها، إذ يختلف معنى الدرجة باختلاف سهولة أو صعوبة مفردات الاختبار، وباختلاف ضيق أو اتساع مدى الاختبار. لذلك إذا أردنا المقارنة بين مستوى أداء فردين، أو مقارنة أداء الفرد نفسه في مواقفين مختلفين، فإننا يجب أن نستخدم الاختبار نفسه في أي من الحالتين. ولو قمنا ببناء اختبارين يشتملان على نفس العدد من المفردات التي تقيس سمة أو قدرة معينة لدى مجموعة من الأفراد، فإن درجة كل فرد سوف تختلف باختلاف صعوبة مفردات كل من الاختبارين؛ لذلك يُقال أن الاختبارات التي تبني بالطرق السيكومترية الكلاسيكية محدودة بعينة المفردات - Item bounded، وبأفراد العينة التي استخدمت في تقييم الاختبار Sample - bounded. وهناك مجموعة أخرى من المشكلات في هذا الشأن نوجزها فيما يلى:

(١) يفترض أن الاختبار يقيس متغيراً أحادى البد، وربما لا يكون ذلك صحيحاً في كثير من الأحيان.

(٢) يفترض أن الدرجات التي يحصل عليها فرد في مفردات الاختبار يمكن جمعها كما لو كانت تمثل ميزاناً خطياً Linear Scale، وأن المفردات المتعلقة بالمتغير

المراد قياسه تحمل المعنى نفسه لدى جميع المختبرين . والحقيقة أن هذا الميزان يكون منحنيا ، فالفارق الثابت بين درجتين من درجات الاختبار يختلف معناه بالنسبة لدرجة السمة أو القدرة المقاسة بحسب موقع الدرجات على متصل السمة أو القدرة .

(٣) يُفترض أن درجات الاختبار التي تمثل السمة أو القدرة المقاسة دالة خطية مطردة Monotone Function، بمعنى أنه كلما زادت درجة الفرد في الاختبار ، دل ذلك على زيادة مقدار السمة أو القدرة لديه . غير أنها نلاحظ أن بعض الأفراد من ذوى القدرات المرتفعة يحصلون أحيانا على درجات منخفضة في الاختبارات ، وربما يحدث العكس بالنسبة لذوى القدرات المنخفضة .

(٤) تغير تكوين الاختبار بمرور الزمن Test Monolithic Structure . ونعني بذلك أن تكون ومعنى مفردات الاختبار يتغير من زمن إلى آخر بالنسبة لمجموعات الأفراد التي أعد لها الاختبار . فالظروف البيئية تتغير ، والمواصفات الاختبارية ليست دائما مقننة ، وحذف أو تغيير أي مفردة من مفردات الاختبار يؤدي إلى تغير في درجات الأفراد بطريقة يصعب التنبؤ بها ، وبذلك يصبح الاختبار أداة قياس وحيدة . وعندما يتغير قصوره وعدم جدواه استخدامه بسبب الشك في صدق البيانات المستمدة منه ، وقدم معاييره؛ يستبعد الاختبار ويحل محله اختبار مقنن آخر حتى تدور عليه الدورة وهكذا .

ولقد حاول علماء القياس التغلب على هذه المشكلات وغيرها ، وتطوير نظم القياس النفسي والتربوي . فعملية القياس هي نتاج تفاعل بين الأشياء أو الخصائص المراد قياسها وبين أدوات القياس . ونتائج القياس يجب أن تكون مستقلة عن أداة القياس المستخدمة . ويرى علماء القياس المعاصرین (DeGruijter, 1984; Hambleton, 1983) أن بناء نظام موضوعي لتحويل الملاحظات إلى قياسات يتطلب أن تتحرر تداريج أدوات القياس من خصائص الأشياء التي تقيسها هذه الأدوات Object - Free Instrument Calibration وأن يتحرر قياس الخصائص أو الأشياء من خصائص أدوات القياس التي استخدمت في هذا القياس Instrument - Free Object Measurement .

وهذا يتطلب أن يشارك كل من قياس الفرد Person Measurement وتعير المفردات Item Calibration في وحدة قياس ثابتة على مدى متصل المتغير . أي يكون تدريج هذا المتصل خطيا (مستقيما) ووحداته متساوية .

وبذلك يسمح هذا النظام بعمميم القياسات خارج حدود الأداة التي أجري بها القياس، كما يسمح بالموازنة بين خصائص الأشياء المراد قياسها بأدوات قياس متشابهة ولكن غير متطابقة، ويمكننا من ضم أو تجزئة أدوات القياس بحيث تناسب متطلبات قياس جديدة كما هو الحال في القياس الفيزيائي. فالنظرية الكلاسيكية للاختبارات تؤدي إلى بناء اختبارات غير مرنة Traditional Fixed Tests ، بينما وجه علماء القياس جهودهم نحو بناء نظام قياس موضوعي يُركز على المفردة الاختبارية ذاتها Individual Item وليس على الاختبار ككل. وهذا النظام يسمح بإضافة مفردات أخرى إلى الاختبار أو حذف بعضها دون أن يتأثر الاختبار ككل. وهذا يعني بناء اختبارات مرنة Flexible Tests ، أي يمكن أن يكون الاختبار قصيراً أو مفردة كثيرة، أو ضيق المدى أو متسعاً، أو يتميز بالصعوبة أو السهولة ومع هذا يظل نظام تعبيره أو تداريجه مستقراً.

وقد أدت هذه الجهود إلى تطوير نظرية معاصرة في القياس النفسي والتربوي أطلق عليها نظرية الاستجابة للمفردة الاختبارية (IRT) ، Item Response Theory (IRT) ، و كان يطلق عليها من قبل نظرية السمات الكامنة (LTT) ، Latent Trait Theory (LTT) ، ويندرج تحتها مجموعة من النماذج تعرف باسم نماذج الاستجابة للمفردة الاختبارية . Item Response Models

فهذه النماذج تسمح بالقياس الموضوعي الذي أشرنا إليه، كما يمكن باستخدامها مواجهة كثير من مشكلات القياس النظرية والتطبيقية التي عجزت عن مواجهتها النظرية الكلاسيكية .

نظرية الاستجابة للمفردة الاختبارية من منظور تاريخي:

يرجع الفضل في تقديم أسس نظرية الاستجابة للمفردة الاختبارية للمهتمين بالقياس النفسي والتربوي إلى لورد (Lord, 1952, 1953). وقد تطورت هذه النظرية ومايتعلق بها من نماذج سيكومترية منذ ذلك الوقت بهدوء وثبات ، وظلت أعمال لورد Lord وغيره من علماء القياس في هذا المجال حتى السبعينيات من هذا القرن لايلتفت إليها إلا عدد محدود من خبراء القياس في الولايات المتحدة الأمريكية وبعض الدول المتطرفة. غير أنه حدث تقدم سريع في الحركة البحثية والتطبيقية لهذه النظرية ابتداء من عام ١٩٦٨ ، وهو العام الذي نشر فيه لورد Lord كتابه «النظريات الإحصائية لدرجات الاختبارات العقلية Statistical Theories of Mental Test Scores » (Lord, 1968) ، الذي جمع بين غلافية لأول مرة أسس النظرية السيكومترية الكلاسيكية، وأسس نظرية الاستجابة للمفردة الاختبارية .

وقد اهتم جورج راش G. Rasch وهو أحد علماء الرياضيات والإحصاء بجامعة كوبنهاغن بالدنمارك بالقياس النفسي منذ عام ١٩٤٥ ، وتوصل إلى نموذج سيكومترى يمكن أن يندرج تحت نظرية الاستجابة للمفردة الاختبارية، ويعرف بنموذج راش Rasch Model واشتعل بالإحصاء الرياضي والإحصاء الحيوى Biometrics منذ ذلك الوقت.

وقد قام راش بزيارة الولايات المتحدة الأمريكية عام ١٩٤٧ للمشاركة في تأسيس مجتمع القياس الحيوى. وأثناء زيارته لجامعة شيكاغو قابل سافاج J. Savage وهو أحد علماء الإحصاء البارزين بهذه الجامعة حيث دعاه لتكرار زيارة الجامعة عام ١٩٦٠ لكنه ينتهي من إعداد كتابه الرائد «النماذج الاحتمالية لبعض اختبارات الذكاء والتحصيل» ويلقى عدّة محاضرات في هذا المجال. وقدّم راش العديد من الأوراق منذ عام ١٩٦٠ حول القياس الموضوعي وبخاصة ورقته حول «القوانين العامة ومعنى القياس في علم النفس» الذي قدمها في مؤتمر الإحصاء الرياضي بجامعة كاليفورنيا عام ١٩٦١ ، وورقتة «تحليل المفردات الذي يأخذ الفروق الفردية بعين الاعتبار» عام ١٩٦٦ . وقد التقى راش بعالم النفس بنجامين رايت B. Wright بجامعة شيكاغو عام ١٩٦٠ وقدّمه لهذا الفكر السيكومترى الجديد الذى يتبنّاه رايت Wright بعد ذلك. وقدّم ورقة حوله عام ١٩٦٧ بعنوان «تغيير الاختبار وقياس الأفراد متحرراً من العينة»، وتلى ذلك نشاط بحثي كبير قامت به جامعة شيكاغو بإشرافه استناداً إلى النموذج السيكومترى الذي توصل إليه جورج راش مستقلاً عن نماذج الاستجابة للمفردة الاختبارية الأخرى التي توصل إليها بيرنبو姆 Birnbaum ، ولورد Lord في الولايات المتحدة الأمريكية ، والنماذج الأخرى التي توصل إليها بعض خبراء القياس في أستراليا ونيوزيلندا. غير أنه يرجع الفضل في انتشار هذه النماذج ، وبخاصة نموذج راش وتوسيع مجالات تطبيقاته واستخداماته إلى كل من رايت Wright ، وهامبتون Hambleton ، وشوين Choppin

وتناول هذه النماذج في وقتنا الحاضر اهتماماً بعثياً كبيراً يمكن ملاحظته بوضوح في أدبيات القياس النفسي والتربوي المعاصر. وتحاول هذه البحوث التغلب على العديد من المشكلات المتعلقة بتقدير برامراتها وتقدير حسن مطابقة البيانات لهذه النماذج ، وإمكانية استخدامها عندما تكون المتغيرات التي تقيسها الاختبارات متعددة الأبعاد ، وغير ذلك من القضايا السيكومترية والتطبيقية.

ويرجع سبب عدم انتشار مبادئ وأسس هذه النظرية المعاصرة انتشاراً واسعاً إلى عدة أسباب نذكر منها:

(١) تمثل هذه النظرية فرعًا معدداً من فروع نظرية الاختبارات، واستيعاب البحث السيكومترية التي تُجرى في هذا المجال وتنشر في الدوريات العلمية المتخصصة يحتاج إلى قدر كبير من الرياضيات العالية، والإحصاء متعدد المتغيرات، وإحصاء بيزي.

(٢) معظم خبراء القياس الذين تناولوا هذا الموضوع وجّهوا دراساتهم وبحوثهم لمنظري القياس، وليس للذين سيقومون بتطبيقه في بناء الاختبارات وتحليل وتفسير نتائجها، وفي غير ذلك من التطبيقات السيكلوجية والتربوية.

(٣) لم يكن هناك برامج حاسوب فعالة ومناسبة لتقدير بaramترات نماذج السمات الكامنة التي تتضمنها هذه النظرية.

(٤) صعوبة تحقق بعض فروض النماذج المتعلقة بهذه النظرية في البيانات الفعلية المستمدة من الاختبارات العقلية والتحصيلية.

لذلك اتجه علماء القياس في الآونة الأخيرة إلى مواجهة هذه الصعوبات مواجهة إيجابية، حيث أجريت كثير من الدراسات لتحويل هذه النماذج من الجانب السيكومترى النظري البحث إلى الجانب العملى التطبيقي وتبسيطها حتى يتمكن المهتمون بالتطبيق من استخدامها دون عناء. وأعدّت كثير من برامج الحاسوب التى يمكن الاستعانة بها فى تقدير بaramترات هذه النماذج. (Mislevy & Bock, 1985; Wingersky, Barton & Lord, 1982; Wright & Mead, 1975)

ونتيجة لهذه التطورات اتجهت كثير من المؤسسات والهيئات الأمريكية المسئولة عن إعداد الاختبارات ونشرها إلى استخدام هذه النماذج فى تصميم وبناء مختلف أنواع الاختبارات والمقياس العقلية والتحصيلية، وموازين التقدير، ومقاييس الاتجاهات.

كما استعانت بها إدارات التقويم التربوى بمختلف الولايات فى إعداد الاختبارات الصحفية فى المواد الدراسية المختلفة. وحدثت بعض دول العالم مثل كندا ، وانجلترا، وأستراليا، ونيوزيلنده، حذوا الولايات المتحدة الأمريكية فى هذا الشأن.

الأنواع الرئيسية لنماذج الاستجابة للمفردة الاختبارية:

تهدف نماذج الاستجابة للمفردة الاختبارية لتحديد علاقة بين أداء الفرد في اختبار معين، وهو ما يمكن ملاحظته ملاحظة مباشرة، وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره. وبالطبع تختلف السمات الكامنة Latent Traits باختلاف ما يقيسه الاختبار. فالسمة الكامنة التي تنتوى عليها استجابات الأفراد لاختبار

فهم معانى الكلمات مثلاً تختلف عن تلك التى تنطوى عليها استجاباتهم لاختبار العمليات الحسابية، أو اختبار الإدراك المكانى. وتُعد هذه النماذج دوال رياضية احتمالية تختلف باختلاف عدد بارامتراتها أو مكوناتها.

وتنقسم هذه النماذج إلى مجموعتين رئيسيتين كما يلى:

النماذج الاستاتيكية Static Models: وتهتم هذه النماذج بالقياس فى وقت معين، كما تهتم بتحديد العمليات التى ينطوى عليها الأداء فى الاختبارات التربوية والنفسية. فمثلاً يمكن نمذجة الأداء فى اختبار للاستعداد الدراسي أو التحصل على بواسطة هذا النوع من النماذج. ومن أمثلة هذه النماذج نماذج راش، ونموذج لورد، ونموذج بيرنوب وغيرها.

النماذج الديناميكية Dynamic Models: وتهتم هذه النماذج بمشكلة قياس التحسن أو التغير الذى يحدث فى السمات النامية المختلفة فى فترات زمنية متباينة. وبعض هذه النماذج يعتبر التغير النمائى عملية تدريجية، أى أن التغير متصل، بينما يعتبر البعض الآخر أن هذا التغير انتقالى من حالة إلى أخرى، أى تغير منفصل، مثل الانتقال من حالة عدم التمكن إلى حالة التمكن فى الاختبارات المرجعية المحك. وتركزت غالبية هذه النماذج على الحالة الأولى، مثل نموذج بوك (Bock, 1972) الذى يتعلق بمنحنيات قياس التحصل على ضوء نموذج بيرنوب الاستاتيكى، أما نموذج فيشر (Fischer, 1973) فقد اعتمد على نموذج راش.

ويرى أصحاب كلٌّ من نوعى النماذج الكلاسيكية فى بناء الاختبارات (CTM) تناسب قياس المتغيرات الناجمة عن ملاحظات أو أحکام منفردة Single Measurement، مثل تقدير المعلم لطالب فى سمة معينة أو قياس الطول أو الوزن، ولكن يصعب تطبيقها فى حالة المتغيرات الناجمة عن قياسات مركبة Composite Measurement من ملاحظات متعددة غير متكافئة، مثل المجموع الموزون لدرجات مجموعة من المفردات يشتمل عليها اختبار معين. فهذه المفردات تكون عادة متباعدة فى صعوبتها وصدقها فى قياس السمة أو السمات التى يفترض أنها تنطوى عليها. ونماذج الاستجابة للمفردة الاختبارية (IRM) تُطبق بدرجة أفضل فى هذه الحالة.

ولكن بعد تكوين «المتغير» فى الحالة الثانية بطريقة مناسبة يصبح متغيراً من الحالة الأولى، وهنا يمكن استخدام النماذج الكلاسيكية.

الأساس المنطقي لنماذج الاستجابة للمفردة الاختبارية:

يمكن إيجاز الفكرة الأساسية لنماذج الاستجابة للمفردة الاختبارية في أن هذه النماذج تحاول اشتغال قياسات أو قيم تقديرية للسمة أو السمات التي تتطوى عليها مجموعة من الاستجابات لمجموعة من المفردات الاختبارية. وهذه القياسات تميز بخصائص تفسيرية تتخطى حدود المجموع الموزون للدرجات الاختبار. أي أن هذه النماذج تفسر استجابة الفرد للمفردة الاختبارية التي تتطوى على السمة التي تقيسها هذه المفردة. وعادة يفترض أن السمة المقاسة هي قدرة معينة، أو خاصة من خصائص الفرد الذي يختر بها. بحيث إنه توجد علاقة منتظمة بين مستويات السمة المقاسة لدى أفراد مختلفين وبين احتمالات الاستجابة الصحيحة لمفردات مختلفة.

لذلك فإن هذه النماذج تعد نماذج إحتمالية Probabilistic Models، وتستند جميعها إلى دالة الترجيح اللوغاريتمي Logistic Function بدلاً من دالة الكثافة الاحتمالية الاعتدالية Normal Density Function التي تعتمد عليها النماذج الكلاسيكية التي استند إليها جميع الاختبارات والمقاييس التي أوضحتناها في البابين الثاني والثالث. كما تستند إلى فرضيات تختلف عن فرض هذه النماذج الأخيرة سوف نوضحها بعد قليل.

تعريف نماذج الاستجابة للمفردة الاختبارية:

تعدد كل من النماذج الاستاتيكية والنماذج الديناميكية التي أشرنا إليها وفقاً للمتغيرات التي يتضمنها النموذج، والتي يرى أصحاب هذه النماذج أنها تلعب دوراً في الاستجابة للمفردة الاختبارية. غير أنه يمكن تصنيف هذه النماذج في ضوء محركات أربعة هي:

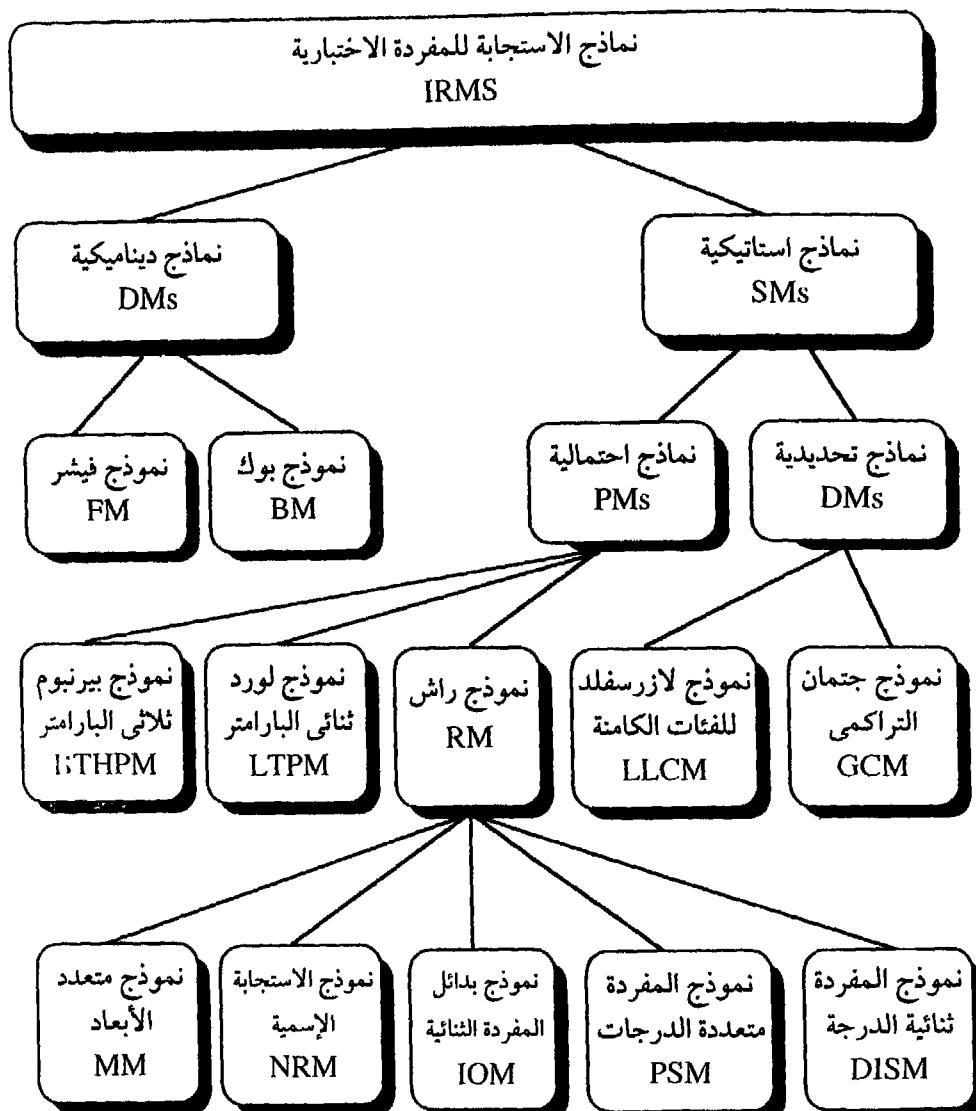
(أ) مستوى الاستجابة Response Level

(ب) التركيب الباراميترى Parametric Structure

(ج) الفرضيات الإحصائية Statistical Assumptions

(د) أبعاد الفضاء الكامن Latent Space Dimensionality

ويقترح المؤلف التصنيف التالي الذي يوضحه الشكل التخطيطي (١٤ - ١):



(من تصميم المؤلف)

شكل (١٤ - ١) يوضح تصنيف معمّر للتنوع الرئيسي لنماذج الاستجابة للمفردة الاختبارية

ويعتمد تصنيف نماذج الاستجابة للمفردة الاختبارية الموضّع بشكل (١٤ - ١) على أربعة

محكّمات كالتالي :

(١) **مستوى الاستجابة**: ويعد أهم المحركات التصنيفية لهذا النوع من النماذج، لأنها يتعلق بطبيعة البيانات المراد تحليلها. ويمكن التمييز بين ثلاثة مستويات للاستجابة لمفردات الاختبارات والمقاييس:

(أ) **مستوى الاستجابة الثنائية** Dichotomous Response Level: ونقصد بذلك أن تكون الاستجابة إما صحيحة أو خطأ، ويحدث ذلك عندما تكون مفردة الاختبار من نوع الاختيار من متعدد، أو الصواب أو الخطأ، أو ما شابه ذلك. غير أنه إذا كانت الاستجابة المحتملة لنوع ما من أنواع المفردات أكثر من استجيبتين، فإنه يمكن احتزازها إلى استجابة ثنائية. وكثير من الاختبارات العقلية والتحصيلية تتطلب هذا المستوى من الاستجابة.

(ب) **مستوى الاستجابة المتعددة** Polytomous Response: وهنا لانستطيع احتزاز الاستجابة إلى قسمين كما سبق أن أشرنا، فالاستجابة في هذا المستوى تشتمل على عدة أقسام. ومثال ذلك الاستجابة على بعض الاستبيانات، وموازين التقدير، ومقاييس الشخصية، وأسئلة المقال في اختبارات التحصيل.

(ج) **مستوى الاستجابة المتصلة** Continuous Response Level: وفي هذا المستوى يفترض أن عدد أقسام الاستجابة لانهائي. وهذا النمط من الاستجابات يتطلب أنواعاً من مفردات الاختبار تختلف اختلافاً ملحوظاً عن تلك التي تتطلبها كل من الاستجابات الثنائية والمتعددة. ويفيد النموذج المتعلق بهذا المستوى من الاستجابة عندما يطلب من الفرد مثلاً الحكم على صحة قضايا معينة حكماً احتمالياً ذاتياً.

(٢) **التركيب البارامترى**: ونقصد بذلك عدد وطبيعة بارامترات كل من هذه النماذج الاستاتيكية. وهذه البارامترات تتعلق بالافتراضات الخاصة بالعمليات السيكلوجية التي تتطوّر عليها استجابات الفرد لمفردة اختبار.

وتوجد في الوقت الحالى ثلاثة نماذج من النوع الثنائى الذى أشرنا إليه تختلف فى عدد بارامتراتها. فهناك النموذج أحادى البارامتر الذى يعرف باسم نموذج راش Rasch Model (RM)، وهو ينطبق على البيانات المستمدّة من الاختبارات إذا اختلفت مفردات كل منها في الصعوبة فقط. وبذلك توصف المفردة ببارامتر واحد هو بارامتر الصعوبة. وكذلك نماذجه المنشقة عنه الموضحة بشكل (١٤ - ١). وهناك النموذج ثانوى البارامتر (LTPM) Two - Parameters Model (Lord, 1968)، وهو ينطبق على البيانات المستمدّة من الاختبارات التي تختلف مفرداتها في كل من الصعوبة والتميز، وبذلك توصف المفردة ببارامترتين.

وقد أضاف بيرنباوم (1968) بارامترًا ثالثاً إلى نموذج لورد يتعلّق بإمكانية التوصل إلى الإجابة الصحيحة عن مفردة اختبار عن طريق التخمين الشوائى، أى بارامتر التخمين، أو مايعرف ببارامتر السخط التقاربى الأدنى Lower Asymptote Three - Parameter Model . وبذلك يكون لدينا نموذج ثالثى البارامتر (BTHPM) . وسوف نتناول هذه النماذج الثلاثة بمزيد من التفصيل بعد قليل.

كما يوجد نموذجان من النوع المتعدد الاستجابة: أحدهما النموذج متدرج Nominal-Graded-Response Model ، والأخر النموذج نوعي الاستجابة Nominal-Response Model .

وينطبق النموذج الأول عندما تكون الاستجابات المختلفة مرتبة، ويكون لكل قسم من أقسام الاستجابة لمفردة نفس القدر من التمييز. ويمكن أن يحدث ذلك عندما تنطوى الاستجابة على عملية مرحلية أو تجمعية. وعندئذ يستخدم النموذج المتدرج الاستجابة في التنبؤ بالاستجابة المحتملة للفرد التي تسق مع مستوى النماءى. ومثال ذلك الاستجابات على استبيانات النمو الخلقي الذى يفترض أنه يحدث من خلال مراحل متتابعة مرتبة أثناء نمو الأطفال. أما فى حالة النموذج الثانى لأن تكون الاستجابات مرتبة، ولا يفترض وجود تتابع مرحلى تنطوى عليه الاستجابة. أى أن ميزان الاستجابة لا يكون رتيباً، بل اسمياً أقسامه شاملة ومتناهية مثنى مثنى، كما أنه لا يفترض في هذه الحالة أن يكون لكل قسم من أقسام الاستجابة القدر نفسه من التمييز .

(٣) **الفرضيات الإحصائية**: تتعلق بطبيعة توزيع الباقي Residuals ، ونقصد بذلك الأخطاء الناجمة عن عدم مطابقة البيانات لأحد هذه النماذج. فإذا كان توزيع الباقي اعتداليا Normal تُستخدم النماذج الكلاسيكية المعتادة، أما إذا كان التوزيع لوغاريتmic Logistic تُستخدم نماذج الاستجابة للمفردة الاختبارية .

ومن الجدير بالذكر أن هذين النوعين من التوزيعات يمثلان أيضاً الصيغتين الرياضيتين لما يُطلق عليه «المتحنى المميز للمفردة Item Characteristic Curve (ICC)» الذي يُعد من المفاهيم الأساسية في هذه النظرية المعاصرة.

(٤) **أبعاد الفضاء الكامن**: يمكن التمييز بين نماذج الاستجابة للمفردة الاختبارية في ضوء أبعاد الفضاء الكامن Latent Space للاستجابة. فمعظم النماذج الموضحة بشكل (١٤ - ١) تفترض وجود سمة أو قدرة واحدة تفسر أداء الفرد في الاختبار؛ لذلك تسمى نماذج أحادبة البعد Unidimensional Models ، ومن أمثلتها نموذج جتمان التراكمي Gutman Cumulative Model (GCM) الذي سبق أن أوضحناه في

الفصل الثاني عشر، وكذلك نموذج لازرسفلد للفئات الكامنة Lazarsfeld Latent Class Model (LLCM)، ويُعدان من النماذج التحديدية Deterministic Models وليست احتمالية. أما نماذج راش، والنموذج ثالثي البارامتر، والنماذج ثلاثي البارامتر فتُعد نماذج احتمالية أحادبة البُعد. وهناك نماذج أخرى يفترض أن أداء الفرد في الاختبار يمكن أن ينطوي على سمات أو قدرات متعددة وليس سمة واحدة، لذلك تسمى نماذج متعددة الأبعاد Multidimensional Models (MM). ويقوم خبراء القياس بإجراء دراسات وبحوث متعددة لتوسيع نموذج راش بحيث يمكن استخدامه في تصميم الاختبارات متعددة الأبعاد وتحليل مفرداتها وبخاصة في مجال قياس القدرات. كما توجد نماذج أخرى لم يتمكن بناؤها، كما لم تتحدد فرضياتها الإحصائية تحديداً دقيقاً لأنها لا تزال موضع البحث والتجريب (Whitely, 1980). لذلك فإن النماذج أحادبة البُعد هي الشائعة الاستخدام في القياس النفسي والتربوي في وقتنا الحاضر.

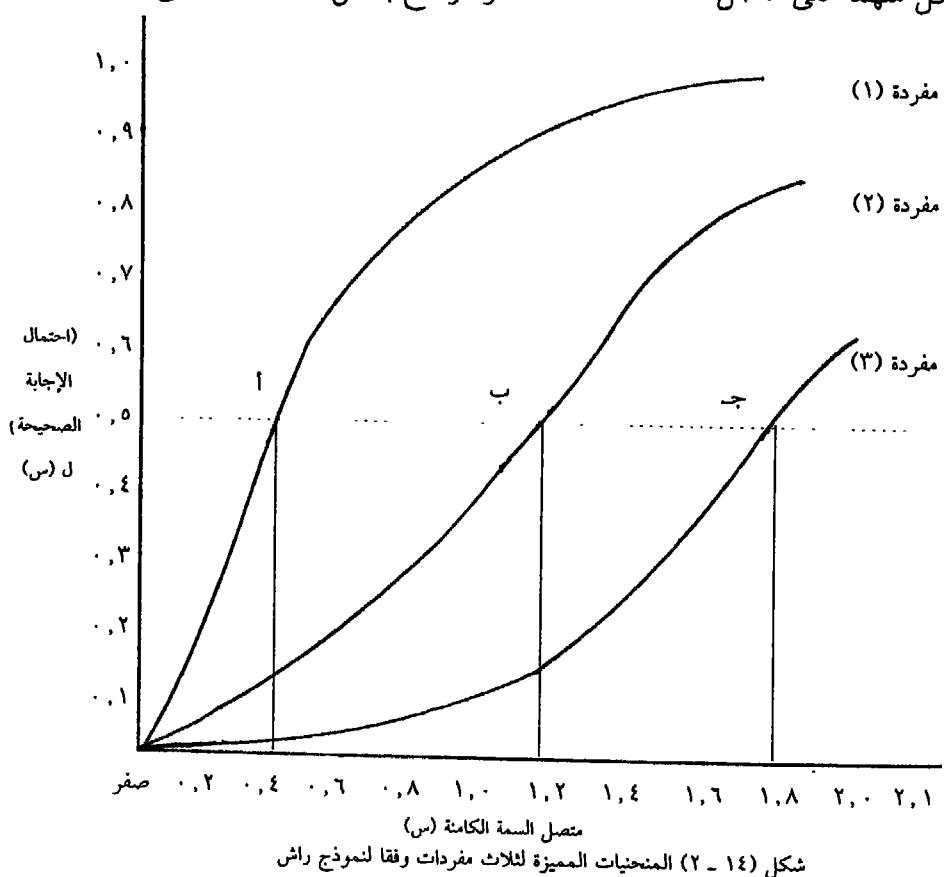
مفهوم المنحنى المميز للمفردة (ICC):

بعد مفهوم المنحنى المميز للمفردة من المفاهيم المهمة في نماذج الاستجابة للمفردة الاختبارية التي تتناول سمة كامنة متصلة يتم قياسها بمفردات اختبارية من النوع الثنائي (أى تتطلب إجابة صحيحة أو إجابة خطأ). ويمثل هذا المنحنى بدالة رياضية تربط بين احتمال نجاح الفرد في إجابة المفردة، وبين السمة أو القدرة التي تقيسها مجموعة مفردات أو يقيسها اختبار معين. أى أنها دالة انحدار درجة المفردة على السمة الكامنة التي يقيسها الاختبار، وهذه الدالة غير خطية Non - Linear. ويمكن التمييز بين هذه النماذج وفقاً للصيغة الرياضية للمنحنى المميز للمفردة، حيث إن هذه الصيغة تمثل النموذج تمثيلاً رياضياً.

ويمكن أن يتضح مفهوم المنحنى المميز للمفردة إذا أجرينا مقارنة بين كل من النظرية الكلاسيكية للاختبارات ونظرية الاستجابة للمفردة الاختبارية. فكل منها يعتمد على فرضية مؤداها أن موقع الفرد على متصل سمة كامنة معينة يحكم احتمال إجابته إجابة صحيحة عن مفردة اختبار يقيس هذه السمة، ولكنها يختلفان في كيفية تحديد هذا الموقع. وارتباطه باحتمال الإجابة الصحيحة عن المفردة. أى أن الفرق بينهما يتعلق بشكل وخصائص الدالة التي تحدد هذا القيد. فالنظرية الكلاسيكية تفترض أن منحنى الدالة المميزة للمفردة لا يختلف باختلاف مفردات الاختبار، وتتمثل هذه الدالة بالمنحنى الاعتدالي التراكمي لمتغير عشوائي. ويفترض في هذا المنحنى أن احتمال

الإجابة الصحيحة عن مفردة اختبارية دالة تزايدية مطردة لموقع الفرد على متصل السمة الكامنة.

ونظرية الاستجابة للمفردة الاختبارية لا تتفق مع النظرية الكلاسيكية في افتراضها تساوي احتمالات الإجابة الصحيحة عن جميع مفردات الاختبار لفرد ذي قدرة معينة. فمن الممكن أن تكون بعض المفردات أكثر صعوبة من غيرها بالنسبة لهذا الفرد؛ لذلك فإن احتمال الإجابة الصحيحة عن مجموعة من المفردات يعد دالة لموقع كل من الفرد والمفردة على متصل السمة الكامنة التي تقيسها هذه المفردات. أى أن هذه النظرية تفترض وجود دالة مميزة Characteristic Function تعتمد على موقع الفرد والمفردة معا على متصل سمة كامنة معينة، ويتحدد هذا المنحنى أو هذه الدالة إذا علمنا هذا الموقع. لذلك يطلق على هذا القياس «القياس المتعدد أو المقترن Conjoint Measurement». ولتوسيع ذلك نفترض أن لدينا ثلاثة مفردات اختبارية يختلف موقع كل منها على متصل السمة الكامنة كما هو موضح بشكل (١٤ - ٢) التالي:



ويتبين من شكل (١٤ - ٢) أن المفردات الثلاث تقع عند النقط ٤، ٠، ٢، ٢ على متصل السمة الكامنة، فالأفراد الذين يقعون عند إحدى هذه النقط على نفس المتصل يكون احتمال إجابتهم إجابة صحيحة عن المفردة المنشورة لنفس الموقع . . . لذلك فإن صعوبة المفردة في إطار نماذج الاستجابة للمفردة الاختبارية ليست نسبة عدد الأفراد الذين يجيبون عنها إجابة صحيحة كما هو الحال في النظرية الكلاسيكية، وإنما هي النقطة التي على متصل السمة الكامنة التي تتوقع عندها أن يكون احتمال إجابة فرد عنها إجابة صحيحة بدون تخمين . . .

فالفرد الذي يقع عند النقطة ٤، ٠ على متصل السمة الكامنة يكون احتمال إجابته إجابة صحيحة على المفردة (١) ٥٠، ٠ بينما لا يكون الحال كذلك عند محاولته الإجابة عن أي من المفردتين (٢) أو (٣) مثلاً، وذلك لأن المفردة (١) أقل صعوبة من كل من المفردتين (٢)، (٣) بالنسبة لدرجته في السمة الكامنة. وهذا يدل أيضاً على أن المفردات الثلاث تختلف في تميزها بين المستويات المختلفة للسمة. فتميز المفردة في إطار نظرية الاستجابة للمفردة الاختبارية يختلف أيضاً عن نظيره في النظرية الكلاسيكية. فلكل نحدد معامل تميز كل من المفردات الثلاث في شكل (٢) يجب أن نرسم خطأ أفقياً عند النقطة ٥، ٠ على المحور الرأسى التي تناظر الإجابة الصحيحة بحيث يقطع المنحنيات المميزة لهذه المفردات. ومعامل تميز كل مفردة عندئذ هو ميل المنحني المميز لها عند نقطة التقاطع. فإذا نظرنا إلى الشكل نجد أن المفردة (١) أكثر تميزاً من المفردة (٢)، والمفردة (٢) تميزاً من المفردة (٣)، وذلك لأن ميل المنحنيات الثلاث عند النقط ١، ب، ج يزيد عن الواحد الصحيح بالنسبة للمفردة (١)، ويقل عن الواحد الصحيح بالنسبة للمفردة (٢). فمعامل التمييز المرتفع يدل على أن الفرد الذي تفوق سنته الكامنة صعوبة المفردة على نفس متصل السمة يكون احتمال إجابته إجابة صحيحة عن هذه المفردة كبيراً، والفرد الذي تقل سنته الكامنة عن صعوبة المفردة يكون احتمال إجابته إجابة خطأ عن المفردة كبيراً. ومعامل التمييز المنخفض يدل على أن احتمال الإجابة الصحيحة عن المفردة منتظم، وأن المفردة لا تميز جيداً بين مستويات السمة المعينة.

ومن هذا يتضح أن لكل من النظرية الكلاسيكية للاختبارات، ونظرية الاستجابة للمفردة الاختبارية افتراضات مختلفة حول شكل وخصائص الدالة المميزة للمفردة. في بينما تفترض النظرية الكلاسيكية أن هذه الدالة تأخذ شكل المنحني الاعتدالى التراكمي، تفترض نظرية الاستجابة للمفردة الاختبارية أن الدالة تأخذ شكل المنحني اللوغاريتمي التراكمي، وأن لكل مفردة منحنى خاصاً بها يختلف في البارامترات المحددة له، من مثل ذلك الصعوبة والتميز.

النماذج الاستاتيكية اللوغاريتمية أحادية البعد:

نظراً لشيوخ استخدام النماذج الاستاتيكية أحادية البُعد في تصميم وبناء الاختبارات والمقاييس التربوية والنفسية في وقتنا الحاضر، فسوف نلقي الضوء على بعض أهم هذه النماذج التي أوضحتها في شكل (١٤ - ١). وبالطبع لن نستطيع في هذا الكتاب تناول هذه النماذج بالتفصيل أو عرض صيغها الرياضية نظراً لأن كلّ منها يحتاج إلى أكثر من مجلد. وقد ناقش المؤلف كثيراً من هذه النماذج بمزيد من التفصيل في مؤلفين آخرين، وفي عدد من الدراسات التي أجرتها في هذا المجال.

نموذج راش Rasch Model

يُطلق على هذا النموذج «نموذج البارامتر الواحد One - Parameter Model»، وقد اقترحه جورج راش G. Rasch كما أوضحنا من قبل. ويهتم بتحديد موقع المفردة الاختبارية على ميزان صعوبة جميع المفردات التي تشكل الاختبار (بارامتر الصعوبة)، كما يهتم بتدرج مستويات قدرة الفرد باختبار معين على نفس ميزان تعيير المفردات . لذلك يعتمد النموذج على ما يسمى القياس المستحد Conjoint Measurement ويفترض النموذج تساوى جميع مفردات الاختبار في التمييز بين مستويات القدرة المقاومة، وهي فرضية تبدو غير واقعية في تصميم الاختبارات وبنائها، كما يفترض النموذج في أبسط صورة أن تكون القدرة المقاومة أحادية البُعد، أي أن جميع مفردات الاختبار التحصيلي مثلاً تقيس قدرة واحدة مثل القدرة على فهم المقروء، أو رسم الأشكال الهندسية... إلخ، وأن تكون جميع مفردات الاختبار من النوع ثنائي الدرجة، مثل مفردات الاختبار من متعدد ذات الإجابة الصحيحة الواحدة، وكذلك يفترض النموذج أن الفرد لا يلتجأ إلى التخمين العشوائي في إجابته عن مفردات الاختبار .

ويوجد نموذجان أكثر عمومية ولكنهما يتبعان إلى عائلة نموذج راش في القياس التربوي والنفسى: أحدهما يسمى «نموذج التقدير المتدرج Partial Credit Model» ويعُد امتداداً لنموذج راش ثنائي الدرجة (صفر، ١). ويُستخدم هذا النموذج إذا كانت الدرجات تقع في أكثر من قسمين مرتبتين Polytomus Scores كما في درجات أسئلة المقال وحل المشكلات، فهنا لا تكون الإجابة إما صحيحة أو خطأ وإنما تقدر درجات للمعلومات الجزئية المتباعدة ، والآخر يسمى نموذج راش متعدد الأبعاد Multidimensional Rasch Model واحد.

نموذج البارامترین Two-parameters Model

ويسمح بأن تختلف مفردات الاختبار في صعوبتها وتميزها وهو ما يلاحظ عادة في بناء الاختبارات. إذ من الصعب إيجاد مجموعة من المفردات تميز بدرجة واحدة بين مستويات السمة أو القدرة التي يقيسها اختبار معين، وهو الافتراض الذي اعتمد عليه نموذج راش.

نموذج البارامترات الثلاثة Three- Parameters Model

وهو يعرف بنموذج بيرنبوم Birnbaum model، فقد أضاف بيرنبوم بارامترًا ثالثاً وأطلق عليه بارامتر الخط التقاربي الأدنى Lower Asymptote Line أو بارامتر التخمين Guessing Parameter فيما يتعلق بالمفردات الاختبارية التي تتطلب الاختيار بين بدائل معطاة. وهذا البارامتر يحدد احتمال أن يجب فرد يفترض أن مستوى قدرته منخفضاً انخفاضاً لا نهائياً، ومع هذا يجب إجابة صحيحة عن غالبية مفردات الاختبار عن طريق التخمين. لذلك تبدو أهمية هذا البارامتر عند مطابقة بيانات مستمدة من مجموعات من مفردات الاختبارية من نوع الاختيار من متعدد لهذا النموذج .

ومن الجدير بالذكر أن هناك نماذج أخرى انبثقت عن هذه النماذج الثلاثة تتعلق بتحليل مفردات موازين التقدير Rating Scales، ونموذج الاستجابة الاسمية Nominal Response Model، ونموذج البدائل التي تشتمل عليها مفردات الاختبار من متعدد Item Options Model، والنموذج رباعي البارامتر Four-parameters Model.

ولكل من النماذج السابقة صيغ رياضية لامجال لذكرها هنا، وهذه الصيغ عبارة عن دوال رياضية ل OGARITMIE احتمالية تربط بين بارامترات النموذج. وتُستخدم أساليب التحليل العددي Numerical Analysis في تقدير بارامترات كل نموذج، مثل بارامتر صعوبة المفردة، وبارامتر تميز المفردة، وبارامتر التخمين، وكذلك التتحقق من حسن مطابقة البيانات للنموذج باستخدام بعض الأساليب الإحصائية. ويطلب تنفيذ هذه الأساليب برامج خاصة للحاسوب. ويوجد في الوقت الحاضر عدد من برامج الحاسوب المتوفّرة التي يمكن باستخدامها التوصل إلى قيم تقديرية لبارامترات النموذج الذي يرى الباحث ملاءمته لاختبار أو مقياس معين. ومن أمثلة هذه البرامج الشائعة الاستخدام مايلي :

- (١) برنامج RASCAL (WINDOWS) : يُجرى تحليل مفردات الاختبارات باستخدام نموذج راش ، ويُعد برنامجاً قوياً وسهل الاستخدام ويعمل على بيئة النافذة . ويمكن استخدامه في تعديل مفردات اختبارية يبلغ عددها ٧٥٠ مفردة بدون حد أقصى لعدد الأفراد الذين يُختبرون بهذه المفردات .
- (٢) برنامج RSP (1.0) : يُجرى تحليل المفردات باستخدام نموذج راش أيضاً ، ولكن في التصميمات غير الكاملة التي تشمل على عدد من الاختبارات يتكون كل منها من عدد من المفردات تُطبق على مجموعة من الأفراد . ويجب أن تلتزم الاختبارات معاً بمفردات مشتركة أو بتوزيعات مشتركة للسمات المراد قياسها . ويستوعب البرنامج ٣٢ اختباراً ، ويتكون كل منها من ٩٦ مفردة على الأكثر .
- (٣) برنامج XCALIBRE (WINDOWS) : يُجرى العمليات اللازمة لتقدير بارامترات النموذجين ثنائي وثلاثي البارامتري . ويستوعب ٧٥٠ مفردة اختبارية بدون حد أقصى لعدد الأفراد الذين يُختبرون بهذه المفردات .
- (٤) برنامج ASCAL : يُجرى أيضاً تقديرات بارامترات النموذجين ثنائي وثلاثي البارامترا ولكن باستخدام إحصاء بيز Bayesian Approach ، ويستوعب ٢٥٠ مفردة اختبارية يُختبر بها ١٥٠ فرد على الأكثر .
- (٥) برنامج MSP 3.0 : يُجرى تحليل مفردات متعددة الاستجابات Polytomous Items وكذلك المفردات ثنائية الاستجابة Dichotomous Items ، وذلك لتصميم اختبارات أحادية البُعد تطبق البيانات المستمدة منها على أحد نماذج الاستجابة للمفردة الاختبارية . ويستوعب ١٠٠ مفردة اختبارية على الأكثر يُختبر بها عدد لا يزيد عن ٣٢٠٠ فرد ، ويُجرى تقديرات بارامترات النموذج باستخدام أساليب لبارامتريا .
- (٦) برنامج PARELLA 1.1 : يُجرى تحليل لاستجابات الأفراد لمجموعة من فقرات أو عبارات موازين الاتجاهات والتفضيلات ، حيث يتم تمثيل موقع كل فرد بالنسبة للفقرة التي يوافق عليها بأقصى مسافة بينهما في بُعد واحد . ويستوعب ٦٠ فقرة يستجيب لها ٣٠٠ فرد على الأكثر .

ميزاً وعيوب النماذج الإستاتيكية:

مما سبق يتبيّن تعدد النماذج الإحصائية السيكومترية التي يمكن استخدامها في القياس النفسي والتربيوي وتوعتها وتبنيها . ونظراً إلى أن النماذج الإستاتيكية المتعلقة بالاستجابة للمفرد الاختبارية (SIRMs) نالت اهتماماً كبيراً من المشتغلين ببناء

الاختبارات والمقاييس، فإنه يجب على المهتمين بذلك أن يتعرفوا مزايا كل من هذه النماذج وعيوبها، وأن يختاروا النموذج الذي يتميز بالبساطة والواقعية وإمكانية التنفيذ، ومدى تحقق فرضياته في البيانات، وسهولة تقدير بaramتراته، ومنطقية تفسير نتائجه.

وعلى الرغم من أن نموذج راش هو أكثر هذه النماذج استخداماً في بناء الاختبارات والمقاييس لما يتوافر فيه من بعض هذه المزايا، إلا أنه يجب على من يستخدم هذا النموذج في تعديل المفردات الاختبارية أن يكون مدركاً لجوانب القصور فيه وحدود تطبيقاته، حتى لا يُساء استخدامه أو تعمم نتائجه تعميماً خطأ.

ولعل أهم مزايا وعيوب هذا النموذج ما يلى:

(١) نموذج راش هو النموذج الوحيد من بين النماذج الاحتمالية التي تتوافر فيه خاصية استقلالية بaramتراته، أي عدم تباين Invariance تقديرات قدرة الأفراد بتباين عينة المفردات التي يُختبرون بها، وعدم تباين الخصائص السيكومترية للمفردات بتباين عينة الأفراد التي أجري عليها الاختبار. غير أن هذا يتطلب التتحقق بصورة دورية من عدم حدوث تغيرات ربما تطرأ على القيم التقديرية لبارامترات النموذج نتيجة لتغير خصائص أفراد مجتمع المختبرين، أو تغير طبيعة أو محتوى أو معنى المفردات الاختبارية بالنسبة إليهم بمرور الزمن.

(٢) يعد نموذج راش أبسط النماذج الإحصائية الاحتمالية التي أشرنا إليها، حيث يشتمل على بaramتر واحد يتعلق بمفردات الاختبار وهو صعوبة المفردة Item Difficulty، فعلى الرغم من تباين صعوبة مفردات الاختبار فإن النموذج يفترض أن جميع المفردات تساوى في قدرتها على التمييز. وهذا القيد يُضيف قيداً آخر على انتقاء المفردات التي يشتمل عليها الاختبار. فالمفردات لا يجب أن تميز بين مستويات القدرة المقاسة تميزاً دالاً فحسب، بل إن جميع مفردات الاختبار يجب أن تكون قدرتها على التمييز متساوية تقريباً.

ويرى مؤيدو هذا النموذج أن الميزة التي تنجم عن ذلك هي أن المجموع غير الموزون للدرجات مفردات الاختبار يعد إحصاءً كافياً Sufficient Statistic لتقدير قدرة الفرد المختبر. كما أنه يترب على ذلك أيضاً ميزة أخرى تتعلق بوصل مفردات الاختبارات على شكل حلقات متتابعة Item Linking لتحقيق عملية تكافؤ الاختبارات Test Equating وهو أمر ضروري بل وأساسى في بناء نظم بنوك الأسئلة من أجل إمكانية تشكيل العديد من صيغ الاختبارات المختلفة، ولكن تؤدي إلى درجات كلية على ميزان موحد. فعملية التوصل إلى درجات متباينة للصيغ المختلفة من الاختبار

تسمى عملية التكافؤ . وعلى الرغم من أن تكافؤ الدرجات الحقيقة سواء التكافؤ الأفقي أو الرأسى يمكن التوصل إليه بشروط متعددة ، إلا أن نماذج الاستجابة للمفردة الاختبارية وبخاصة نموذج راش ربما ييسر ذلك . غير أن هذا يتطلب التتحقق من فرضية تساوى المفردات فى قدرتها على التمييز ، وذلك بفحص أنماط البواقي المعيارية التى تنجم عن تقييم حسن مطابقة البيانات للنموذج .

(٣) يؤدى نموذج راش إلى منحنيات مميزة للمفردات (ICCs) تكون متوازية أى لا تتقاطع ، وكذلك المنحنيات المميزة للأفراد المختبرين (PCCs) . وهذا يسر عملية التقدير الإحصائى لبارامترات النموذج Parameters Estimation والوصول إلى ما يسمى بالتقارب Convergence . فالمنحنى المميز للمفردة هو دالة رياضية تربط بين احتمال الإجابة الصحيحة لفرد ما عن مفردة اختبارية ، وبين القدرة التى يقيسها الاختبار . والمنحنى المميز للفرد هو تلك الدالة التى تربط بين احتمال الإجابة الصحيحة لمجموعة من الأفراد المتساوين تقريبا فى القدرة المقاسة وبين صعوبات المفردات .

غير أن توافق هذه المنحنيات لا يحدث إلا إذا حققت البيانات الفرضيات التى يستند إليها النموذج ، وبخاصة فرضية بعد الأحادي للاختبار ، وتساوى تميز مفرداته ، وعندئذ يمكن تفسير مستويات وأنماط القدرة المقاسة . علما بأن التتحقق من خاصية بعد الأحادي يشوبه كثير من المشكلات الإحصائية والسيكومترية ، كما أن كثيرا من الاختبارات ربما لا تقياس سمات أحادية البعد .

(٤) في حين تميز النماذج الاستاتيكية الأخرى وبخاصة النموذج ثلاثي البارامتر بمزايا تفوق نموذج راش ، حيث إنه يحدد بارامتر الموضع Location Parameter وبaramتر ميل المنحنى المميز Slope Parameter ، وبaramتر التخمين Guessing لكل مفردة وبخاصة مفردات الاختيار من متعدد الذى يعتمد عليها نموذج راش مما يؤدى إلى نتائج أكثر واقعية في العملية الاختبارية وبناء المقاييس ، إلا أن هذا يؤدى إلى تقاطع المنحنيات المميزة للمفردات ، وبالتالي يؤدى إلى صعوبات فى عمليات التقدير الإحصائى Parameters Estimation لهذه البارامترات . لذلك نجد أن المنحنيات المميزة في حالة استخدام نموذج راش تتباين بعض اللينينات والتعرجات بالنسبة إلى الأفراد منخفضى القدرة الذين يكون احتمال توصلهم للإجابة الصحيحة عن مفردة ما من مفردات الاختيار من متعدد الذى يشتمل عليها الاختبار أقل مما هو متوقع عن طريق الصدفة ، نظرا إلى أنهם ينجذبون إلى المشتتات غير الصحيحة للمفردة ، وذلك

لعدمأخذ الأخطاء الناجمة عن التخمين بعين الاعتبار في النموذج، وبخاصة أنه يعتمد اعتماداً كلياً على مفردات الاختيار من متعدد. لذلك اتجهت البحوث في الآونة الأخيرة إلى بناء نماذج بديلة لنموذج راش تعالج هذه الظاهرة. وهذا أمر ضروري نظراً إلى أن مشتتات مفردات الاختيار من متعدد Distractors تبني من أجل الكشف عن أمثل هؤلاء الأفراد.

أما إذا اشتمل الاختبار على أسئلة من النوع الذي يتطلب إجابة حرة فإنه يجب اختيار نموذج آخر مناسب لهذا الغرض، مثل نموذج التقدير المتدرج Partial Credit Model.

غير أن كثيراً من خبراء القياس المعاصرين يرون أن هذه النماذج سوف تؤدي إلى زيادة المرونة والدقة في بناء الاختبارات التحصيلية. فمثلاً يمكن تجميع عدد كبير من المفردات الاختبارية التي يتم تعبيّرها Calibrated Items وفق أحد هذه النماذج، مما يجعل الخصائص الإحصائية لهذه المفردات مستقلة عن عينة الطلاب التي تم تجريب المفردات عليها. وهذا وبالتالي يجعل من الممكن بناء اختبارين يسُوديان إلى درجات متكافئة دون الحاجة إلى تطبيقهما معاً، أو تصميم اختبار ذي خصائص سيكومترية محددة مسبقاً، أو تحديد ما إذا كان أحد الطلاب قد حقق مستوى معندي من مستويات الإتقان، وهذا المستوى يكون مستقلاً عن مجموعة المفردات التي أجاب عنها الطالب. أي إذا اختبر الطالب باختبار آخر يشتمل على عينة من المفردات المسحوبة من نطاق هذه المفردات Item Domain والتي تم تعبيّرها وفق أحد هذه النماذج، فإنه يتوقع أن يتحقق الطالب مستوى الإتقان السابق نفسه تقريباً في حدود قدر ضئيل من الخطأ.

ومما لا شك فيه أن تقنيات الحاسوب أثّرت استخدام نماذج الاستجابة للمفردة الاختبارية وأدت إلى مزيد من تطبيقاتها في مختلف المجالات، مثل بناء الاختبارات مرجعية المحك، وتفريد القياس، وبناء نظم بنوك الأسئلة، والتكافؤ الأفقي والرأسي للدرجات الاختبارات، وغير ذلك. وسوف نلقى الضوء على بعض هذه التطبيقات في الفصل التالي.

(ثانياً) نظرية الثبات متعدد الأوجه Generalizability Theory

أوضحنا في الجزء السابق بعض جوانب قصور النظرية السيكومترية الكلاسيكية للختبارات (CTT)، وكيف حاولت نظرية الاستجابة للمفردة الاختبارية (IRT) التغلب على ذلك. وعلى الرغم من أن كلاً من النظريتين تعالج تباين أخطاء القياس معالجة

مختلفة، إلا أن كليهما ينظر إلى مصدر هذه الأخطاء نظرة ضيقة ومحدودة. فالمعلومات أو البيانات المستمدة من أي إجراءات قياس سواء في العلوم الطبيعية أو العلوم السلوكية تكون مشوبة بدرجة معينة من الخطأ. فقياس الطول مثلاً يختلف وفقاً لظروف متعددة متعلقة بهذا القياس، مثل المسطرة المستخدمة والمادة المصنوعة منها، والفرد القائم بالقياس، وتسجيل القيمة العددية المناظرة للطول، وظروف الإضاءة، وغيرها ذلك. ولزيادة دقة القياس يمكن إيجاد متوسط قياسات متعددة للطول في ظروف محددة للتوصيل إلى قيمة تقديرية للقياس المثالي، أي القياس الافتراضي الذي يمكن الحصول عليه إذا استطعنا إيجاد المتوسط في جميع الظروف المحتملة الممكنة. غير أن هذا يتطلب معرفة عدد مرات القياس والظروف المختلفة المتعلقة به للتوصيل إلى درجة مقبولة من الدقة. فإذا رأينا أن اختيار مسطرة مصنوعة من مادة معينة لا يؤثر على قياس الطول، ولكن ربما يؤدى اختلاف القائمين بالقياس إلى تباين ملحوظ في القيم العددية المناظرة للطول، فإنه ربما يفضل في هذه الحالة إيجاد متوسط القياسات باستخدام عدد قليل من المساطر، وزيادة عدد القائمين بالقياس. وعلى الرغم من أنه يمكننا استخدام مسطرة واحدة في جميع القياسات، وبذلك تقل الأخطاء وتزيد دقة القياس، إلا أن هذا يضع قيداً على نطاق ظروف القياس الذي يمكن التعميم عليهما، مما يجعل تفسير القياسات الناتجة محدوداً، أي أن أخطاء القياس من هذا المنظور ليست متضمنة في عملية القياس، وإنما تعتمد على تحديد أي ظروف القياس التي تسهم في الخطأ، لكنه يتسعى تعميم القياس عليها. ولعل هذا هو ما تتناوله نظرية الثبات متعدد الأوجه.

بعض أوجه قصور المفهوم الكلاسيكي للثبات والصدق:

يعد كل من مفهوم الثبات والصدق من المفاهيم الأساسية في القياس التربوى والنفسى وبخاصة الاختبارات والمقاييس. فالبيانات المستمدة من هذه الأدوات يجب أن تميز بالثبات والصدق مهما اختلف الغرض من استخدامها. فالثبات من وجهة النظر الكلاسيكية يعني أن تعكس هذه البيانات الجوانب الحقيقية للسمة أو القدرة المقاسة، أي المصادر المتتظمة للتباعين، ولاتعكس عوامل الصدفة أو العشوائية، أي تباين الخطأ.

أما الصدق فيعني مدى الاستفادة من المقاييس في صنع قرارات تتعلق بأهداف معينة، أو مدى ملاءمة الاستدلالات من درجات الاختبارات أو غيرها من أدوات القياس. فيقدر ماتكون المعلومات التي تمدنا بها هذه الأدوات مضللة أو غير كافية لصنع قرارات معينة، تكون البيانات المستمدة منها غير صادقة مهما كانت البيانات ثابتة أو متسبة. فأدوات القياس تكون قليلة الجدوى إذا لم تكن صادقة لغرض معين.

والنظرية الكلاسيكية للاختبارات تعد من النظريات الواهنة والمحدودة Weak Score Theory في معالجتها لمفهوم الثبات والصدق نظراً لأنها لا تميز بين اختياء القياس، وإنما تعطى قيمة تقديرية كلية لمصادر اختياء القياس المتعددة المتعلقة بإحدى صيغ الاختبار المستخدمة بواسطة مختبر أو فاحص معين، وفي ظروف معينة؛ لذلك تعدد أساليب تقدير الثبات كما أوضحتنا في الفصل الرابع.

وقد اهتم كرونباك Cronbach وتعاونه بمراجعة وعميق تعريف وتفسير مفهومي الثبات والصدق ومعالجتهما معالجة أكثر شمولية من منظور متعدد الأبعاد. فقد أعاد كرونباك تفسير نظرية الثبات، وقام بوضع أساس بناء نظرية جديدة تهتم بإمكانية تعميم عينة القياسات المستمدة من مجموعة من الملاحظات على النطاق الشامل لهذه الملاحظات Universe of Observations، وأطلق على هذه النظرية «نظرية إمكانية التعميم Generalizability Theory». وقد أشرنا إلى هذه النظرية في الفصلين الرابع والخامس.

ويرى كرونباك أن ضعف النظرية الكلاسيكية للاختبارات يرجع إلى تبسيطها الشديد لمفهومي الثبات والصدق، واستخدام الباحثين لهذين المفهومين استخداماً آلياً، مما أدى إلى غموض معناهما الحقيقي واستخدامهما الفعلى، ويؤكد أن معظم جوانب النظرية الكلاسيكية تعد حالات خاصة من نظرية العامة الجديدة.

نظريّة إمكانية التعميم تزوّدنا بإطار فكري مرجعي لكيفية تقدير ثبات القياسات السلوكية التي نحصل عليها بواسطة إجراءات أو أدوات القياس المختلفة، مثل الاختبارات، وقوائم الملاحظة أو المراجعة، وموازين التقدير، ومقاييس الميول والاتجاهات والشخصية التي تُستخدم في البحوث التربوية والنفسية وفي اتخاذ القرارات في المجالات التطبيقية.

ونظرية إمكانية التعميم لا تعد امتداداً للنظرية الكلاسيكية فحسب، بل هي أسلوب جديد يساعد الباحث ومصمم المقاييس في معرفة ما إذا كانت قياسات معينة تحقق الهدف الذي ينشده من القياس، مثل تقويم مقرر دراسي، أو انتقاء الأفراد، أو تقويم بنائي أو تجمعي لبرنامج معين، أو الحصول على أدلة تتعلق بصدق التكوين الفرضي بعامة.

بعض المفاهيم الأساسية لنظرية إمكانية التعميم:

عرض كرونباك ومعاونه (Cronbach, Gleser, Nanda & Rajaratnam, 1972) هذه النظرية بالتفصيل في مؤلف واحد بعنوان «اعتمادية القياسات السلوكية».

The Dependability of Behavioral Measurements : Theory of Generaizability for Scores and Profiles ».

وقد اهتم كثير من خبراء القياس بعد ذلك بتطوير النماذج الإحصائية متعددة المتغيرات التي تستند إليها هذه النظرية (Brennan, 1980; Cardinet, 1981, webb, 1981). إذ إن هذه النظرية تهتم بتحليل المكونات المتعددة للبيان في مجموعة من القياسات أو الملاحظات. وقد لاحظنا في الفصل الرابع أن بعض أساليب تقدير الثبات في إطار النظرية الكلاسيكية تستند إلى طرق تحليل البيانات Analysis of Variance غير أن هذه الطرق لا تستخدم في إطار نظرية إمكانية التعميم كما هي مستخدمة في التصميمات التجريبية، حيث إنها تختلف عنها في المصطلحات وأنواع التصميمات، كما تختلف عنها في المنظور والتوجه. ويمكن استخدام هذه النظرية في مناقشة قضايا سيكومترية متعددة بطريقة منتظمة ربما أغفلتها النظريات السيكومترية الأخرى أو عالجتها معالجة سطحية.

غير أن ما تتطلبه هذه النظرية من تصورات فكرية وأساليب إحصائية معقدة، بما لا يكون عائقاً إذا أردنا الإفاده من إسهاماتها المتعددة.

فقد أوضحنا في الفصل الرابع أن النظرية الكلاسيكية في تفسيرها للثبات تعتمد على المقاييس المتوازية أو المتكافئة Parallel Measures. فهي تفترض وجود مقاييس متكافئين أو أكثر لقياس سمة أو قدرة معينة. ونقصد بتكافؤ المقاييس تماثل محتواها، وتساوي متوسطات وبيان درجاتها، وكذلك تساوى الارتباطات البيانية للدرجات مفرداتها المتاظرة.

ويتضح استخدام هذا المفهوم في بناء الاختبارات المفتوحة التيتناولناها في الفصل الثامن وغيره من الفصول بدرجة أفضل من استخدامه في الأنواع الأخرى من المقاييس التي تعتمد على التقديرات Ratings كما في قياس الانجذابات والشخصية. فعلى الرغم من أن مثل هذه الأنواع تكون متوسطات التقديرات، وانحرافاتها المعيارية، والارتباطات البيانية غير متساوية، إلا أنه يجب التأكد من ثبات هذه التقديرات بطريقة ما؛ لذلك كان من الضروري ابتكار مفاهيم أخرى لا تعتمد على هذه النظرية

الكلاسيكية. ومن بين هذه المفاهيم، مفهوم النطاق الشامل Universe Domain . فتحن نهتم في التقديرات، وفي القياس بعامة، بإمكانية تعميم هذه التقديرات أو القياسات من موقف إلى آخر، ومن عينة من الأفراد إلى عينة أخرى. فالملحوظات (ولتكن درجات مجموعة من الأفراد في اختبار أو مقياس ما) التي يجمعها الباحث في موقف معين، تعد بمثابة عينة من «الملحوظات» التي يشتمل عليها نطاق شامل يتضمن «ظروف أو شروط جمع الملاحظات Conditions of Observations»، أي المثيرات أو المواقف المعينة التي يتم في ضوئها جمع الملاحظات أو الحصول على القياسات. وبالطبع يمكن أن تتمي ملاحظات أو قياسات معينة إلى نطاقات شاملة مختلفة. لذلك إذا أراد الباحث أن يتأكد من دقة تعميم الملاحظات أو القياسات على نطاق شامل معين، فعليه أن يحدد هذا النطاق تحديداً دقيقاً.

فالدرجة الحقيقة في إطار النظرية الكلاسيكية هي متوسط درجات الفرد التي يحصل عليها إذا طبق عليه عدد لا نهائي من الاختبارات المتوازية، أي أن الدرجة الحقيقة هنا تعتمد على نطاق شامل واحد للقياس. أما الدرجة الشاملة Universe score التي يقترحها كرونياك في نظريته، فتعتمد على فكرة تعدد النطاقات الشاملة التي يود صانع القرار تعميم قياساته عليها.

وهذا يبين أنه ليس هناك درجة حقيقة واحدة للفرد في الاختبار True score ، كما هو معروف في النظرية السيكومترية الكلاسيكية، وإنما يكون له درجة شاملة Observed Universe score ، يمكن تعريفها بأنها القيمة المتوقعة لدرجاته الملاحظة Scores التي يحصل عليها في مختلف المواقف التي تتسمى إلى نطاق الشامل المطلوب. فالدرجات التي تتسمى إلى نطاق شامل معين ربما تباين في أكثر من جانب يطلق على كل منها «وجه أو بعد Facet». فمثلاً ربما تكون الأوجه أو الأبعاد هي حجم عينة الأفراد الذين يطبق عليهم اختبار أو مقياس معين، أو أنواع التدريبات التي يتلقونها قبل اختبارهم، أو صيغة الاختبار التي تجري عليهم، أو ظروف إجراء الاختبار، وغير ذلك. وتعد كل هذه الأوجه أو الأبعاد مصدراً هاماً من مصادر التباين التي يجب أن تؤخذ بعين الاعتبار. لذلك ينبغي أن يحدد الباحث مصدر أو مصادر التباين التي يهتم بها. فمثلاً ربما يميز الباحث بين عينات من الأفراد تشتمل على أعداد تترواح بين ١ ، ١٥ ، ٣٠ فرداً، وأكثر من ٣٠ فرداً. أو ربما يكون الأفراد الذين يُجرى عليهم الاختبار طلاباً أو معلمين أو فئات أخرى، وهكذا.

لذلك يوصى كرونباك Cronbach القائمين ببناء الاختبارات أو نشرها تحديد «الشروط أو الظروف Conditions» التي تُمكّن مستخدمي الاختبار من الحصول على نتائج متكافئة من الصيغتين (أ)، (ب) مثلاً من صيغ اختبار معين. إذ ربما يكون أحد هذه الشروط أن تطبق صيغتا الاختبار على عينة تشتمل على عدد لا يقل عن ٣٠ فرداً لكي يحصل على نتائج متكافئة، أو ربما يجب عليه أن يُطبق الاختبار على أفراد تلقوا قدرًا معيناً من التدريب، وما إلى ذلك.

والحقيقة أن نظرية إمكانية التعميم لم تهتم فقط بكيفية بناء أدوات قياس تكون نتائجها متكافئة في ظروف ومواصفات اختبارية مختلفة، وإنما اهتمت أيضاً باستخدامات أدوات القياس في صنع قرارات معينة. لذلك تميز بين نوعين من الدراسات التي تتضمنها: النوع الأول هو دراسات إمكانية التعميم Generalizability Studies ويرمز لها بالحرف الأول (G)، والنوع الثاني هو دراسات القرارات Decision Studies ويرمز لها بالحرف الأول (D). وهذا التمييز ليس واضحًا في النظرية الكلاسيكية للختبارات.

دراسات إمكانية التعميم (G):Generalizability Studies

تعتبر النظرية الكلاسيكية للختبارات أن درجة الفرد في اختبار أو مقياس ما تتكون من جزأين، هما الدرجة الحقيقية وتمثل الأداء أو القدرة الحقيقية، ودرجة وحيدة للخطأ وهي تعكس عوامل الصدفة. والدرجة الحقيقة للفرد هي القيمة المتوقعة لدرجاته الملاحظة في جميع المقاييس المتوازية. ونظراً لأن يفترض أن الدرجات الحقيقة تكون مستقلة عن بعضها البعض، فإن تباين الدرجات الملاحظة يساوي مجموع تباين الدرجات الحقيقة مضاعفاً إليه تباين الخطأ. ويعرف الثبات في إطار هذه النظرية، بأنه مربع معامل الارتباط بين الدرجات الملاحظة والدرجات الحقيقة، وهذا يساوى النسبة بين التباين الحقيقي والتباين الملاحظ. ونظراً لأننا لا نعرف الدرجة الحقيقة للفرد، وبالتالي لا يمكن معرفة تأثير عوامل الصدفة في تحديد الدرجة الملاحظة. ولكن يمكن تقدير تأثير هذه العوامل على القياس عموماً. وتوجد عدة طرق لتقدير معامل الثبات بهذا المعنى التقليدي كما سبق أن ذكرنا إحداثاً إيجاد معامل الارتباط بين درجات الأفراد في اختبارين متوازيين، أي صيغتين متكاففتين للاختبار يطبقان عليهم في الوقت، أو استخدام طريقة كيودر - ريتشاردسون (الصيغة ٢٠)، أو معامل ألفا لكرونباك Cronbach's Alpha Coefficient.

وهاتان الطريقتان تتطلبان تطبيق الاختبار نفسه مرة واحدة. كما توجد طرق أخرى تُستخدم في الحصول على أنواع مختلفة من معاملات الثبات، مثل معاملات

الاستقرار Stability and Coefficients، ومعاملات الاستقرار والتكافؤ Reliability of Ratings. Equivalence Coefficients ولكل نوع من هذه الأنواع من المعاملات تباين خاص بخطأ أحادى البعد Unidimensional Error Variance . ونقصد بذلك أنه يوجد مصدر واحد أو تباين واحد غير متميز للخطأ خاص بكل منها. فكل طريقة من هذه الطرق تفترض فروضا مختلفة عن مكونات مصدر تباين الخطأ في القياس.

وهذا في الحقيقة يعد نقطة ضعف أخرى في النظرية الكلاسيكية للاختبارات. فاختلاف طرق تقدير الثبات يؤدي إلى غموض تفسير نتائج الاختبارات والمقاييس. ويرجع سبب ذلك كما ذكرنا إلى البساطة الشديدة التي تتميز بها النظرية الكلاسيكية. وهذا هو ما دعا كرونباك وزملاءه إلى بناء نموذج متعدد الأوجه Multifaceted Approach يتصف بالعمومية لمواجهة نقط الضعف السابقة في النظرية.

فنظرية إمكانية التعميم تأخذ بعين الاعتبار المصادر المتعددة لأنواع الخطأ القياسي، وتهتم دراسات إمكانية التعميم Generalizability Studies (G) بتحديد درجة تكافؤ نتائج الاختبارات التي نحصل عليها في ظروف اختبارية مختلفة وتحت شروط متباعدة، أي تهتم بدراسة تأثير معابينات هذه الشروط أو الظروف على الدرجات من أوجه مختلفة Facets . ويمكن إجراء ذلك عن طريق جمع البيانات من الأفراد المختبرين في ظروف محددة (أي مستويات مختلفة لأوجه محددة)، وتقدير مكونات التباين التي ترجع إلى اختلاف مستويات الأوجه، وكذلك تقدير تفاعل هذه المكونات باستخدام طرق تحليل التباين Analysis of Variance ، وبذلك يمكننا الحصول على معاملات تسمى معاملات إمكانية التعميم Coefficients of Generalizability . ومعامل إمكانية التعميم هو النسبة بين تباين الدرجة الشاملة Universe Score إلى تباين الدرجة الملاحظة Observed Score . أي أنه يمكننا اعتبار معامل إمكانية التعميم بدليلاً عن معامل الثبات الذي نحصل عليه باستخدام النظرية الكلاسيكية للاختبارات. وكل اختبار أو مقاييس يكون له أكثر من معامل إمكانية تعميم واحد على حسب الأوجه التي يتم أخذها بعين الاعتبار في دراسة إمكانية التعميم study - G . ولذلك فإن تحليل التباين في هذه الدراسات يحتاج إلى تصميمات تجريبية Experimental Designs أكثر تعقيداً من التصميمات المستخدمة في التحليل الكلاسيكي لبيانات الاختبارات.

دراسات القرارات Decision Studies:

من المعلوم أن أدوات وأساليب القياس تُستخدم في جمع بيانات تفيد في صنع القرارات، أو في التوصل إلى استنتاجات معينة. ومثال ذلك قبول الطلاب في برامج تعليمية مختلفة، أو انتقاء الأفراد، أو التعرف على الأطفال المتفوقين في التحصيل الدراسي وغير ذلك. وقد تبين لنا مما سبق أن المعلومات التي تمدنا بها دراسات إمكانية التعميم تُستخدم في جمع بيانات تفيد في تقدير مكونات تباين القياسات التي نحصل عليها بطريقة ما من الطرق، وهذا يفيد في تطوير هذه القياسات. أما دراسات القرارات فتُستخدم في جمع بيانات تفيد في صنع العديد من القرارات، والتوصيل إلى كثير من التفسيرات والاستنتاجات بناء على النتائج التي توصل إليها دراسات إمكانية التعميم G - Studies. ويمكن أن نميز بين أربعة أنواع من القرارات على الأقل هي:

(١) القرارات المطلقة:

عندما نود تصنيف أداء أحد الأفراد بطريقة معينة ينبغي تحديد مستويات الأداء التي يتم تصنيفه على أساسها. وقد عرضنا لهذا النوع من القرارات في الفصل الثامن عند مناقشتنا للاختبارات المرجعية المحك. فنحن نهتم في هذا النوع من القياس بموازنة أداء الفرد بمستوى أداء مطلق محدد مسبقاً، مثل النسبة المئوية لعدد مفردات الاختبار التي ينبغي أن يجib عنها الفرد إجابة صحيحة، أو عدد الأخطاء، أو الزمن اللازم لأداء عمل معين، وما شابه ذلك. وهنا يكون القرار خاصاً بفرد معين دون الاهتمام بغیره من الأفراد. ويفيد مثل هذا القرار في تصميم البرامج التعليمية والتدريبية، وفي تشخيص الصعوبات التي يواجهها الفرد أثناء التدريب أو التعليم.

(٢) القرارات المتعلقة بالموازنة بين أسلوبي نشاط للفرد:

يتّخذ هذا النوع من القرارات أيضاً لكل فرد على حدة. ويفيد في التوجيه التعليمي، حيث تستخدم درجات الفرد في اختبارات القدرات المرتبطة ببرامج تعليمية مختلفة في توجيهه إلى البرنامج المناسب لقدراته.

(٣) القرارات المتعلقة بالموازنة بين الأفراد:

وهنا يهتم صانع القرار بالفارق بين درجات الأفراد أو المجموعات. فمثلاً يمكن أن يستخدم اختبار أو اختبارات في انتقاء أفضل الأفراد الذين يتحقون بعمل معين، أو يحصلون على منحة دراسية معينة. ويمكن أيضاً استخدام أحد المتغيرات التابعة في المقارنة بين المجموعات في إحدى التجارب.

(٤) استنتاج العلاقة بين أزواج المتغيرات،

أحياناً يهتم صانع القرار ببحث العلاقة بين أزواج المستويات، من مثل ذلك العلاقة بين القدرة على حل مشكلات تتضمن أشكالاً هندسية، والقدرة على تكوين المفاهيم. أي أن نظرية إمكانية التعميم تراعي اختلاف طرق استخدام درجات أداة القياس نفسها باختلاف القرار المطلوب صنعه. فاحياناً يكون الاهتمام منصباً على الكشف عن الفروق الفردية (أي الموازنة بين أداء الأفراد)، وأحياناً أخرى ينصب الاهتمام على استخدام الدرجة الملاحظة لتقدير Estimate الدرجة الشاملة للفرد (أي تفسير الدرجات تفسيراً مرجعياً للمحك Criterion -Referenced)، أو ربما تستخدم معادلة الانحدار في تقدير الدرجة الشاملة بمعلومية الدرجة الملاحظة.

والمميز بين دراسات إمكانية التعميم ودراسات القرارات لا يعني أن كلاً منها منفصل عن الآخر. فالحقيقة أن التصميم التجاري الأمثل الذي تبني عليه دراسة إمكانية التعميم يعتمد على القرار الذي يهم الباحث صنعه. فإذا كان الباحث مهتماً بال النوع الأول من القرارات مثلاً، أي القرارات المطلقة الخاصة بفرد معين دون اعتبار لمعايير الجماعة المرجعية التي يتمتع بها الفرد، يمكنه استخدام تصميمات تجريبية بسيطة تمثل فيها معايير مشتركة Joint Sampling من الشروط المختلفة للأبعاد أو الأوجه Facets.

وهذا التمييز يدل فقط على أن بعض الدراسات تجري أثناء عملية تصميم طريقة قياس معينة، وبالتالي يمكن الاستفادة من هذه الطريقة في دراسات تالية.

فمثلاً إذا أجرى باحث تحليلًا عاملياً لمفردات استبيان بغرض اتخاذ قرار خاص بتجميع المفردات في أبعاد معينة على أساس تشبّعاتها بهذه الأبعاد، فيمكننا اعتبار أنه يقوم بدراسة استطلاعية تشبه دراسة إمكانية التعميم G - Study.

إذاً أراد باحث آخر التوصل إلى أفضل تصميم لبطارية اختبارات مستخدمة في أغراض التنبؤ، فإنه ربما يبدأ بدراسة استطلاعية يجمع فيها بيانات باستخدام بطارية مبدئية، ثم يقوم ببناء بطارية جديدة اعتماداً على معاملات الارتباط بين درجات كل اختبارين من الاختبارات التي تشتمل عليها البطارية المبدئية، وبين درجات كل اختبار ودرجات المحك، وكذلك معاملات ثبات درجات كل منها. ويمكن عندئذ استخدام هذه البطارية الجديدة في جمع بيانات خاصة بدراسات القرارات D - Studies. ومن هذا يتضح أن دراسات إمكانية التعميم (G) تُعد جزءاً من بناء أدوات القياس. ولذلك يجب أن تجري هذه الدراسات قبل عملية جمع البيانات الخاصة بدراسات القرارات (D).

والميزة بين دراسات إمكانية التعميم ودراسات القرارات، له أهمية خاصة في البحوث متعددة الأوجه أو الأبعاد Multifacet Investigations. فتعدد الأوجه يؤدي إلى تنوع التصميمات التجريبية التي يمكن استخدامها في هذه البحوث. ولذلك يجب أن تكون التصميمات التجريبية المستخدمة في دراسات إمكانية التعميم مختلفة عن التصميمات المستخدمة في دراسات القرارات. فالباحث الذي يجري دراسة إمكانية تعميم (G) باستخدام تصميم تجاري معين، ويرى أنه من الممكن استخدام التصميم نفسه في دراسة القرارات (D)، يجب أن يدرك أن باحثاً آخر ربما يستخدم تصميماً مختلفاً للحصول على بيانات مماثلة لبياناته.

لذلك يوصي كرونباك Cronbach بأن يصاغ تقرير دراسة إمكانية التعميم بصورة تُمكّن الباحثين المختلفين من تصميم دراسات القرارات التي يهتمون بها، وتقدير خطأ التعميم الخاص بالتصميم المستخدم.

تحقيق النطاق الشاملة في دراسات إمكانية التعميم

Universes of generalization

إن قياس السلوك الذي نحصل عليه يكون بمثابة عينة من النطاق الشامل للقياسات الممكنة. واهتمامنا بالدرجات التي نحصل عليها من عينات من القياسات يرتبط ب مدى تمثيلها للنطاق الشامل للقياسات. فلو أن الباحث استطاع قياس سمة معينة لدى الفرد بكافة الطرق الممكنة، وكانت الدرجة التي يحصل عليها الفرد في هذه الحالة هي متوسط جميع هذه القياسات. ويطلق على هذه الدرجة في النظرية الكلاسيكية للختارات «الدرجة الحقيقة True Score»، ولكن يُفضل كرونباك Cronbach أن يطلق عليها «الدرجة الشاملة Universe Score» في إطار نظرية إمكانية التعميم، لأن هذا المفهوم الجديد يتضمن عملية استدلال على النطاق الشامل للقياسات أو الملاحظات من عينة القياسات أو الملاحظات التي تجريها. كما يؤكد أن إمكانية التعميم يمكن أن تنسحب على أكثر من نطاق شامل Universe. إذ يمكن اعتبار أن الفرد عضو في مجتمعات مختلفة Population، مثل مجتمع الأفراد الذين يقيمون في مدينة معينة، أو مجتمع العمال، أو مجتمع الأفراد الذين يزيد دخلهم الشهري عن مبلغ معين، أو أعضاء حزب سياسي معين ... الخ.

وكذلك يمكن اعتبار «ملاحظة» معينة أو «قياس» معين بمثابة عينة من النطاق الشامل للملاحظات أو القياسات. ولذلك إذا قلنا أن «تقدير الدرجة الشاملة هو ٧٥»، لا يكون لهذا القول معنى إلا إذا حددنا النطاق الشامل لهذه الدرجة

ويتم وصف الملاحظات أو القياسات عن طريق «الشروط أو الظروف Conditions» التي أجريت فيها، ونقصد بذلك اليوم الذي جُمعت فيه الملاحظات أو الذي أجريت فيه القياسات، وطبيعة الموقف الذي أجريت فيه، والشخص القائم بعملية الملاحظة أو الفرد المختبر، وكذلك المظاهر الأخرى للعمليات التي يؤديها الفرد في الموقف الاختباري. وقد سبق أن ذكرنا أن المصطلح العام الذي يشير إلى شروط أو ظروف معينة يعرف باسم «البعد أو الوجه Facet». لذلك فإن «الملاحظات» أو «القياسات» يمكن أن تُصنف بالنسبة إلى أبعاد أو أوجه مختلفة. وهناك بعد أو وجه يُمثل المثيرات التي تعرض على الفرد، وبعد يمثل الوقت الذي جمعت فيه الملاحظات أو أجريت فيه القياسات، وبعد يمثل طبيعة الموقف الاختباري، وهكذا.

وكل هذه الأبعاد أو الأوجه منفردة أو مجتمعة تحدد النطاق الشامل، فإذا طلبنا من طفل في يوم ما أن يرسم شجرة، فإن هذا الرسم يتسم إلى نطاق شامل من الأشجار التي يرسمها في مختلف الأيام، وإلى نطاق شامل من الأشجار التي يرسمها غيره من الأطفال، وإلى نطاق شامل من الأنماط المختلفة من رسومات الأطفال في ذلك اليوم. فلاستفسار عن النطاق الشامل المطلوب إذن مرتبط بالاستفسار من الباحث عن هدفه من تفسير القياس.

ويرى كرونباك أن «النطاق الشامل للملاحظات أو القياسات» يمكن أن يتميز بـ *بعد* أو وجه واحد، أو وجهين أو أكثر، فهو يفترض أن الرابط أو الجمع بين أحد شروط الوجه الأول وشرط معين من شروط كل وجه آخر، هو الذي يحدد إمكانية تعريف «ملاحظة معينة»، أو «قياس معين». ففي المثال السابق ربما يكون هناك وجهان هما: أنماط الرسم، والأطفال الذين يطلب منهم هذا الرسم. والرابط بين أي من هذه الأنماط وبين الطفل الذي يطلب منه رسم هذا النمط هو الذي يحدد تعريف «الملاحظة» المطلوبة. وإضافة وجه ثالث، ولتكن الوقت الذي يطلب فيه الرسم من الطفل، يجعل «الملاحظة» تتعدد عن طريق الرابط بين النمط والطفل والموقف.

والحقيقة أن النطاق الشامل الذي تعمم عليه «ملاحظة» معينة أو «قياس» معين يعتمد على الاعتبارات العملية أو النظرية التي تهم صانع القرار. فمثلاً يختلف تقدير أحد المشرفين على العمل لأحد العمال من موقف إلى آخر. فالحالة المزاجية للمشرف وخبرته السابقة عن العامل يكون لهما تأثير على تقديره الحالى. والباحث المهتم بدراسة كفاءة العمال، يود بالطبع تعميم هذا التقدير على مجموعة التقديرات التي يمكن الحصول عليها من مواقف مختلفة، فهو يستطيع التعميم على مدة ولتكن شهراً إذا كان

التقدير عبارة عن كفاءة العامل في نهاية العام، فما وقت خلال هذا الشهر ربما يكون مناسباً لتقدير كفاءة هذا العامل.

أما إذا كانت البيانات المستعملة من التقدير سوف تستخدم في إجراء دراسة للتغيرات التي تطرأ على اتجاهات المشرف نحو العمال أثناء تفاعلهم معهم من أسبوع إلى آخر، فإن الباحث يستطيع التعميم على يوم واحد فقط.

وإذا استُخدم التقدير كمحك يوازن في ضوئه صدق اختبار يقيس الاستعدادات، فإن الباحث يحتاج إلى التعميم على كل من المشرفين والموافق.

ويصبح النطاق الشامل للمشرفين غير مطلوب إذا اقتصر الهدف على دراسة التفاعل المباشر بين العامل والمشرف أثناء أداء العمل. لذلك فإن الباحث الذي يقوم ببناء أداة قياس، أو الذي يجري دراسة إمكانية تعميم G - Study، يجب أن يأخذ في اعتباره بعض «الأبعاد أو الأوجه Facets»، وما يندرج تحت كل منها من «شروط Conditions». وتسمى «الملاحظات Observations» التي تحصل عليها من مختلف تجمعات الشروط Conditions الممثلة في دراسة إمكانية التعميم «النطاق الشامل للملاحظات المقبولة The Universe of Admissible Observations»، كما أن لكل بُعد أو وجه Facet «نطاق شامل من الشروط المقبولة The Universe of Admissible Conditions». فعند صنع القرارات استناداً إلى أسلوب قياس معين ينصب الاهتمام على تعميم هذه القرارات على «نطاق شامل من الشروط» التي تتوصل منها إلى المعلومات نفسها، ويسُمى هذا النطاق «النطاق الشامل للتعميم Universe of Generalization». ويمكن أن تُفيد دراسة إمكانية التعميم G - Study صانعي القرار الفائدة المرجوة إذا كان النطاق الشامل الذي يستخدم في دراسة إمكانية التعميم مطابقاً للنطاق الشامل للشروط المقبولة أو على الأقل يكون جزءاً منه.

فالنطاق الشامل للتعميم يختلف باختلاف صانع القرار. فكلما اهتم صانع القرار في دراسة إمكانية التعميم بتحديد النطاق الشامل للملاحظات المقبولة تحديداً أكثر اتساعاً بحيث يتضمن الأنشطة المختلقة المحمولة للتعميم - أفادته الدراسة وأفادت غيره من المهتمين بصنع القرارات. ويتحدد النطاق الشامل للتعميم بناء على الهدف من القرار المطلوب اتخاذة.

فعند بناء أداة قياس يجب أن يجري الباحث دراسة إمكانية تعميم G - Study تتضمن الأبعاد أو الأوجه المختلفة التي يمكن أن تُفيد في تعميم نتائج استخدام هذه الأداة. وبذلك يستطيع غيره من الباحثين تفسير هذه النتائج في ضوء الأبعاد أو الأوجه التي يهتمون بها.

ويقترح كرونباك Cronbach ثلث طرق مختلفة لمعالجة «بعد أو وجه Facet معين في دراسة إمكانية التعميم هي:

(١) تمثيل «البعد أو الوجه» عن طريق معينة Sampling «شروط Conditions» أو أكثر لهذا البعد أو الوجه.

(٢) استخدام «شرط واحد ثابت Condition» للبعد أو الوجه في جميع «اللاحظات» التي تهتم الدراسة بجمعها.

(٣) تغيير «شروط البعد أو الوجه» دون استخدام أساليب الضبط التجربى. ويمكن توضيح هذه الطرق بالرجوع إلى المثال السابق الخاص بدراسة تقديرات المشرفين على العمل. فربما يعتبر الباحث مواقف العمل والمشرفين عليه وجهين من النوع المذكور في الطريقة الأولى التي اقترحها كرونباك. ولذلك ربما يطلب الباحث من مشرف أو أكثر أن يقوم بتقدير كفاءة كل عامل في موقفين أو أكثر من مواقف العمل. أو ربما يعتبر أن استمارة التقدير ثابتة، أي يعتبرها «شرطًا» من النوع المذكور في الطريقة الثانية، ثم يقوم بجمع البيانات الخاصة بدراسة إمكانية التعميم باستخدام هذه الاستمارة فقط، وهذا يمكّنه من معرفة ما سوف يحدث في حالة تغيير عبارات أو محتوى أو أشكال استمارات التقدير. وبالطبع لا يستطيع التعميم على جميع محتويات أو أشكال استمارات التقدير، كما لا يستطيع ضبط جميع الأوجه Facets الممكنة في إحدى دراسات إمكانية التعميم. ففي المثال السابق يصعب عند دراسة تقديرات المشرفين على العمل، التحكم في المدة التي يشرف فيها مشرف معين على أحد العملاء. ويمكن في هذه الحالة معالجة هذه الأوجه باستخدام الطريقة الثالثة.

فصانعو القرارات يهتمون عادة بتعيم النتائج على أوجه Facets يصعب التحكم فيها، وتسهم هذه الأوجه بلا شك في أخطاء التباين عند تحليل نتائج دراسة إمكانية التعميم. ولذلك اقترح كرونباك بعض التصميمات التجريبية التي يمكن أن تستخدم في معالجة مصادر التباين التي يصعب التحكم فيها.

تحليل مكونات التباين في دراسات إمكانية التعميم:

تهدف دراسة إمكانية التعميم إلى جمع بيانات تفيد في تقدير مكونات التباين في الدرجات الملاحظة التي تحصل عليها باستخدام أسلوب معين من أساليب القياس. وقد اقترح كرونباك عدداً من التصميمات التجريبية التي يمكن أن تُستخدم في تحليلاً

هذه المكونات. وتحتختلف هذه التصميمات باختلاف عدد الوجه والشروط التي يفترضها الباحث في دراسة إمكانية التعميم.

فإذا كان النطاق الشامل Universe يتكون من بُعد أو وجه Facet واحد، تستخدم تصميمات تجريبية أحادية الوجه Single - Facet Designs، أما إذا اشتمل على أكثر من وجه واحد، فإنه يمكن استخدام تصميمات متعددة الأوجه Multiple Facets Designs، وقد يشير مكونات التباين في هذه التصميمات التجريبية يعتمد على الأساليب الإحصائية، وتحتختلف هذه الأساليب في درجة تعقدتها وفقاً لنوع التصميم التجريبي المستخدم. وفيما يلى بعض أنواع هذه التصميمات:

(١) التصميمات أحادية الوجه :Single - Facet Designs

يشتمل «النطاق الشامل للملاحظات» في هذا النوع من التصميمات على وجه واحد. وسوف نرمز للأفراد بالرمز (ف)، ولشروط الوجه الواحد (مفردات اختبار أو مقاييس مثلاً) بالرمز (ش)، ولتقاطع الأفراد مع الشروط بالعلامة (خ)، ولتشابك أو تداخل الأفراد مع الشروط بالعلامة (:). وسوف نوضح بإيجاز بعض أنواع هذه التصميمات:

(أ) التصميم (ف × ش): يُعد أكثر التصميمات أحادية الوجه استخداماً، وفيه يطبق على كل فرد من الأفراد الذين عددهم (ن)، مفردات عددها (ن)، وهنا يكون الأفراد هم موضع القياس، حيث لا تُعدهم مصدراً من مصادر الخطأ، وبالتالي فلا يمثلون «وجه Facet»، وإمكانية التعميم تكون على جميع الصيغ الممكنة لاختبار حيث يشتمل كل منها على عينة عشوائية من مفردات مسحوبة من نطاقها الشامل.

(ب) التصميم (ش : ف): يُعد من التصميمات أحادية الوجه التي تشابك فيه أو تتدخل الشروط مع الأفراد Nested Design. وهنا يطبق على كل فرد مجموعة مختلفة من المفردات.

(٢) التصميمات متعددة الأوجه :Multiple Facets Designs

يشتمل «النطاق الشامل للملاحظات» في هذا النوع من التصميمات على أكثر وجه واحد. وتتنوع هذه التصميمات تنوعاً كبيراً، ولكنها تعتمد على المفاهيم والأسس التي تعتمد عليها التصميمات أحادية الوجه. غير أنها تتطلب تصوراً فكريّاً أكثر عمقاً، كما تتطلب عمليات جبرية ورمزيّة أكثر تعقيداً عند تقدير مكونات التباين في دراسات إمكانية التعميم.

وفيما يلى بعض أنواع هذه التصميمات:

(ا) التصميم ($F \times S \times J$): ويُعد من التصميمات متعددة الأوجه الشائع الاستخدام، حيث (J) ترمز إلى المواقف، أو الاختبارات الفرعية، أو إلى القائمين بعملية التقدير. ويمكن باستخدام هذا التصميم تقدير جميع مكونات التباين ولكن كل منها مستقل عن الآخر لأن التصميم متقطع تقاطعاً كاملاً Completely Crossed، وذلك لأنه يجب أن تطبق مجموعة المفردات على جميع الأفراد في المواقف المختلفة. لذلك يستخدم هذا التصميم في تقدير استقرار درجات الاختبارات وتكافؤها، وذلك بتحديد النطاق الشامل لإمكانية التعميم تحديداً مناسباً. كما يمكن استخدامه في إثبات ثبات تقديرات المحكمين إذا كانت (S) ترمز إلى المحكمين.

(ب) التصميم ($F \times (S:J)$): ويُعد من التصميمات المتشابكة أو المتداخلة Nested Design Effects، حيث تتشارك أو تداخل فيه بعض مكونات التباين أو التأثيرات.

فمثلاً يمكن أن تتشارك أو تداخل مفردات الاختبار (S) في الموقف (J)، أي تطبق على جميع الأفراد مجموعة المفردات نفسها في أحد المواقف، ويُطبق عليهم مجموعة أخرى من المفردات في موقف آخر، وهكذا ويُستخدم هذا التصميم في تقدير تكافؤ درجات الاختبارات في إطار النظرية الكلاسيكية للختبارات. وإذا كانت (J) ترمز إلى أحد الاختبارات الفرعية التي تشتمل عليها بطارية من الاختبارات، فإنه يمكن استخدام هذا التصميم أيضاً في دراسة إمكانية التعميم من درجات اختبارات البطارية.

وقد تناول المؤلف هذه التصميمات المشعّدة بمزيد من المناقشة في ضوء النماذج الإحصائية في كتابه «تطورات معاصرة في القياس النفسي والتربوي».

ومن الجدير بالذكر أن كريك، وبرينان (Crick & Brennan, 1982) أعداً برنامجاً للحاسوب أطلقوا عليه GENOVA يمكن باستخدامه إجراء جميع التحليلات الإحصائية التي تتطلبها نظرية إمكانية التعميم عندما تكون التصميمات من النوع الكامل المتوازن Completely Balanced Designs. ويجري البرنامج تحليل دراسات إمكانية التعميم G - Studies Analysis باستخدامه على متوسطات مستويات الأوجه، ومتوسطات توفيقات مختلفة من مستويات الأوجه، ومجموع المربعات، ومتوسطات مجموع المربعات كما في أساليب تحليل التباين، وقيم تقديرية لمكونات التباين، والأخطاء المعيارية لهذه القيم، ومصفوفة التباين والتغير. كما يمكن الحصول على قيمة تقديرية لمكونات التباين، والأخطاء المعيارية.

لهذه القيم في دراسة اتخاذ القرارات D - Study Analysis استناداً إلى مكونات التباين التي يتم تقديرها في دراسة إمكانية التعميم. ويتميز هذا البرنامج بالسرعة الفائقة في تجهيز البيانات ومعالجتها حيث يستطيع معالجة ٣٥٠٠ ملاحظة في الثانية.

العلاقة بين نظرية إمكانية التعميم ونظرية الصدق:

تناولت نظرية إمكانية التعميم العلاقة بينها وبين نظرية الصدق Validity Theory يأيجر شديد، كما ألقى كل من كامب Kamp ، وماكدونالد McDonald ضوءاً خافتاً على هذه العلاقة. غير أن أولى المحاولات المنظمة للتكامل بين النظريتين هي دراسة كين (Kane, 1980) ، بيد أن هذه الدراسة تحتاج إلى كثير من التطوير النظري. فالتحقق من صدق إجراءات قياس معين يتطلب تحديد ووصف هذه الإجراءات من حيث مجموعة المثيرات التي يستجيب لها الفرد، والتعليمات التي تُعطى له، والقواعد التي يستند إليها في تقييم استجاباته. فمجموععة الشروط التي تحصل عليها من دراسة الصدق تعد واحدة من المجموعات المتعددة لهذه الشروط. لذلك ينبغي تحديد «النطاق الشامل للإجراءات المقبولة» التي تتمنى إليها هذه المجموعة من الشروط، والإفادة من دراسات إمكانية التعميم G - Studies في التتحقق من صدق التكوين الفرضي للاختبارات. فالتكوينات الفرضية التي تفسر تباين الأداء في الاختبارات يمكن دراستها عن طريق تحليل مكونات التباين وفقاً لأحد التصنيمات المناسبة.

ويرى شيفلسون، ووب (Shavelson & Webb, 1981) أن تطوير هذا المجال سوف يكون بداية مرحلة جديدة في التنظير السيكومترى في العقود القادمة.

(ثالثاً) إسهام علم النفس المعرفي في تطوير أساليب القياس والتقويم:

حدثت تطورات كبيرة في علم النفس المعرفي Cognitive Psychology يعد من العلوم المتداخلة، حيث يشارك علماء النفس، وعلماء الحاسوب، والفلسفه، وعلماء اللغة، وعلماء الأجناس البشرية، وعلماء التربية في دراسة وتحليل العمليات المعرفية الإنسانية. فمحاولة تدريس مجال تعليمي معين تستند عادة إلى نظرية أو افتراضات معينة تتعلق بالأداء الذي يتطلب عمليات معرفية عليا في هذا المجال، وكذلك إلى نظرية صريحة أو ضمنية تتعلق بعملية التعلم. فعلم النفس المعرفي المعاصر يهتم بالبحث التقييمي للأداء والتعلم، وأصبحت الإسهامات النظرية التي يقدمها تفوق إسهامات ما قدمته نظريات التعلم في العقود السابقة، كما أصبحت هذه الإسهامات أكثر ارتباطاً بدراسات التعلم في الفصل المدرسي. فنظريات علم النفس

المعرفي تعد محاولات لتفسير عمليات الفكر ، مما يتطلب نوعا من التمثيل Representation يختلف عن التمثيل الرياضي الذي يناسب العمليات الاختبارية.

ولتوضيح ذلك نفترض أنه طلب من فرد محاولة حل بعض المشكلات اللفظية والرياضية ، فعدد المشكلات التي قام بحلها يمكن التنبؤ به إذا علمنا درجات قدرته اللفظية والمكانية والعددية. غير أنها لا تستطيع باستخدام معادلة التنبؤ معرفة أي المشكلات استطاع أن يحلها فرد معين ، نظرا لأن عمليات حل المشكلة هي ما ينبغي أن تتناوله النظرية المعرفية. فمن الممكن أن يجيب مجموعة من الأفراد إجابات واحدة بعمليات استدلال مختلفة .

وفحص هذه العمليات يزودنا بمعلومات عن الفرد أكثر مما نحصل عليه من مجرد تقدير درجة كلية لإجاباته. فعندما تقدم مشكلة معينة فإنه يجري وصفها داخليا في ضوء المعلومات المخزنة بالفعل في الذاكرة. وتشتمل الذاكرة على نوعين منفصلين من المعرفة: أحدهما معرفة تقريرية للحقائق، وثانيهما معرفة إجرائية عما يمكن فعله. وهذه المعرفة الإجرائية Procedural Knowledge تعد بمثابة «برامج» تُخبر العقل بوظائف تجهيز المعلومات المناسبة التي يمكن تنفيذها في مواقف مختلفة. ويمكن توضيح ذلك بعملية فهم المقصود ، فكثير من الطلاب يتعلمون أن الجملة أو الجملتين الافتتاحيتين في مقال معين تحددان موضوع المقال بعامة ، لذلك فإنهم يستخدمون المعلومات التي تتضمنها هاتان الجملتان لتنشيط مخطط يوجه فهمهم لبقية المقال.

وهذا يؤكد أهمية العمليات التي يستخدمها الفرد في مواقف حل المشكلة بخلاف من التنبؤ بأدائه في هذه المواقف. وللتوصيل إلى تفسير لهذه العمليات ينبغي أن يكون لدينا قدر كبير من فهم الموقف المعين ، وهذا يعد أكثر صعوبة من مجرد التنبؤ بالأداء. فمثلا ينبغي معرفة سرعة التجهيز المعرفي المناسبة للموقف ، والمعرفة التقريرية والإجرائية المطلوبة لحل المشكلة ، وبدائل الأساليب التي يمكن باستخدامها تجهيز المعلومات المتعلقة بالمشكلة ، وكيفية تحديد العمليات الأفضل لموقف معين. ومعرفة ذلك يمكن منظري علم النفس المعرفي من تناول العلاقة بين أداء الفرد وامتلاكه المعرفة المناسبة ، ومدى اعتماد الأداء على وظائف معينة لتجهيز المعلومات ، مثل قدرات التعميم ، وتعريف الترميزات اللفظية ، وتناول التصورات العقلية.

علاقة النظرية المعرفية بالعملية الاختبارية:

تهتم الدراسات المتعلقة بالعمليات المعرفية بوظائف تجهيز المعلومات، والمهارات العامة لحل المشكلات، أو أساليب استخدام المعرفة في مواقف محددة لحل المشكلة. غير أن دراسة التفاعل بين المستويات المختلفة للمعرفة لم يحظ إلا بقدر ضئيل من البحث مما أدى إلى عدم توظيفها في بناء الاختبارات، وذلك لأن العملية الاختبارية بطبعتها تتناول الفرد ككل. غير أنه يمكن الإفادة من بحوث العمليات المعرفية في مجال الاختبارات، وذلك بضم المستويات الثلاثة للمعرفة التي أشرنا إليها مع الاهتمامات المتعلقة بمجال الاختبارات من حيث تطوير المقاييس المتوفرة، وبناء مقاييس جديدة، ومزيد من فهم مواقف حل المشكلة وغيرها من المواقف التي تتطلب نشاطاً معرفياً.

وقد أوضح هنت (Hunt, 1985) أن كلّاً من المستويات البحثية يمكن أن يسهم إسهاماً معيناً في مجال الاختبارات. فعلى مستوى تجهيز المعلومات ينبع التمييز بين تجهيز المعلومات اللغوية، وتتجهيز المعلومات غير اللغوية. وتوجد تقنيات جديدة تعتمد على الحاسوب يمكن توظيفها في بناء اختبارات تقدير وظائف تجهيز المعلومات اللغوية. فمثلاً يمكن بناء مقاييس مختبرية تحديد سرعة الفرد في تعرّف الكلمة مألوفة سواء بمفردها أو في مضمون معين، حيث إن هذا يعد من الوظائف الأساسية في القراءة. وتنيد هذه المقاييس في بناء نظرية معرفية تتعلق بتفسير تباين سرعة الأفراد في التجهيز اللغوي Language Processing. وكذلك يمكن تطوير أساليب لدراسة كيفية استجابة الأفراد لأشياء متحركة، وذلك لقياس عمليات الاستدلال البصري المكاني، حيث إن هذا يختلف عن الاختبارات التقليدية لقياس القدرة على الإدراك المكاني، نظراً لأن الأشكال التي تشتمل عليها هذه الاختبارات تكون استاتيكية وليس ديناميكية كما في الأساليب المقتربة. إذ يمكن بناء اختبارات تقدير قدرة الفرد على إضافة عناصر لمثير بصري يتصوره الفرد عقلياً، أو تقدير قدرته على إسقاط مسارات مثيرات متحركة للأمام بالنسبة للزمن والموقع، ويتم التحكم في عرض هذه المثيرات وتسجيل الاستجابات باستخدام الحاسوب، وذلك لصعوبة قياس هذه القدرة باستخدام اختبارات الورقة والقلم.

كما يصعب استخدام اختبارات الورقة والقلم في تقييم وظائف تجهيز المعلومات، ومثال ذلك قياس سرعة الفرد في الانتقاء الإدراكي المركزي، أي يقرر مثلاً أي من المثيرين الضوئيين حدث أولاً.

كذلك يمكن توسيع مجال الاختبارات بقياس القدرة على «توجيه الانتباه» حيث اعتبر كثير من علماء النفس هذه القدرة جزءاً من الكفاءة العقلية. وأكملت البحوث المعاصرة في هذا المجال أن قدرة الفرد على توجيهه انتباذه لجزء معين من مثير مركب تعد أحد المكونات الرئيسية للفرق الفردية. ويمكن قياس هذه القدرة بأن يطلب من الفرد تحويل انتباذه من موضع معين في المجال البصري إلى موضع آخر. كما يمكن توسيع نطاق هذه البحوث لدراسة الفرق الفردية في القدرة على تحويل الانتباه من صيغة معينة لتجهيز المعلومات إلى صيغة أخرى، ومثال ذلك التحول بسرعة من إجراء عمليات جمع إلى إجراء عمليات ضرب، وهكذا.

وبالطبع توجد أنواع أخرى متعددة من وظائف تجهيز المعلومات يمكن قياسها، ولكنها تتطلب عمليات تحكم باستخدام الحاسوب.

تصميم الاختبارات في بيئة النظرية المعرفية:

يرى كثير من علماء النفس المعرفي وخبراء القياس أن عمليات القياس وتصميم الاختبارات وعلاقتها بالعملية التعليمية سوف تختلف في القرن الحادى والعشرين توجهاً جديداً يختلف اختلافاً جوهرياً عما هو سائد في وقتنا الحاضر (Bejar et al., 1993, Hunt, 1985; Ward, 1985) . فالتطورات التي حدثت في علم النفس المعرفي، وعلم القياس النفسي والتربوي، وتقنيات المعلومات والاتصالات، وما تفرضه طبيعة هذا القرن من تحديات، تتطلب إعادة النظر في ممارسات التعليم، ونظم وأساليب القياس والتقويم لمجتمع المتعلمين.

وتحقيق ذلك يتطلب توافر معلومات تختلف عما تزودنا به أساليب القياس الكلاسيكي التي تقتصر على تقديم مؤشرات عامة تتعلق بمجموعات من الأفراد بهدف المقارنة بينهم، أو التنبؤ بأدائهم الأكاديمي. فعملية التقويم سوف تصبح جزءاً لا يتجزأ من عملية التعليم، أى أنهما سيصبحان عملية واحدة متكاملة. فالاختبارات الحالية التي تستخدم فقط في أغراض انتقاء الطلاب للالتحاق بمراحل التعليم المتتالية، ولا تقدم معلومات يمكن الاستفادة منها في التقدم التربوي سوف تكون قليلة الجدوى، وذلك لأننا لا نستطيع باستخدامها قياس نمو وتطور الكفايات الإنسانية. وهذا يتطلب التتحقق من الفروق الجوهرية بين الأداء المتميز والأداء الضعيف عن طريق تقييم بنية المعرفة والعمليات المعرفية التي تكشف عن درجات كفايات الفرد في مجال دراسي معين.

فمعظم تقييمات القياس في وقتنا الحاضر صُممت لكي تُستخدم بعد بناء مفردات الاختبارات. فقد لاحظنا في الأبواب السابقة أن تحليل مفردات الاختبارات، والتحقق من صدقها وثبات درجاتها ، وإيجاد معايرها ، وتحليل أبعاد الاختبار ومكوناته العاملية يتم بعد بناء المفردات. بينما تشير التوجهات المعاصرة إلى أنه ينبغي الاهتمام بإسناد تصميم المفردات إلى نظريات معرفية تتعلق بخصائص الكفاءة المرجوة، والبني والعمليات التي تكتسب وتنمو لدى الفرد المبتدئ حتى يصبح لديه كفاءة عالية نتيجة تعلمها معرفة واستراتيجيات متقدمة. وبذلك يمكن باستخدام هذه الاختبارات تحديد إنجازات الفرد في مختلف مستويات الالكتساب. وهنا لا يكون التركيز منصبا على المحتوى وإنما على البنى والعمليات المعرفية التي تيسر اكتساب ونمو الكفاءة المرجوة (Messick, 1984) . فالنظريات السيكومترية المتعارف عليها ركزت على قياس عمليات عامة ، فالقدرات العقلية تقاس في عموميتها ، مثل القدرة اللغوية أو العددية أو المكانية ، وكذلك الاستعدادات الخاصة بأنواعها المختلفة التي يتبيّن ارتباطها بالنجاح الأكاديمي . كما أن نظريات التعلم المتعارف عليها هدفت لتقديم أدلة تتعلق بعمليات والشروط العامة للتعلم ، والقوانين التي تفسّر أنماط التعلم المختلفة ، مثل أنواع الاشتراط ، وطبيعة التعزيز ، والتغذية الراجحة نتيجة التعلم ، والعمارة الموزعة والمركزة .

وعلى الرغم من أن هذه الأدلة قد أسهمت في تفسير مبادئ عامة مهمة في التعلم الإنساني ، إلا أن جلizer (Glaser, 1984) أوضح أن الاهتمام أصبح موجها نحو دراسة أنماط معينة من المعرفة والمهارات التي يكتسبها الأفراد ويواجهونها في سياق حياتهم . وهذا التوجه الجديد يؤكّد التعلم في نطاقات معرفة ثرية تتعلق بالمجالات الدراسية والأنظمة الأكاديمية مما يمكننا من تحديد جوانب القوة والضعف في أداء المتعلمين لمهام أكاديمية . فبدلا من محاولة تحديد جوانب ضعف عامة لديهم يتم توجيه الاهتمام إلى معاونة كل منهم في التعرّف بدقة على ما لديه من معرفة جزئية أو غير مكتملة لكي تتحاول عملية التعليم التركيز عليها مباشرة .

وأشار جلizer إلى بعض التطورات التي حدثت في هذا الاتجاه ، من بينها تحليل الأداء استنادا إلى أساليب معالجة المعلومات للإسهام في تنمية كفاءة المتعلمين في المهام الأكademية ، وكذلك الفهم المستبصر لطبيعة الأداء الكفاءة نتيجة دراسة خصائص الفرد الكفاءة والفرد المبتدئ في النطاقات المختلفة للسلوك الإنساني . ومثال ذلك اكتساب الأطفال الحقائق الحسابية ، والمعرفة والأساليب التي يتبعونها في حل المسائل الكلامية ، والمبادئ التي ينطوي عليها مفهوم العلامات العشرية ، وغير ذلك . وهذه

الدراسات المتعلقة بتشخيص وتصنيف أنماط الأخطاء في أداء الأطفال في الحساب تعد أساساً للتشخيص الدقيق لفهم الطفل المبتدئ في تعلم الحساب وكذلك سوء فهمه. وأوضح ريزنك (Resnick, 1984) أن هذا ييسر تحديد المعرفة والإجراءات غير المكتملة التي تسهم في ضعف أداء الأطفال في الحساب بحيث يمكن علاج هذا الضعف أثناء عملية التعليم.

فالأطفال الذين يخطئون في إجراء العمليات الحسابية يفترض أنهم لم يتبعوا إجراءات معينة بدقة وكفاءة. لذلك فإن علم النفس المعرفي يمكن أن يزيد من فاعلية عملية التعليم ليس فقط ببحث إجابات الطفل عن مجموعة من المشكلات، وإنما أيضاً ببحث العمليات التي توصل عن طريقها إلى الإجابة بحيث يمكنه تعلم العمليات الصحيحة. ونظام تقدير الدرجات الذي يحدد الأخطاء المتتظمة له انعكاساته في العملية الاختبارية، وذلك لأن تقويم الأطفال لا يتم على أساس عدد الأخطاء في إجابتهم في الاختبارات، وإنما على أساس سوء فهمهم أو استخدامهم قواعد غير مكتملة مما يؤثر في أدائهم. فتشخيص الأداء بهذا الشكل يربط بين عملية التقويم وعملية التعليم.

كذلك صمم بعض الباحثين نماذج رياضية تتعلق بكيفية تعلم الأطفال مهارات القراءة والكتابة استناداً إلى مقارنات بين الأساليب والإجراءات التي يتبعها الأطفال المبتدئين، والأطفال الأكثر تقدماً. وأوضحت الدراسات التي تناولت الأداء الكفاءة أن قدرة الإنسان على أداء كثير من المهام التي تتطلب الانتباه تكون محدودة. فإذا كان التجهيز الآنى Simultaneous Processing لهذه المهام التي يتضمنها نشاط معقد يتطلب انتباهاً يقظاً، فإنه تكون هناك صعوبات نظراً لأن الانتباه ينبغي أن يتحول من مهمة إلى أخرى. ولكن إذا أصبح أداء بعض هذه المهام آلية من خلال الممارسة ولا يتطلب عندئذ انتباهاً يقظاً، فإن الفرد يمكن أن يكرس جهده لأداء المهام الأعلى (Glaser, 1984; Greeno, 1984; Kintsch, 1985) ويوضح هذا التناقض بين مكونات المهام عند دراسة مهارات القراءة، وبخاصة العلاقات بين مهارات القراءة المتعلقة بالكلمات، وكذلك عمليات فهم المقرء. فانتباها القارئ يتحول بين مهارات فك الترميز المتعلق بتعريف الكلمات، ومهارات فهم المقرء التي تعمل على تكامل الأفكار التي يتضمنها موضوع القراءة في الذاكرة. ويمكن ملاحظة التحول في الانتباه إذا كان الفا، د، مبتدئاً حيث يكون تركيزه متبدلاً بين معرفة نطق الكلمة وبين معنى الكلمة في سياق المفروء، والانتباه لكل من هذه العمليات يستغرق وقتاً من القارئ، وبطء تفاعلها مع

العمليات الأخرى يمكن أن يؤدي إلى عدم الكفاءة في القراءة. لذلك فإن الأداء المنخفض في القراءة يعكس في كثير من الأحيان تداخل تأثير بطيء وعدم فاعلية فك الترميز في تنفيذ فهم مهام من مستويات أعلى. وهذا التداخل له تضميناته في تقييم التعلم.

بعض العمليات تتطلب وصولها إلى مستوى معين من الكفاءة لدى الفرد بحيث يمكن تنفيذ عمليات أخرى معاً وبطريقة متسقة. ولكن يكون التعلم فاعلاً ينبغي تقييم مستوى فاعلية المهارات الأساسية المطلوبة لتقليل تداخل هذه التأثيرات مع العمليات التي من مستوى أعلى، بحيث تيسر كل عملية العملية الأخرى. وهذا يتطلب تصميم أساليب تشخيص الكفاءة في المهارات الأساسية بحيث يتبع نجاحها في تركيز الانتباه للمستويات العليا. وبينما علماء النفس المعرفي جهوداً كبيرة في محاولة فهم البنية المعرفية والقدرات لدى الأفراد الذين يتميزون بمهارة معينة، وتحليل العمليات المتضمنة في تحويل الأفراد المبتدئين إلى متميزين تدريجياً. ففهم طبيعة التميز أو الكفاءة سوف يمكن خبراء القياس من تطوير أساليب لتقدير الأداء في مستويات كفاءة متعددة.

وقد أوضحت نتائج الدراسات في هذا المجال أن الأفراد المتميزين يعملون على تنمية تظميمات للمعرفة تمكّنهم من إدراك أنماط ذات معنى بسرعة في ذاكرتهم وهذا يسمح لهم بتكون تمثيلات Representations للمشكلات تمكّنهم من التصرف الجيد المناسب. بينما يقوم الأفراد المبتدئون بتمثيل المشكلات بطريقة مختلفة اختلافاً نوعياً وأصطناعياً مما يجعل الموقف المشكلة أكثر صعوبة في حلها.

فتعزّز الأنماط Pattern Recognition، وتمثيل المشكلات Problem Representation تعد بمثابة مؤشرات للكفاءة، وينبغي تضمينها في تقييم نمو مهارات المبتدئين ليصبحوا متميزين في مجال معين، وكذلك ينبغي الاستناد إليها كأساس لبناء مفردات الاختبارات قبل إجراء العمليات السيكومترية، مثل تحليل المفردات أو تقدير الثبات. ولعل هذا يعد من التوجهات المهمة التي ينبغي أن تناول قدرًا وافياً من البحوث المستقبلية بحيث لا يعتمد تقييم التعلم أو الأداء على مجرد درجة كلية أو تقدير عام أو مئيني، وإنما يعتمد على نظام تقدير تعليمي Instructional Scoring يوجه التعلم ويوضح متطلبات رفع مستوى الكفاءة لدى المتعلمين.

خلاصة

تناولنا في هذا الفصل بعض أهم التطورات المعاصرة في مجال القياس والتقويم التربوي والنفسى التي أصبح خبراء القياس يركزون عليها اهتمامهم ويولونها مزيداً من البحث والدراسة، وتنعدد حولها العديد من المؤشرات المتخصصة في كثير من الدول المتقدمة. ويتعلق أحد هذه التطورات بالنظرية السيكومترية المعاصرة التي يطلق عليها نظرية الاستجابة للمفردة الاختبارية IRT وما تضمنه من نماذج، مثل نموذج البارامتر الواحد (نموذج راش)، والنموذج ثنائى البارامتر (نموذج لورد)، والنموذج ثلاثي البارامتر (نموذج بيرنوبم)، وغيرها من النماذج التي انبثقت عنها.

وتحاول هذه النماذج التغلب على أوجه قصور النظرية الكلاسيكية للاختبارات التي سيطرت على الفكر التربوي والنفسى ما ينوف عن نصف قرن، وقد أوضحتنا في هذا الفصل بعض أهم هذه الأوجه. فهذه النماذج الجديدة ترتكز على المفردة الاختبارية وتأثيرها في الاختبار ككل. وكل منها يعبر عن التوزيع الاحتمالي للاستجابات الممكنة للفرد لمفردة اختبار معينة، وتشتق خصائص الاختبار ككل من خصائص المفردات التي يشتمل عليها، على عكس النظرية الكلاسيكية التي ترتكز على الاختبار بعامة. لذلك تتميز هذه النماذج بميزات متعددة، فإذا تم تعديل مجموعة من المفردات (أى تقدير بارامتراتها) وفقاً لنموذج راش مثلاً، فإنه يمكن بسهولة تحويل درجات الاختبارات التي يشتمل كل منها على مجموعة فرعية من المجموعة الأصلية إلى ميزان مشترك Common Scale وهذا ييسر عملية تكافؤ الاختبارات Test Equating، وبناء اختبارات مفصلة أى موائمة لمستوى قدرة الفرد الذى يُختبر بها كما سنوضح في الفصل التالي. كما أن هذه النماذج تتميز بأنها تزودنا بأساس إحصائى سيكومترى يستند إليه في تصميم الاختبارات.

والتطور الثانى يتعلق بنظرية الثبات متعدد الأبعاد التي يطلق عليها نظرية إمكانية التعميم Generalizability Theory . واهتمت هذه النظرية أيضاً بمعالجة بعض أوجه قصور النظرية الكلاسيكية للاختبارات فيما يتعلق بمفهومي الثبات والصدق، وتوسيع نطاق هذه النظرية من أجل تعميق وتفسير كل من هذين المفهومين، ومعالجتهما معالجة أكثر شمولية. وهذه النظرية المعاصرة تسمح بمعالجة مصادر متعددة للخطأ،

كما تسمح بتعريفات متعددة للدرجات الحقيقة. فالنظرية الكلاسيكية لم تميز بين مصادر الخطأ المحتملة في هذه الدرجات الحقيقة. لذلك تهتم نظرية إمكانية التعميم بمدى تعميم عينة القياسات المستمدة من مجموعة من البيانات أو الملاحظات على المجموعة الشاملة لهذه البيانات. وتحاول دراسات إمكانية التعميم G-Studies تحديد تأثير معاینات شروط أو ظروف مختلفة على درجات الاختبارات من أوجه Facets المختلفة، وتقدير مكونات التباين التي ترجع إلى اختلاف هذه الأوجه، وكذلك تقدير تفاعل هذه المكونات باستخدام أساليب تحليل التباين Analysis of Variance وفقاً لتصنيفات مناسبة أشرنا إليها. وبذلك يمكن الحصول على معامل إمكانية التعميم الذي يدل على النسبة بين تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة، ويُعد هذا المعامل بدليلاً عن معامل الثبات الذي تتناوله النظرية الكلاسيكية للاختبارات. وكذلك يمكن الإفادة من دراسات القرارات D - Studies في جمع بيانات تفيد في اتخاذ العديد من القرارات، والتوصيل إلى تفسيرات واستنتاجات بناء على نتائج دراسات إمكانية التعميم D - Studies أو أوضحنا بعضها بأمثلة، كما أوضحنا علاقة هذه الدراسات بنظرية الصدق.

أما التطور الثالث فيتعلق بإسهامات علم النفس المعرفي في تصميم الاختبارات بحيث يتحول الاهتمام من استخدام الاختبارات في تقويم التحصيل العام للمتعلمين فقط إلى الإفادة من البيانات المستمدة منها في تحسين وإثراء عملية التعليم وتقويم المعلمين والأساليب التعليمية. وقد أدى هذا التوجه إلى توظيف نتائج بحوث علم النفس المعرفي في تصميم الاختبارات من منظور جديد يركز على العمليات المعرفية المتضمنة في الأداء الكفاء في مختلف المجالات الدراسية. وسوف يؤدي ذلك إلى بناء اختبارات تختلف اختلافاً ملحوظاً عن الاختبارات المستخدمة في وقتنا الحاضر من حيث التصميم والشكل. فالجيل الجديد من الاختبارات سوف يركز على قياس عمليات تجهيز المعلومات ومعالجتها وما تتطلبه من انتباه يقظ، وترميز، وفك الترميز، وتعارف الأنماط، وتمثيلات معرفية مختلفة، وغير ذلك من العمليات التي يستخدمها المتعلم في إجابته عن مفردات الاختبار. وسوف يساعد ذلك في تنمية مهارات المبتدئين لكنه يصبحوا مقتدرین في مجال معين من مجالات المعرفة استناداً إلى أساليب علمية، وتقويم ذلك تقويمًا تشخيصياً متضمناً في عملية التعليم ومتكملاً معها. وسوف يؤدي ذلك إلى تغيرات جوهرية في نظريات القياس وممارسات التقويم.

الفصل الخامس عشر

بحجز التوجّهات التطبيقية

* مقدمة

- * تطوير أساليب قياس وتقدير التحصيل الدراسي
 - (أولا) استخدام الاختبارات مرجعية المحك في تقدير التحصيل
 - (ثانيا) تقنيات تفريذ قياس التحصيل باستخدام الاختبارات المفصلة
- * الاستراتيجية ذات المرحلتين
- * الاستراتيجية متعددة المراحل
- * نموذج التفريغ الثابت
- * نموذج التفريغ المتغير
- * استراتيجية الاستعانة بالحاسوب
 - (ثالثا) بناء نظم بنوك الأسئلة
- * مراحل إنشاء نظام بنك أسئلة مقترن
- * البنك الدولي للأسئلة، وتقدير التحصيل التربوي من منظور دولي
 - (رابعا) التقييم التوثيقى الشامل للتحصيل والكميات
- * مكونات الصحف الوثائقية Portfolios
- * صيغ الصحف الوثائقية
- * بعض مشكلات تقييم التحصيل باستخدام الصحف الوثائقية
- * خلاصة

مقدمة

أسهمت التطورات النظرية السيكومترية التي أوضحتها بایجاز في الفصل السابق وغيرها من التطورات المعاصرة في إثراء تطبيقات القياس والتقويم التربوي النفسي وتوسيع نطاقها في مجالات متعددة. وعلى الرغم من أن بعض هذه الإسهامات لا تزال في مرحلة التجريب، والبعض الآخر أصبح واقعاً مستخدماً في تصميم وبناء العديد من الاختبارات والمقاييس المنشورة في وقتنا الحاضر، إلا أن هناك جهوداً متزايدة من جانب خبراء القياس لتذليل العقبات التي تحول دون الإفادة المثلثة من هذه الإسهامات في مختلف الجوانب العملية التطبيقية، وتبسيطها للممارسين والدارسين في هذا المجال. فتصميم الاختبارات والمقاييس من منظور جديد أصبح يستند إلى التطورات التي حدثت في علم النفس المعرفي، وما أسهمت به في إلقاء مزيد من الضوء على العمليات المعرفية التي ستؤدي إلى توجه جديد في تصميم عملية التعليم والأنظمة الاختبارية.

كما أن التطورات التقنية وبخاصة تقنيات الحاسوب منخفضة الكلفة سوف تلعب دوراً مهماً في الصناعة المتعلقة بالاختبارات Testing Industry . وسوف يكون لذلك انعكاساته الواضحة على نظم وأساليب القياس والتقويم التربوي النفسي في هذا القرن.

لذلك سوف نتناول في هذا الفصل بعض أهم التوجهات التطبيقية، والقضايا المتعلقة باستخدام الاختبارات والمقاييس من هذا المنظور الجديد.

تطوير أساليب قياس وتقويم التحصيل الدراسي:

ينظر الكثيرون إلى التحصيل الدراسي نظرة تقليدية محدودة، حيث يرون أن تحصيل الطلاب يتعلق ببعض الجوانب المعرفية التي يكتسبونها من تعلم المحتوى الدراسي المقرر والتي يسهل على المعلم قياس وتقويم مدى تحققه. وفي إطار هذا المفهوم الضيق أصبحت تقصر الاختبارات التحصيلية الصافية وال العامة على قياس قدرة الطلاب على استرجاع الحقائق والمعلومات المستقرة المتعلقة بالمحتوى الدراسي باستخدام أسئلة ومفردات اختبارية بسيطة. وتبني هذه الاختبارات في غالبية الأحيان دون الاستناد إلى نظرية أو نموذج متتطور في القياس التربوي، ونفس درجاتها دو

تحديد واضح للإطار المرجعى الذى يعطى معنى ودالة لهذه الدرجات، وبذلك لا تشكل هذه الأسئلة والمفردات فى مجلتها اختبارا تحصيليا بمفهومه العلمي، نظرا لأنها لا تمثل النطاق السلوكي الشامل للمعارف والمهارات التى ينطوى عليها المحتوى الدراسى، إذ أن إجابتها لا تتطلب عادة قدرًا كافيا من التفكير، أو إدراك العلاقات، أو تقييم الأدلة، أو اقتراح بدائل وتوجهات، أو اتخاذ القرارات، أو حل المشكلات التي يمكن أن يواجهها الفرد المتعلم فى تعامله مع معطيات حياته ومتطلباتها الملحة. فكيف والحال كذلك يمكن الإفادة من نتائج مثل هذه الاختبارات التحصيلية والاستناد إليها فى اتخاذ قرارات صائبة تتعلق بالمتعلم كإنسان مفكر ومبعد، وبالتعلم كمصمم ومهندس تعليمى، وباستراتيجيات التعليم كمنهجية وأسلوب علمي، وبالبرنامج التعليمى كمنظومة متفاعلة ومتعددة.

فالاختبارات التحصيلية بوضعها الراهن تُبنى على فلسفة تربوية تؤكد بل وتشجع على إبراز الفروق الفردية، وتحث على التعليم التنافسى من أجل حصول الفرد على مركز نسبي متوفّق بين أقرانه دون محاولة تحديد ما يمتلكه الفرد من مهارات وظيفية وأخلاقيات وسلوكيات إيجابية بناءً أو أخذ المصالح المشتركة للجماعة كفريق ينبعى أن يعمل متأللاً لخير المجتمع ورقى بعين الاعتبار.

فهذه النظرة الضيقية والمحدودة لمفهوم التحصيل الدراسي والفلسفة التربوية التي تهتم بالفارق بين الأفراد والموازنة بينهم وما ترتب على ذلك من أساليب قياس وتقويم للتحصيل تركز على ما اختزنه المتعلم في ذهنه من معلومات محددة لم تعد تناسب المتطلبات المستقبلية للتربية واحتياجاتها المتغيرة في هذا القرن الذي يتميز بالتفجر المعرفي، والتقدم التكنولوجي، وثورة المعلومات والاتصالات. فال التربية سوف تقود التغيير وتكون في مقدمته، وسوف تلعب دورا رئيسا في تنمية المجتمعات المستقبلية التي نظمت إليها. كما أن القضايا والمشكلات التربوية الحيوية المتعلقة بالتحديات الثلاثة المرتبطة والمتغيرة التي تواجهها البشرية وهي: التنمية، والحفاظ على البيئة وحمايتها، والسلام سوف تصبح المحور الرئيسي للسياسات الدولية والقومية، هذا بالإضافة إلى أن المتغيرات المتعلقة بالنمو السكاني، والتطورات العلمية والتكنولوجية، وأنماط الوظائف والأعمال المستحدثة يتوقع أن تصبح من متغيرات التربية والتعليم مستقبلا (Klchinkoff, 1922; Naisbitt 1984; UNESCO, 1989; USIA, 1991) ..

ففى إطار هذه الأهداف المستقبلية للتربية وفي ضوء هذا المنظور الربح للتحصيل الدراسي لا تكون عملية التقويم غاية فى ذاتها لتحديد نجاح المتعلمين

وانتقالهم إلى الصفوف الأعلى أو المراحل التعليمية اللاحقة أو رسوبيهم، وإنما تصبح جزءاً متكاملة من عملية التعليم توجهها وتعززها وتصحح مسارها بما يتحقق هذه الأهداف. وهذا يتطلب التحول من أساليب ونظم الامتحانات التقليدية السائدة والدرجات والتقديرات الاعتبارية التي تقصر على الموازنة بين أداء المتعلم وأداء آقرائه في اختبارات تهم يلقي الضوء على الفروق الفردية وتؤكدتها إلى أساليب ونظم تؤكد الشخصية المتكاملة والمتوازنة للمتعلم وما يمتلكه من كفايات ومهارات وظيفية، والفهم العميق والممتد لمضمون المواد الدراسية الذي يكتسبه المتعلم من خلال أساليب التعلم الذاتي وحب الاستطلاع بما يُمكّنه من التعامل مع بيئته وإثرائها.

وقد توصل كروكس (Crooks, 1988) في دراسة شاملة حول أثر التقويم داخل الصنف على المتعلمين إلى وظائف متعددة لعملية تقويم التحصيل الدراسي على مستوى المدرس، ومستوى المقرر ، وعلى المدى الأبعد. ويمكن إيجاز بعض الوظائف فيما يلى :

على مستوى الدرس:

- تشجيع وتعزيز المعارف والمهارات الالزمة للبدء في دراسة موضوع دراسي جديد .
- تركيز الانتباه على الجوانب المهمة في موضوع الدرس.
- حفز المتعلمين على استخدام استراتيجيات تعلم نشطة .
- إتاحة الفرص الكافية للمتعلمين للتدريب على المهارات تدريباً كافياً يسمح بعمق التعلم .
- تعريف المتعلمين بنتائج جهودهم وإرشادهم إلى كيفية توجيه هذه الجهود وتصحيح مسارها .
- إرشاد المتعلمين إلى كيفية متابعة تقدمهم الدراسي بأنفسهم وتنمية مهاراتهم في التقويم الذاتي بما يؤهلهم لانتقاء الأنشطة التعليمية والتعلمية التالية من أجل زيادة درجة تمكّنهم واقتانهم .
- تنمية حس الإنجاز لدى المتعلمين .

على مستوى المقرر أو الموديول التعليمي:

- التأكيد من تحقيق المتعلمين للمعارف والمهارات المطلوبة للتعلم الفاعل للمقررات الدراسية.
- حت المتعلمين على دراسة المقرر وتنمية إدراكيهم لقدراتهم في دراسة موضوعاته.
- تعريف المتعلمين بالأهداف الأساسية المتکاملة للمقرر، ومستويات الأداء المرجو تحقيقها.
- تشجيع المتعلمين على انتقاء وتطوير استراتيجيات وأساليب استذكار تثري تعلمهم.
- تحديد إنجازات المتعلمين في المقرر بما يسهم في نشاطاتهم المستقبلية.

على المدى الأبعد:

- تعزيز قدرة المتعلمين على استبقاء المعلومات المعلمة، وتوظيفها في مواقف متعددة وبطرق متعددة.
- تعزيز مهارات التعلم وأساليب الاستذكار النشط لدى المتعلمين.
- الحفز المستمر لدافعية المتعلمين من أجل استمرارية التعلم.
- تنمية إدراك المتعلمين لجوانب التميز في شخصياتهم وكفاءاتهم.

وتحقيق هذه الوظائف يتطلب ضرورة الأخذ بالمفهوم العلمي لعملية التقويم في إطارها الشامل كما أوضحته في الفصل الأول باعتبارها عملية منهجية تتطلب جمع بيانات موضوعية ومعلومات صادقة باستخدام أدوات مقاييس متنوعة في إطار مجموعة من الأهداف المحددة للوصول إلى تقديرات كمية وأدلة كيفية يستند إليها في صنع القرارات وإصدار الأحكام المناسبة التي تتعلق بالمتعلمين وبعملية التعليم، وذلك لتحسين نوعية الأداء ورفع درجة الكفاءة بما يساعد على تحقيق هذه الأهداف.

(أولاً): استناد الاختبارات مرجعية المحكء في تقويم التدريس

: Criterion - Referenced Evaluation

يتضح مما سبق أن الأساليب السيكومترية الكلاسيكية في القياس التربوي وما تستند إليه من فلسفة تربوية تحت على التنافس الفردي من أجل التفوق على الأقران

تجعل عملية تقويم التحصيل محدودة الفائدة وذات تأثير سلبي على المتعلمين والمناخ المدرسي، وبذلك لا تهم في تحقيق عملية التقويم لوظائفها المرجوة التي أشرنا إليها.

فالاختبارات التحصيلية التي تبني وفق هذا التوجه التنافسي تستهدف التميز والتفاصل بين الطلاب من أجل وضع تقديرات لهم في نهاية المدة الدراسية أو اتخاذ قرارات الانتقاء. لذلك يحاول الطالب جاهداً توقيع الأسئلة التي سوف يستعمل عليها الاختبار، واسترجاع ما اختزنه في ذاكرته من حقائق وتعليمات، والعمل منفرداً بغاية النجاح.

ويفترض في هذا التوجه أن تعلم الطالب يعتمد اعتماداً أساسياً على قدراته والجهد الذي يبذله، وعلى العكس من ذلك يفترض أن إخفاق الطالب يرجع إلى انخفاض قدراته أو سلبية اتجاهه نحو التعلم أو المادة الدراسية. أي أن الفروق الفردية في إطار هذا التوجه تعد بمثابة معطيات. وبذلك يحظى الطالب الأكثر قدرة على رتبة أعلى من الطالب الأقل قدرة. غير أن هذا يتربّط عليه آثار ضارة بكل منهما، فالطلاب الناجحون يجدون أن التعليم الذي يتلقونه يغفل تنمية مهاراتهم في التفكير وحل المشكلات التي يمكن أن يواجهونها. أما بالنسبة للطلاب الراسبين فالضرر أكبر حيث إنهم يجدون أن التعليم الذي يتلقونه غير مناسب لهم وربما يحرمون من فرص متابعة دراستهم.

ومما يؤخذ على هذا النوع من الاختبارات التحصيلية التي يُطلق عليها «الاختبارات مرجعية الجماعة أو المعيار Referenced Tests» أن التأكيد الأساسي فيها يكون على الدرجات الكلية، وبذلك يقل الاهتمام فيها على دقائق المحتوى بينما يزيد التركيز على التجريدات الأكثر عمومية. ولهذا فإن الاختبار يحدد الدرجة الكلية في العمليات الحسابية مثلاً ولكنها لا يوضح شيئاً عن نوع العملية الحسابية التي يستطيع الطالب أو لا يستطيع إجراءها. فالدرجة الكلية يصعب استخدامها في تحديد مواطن الضعف أو الأخطاء الشائعة أو فجوات التحصيل. لذلك لا نعجب إذا وجدنا أن المعلمين أو الطلاب لا يستفيدون شيئاً من نتائج هذه الاختبارات التحصيلية سواء الصافية أو المقننة في إثراء عملية التعليم والتعلم. كما أن انتقاء مفردات هذه الاختبارات يعتمد بدرجة كبيرة على الأساليب الإحصائية في تحليل مفرداتها حيث تستيقن فقط المفردات التي تميز بدرجة أفضل بين الطلاب وتسمح بترتيبهم بالنسبة لبعضهم البعض.. لذلك فإن الاختبار في صورته النهائية يكون مشتملاً على مفردات

متوسطة الصعوبة وترتبط درجاتها ارتباطاً مرتقاً بالدرجة الكلية في الاختبار . وانتقاء المفردات التي يتكون منها الاختبار التحصيلي على هذا الأساس يقلل من صدق محتوى الاختبار ، أي مدى تمثيل مفرداته لنطاق المحتوى الذي يقيسه ، ويحد بالاختبار عن الغرض منه كأداة قياس نظراً لما يعترى نطاق المحتوى من تغيرات غير معلومة أو محددة أثناء عملية تحليل المفردات . لذلك فإن إضافة أو حذف المفردات بهذا الشكل من أجل التحكم في مستوى صعوبتها وتميزها يؤثر تأثيراً بالغاً في صدق الاختبار . هذا بالإضافة إلى أن كثرة عدد مثل هذه المفردات التمييزية يجعل الاختبار أقرب إلى اختبارات الاستعدادات منه إلى اختبارات التحصيل . وعندئذ لا يصبح الاختبار أداة قياس للتحصيل ، وإنما أداة للتمييز بين الطلاب . فقياس التحصيل ربما يؤدي إلى تمييز جيد بين الطلاب أو ربما لا يؤدي إلى ذلك (Izard, 1990) . فالهدف من قياس التحصيل كما حدده جلizer (Glaser, 1963) بوضوح هو «تقديم معلومات تفصيلية محددة عما يستطيع المتعلم أداءه وما لا يستطيع» . فعملية قياس تحصيل الطلاب إذن تعد عملية تحليلية تقنية ، بينما عملية التمييز بينهم تعد عملية سيكومترية . فالاختبار الذي يصمم بهدف قياس التحصيل ينبغي أن يكون تركيزه الأساسي على محتوى المفردات وليس على خصائصها الإحصائية التي تناسب التوزيع الاعتدالي . فاختبار التحصيل يهدف لتقرير مستوى أداء الفرد بالنسبة لنطاق سلوكي معرف تعريفاً جيداً ومحدوداً وليس مجرد توضيح أن الطالب قد حصل على درجة كلية أعلى أو أقل من معظم أقرانه . وتصبح الخطوة الأساسية في بناء اختبارات التحصيل من هذا المنظور هي تحديد محتوى المفردات تحديداً تفصيلياً واضحاً ، بل وينبغي أن يكون نطاق المحتوى كاملاً بقدر المستطاع .

ونعني بذلك أنه ينبغي حصر جميع مفردات النطاق الممكنته صراحة أو ضمنياً . وهذا يتطلب تحديد مواصفات النطاق المعين بدلاً من الاعتماد على جدول مواصفات الاختبار الشائع الاستخدام . فإذا ما تم تحديد نطاق المحتوى Content Domain تحديداً كاملاً لا تصبح عملية تحليل المفردات غير ضرورية فحسب بل وغير مرغوبية وذلك من أجل الحفاظ على صدق معايير هذا النطاق Content Sampling Validity ، ويكون الحكم على نوعية الاختبار كمقاييس للتحصيل في ضوء نطاق المحتوى وليس لكل مفردة على حدة . غير أن التحديد الكامل لنطاق المحتوى يكون ممكناً بدرجة أكبر في الجوانب النفسية الحركية وبعض المجالات الدراسية كالرياضيات والعلوم ، أما في المجالات الوجدانية ومجالات العلوم الإنسانية ربما يصعب ذلك ، ويقتصر على تحديد الأهداف أو النواتج السلوكية المرجوة من دراسة المحتوى ، ووضع مواصفات تفصيلية

لهذه الأهداف ، وما يجب أن تشمل عليه مفردات الاختبار . فعندئذ يمكن استخدام أساليب تحليل المفردات في التعرف على المفردات الجيدة والمفردات التي تحتاج إلى مراجعة . غير أنه في هذه الحالة لا ينبغي استبعاد المفردات التي يتبيّن أن معامل تميّزها منخفض أو سلبي وإنما يجري فحصها لمعرفة جوانب القصور فيها (Berk, 1978) . Popham, 1980)

ومما لا شك فيه أن التركيز على نطاق المحتوى وخصائصه وليس على الخصائص الإحصائية لمفردات الاختبار يتطلّب إعادة النظر في الممارسات السائدة في تصميم وبناء الاختبارات التحصيلية الصفيّة والعامّة والفلسفه التربويّة التي تستند إليها هذه الممارسات . فنطاق المحتوى يصبح الإطار المرجعي أو المحك Criterion الذي تُنسب إليه درجات الاختبار وليس معيار الجماعة Norm كما هو الحال في الاختبارات التحصيلية الكلاسيكية . وبذلك لا يقتصر تقرير نتائج الطلاب في الاختبار على درجة كلية وإنما يكون التقرير وصفياً تفصيلياً يحدد إنجارات الطالب وما استطاع أن يحققه أو يتقنه من معارف ومهارات وسلوكيات متضمنة في نطاق المحتوى الذي يقيسه الاختبار الذي يكون عندئذ مرجعى المحك Criterion - Referenced وليس مرجعى المعيار Norm - Referenced .

فالاختبارات مرجعية المحك - كما أوضحنا في الفصل الثامن - تهتم بقياس مدى تحقق مجموعة متكاملة محددة من المهام وتزود المتعلم بأدلة عن درجة تمكنه منها ، كما توضح تقدّمه أثناء دراسة مطلب تعليمي محدد وما يترتبه من صعوبات تعلق هذا التقدّم ، وبالتالي يمكنه أن يجري تعديلاً في بيته تعلمه ونوعية تعليمه مستفيداً في ذلك من مقتراحات معلمه وتعليقاته الإرشادية التي يوضّحها على ورقة إجابة الاختبار . فالمعلم يسعى إلى تمكين جميع الطلاب من التحسّن وذلك بتحديد متطلبات يمكنهم تحقيقها ويكون دور الطالب هو مشاركة المعلم في تحضير الأنشطة والعمل بجد لاستكمال هذه المتطلبات ، وإبراء وجهة نظره فيما تعلمه في إطار التعاون والمشاركة النشطة مع الأقران في الفصل المدرسي الذي يشتمل على طلاب من ذوي قدرات متفاوتة . حيث إن مستويات التحصيل لا يتّحد غالباً في ضوئها تجمّع الطلاب في مجموعات .

فالقدرة يُنظر إليها على أنها مرتبطة ارتباطاً وثيقاً بنوعية التعليم وبيئته وطبيعة مهامه وملاءمتها للفرد ودأعيته . وهذه المنهجية الجديدة في قياس التحصيل تؤدي إلى

تكامل عمليتي التعليم والتقويم وهو هدف مهم تسعى إليه الأنظمة التربوية
(Cronbach & Snow, 1977; Snow, 1980).

لذلك فإن الاختبارات مرجعية المحك أثرت تأثيرا ملحوظا في تطوير مفهوم التحصيل حيث أصبح ينظر إلى قياس التحصيل كما سبق أن ذكرنا على أنه تقويم الكفايات Competency Assessment، أي قدرة المتعلم على القيام بمهام محددة ذات طابع تكاملي وظيفي بنجاح. فالتحصيل من هذا المنظور يعد بمثابة أداء Performance أي القياس بنشاط محدد. وهذا يتطلب أن يكون النطاق السلوكي لهذه الكفايات أو المهام معرفاً دقيقاً بحيث يصبح هذا النطاق هو المحك الذي تُنسب إليه درجة الفرد في الاختبار. فتحديد النطاقات السلوكية يعد الخطوة الرئيسية في بناء هذا النوع من الاختبارات. فقد اهتم علماء النفس في الماضي ببناء نماذج سيكومترية تتعلق بدرجات الاختبارات وخصائصها الإحصائية، ولم يهتموا كثيراً ببناء نظريات يمكن الاسترشاد بها في تحديد النطاقات السلوكية التي تقيسها الاختبارات التحصيلية. لذلك تبيّن الحاجة الماسة إلى نظرية معرفية متكاملة Cognitive Theory يسترشد بها في تعريف نطاق الأهداف التعليمية تعرضاً يُمكن من تكوين اختبارات محاكمة تسمح بالتفصير الدقيق للمطالب والمهام التي أتقنها الطالب، كما تبيّن الحاجة إلى برمجيات للحاسوب يمكن باستخدامها ميكنة عملية تكوين المفردات الاختبارية Computerized Item Generation لتقليل عنصر الذاتية إلى أقل حد ممكّن وقياس النطاق السلوكي الشامل للمحتوى التعليمي. وُوجه خبراء القياس في الدول المتقدمة جانباً كبيراً من جهودهم لتطوير هذه التقنية من خلال أساليب متعددة أشرنا إليها في الفصل الثامن وهي: برمجيات بناء المفردات Item generation Programs ، والجمل الراسمة Mapping Sentences ، وصيغ المفردات Item Forms ، والتحويلات اللغوية للمفردات Item Linguistic Transformations.

وتعتمد معظم هذه الأساليب على تخزين برامج خاصة للحاسوب يتيح عن تنفيذها تكوين عدد كبير من المفردات الاختبارية وفق قواعد ومواصفات محددة للنطاق السلوكي المراد قياسه. وعلى الرغم من التطورات التي حدثت في هذا المجال، إلا أن هناك العديد من المشكلات المنهجية التي تعيق تنفيذ هذه الأساليب وتوسيع نطاق تطبيقاتها في مختلف المجالات الدراسية. لذلك أصبحت هذه المشكلات مجالاً خصباً لدراسات وبحوث القياس التربوي المعاصر.

(ثانياً) تقنيات تفريط قياس التحصيل باستخبارات الاختبارات

المُفْرَطَة : Tailred testing

إن التطورات العلمية والتكنولوجية والحركة البحثية الحادثة في مجال علم النفس المعرفي المعاصر - كما أوضحتنا في الفصل السابق - سوف تؤدي إلى تطوير ملحوظ في أساليب قياس التحصيل الدراسي والعمليات الاختبارية وما تسهم به في تغيير الإطار التقليدي لبيئة التعلم وبخاصة عندما تتوافر الحاسوبات الآلية المصغرة ويسهل اقتناها. فالتطورات التكنولوجية التي حدثت منذ أواخر الثمانينيات المتعلقة بتصميم شبكات الحاسوبات، والتعقد المتزايد في وحدات هذه الحاسوبات سواء استخدمت منفردة أو متصلة ببعضها كشبكة سوف تؤثر تأثيراً ملحوظاً في البنية التحتية للمؤسسات التعليمية عندما يتوافر لكل من المعلم والمسئول الإداري والطالب حاسوب شخصي Personal Computer سواء في المدرسة أو المنزل بحيث ترتبط جميع هذه الحاسوبات من خلال شبكة بحيث يمكن أن يتصل كل منهم بالآخر والمشاركة في استخدام البرامج والبيانات. وسوف يتبع هذا النظام التقني إمكانات متعددة جديدة لمدرسة المستقبل، حيث يؤدي إلى زيادة مصادر المعلومات نتيجة توسيع حلقة الاتصال بالعالم الخارجي سواء عن طريق القمر الصناعي، أو الوسائل الخاصة بالمقررات الدراسية ، أو الاتصال بين الطلاب في شتى الأماكن أو ربما بينهم وبين طلاب الدول الأخرى عن طريق شبكات الإنترنت . كما يؤدي إلى تجميع المعلومات من مصادر متعددة في إطار متكامل بحيث يشكل قاعدة بيانات ، وكذلك جعل الطلاب والكبار يتعلمون في المنزل كما لو كانوا في المدرسة (Sleeman & Brown, 1982).

وتبرز هذه التطورات التي سوف تحدث في البيئة التربوية وما سوف تقدمه التقنيات الجديدة من خبرات ثرية للطلاب العديد من التساؤلات حول الأنظمة الاختبارية فإذا تعلم الطالب الهندسة الفراغية مثلاً من المنظور الديناميكي الذي يتميز بالشكل والحركة لموقع مجسم معقد يوضحه الحاسوب ، فكيف يمكن قياس تمكنه من هذه المعلومات باستخدام اختبار ورقة وقلم. إن هذا سوف يتطلب أدوات قياس من نوع جديد تحدد للطالب إنجازه المباشر في المهمة التي بين يديه من خلال استجابته لمثير أو مثيرين بدلاً من اختبار يستغرق ساعة أو ساعتين بحيث يمكنه التعرف مباشرة على مدى تمكنه من المهمة المعينة بما يتناسب وقدرته وسرعة تعلمها . ونظراً لأن المهام تختلف في طبيعتها فإن هذه الأدوات ينبغي أن تتتنوع بتنوع المهام، مما يتطلب

تقديم هذه المثيرات باستخدام الحاسوب. لذلك اتجهت بحوث القياس المعاصر في الآونة الأخيرة إلى دراسة كيفية تفريذ قياس التحصيل الدراسي وتطوير تقنياته مما سيؤدي إلى ابتكار أنواع جديدة من اختبارات التحصيل تختلف اختلافاً ملحوظاً عن الاختبارات المتعارف عليها.

فالقياس التقليدي للتحصيل الدراسي يتم بشكل جماعي، حيث يُطبق على جميع الطلاب الاختبار نفسه أو صيغ مكافأة له على الرغم من تباين مستويات قدراتهم. فلكل تكون للاختبارات الجماعية المقنة فائدة عملية في إطار النظرية الكلاسيكية، يجب أن تشتمل على عدد كبير من المفردات المناسبة للأفراد من ذوى القدرة المتوسطة. ولكن يترتب على ذلك عدم فاعليتها في قياس جميع مستويات قدرة الأفراد بالدقة نفسها. وربما يظن البعض أن إضافة بعض المفردات السهلة والمفردات الصعبة يعد حللاً لهذه المشكلة، ولكن هذا ليس صحيحاً. إذ إن المفردات السهلة ليس لها فائدة في تقدير مستوى الأفراد من ذوى القدرة المرتفعة، والمفردات الصعبة لا تفي في تقدير مستوى الأفراد من ذوى القدرة المنخفضة. فتأثير عامل الممل في الحالة الأولى، وعامل التخمين في الحالة الثانية يزيد من الأخطاء العشوائية للفياس الناجمة عن إضافة هذه المفردات. كما أن اعتماد الخصائص السيكومترية لمثل هذه الاختبارات على خصائص عينة الأفراد المختبرين يجعل الموازنة بينهم ممكناً إذ اختبر كل منهم باختبار مختلف يقيس القدرة نفسها.

وتفريد قياس التحصيل الدراسي يتطلب أن تكون صعوبة مفردات الاختبار مناسبة لمستوى قدرة الفرد الذي يختبر بها. وهذا يتطلب توافق مجموعة كبيرة من المفردات الاختبارية ذات خصائص سيكومترية معلومة بحيث يمكن سحب عينات منها تناسب المستويات المختلفة لقدرة الأفراد. وقد تمكّن علماء القياس المعاصر من تطبيق أسس نظرية الاستجابة للمفردة الاختبارية (IRT) Item Response Theory التي أوضحتها بإيجاز في الفصل السابق في مجالات تطبيقية متعددة عجزت النظرية الكلاسيكية للاختبارات عن مواجهتها. وسوف نوضح فيما يلى أهم هذه المجالات، وهي نظام الاختبارات المفصّلة Item Banking ، ونظام بنوك الأسئلة Tailored Testing .

فالطريقة المثلثى لقياس قدرة الأفراد هي أن تراوّج بين صعوبة مفردات الاختبار ومستوى الدرجة الحقيقية True - Score level ، أو القدرة التي يقيسها الاختبار، أو بعبارة أخرى أن نختبر الأفراد من ذوى القدرة المرتفعة بعدد أكبر من المفردات الصعبة، ونختبر ذوى القدرة المنخفضة بمفردات سهلة. وذلك لأنه إذا كانت مفردات الاختبار

صعبه جداً، أو سهلة جداً بالنسبة لفرد ما، فإننا لن نستفيد شيئاً من هذه العملية الاختبارية سوى الحصول على بعض المعلومات عن الحد الأعلى، أو الحد الأدنى لقدرة هذا الفرد. فإذا أجاب فرد ما إجابة خطأً عن مفردة متوسطة الصعوبة، لا يجوز لنا بعد ذلك مباشرة أن نختبره بمفردات صعبة جداً. لذلك يجب أن نختبر الفرد بمفردات اختبار تناسب مستوى قدرته، لكي نحصل على أكبر قدر من المعلومات عن هذا المستوى. وأفضل المفردات هي تلك التي يكون مقدار احتمال إجابة الفرد عنها إجابة صحيحة دون أن يلتجأ إلى التخمين يساوي ٥٥٪، فيليس من المعقول إذن أن نختبر مجموعة من الأفراد المختلفة في قدرتهم بالاختبار نفسه. ولكن الأرجى أن نختبر هؤلاء الأفراد بمفردات مختلفة. ومن هنا نشأت الحاجة إلى الاختبارات المفصلة ، Individualized Tests ، أو الاختبارات المناسبة لقدرة الفرد Tailored Testing ، وأحياناً تسمى الاختبارات التفريدية Adaptive Testing . والمشكلتان المرتبطتان بهذا النوع من الاختبارات تتلخصان فيما يلى :

- (١) اختيار مفردات تناسب كل فرد.
- (٢) الحصول على تقدير لقدرة الأفراد على ميزان مشترك Common Scale ، على الرغم من اختلاف المفردات التي يختبر بها كل منهم .

وقد اقترح رايت Wright ودوجلاس Douglas طريقة لإعداد هذا النوع من الاختبارات، بحيث يتم تدريج صعوبة مفرداتها باستخدام نموذج راش. وتعتمد هذه الطريقة على إعطاء حرية للفرد في انتقاء المفردات التي يود أن يختبر بها. ثم يطبق عليه اختبار تدرج فيه المفردات من حيث صعوبتها. ويتصفح الفرد بأن يبدأ بالمفردات الأكثر صعوبة والمفردات الأقل صعوبة، إلى أن يقرر بنفسه المفردات المناسبة له. وتقدر درجة الفرد في الاختبار بعدد المفردات التي أجاب عنها إجابة صحيحة، مضافاً إلى ذلك عدد المفردات التي قرر أنها سهلة جداً ولا تتحدى قدرته. لذلك تسمى هذه الطريقة «طريقة إعداد الاختبارات ذات المستوى المرن Flexilevel Testing».

- ويرى لورد (Lord, 1977) أن إعداد نظام الاختبارات المفصلة، يحتاج إلى برنامج حاسوب يؤدي العمليات التالية :

- (١) التنبؤ بكيفية إجابة الفرد عن المفردات التي لم يختبر بها بعد، بمعلومية إجابته عن المفردات التي اختُبر بها بالفعل .

(٢) استخدام هذه المعلومات بطريقة مُجدية في انتقاء مفردة الاختبار التي تطبق عليه في المرحلة التالية.

(٣) تقدير درجة تمثل قدرة الفرد في نهاية الفترة الاختبارية.

وقد حدثت تطورات كثيرة لفكرة الاختبارات المفصلة تناولت بصفة أساسية طبيعة الاستراتيجية التي يمكن ان تستخدم في انتقاء المفردة التالية التي يُختبر بها الفرد. ومن بين هذه الاستراتيجيات:

(١) الاستراتيجية ذات المرحلتين Two - Stage Strategy

(٢) الاستراتيجية متعددة المراحل Multi - Stage Strategy

(٣) استراتيجية الاستعانة بالحاسوب Computer - Assisted Testing Strategy

الاستراتيجية ذات المرحلتين:

اقتصر كل من لورد (Lord, 1971)، وبتز، وويس (Betz & Weiss, 1974) على تطبيق اختبار مبدئي Locator or Routing Test على الفرد، استراتيجية تعتمد على تقييم قدرة الفرد في كل من الاختبار المبدئي Second Stage Test . وهذه تتضمن مجموعة من الاختبارات التي تتوزع صعوداً مفرداً منها على متصل القدرة. وتقدر قدرة الفرد باستخدام درجته في كل من الاختبار المبدئي واختبار المرحلة التالية. وأحياناً تفيد درجة الاختبار المبدئي في توجيه الفرد إلى المستوى الاختباري التالي دون أن تؤخذ هذه الدرجة في الاعتبار عند تقدير قدرته باستخدام درجته في الاختبار التالي. وقد أجرى لورد Lord دراسات مستفيضة باستخدام المحاكاة عن طريق الحاسوب Computer Simulation Studies للبحث عن أفضل طرق تقدير القدرة في هذه الاستراتيجية.

ووجد أن استخدام الدرجتين اللتين يحصل عليهما الفرد في المرحلتين الاختباريتين أفضل من استخدام درجة الاختبار المبدئي فقط في تقدير القدرة. ولكن هذا لا ينطبق في حالة الأفراد من ذوى القدرة المتوسطة.

الاستراتيجية متعددة المراحل:

اعتمدت الاستراتيجية ذات المرحلتين على تفريعة واحدة One Branching، أي انتقال الفرد من الاختبار المبدئي إلى الاختبار التالي الذي يستخدم في قياس قدرته. أما

الاستراتيجية متعددة المراحل فهى تتطلب اتخاذ قرار عقب إجابة الفرد عن كل مفردة، لتوجيهه إلى المرحلة الاختبارية التالية. أى أن الاستراتيجية متعددة المراحل تعتمد على عدة تفريعات.

وتتضمن هذه الاستراتيجية نماذجين: أحدهما يسمى نموذج التفريع الثابت Fixed Branching Model ، والآخر يسمى نموذج التفريع المتغير Variable Branching Model .

نموذج التفريع الثابت:

في هذا النموذج يبدأ جميع الأفراد الإجابة عن مفردة متوسطة الصعوبة، وينتقلون بعدها إلى مجموعة من المفردات المتدرجة في الصعوبة. وبعد الانتهاء من الإجابة عن مجموعة ثابتة من المفردات تقدر قدرة كل منهم إما باستخدام صعوبة المفردة (الافتراضية) التي كان من الممكن أن يُختبروا بها عقب المفردة التي ترتيبها (ن)، أو باستخدام متوسط صعوبة جميع المفردات بما في ذلك المفردة الافتراضية التي ترتيبها (ن + 1) مع استبعاد المفردة الأولى من التقدير.

نموذج التفريع المتخيم:

في هذا النموذج يتم اختيار مفردة من بين مجموعة المفردات Items Pool في كل مرحلة اختبارية يمر بها الفرد إلى أن تتأكد بدرجة كبيرة من الثقة من مستوى قدرته. ويمكن إعادة تقييم قدرة الفرد في كل مرحلة إما باستخدام نظرية بيز Bayes Theory أو باستخدام طرق الأرجحية القصوى Maximum Likelihood Procedures . وقد عرض كل من أوري، (Urru , 1970)، وويس (Weiss,1974)، وجنسما (Jensema, 1974) وود (Wood , 1976) ، ماكرايد Mcbride لهذه الطرق بالتفصيل.

استراتيجية الاستخانة بالحاسوب:

تعد هذه الاستراتيجية أكثر الاستراتيجيات تعقيدا، إذا إنها تعتمد على الحاسوب في اختيار المفردات التي يختبر بها الأفراد أثناء العملية الاختبارية Computer Assisted Testing .

بعد أن يجيب فرد عن مفردة اختبار إجابة صحيحة ، تقدم له مفردة أكثر صعوبة لاختبار طبقاً لصيغة رياضية معينة. أما إذا أجاب عن المفردة إجابة خطأ، عندئذ تقدم له مفردة أقل صعوبة. وانتقاء مستوى صعوبة المفردات التالية يعتمد على «دالة درجية

Step Function». وهذا يعني أن صعوبة المفردة التي ترتيبها ($n + 1$) تساوي صعوبة المفردة التي ترتيبها (n) مضافاً إليها أو مطروحاً منها قيمة معينة تسمى «مدى الخطوة» Step Size.

وتوجد طريقتان مختلفتان لتحديد مدى الخطوة: إحداهما تسمى «طريقة تقليل مدى الخطوة Shrinkage- Step-Size Procedure»، والأخرى تسمى «طريقة مدى الخطوة الثابت Fixed-Step-Size Procedure».

وتبدأ الطريقة الأولى بخطوات مدها متسع، ثم يقل مدى الخطوة كلما اقترب مستوى صعوبة مفردة الاختبار من مستوى قدرة الفرد. ويمكن الحصول على تقديرات Estimates جيدة لقدرة الفرد باستخدام مجموعة معينة من المفردات. إلا أن هذا يتطلب وجود مجموعة كبيرة من المفردات Items Pool تنتقى من بينها المفردات المطلوبة. فإذا كان المطلوب من الفرد الإجابة عن (n) من المفردات، يجب أن يكون عدد مفردات المجموعة $(2)^n$. فمثلاً إذا كانت $n = 6$ يجب أن يكون لدينا 64 مفردة، وإذا كانت $n = 10$ يجب أن يكون لدينا 1024 مفردة. وبالطبع يندر توافر هذا العدد الكبير من المفردات.

أما الطريقة الثانية فهي لا تتطلب إلا عدداً قدره $(n-1)$ من المفردات في المجموعة الكلية. فمثلاً إذا كانت $n = 5$ ، يجب أن يكون لدينا 15 مفردة، وإذا كانت $n = 10$ ، يجب أن يكون لدينا 55 مفردة، وإذا كانت $n = 15$ ، يجب أن يكون لدينا 120 مفردة. ولهذا فإن هذه الطريقة أقل كلفة وأكثر فاعلية من الطريقة الأولى.

وتقدير قدرة الفرد باستخدام هذه الطريقة يتطلب إيجاد متوسط صعوبة المفردات التي تم انتقاها بما في ذلك صعوبة المفردة الأخيرة - أي المفردة التي ترتيبها $(n+1)$ - التي انتقاها الحاسوب، ولكن لم يتمكن الفرد من الإجابة عنها. وتقدر صعوبات هذه المفردات باستخدام أحد نماذج الاستجابة للمفردة الاختبارية (IRMs)، وبخاصة نموذج راش Rasch Model.

غير أن طرق تصميم وبناء هذه الاختبارات لا تزال موضع دراسات وبحوث سيكومترية متخصصة، حيث إن نظام تفرييد قياس التحصيل وما يتطلبه من برامج حاسوب مبتكرة لا يزال في مراحل تطوره الأولى. فهناك العديد من المشكلات الفنية المتعلقة بتنفيذها على نطاق واسع، ويسعى هؤلاء العلماء إلى التغلب عليها لكي يصبح مألفاً لدى المشغلي بالقياس والمربيين في المستقبل القريب.

ومن أهم هذه المشكلات في الوقت الحاضر الكلفة الكبيرة نسبياً التي يتطلبها تنفيذ هذا النظام، وبخاصة ما يتعلق منها بفنون تجهيز عتاد الحاسوب الآلي Hardware وإعداد البرامج المساعدة على تنفيذه Software.

ومن المشكلات المهمة أيضاً التي تشغّل فكر علماء القياس في هذا المجال كيفية تطوير الأسئلة التي تتطلب إجابات مفتوحة وتخزينها في الحاسوب من أجل التخفيف من سيطرة مفردات الاختبار من متعدد على الاختبارات التحصيلية في العقود الماضية، أو إعادة النظر في طريقة بناء وتصحيح هذه المفردات بحيث تقدم معلومات أكثر تفصيلاً عن تحصيل الطلاب. ومن المتوقع مستقبلاً تطوير أنظمة آلية ماهرة Expert systems ذات إمكانية تجهيز باللغات الطبيعية يمكنها تحليل القطع الشريحة وتقدير درجاتها تقديراً ذكياً وأقل كلفة مما هو عليه الآن.

وعندما يتم التغلب على هذه المشكلات فإن هذا النظام سوف يساعد على جعل عملية قياس التحصيل الدراسي أكثر مرونة وفاعلية ودقة بشرط توافر بنك للأسئلة تم تعيير مفرداته Calibrated Items وفقاً لأحد نماذج الاستجابة للمفردة الاختبارية كما سنوضح بعد قليل.

ومع هذا فإن هناك برامج حاسوب متاحة في وقتنا الحاضر لعل أكثرها استخداماً برنامج MicroCAT الذي يعد نظاماً اختبارياً يتم من خلال الحاسوب.

ويمكن للبرنامج أن يكون بنوك أسئلة تشتمل على مفردات نصية Text، وبيانية Graphics وصور متحركة Scanned Images، ويُشكّل من هذه البنوك اختبارات بمواصفات محددة ويعطّلها فردياً بما يتناسب مع مستوى قدرة الفرد الذي يختبر به.

كما يجري البرنامج تحليل مفردات الاختبارات باستخدام النظرية الكلاسيّة ونظرية الاستجابة للمفردة الاختبارية. ويمكن الحصول على البرنامج كنظام متكامل أو كوحدات نسقية منفصلة مثل بنوك الأسئلة، وتفريد الاختبارات، وتحليل المفردات.

(ثالثاً) بناء نظم بنوك الأسئلة : Item Banking Systems

إن نجاح علماء القياس التربوي والنفسى في تصميم تكنولوجيا تساعد على بناء نظام الاختبارات المفصلة Tailored Testing ، أدى إلى زيادة الاهتمام ببناء نظم بنوك الأسئلة. والمقصود بينك الأسئلة هو مجموعة من مفردات الاختبار يكون لها خصائص سيكوفترية مميزة ومعلومة. ويجب أن تودع هذه المفردات في البنك بطريقة تيسّر على المربين سحب مجموعات منها لاستخدامها في بناء اختبارات تناسب الأغراض المختلفة

للقیاس والتقویم التربوی. فتقویم التقدم الدراسي للطلاب يبدأ بالبحث عن أسلوب لقیاس متغيرات القدرات المعرفیة لدى الطلاب. وعلى الرغم من أن معظم نظم التقویم السائدة تعتمد في قیاس هذه المتغيرات على إیجاد العدد الكلی لمفردات الاختبار التي يجب عنها کل طالب إجابة صحيحة، إلا أن هذا الأسلوب يقوم على أساس افتراضات يشوبها الخطأ كما سبق أن أوضحتنا في الفصل السابق.

فمثلاً يفترض أنه يمكن جمع درجات جميع المفردات للحصول على الدرجة الكلية لمتغير واحد، وأنه يمكن الموازنة بين درجات الطلاب في الاختبار، على الرغم من أن مفردات الاختبار التي تمثل المتغير المراد قیاسه ربما لا تحمل المعنی نفسه بالنسبة لكل منهم، وكذلك يفترض أن درجات الاختبار تمثل القدرات تمثيلاً خطياً.

ويعتمد استخدام الدرجات الخام في تعریف «المیزان Scale» الذي يتم على أساسه الموازنة بين الطلاب في عینة المفردات التي استخدمت في القیاس. وإذا حولت الدرجات الخام إلى درجات موزونة، من مثل الدرجات المعيارية أو الثانية، للتغلب على هذه المشكلة، تنشأ مشكلة أخرى وهي اعتماد هذا المیزان على عینة الطلاب المستخدمة في إجراء هذا التحويل.

فالدرجات الخام تختلف عن كل من تدریج المفردات Items Calibration وقیاس قدرة الأفراد Person Measures، لأن هذه الدرجات تكون مقصورة على عینة الأفراد التي يتم اختبارها، في حين أن تدریج صعوبة المفردات يجب أن يكون مستقلًا عن خصائص العینة Sample-Freed. كما أن هذه الدرجات تقتصر على الاختبار الذي استمدت منه، في حين أن قیاس القدرة يجب أن يتصرف بالعمومية بالنسبة إلى المتغير الذي تقیسه المفردات التي تم تدریج صعوبتها.

وقد رأينا فيما سبق أن نماذج الاستجابة للمفردة الاختبارية وبخاصة نموذج راش يفيد في تحويل الدرجات الخام إلى تقدیرات Estimates لموقع الأفراد والمفردات على میزان لوغاریتمي خطی بحيث تكون هذه التقدیرات مستقلة عن كل من خصائص الاختبار، وخصائص عینة الأفراد. لذلك حاول علماء القياس النفسي والتربوی المعاصر بناء بنوك الأسئلة بالاستعانة بهذا النموذج.

فيعد الانتهاء من تعییر المفردات Items Calibration واختبار حسن مطابقتها Items Difficulties لنموذج راش، يصبح من السهل توحید میزان صعوبة المفردات

Scale التي تم تقديرها من عينات مختلفة من الأفراد، وتودع هذه المجموعة من المفردات في بنك الأسئلة. ويمكن عندئذ أن تسحب أي مجموعات جزئية من هذه المفردات لتكوين اختبارات، وإجراء قياسات متكافئة.

والطريقة التقليدية لجعل اختبارين متكافئين، هي أن نطبق كلاً منها على عينة من الأفراد، ونحدد الدرجات المتكافية Equivalent Scores ، أي الدرجات التي تناظر رتبة مئوية واحدة. والطريقة الأخرى هي أن نطبق كلاً من الاختبارين على عينة من الأفراد، ونستخدم نموذج راش في الحصول على تدريج أو ميزان مشترك Cimmon Scale للمفردات. وهذا يعني أن القياسات الناتجة عن استخدام أي من الاختبارين تكون متكافية إحصائياً.

ويمكن ضم الاختبارين معاً ليصبحا اختباراً واحداً، ثم يطبق على عينة الأفراد، واستخدام نتائج الاختبار ككل في الحصول على التدريج المشترك. ولكن هذه الطريقة تصبح غير عملية إذا كنا بقصد بناء بنك للأسئلة يشتمل على عدد كبير من المفردات التي تقيس سمة كامنة أو قدرة معينة. فعندئذ يمكن تكوين اختبارات مختلفة من هذه المفردات، بحيث يشتراك كل اختبارين منها في عدد معين من المفردات يتراوح بين ١٠ ، ٢٠ مفردة. ثم نطبق كلاً من هذه الاختبارات على عينة مختلفة من الأفراد، بحيث يُجبِّب كل فرد عن مفردات اختبار واحد فقط خلال شبكة من الوصلات بين المفردات Network of Links .

ولتوسيع ذلك، نفترض أن لدينا اختبارين يشتمل كل منهما على ٦٠ مفردة، وأردنا أن نجعل هذين الاختبارين متكافئين. فإذا طرق إجراء، ذلك هي أن نطبق كلاً منهما على عينة تتكون من ١٢٠ فرد مثلاً، ونحسب المئويات المترادفة لكل درجة، ثم نحدد الدرجات التي تناظر رتبة مئوية واحدة.

ويمكن إجراء ذلك بطريقة أخرى تعتمد على بيانات أقل، وذلك بتكوين اختبار ثالث يشتمل على ٦٠ مفردة مأخوذة من الاختبارين الأصليين (٣٠ مفردة من كل). ثم نطبق كلاً من هذه الاختبارات الثلاثة على عينة مختلفة تتكون كل منها من ٤٠ فرد. وبذلك تكون قد اختبرنا كل فرد من أفراد العينة الكلية (١٢٠ فرد) باختبار واحد، ومع هذا يمكن الحصول على تدريج مشترك لجميع المفردات (١٢٠ مفردة) من خلال وصلتي المفردات Two Links (عدد مفردات كل منها = ٣٠) التي تربط بين الاختبارات الثلاثة. ويمكن استخدام نموذج راش في الحصول على تقديرين منفصلين

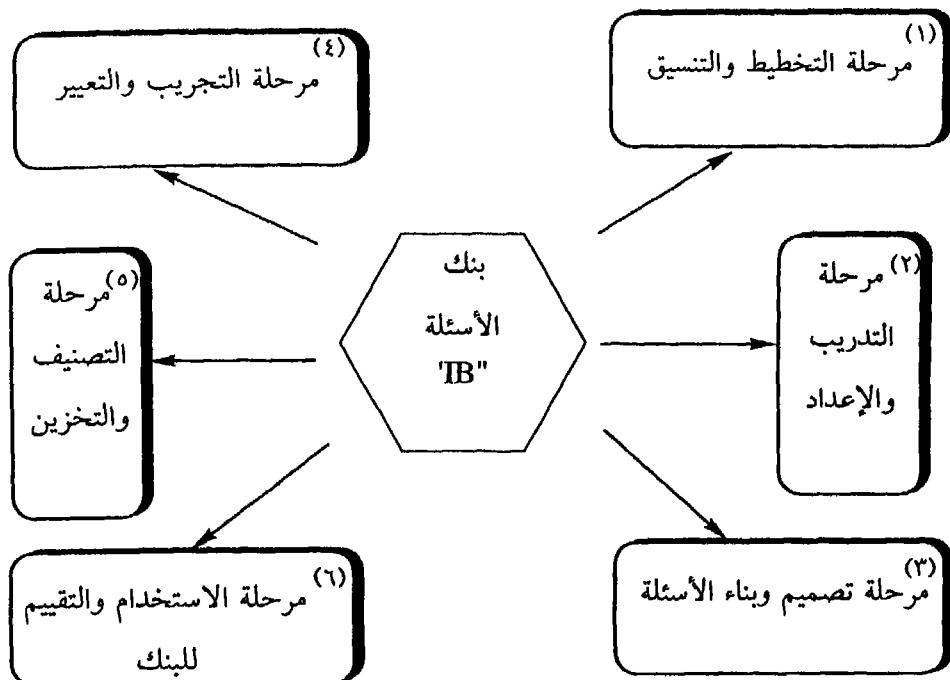
ومستقلين لصعوبة كل مفردة مشتركة بين الاختبارين الأصليين. وكل تقديرین يكونان متكافئین إحصائیاً، ولكنهما يختلفان عن غيرهما بمقدار ثابت يسمى ثابت الانتقال Constant of Translation، ويكون لجميع المفردات التي يشتمل عليها كل من الاختبارين الأصليين المقدار الثابت نفسه. وهذا يمكننا من تقدير ثابت الانتقال بسهولة وإضافته إلى جميع القيم التقديرية لصعوبة مفردات أحد الاختبارين، وبذلك يتم توحيد تدريج ميزان صعوبة كل اختبارين من خلال المفردات المشتركة بينهما.

كما يمكن باستخدام نموذج راش تقدير حسن تكافؤ الاختبارين Goodness of Equating.

وبالطبع تنشأ بعض المشكلات أثناء إجراء هذا التحليل تتعلق بكيفية انتقاء المفردات المشتركة وعدد هذه المفردات، وكيفية التحليل إذا لم يكن الفرق بين القيم التقديرية لصعوبة المفردات المشتركة ثابتاً، ومدى تأثير ذلك على تقدير قدرة الأفراد باستخدام الاختبارات المختلفة التي تشتمل على مجموعات متباينة من مفردات أو أسئلة البنك. هذا بالإضافة إلى ما يتطلبه إجراء هذا التحليل من عمليات إحصائية متعددة، وتتوفر معدات حاسوب Hardware، وبرمجيات جاهزة مساعدة Software لتنظيم شبكات بنوك الأسئلة، وإدارة العملية الاختبارية بأسلوب ييسر توظيف هذه الشبكات توظيفاً فاعلاً.

مراحل إنشاء نظام بنك أسئلة المقترن:

نظراً لأهمية إنشاء نظام بنك أسئلة يسهم في تطوير أساليب قياس وتقديم التحصيل الدراسي. فسوف نقدم تصميمياً يقترحه المؤلف لبناء نظام بنكى للأسئلة متعدد الأبعاد والإجراءات. ويوضح الشكل التخطيطي (١-١٥) التالي المراحل التي يمر بها إنشاء هذا النظام البنكي المقترن:



شكل (١-١٥) مخطط مقترح لمراحل تأسيس وإنشاء نظام بنك الأسئلة

يتضح من شكل (١-١٥) أن نظام بنك الأسئلة المقترن يمر بست مراحل هي:

(١) مرحلة التخطيط والتنسيق:

نظام بنك الأسئلة يتطلب وضع خطة متكاملة ومحكمة قبل البدء في تنفيذ المشروع، كما يتطلب تخطيطاً وتنسيقاً لتحديد المسؤوليات والمهام والإسهامات الواجبة في المشروع. كذلك يتطلب تهيئة المسؤولين الإداريين والمربين والمتعلمين والجمهور بهذه النقلة النوعية في نظم وأساليب التقويم والإمتحانات من خلال اللقاءات والمشاركات.

(٢) مرحلة التدريب والإعداد:

فنظراً إلى أن المشروع يعتمد بدرجة أساسية على توافر كفايات على درجة عالية من التخصص والخبرة فإنه ينبغي وضع خطة تنظيمية لتدريب الأطر الفنية التي ستوكيل إليها مهام تحديد الأهداف والكفايات التعليمية وصياغتها بطريقة تسمح بقياس نواتج

التعلم، وكتابة مختلف أنواع الأسئلة والمفردات الاختبارية التي تقيس الأهداف التي يتم تحديدها. وينبغي التأكيد من امتلاكهم القدرة والمهارة الفائقة في صياغة هذه الأسئلة والمفردات. وبخاصة تلك التي تقيس المستويات المعرفية العليا سواء الموضوعية منها أو المقالية.

(٣) مرحلة تصميم الأسئلة وبناؤها

فالأسئلة والمفردات الاختبارية هي المادة الخام التي تستخدم في بناء الاختبارات، لذلك فإن مرحلة تصميم هذه الأسئلة وبناؤها ومراجعتها مراجعة متأنية يعد من أدق المراحل وأهمها على الإطلاق. لذلك ينبغي أن يقوم بهذه المرحلة خبراء متخصصون وعلى أعلى مستوى من الكفاية في بناء مختلف أنواع الأسئلة والمفردات الاختبارية وبخاصة في المستويات المعرفية والمهارية العليا.

(٤) مرحلة التجريب والتعديل

فأهم ما يميز بنك الأسئلة احتواه على رصيد متجدد من أسئلة ومفردات اختبارية ذات خصائص قياسية معلومة. فالاتهاء من كتابة هذه والأسئلة في ضوء أهداف المحتوى الدراسي ومراجعتها لا يعني أنها أصبحت صالحة للإيداع في البنك، إذ ينبغي أن تخضع لعمليات مستمرة من التجريب الميداني على عينات مناسبة من الطلاب، واستخدام الأساليب السيكومترية والإحصائية لتحليل البيانات المتعلقة بكل سؤال أو مفردة لتقييم فاعليتها، من حيث مدى ملائمتها للمجتمع الذي تمثله هذه العينات من المتعلمين، وقدرتها على التمييز بين المستويات التحصيلية المتباينة، وذلك وفق نموذج سيكومترى مناسب.

فنظام بنك الأسئلة لا يعد نظاماً اختبارياً أو نظام قياس ما لم يحقق خاصيتي المرونة والاتساق. ونقصد بالمرونة إمكانية تشكيل اختبارات ذات خصائص محددة وتناسب أغراضًا معينة من مخزون البنك من الأسئلة والمفردات الاختبارية. أما الاتساق فمعنى به توحيد نظام تعيير أو تدريج جميع مفردات البنك بما يسمح بتكوين اختبارات متكافئة يسهل تفسير درجاتها من اختبار إلى آخر، كما يسمح بالموازنة بين أداء فردان في اختبارين مختلفين في درجة صعوبة مفردات كل منهما ولكنهما يقيسان نطاقاً سلوكياً واحداً.

ولعل استخدام أحد نماذج الاستجابة للمفردة الاختبارية، وبخاصة نموذج راش يمكن أن يحقق ذلك إذا أحسن استخدامه.

(٥) مرحلة التصنيف والتحزبين:

فتظيم بنك الأسئلة يعد من الأمور المهمة، إذ من الصعب أن يتحقق البنك الفائدة المرجوة منه دون تنظيم دقيق وإدارة محكمة. فأساليب الإيداع والسحب من رصيد الأسئلة والمفردات الاختبارية، والتطوير المستمر لمخزون البنك، وعمليات الإضافة والحذف من رصيده، وإجراءات مراقبة هذه العمليات ومتابعتها والإشراف عليها والتحكم في حركتها تتطلب جميعها إجراءات تنظيمية دقيقة، ومحاكمات وضوابط فنية، كما تتطلب أن تكون الهيئة الإدارية المسئولة عن بنك الأسئلة ذات كفاءة عالية.

(٦) مرحلة الاستخدام والتقييم للبنك:

فمحك فاعلية نظام بنك الأسئلة يتحدد في ضوء الاستخدام الميداني الفعلى للأسئلة والمفردات الاختبارية المودعة فيه. وما لا شك فيه أن نتائج الاستخدام وما يترتب عليه تكون بمثابة تغذية مرتبة تعين في تقييم فاعليته وإعادة النظر في مكوناته وأسلوب تنظيمه بصفة مستمرة.

إذ يمكن تشكيل اختبارات كاملة بمواصفات محددة من رصيد البنك يمكن على أساسها تقويم تحصيل الطلاب في المجالات الدراسة المختلفة، والإفادة من التطبيق الميداني في المراجعة المستمرة لأسئلة البنك ومفرداته سواء بالإضافة أو الحذف أو إعادة التعيير، بل وربما يؤدي إلى إعادة النظر في المنهج المدرسي وفقاً لما تسفر عنه نتائج الاستخدام.

كذلك ينبغي تنظيم الخدمات الاختبارية التي يمكن أن يقدمها البنك للمعلمين والطلاب، مثل الاختبارات التدريبية أثناء العام الدراسي، وتصحيح الاختبارات الموضوعية، وبخاصة مفردات الاختيار من متعدد، والصواب أو الخطأ، وإعداد دليل متجدد كل عام يوضح للمعلمين والمعدين بنظام البنك مواصفات ما يتضمن عليه من أسئلة ومفردات اختيارية ومجالاتها. وبذلك يسهم هذا النظام البنكي للأسئلة في تطوير عمليات قياس وتقويم التحصيل الدراسي وجعلها أكثر مرونة وفاعلية ودقة، وكذلك جعلها أكثر تقيناً مما هي عليه في إطار الممارسات الحالية، كما يسهم في تزويد المربين ببيانات تتعلق بالتقدم الدراسي للطلاب في وقت قصير. وبذلك يكون البنك مكونة رئيسة من مكونات المنظومة التعليمية، ويسهم إسهاماً فاعلاً في تطويرها، كما تصبح عمليتنا التعليم والتقويم عملية واحدة متكاملة، وهو ما يسعى إليه خبراء القياس المعاصرون.

البنك الدولى للأسئلة، وتقدير التحصيل التربوى من منظور دولى:

يعد مشروع البنك الدولى للأسئلة International Item Bank من التطورات المهمة الأخرى فى مجال تقويم التحصيل التربوى على الصعيد الدولى. وهو أحد المشروعات الدولية التى بدأتها الجمعية الدولية لقياس التحصيل The International Association for the Evaluation of Educational Achievement (IEA) بالتعاون مع مركز دراسات التقويم بجامعة كاليفورنيا Center for the Study of Evaluation (CSE). وقد اشترك فى هذا المشروع الذى بدأ تنفيذه عام ١٩٨٠ عشرون دولة من الدول الأعضاء فى هذه الجمعية الدولية.

واهتم هذا المشروع ببناء مركز دولى منبع وشبكة بنوك أسئلة Item Banks Network تساعد فى تطوير أساليب وعمليات تقويم التحصيل التربوى فى الدول الأعضاء. ورأى المسؤولون فى هذه الجمعية أن بناء بنك دولى للأسئلة متصل بنوك الأسئلة الخاصة بكل دولة سوف يؤدي إلى ابتكار نظم تصنيفية للأسئلة ولخصائصها السيكومترية على موازين دولية موحدة International Common Scales بحيث يسهل على جميع الدول الأعضاء استخدامها، ويتميز بالمرنة الكافية بحيث يتفق والأنظمة التعليمية فى هذه الدول، كما يساعد فى الحصول على بيانات مقارنة عن جوانب تعليمية معينة فى الدول المختلفة.

وقد قامت هذه الجمعية أيضا بسلسلة من الدراسات المقارنة لتطوير مقاييس معيارية للتحصيل التربوى تصف مستوى التحصيل فى مختلف هذه الدول، وتحدد العوامل التى تفسر الاختلافات عبر الثقافى فى التحصيل. فأجريت مثلا دراسة مقارنة فى الرياضيات فى ٢٠ دولة، وجمعت أيضا بيانات تتعلق بالطالب والمعلم وخصائص المدرسة. وكذلك أجريت دراسة مقارنة فى ست مواد دراسية هي: العلوم، والأداب، وفهم المقروء، واللغتين الإنجليزية والفرنسية كلغتين ثانيتين، والتربية الاجتماعية فى ٢٠ دولة بما فى ذلك بعض الدول النامية. ومن المشكلات التى واجهت مثل هذه الدراسات صعوبة بناء اختبارات تحصيل تتناسب جميع الدول على الرغم من سهولة ذلك إلى حد ما فى الرياضيات، وكذلك مشكلة تباين المناهج وما تركز عليه من دولة إلى أخرى، ومشكلة معاينة الطلاب Students Sampling. وقد أوضحت النتائج أن العوامل التى تُسهم فى رفع مستوى التحصيل تباين بتباين الدول. ومن بين هذه العوامل: مستوى تعليم الوالدين، والمستوى الاجتماعى الاقتصادي، ومصادر التعلم المتنوعة المتوفرة. وهذه الدراسات الدولية تفيد بلا شك فى تقديم معلومات ثرية

يمكن استخدامها في ترشيد القرارات المتعلقة بتطوير التعليم دون الحاجة إلى الاستناد إلى درجات كل طالب على حدة.

(رابعاً) التقييم التوثيق الشامل للتحصيل وكفاياته

: Portfolio Assessment

بعد هذا التوجه في تقييم تحصيل الطلاب وكفاياتهم من الاتجاهات المعاصرة التي بروزت في الميدان التربوي في الآونة الأخيرة، ويختلف اختلافاً جوهرياً عن غيره من التوجهات التي أوضحتها في الجزء السابق. ويرى مؤيدوه أن معظم أساليب القياس والتقويم السائدة تنظر إلى الفرد نظرة جزئية تفتقر إلى التكامل. فالاختبارات التي تعتمد على مفردات الاختيار من متعدد أو غيرها من المفردات ترتكز على معرفة الحقائق وحل مشكلات اصطناعية. لذلك يتبنى هذا التوجه مدخلاً جديداً في التقييم يطلق عليه التقييم البديل Alternative Assessment، أو التقييم التوثيق الشامل Portfolio Assessment، أو التقييم الواقعي Authentic Assessment.

وهذا النوع من التقييم يهدف لقياس إمكانات عقلية عليا ويرتكز على عمليات تعلم مهمة يمكن تنسيتها في إطار العمل المدرسي وخارجه ومتابعة تطورها، ويعتمد في ذلك على صحيفة وثائق تتعلق بالطالب Portfolio. ويرى بعض المريين أن هذه الصحيفة الوثائقية تشتمل على جميع الأنشطة أو الأعمال المدرسية للطالب طوال العام الدراسي، وتُستخدم في تسجيل هذه الأعمال. ويرى البعض الآخر أن هذه الصحيفة تشتمل على عينة مماثلة لأعمال الطالب توضح إنجازاته لكي يطلع عليها ولـ أمره ويتعرف نوعية هذه الإنجازات. كذلك يمكن أن تشتمل هذه الصحيفة على مجموعة متنقاة من الأعمال لغرض معين، مثل عينة من كتابات الطالب أو مهاراته في حل المشكلات، أو مجموعة متنقاة من أفضل أعمال الطالب لعرضها على أترانه أو أسرته أو أفراد مجتمعه المحلي.

ولعل هذا يمثل ما يعرضه الفنانون والمصورون من عينات جيدة لأعمالهم، غير أن عينات أعمال الطالب تختلف عن ذلك من حيث الغرض منها وتنوع محتواها وصيغتها وفقاً لمجال الدراسي أو المهارة المرجوة. وقد استخدمت الصحفاء Protfolios بدرجة أكبر في مجالات المهارات اللغوية، مثل الكتابة والقراءة، والمهارات الفنية الرفيعة، وكذلك في الرياضيات، والعلوم الإنسانية.

أما التقييم التوثيقى الشامل Portfolio Assessment فإنه يهدف للتحقق من جودة ونوعية تحصيل الطالب واكتسابه المعارف والمهارات والقدرات والتوجهات العقلية المحددة في عبارات تشير إلى مستويات الاكتساب، وذلك من خلال فحص عينات من أعمال الطالب التي يختارها بإرشاد من معلمه، ويتم هذا الفحص بواسطة فرد يختلف عن المعلم Audience . ويترشد في تقييمه بمحكمات متقد علىها ومرتبطة بواقع المجتمع. وغالباً ترتكز عملية التقييم على شكل الصحائف الوثائقية للطالب Portfolio ومحتها . ولكن استخدام هذه الصحائف في تقييم التحصيل يمكن أن يفيد فائدة كبيرة في تعميق فكر المربين فيما يتعلق بالتدريس والتعلم.

مكونات الصحائف الوثائقية:

تشتمل الصحائف الوثائقية للطالب Portfolios على ثلاثة مكونات (Gredier, 1995; Storms et al., 1996) هي:

(١) سيرة أعمال الفرد Biographies of Works: وهي توثيق لتطور أو إنتاج مشروع رئيس يقوم به الطالب ويمثل أفكار الطالب في المشروع المختار. وسيرة أعمال الفرد تعبر عن مدى عمق جهوده في مجال دراسي معين .

(٢) أعمال متنوعة للطالب Range of Works: وتتضمن الأنشطة والأعمال التي ينجذبها الطالب، وتوضح مدى عمق إنتاجه . فمثلاً يمكن أن تشتمل صحيفية وثائقية لطالب في مجال المهارات اللغوية على مقالات كتبها، أو شعرًا نظمها، أو حواراً مسجلاً داخل الصدف، أو غير ذلك .

(٣) انعكاسات الطالب Student Reflections: فخلال المدة الدراسية يطلب المعلم من الطالب وصف الجوانب المميزة لأعماله وإنتاجه، وتوضيح مدى تقدمه ونموه، والأهداف التي تتطلب التحقيق. كذلك ربما يتطلب منه القيام بدور الناقد المتخصص أو المؤرخ الشخصي . وتفيد هذه المكونة في إثراء تعلم الطالب ونموه .

صيغ الصحائف الوثائقية:

تنوع صيغ الصحائف الوثائقية Portfolio Formats التي أشارت إليها الأدبيات في هذا المجال & (Gredler, 1996; Scanlan & Heiden, 1996; Valencia Calfee, 1991)

ولعل أهمها ما يلى :

(١) **الصيغة المثالية Ideal Format**: وتهدف لمساعدة الطالب ليصبح قادراً على تقييم سيرة دراسته تقييماً واعياً مستبمراً، وذلك لإثراء نوعية تعلمه وتعريف نموه. وتشمل هذه الصيغة سيرة أعمال الطالب، وأعمال متعددة، وانعكاسات الطالب التي تكون مؤشراً لقدرته على تحليل وتقييم أعماله.

(٢) **صيغة العرض Showcase Format**: وتهدف لعرض عينات من أعمال الطالب على الآباء وغيرهم من أفراد المجتمع المحلي. لذلك تشمل هذه الصيغة على مجموعة من أفضل إنتاج الطالب الذي يختاره بنفسه. فانعكاسات الطالب و اختياراته تعد عنصراً أهم من استخدام مقاييس مفنته.

(٣) **صيغة التوثيق Documentation Format** : وتهدف لتقديم سجل منظم لإنتاج الطالب ونتائج التقويم الكمي والكيفي لأعماله. وتشتمل على سجل منظم مستمر لتقديم الطالب إلى جانب بعض انعكاسات الطالب وتقويم المعلم، ونتائج الملاحظات، وقوائم المراجعة، والسجلات القصصية أو الواقعية، واختبارات الأداء.

(٤) **صيغة التقويم Evaluation Format** : وتهدف لتقويم تقرير مقتني عن تحصيل الطالب وإنتجاهه لولي أمره وللمسئولين في المدرسة. وتشتمل على مجموعة من أعمال الطالب التي يختارها المعلم أو المسؤولين في المدرسة أو الإدارة التعليمية وفق محكمات محددة مسبقاً.

(٥) **صيغة صافية Class Format** : وتهدف لإعلام الآباء والمسئولين في المدرسة بأحكام المعلم عن إنجازات كل طالب. وتشتمل على بطاقات تشخص أداء جميع الطلاب في ضوء أهداف المنهج، وملاحظات المعلم وتعليقاته حول أداء كل طالب، وخطة المعلم التعليمية طوال العام الدراسي مع بعض اقتراحاته المتعلقة بتطوير العمل.

ومن هنا يتضح أن الصحف الوثائقية Portfolios تختلف باختلاف الهدف منها الذي يتراوح بين إثراء نوعية عملية التعلم، وتقديم معلومات مفنته للمسئولين في المدرسة أو الإدارة التعليمية عن إنجازات الطلاب في ضوء أهداف المقررات الدراسية. ففي الحالة الأولى يكون التركيز على نمو كل طالب وتقييم الطالب لأدائه وإنجازاته تقييماً ذاتياً وتحليله لمدخلاته، أما في الحالة الثانية فيكون التركيز على كتابة تقارير تتضمن معلومات مفنته تتعلق بمجموعات الطلاب وذلك لتقديمها للمسئولين.

وعلى الرغم من اختلاف أهداف هذه الصحائف إلا أنها تقدم مؤشرات متعددة عن تحصيل الطالب، كما يمكن أن تفيد في تقييم الأداء الذي يتطلب مدى متسعاً من الوسائل والأنماط، وكذلك تقييم أعمال الطالب التي يستغرق إنجازها مدة زمنية معينة. وربما يصعب إجراء ذلك باستخدام الاختبارات التحليلية التقليدية.

ويعتمد نجاح التقييم التوثيقى الشامل Portfolio Assessment على عوامل متعددة؛ من أهمها تطوير مهارات المعلمين وتغيير اتجاهاتهم نحو أساليب التعليم التي ترتكز على الحفظ والتلقين إلى أساليب تركز على النمو الشامل للطالب وإعطائه الثقة في أهمية تحصيله لأدائه وتقديره الذاتي وتوثيق إنجازاته وتتبع نموه بنفسه، ومشاركة المعلمين والأباء والطلاب في لقاءات دورية لمناقشة الصحائف الوثائقية للطلاب، وتصميم إجراءات التقييم التوثيقى الشامل من حيث كيفية جمع المعلومات والبيانات، وتحليل محتوى الصحائف، وتحديد محكمات تقييمها والحكم على نوعيتها، والإفادة من نتائجها. وبذلك تصبح هذه الصحائف الوثائقية Portfolios أداة فاعلة في التعليم والتقويم.

بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية:

إن تحديد ما يجب تقييمه، أي أهداف نظام التقييم التوثيقى الشامل Portfolio Assessment يعد أحد المشكلات التي تواجه مستخدمي هذا النظام. فقد لوحظ من خلال البرامج التي تنفذها بعض الدول المتطرفة في هذا المجال عدم اتفاق المسؤولين حول ما يجب أو يمكن تقييمه في الصحائف الوثائقية Portfolios . لذلك اعتمد تقييم بعض هذه البرامج على الأهداف والمستويات القومية المحددة، ومرشد المناهج، وتوجيهات خبراء المواد الدراسية، واستنطاب عدد محدود من العناصر الشرعية من هذه المصادر المتعددة للاسترشاد بها في التقييم، وعدم الاقتصار على وضع قائمة بالمهارات المرجوة.

ونظراً لأن هناك تعريفات متعددة للصحائف الوثائقية Portfolios ، فإن هناك طرقاً متعددة لتقييمها وفقاً للظروف المحلية وما يناسب القائمين بالتقسيم. ففي بعض الأنظمة تقدر درجة لكل جزء من الصحيفة الوثائقية للطالب، ويوجد متوسط الدرجات للتوصيل إلى تقيير عام. وبهتم بعض آخر بإصدار أحكمات شاملة في مجالات أداء معينة في ضوء محكمات محددة في دليل معلم. وعلى الرغم من اختلاف هذه الأنظمة، إلا أنها تؤكد أهمية الاستناد إلى محكمات أداء متفق عليها مجتمعاً توضح مستويات الأداء المختلفة، وأحياناً تحدد دليل تقيير درجات ممثلة في خمسة أو ستة مستويات. والمشكلة الثانية تتعلق بدور المعلم في نظام الصحائف الوثائقية. فالمعلم يقوم بتصميم

ما يجب أن يتناوله الطالب في صحيفته، وإرشاد الطلاب أثناء قيامهم بتكوين هذه الصحائف، وتنظيم الجدول الزمني لهذه الصحائف في إطار اليوم المدرسي، وكذلك المشاركة في بناء محكّات لتقدير الصحائف نظراً لأنّه أقدر من غيره على تحديد الأداء المناسب والأداء غير المناسب، وتعرف الظروف الأكثر ملاءمة لتنفيذ هذا النظام داخل غرفة الصف، ومراجعة نتائج التقييم. وهذا يتطلّب بالطبع تنمية مهنية معينة للمعلم للقيام بهذه الأدوار الجديدة.

والمشكلة الثالثة تتعلّق بأساليب تقييم صيغ الصحائف الوثائقية الممتوّعة التي أشرنا إليها. فالصيغة التوثيقية والتقويمية والصفية لا تهتمّ بانعكاسات الطالب أو نمو مهارات تحليله للأداء، وإنما تهتمّ بإنجازاته وتحصيله. وأحياناً تستخدم الصيغة التقويمية في تقديم بيانات تتعلّق بنوعية تحصيل الطالب والإسهام في تطوير عملية التعليم داخل الصف المدرسي. غير أنّ هناك تعارضاً إلى حد ما بين أهداف قياس نوعية التحصيل، وتطوير عملية التعليم. فتطوير أداء البرنامج لكي يحقق هدفاً معيناً ربما يقلّل من فاعلية أدائه في تحقيق هدف آخر، وهذا ينعكس على بنية الصحائف الوثائقية.

والمشكلة الرابعة تتعلّق بصدق وثبات درجات أو تقديرات الصحائف. فالتساؤل الأساسي المتعلّق بصدق الصحائف التي توثّق إنجازات الطالب هو التكوينات الفرضية التي يمثلها إنتاج الطالب. ولكن نظراً لاتساع تعريفات نطاق محتوى المجالات الدراسية مما يؤدي إلى تفسيرات متعددة فإنه يصعب الاستناد إلى هذه التعريفات في التتحقق من صدق تقييم هذه الصحائف.

لذلك اتجه بعض الباحثين إلى إيجاد الارتباط بين درجات الصحائف ودرجات الاختبارات المعتادة في المجال نفسه، غير أنّ القيم الناتجة كانت منخفضة.

وفيما يتعلّق بالثبات فقد استخدم بعض الباحثين ثبات تقديرات المحكمين للصحائف، ولكن القيم الناتجة كانت منخفضة أيضاً في بعض المجالات. وربما يرجع ذلك إلى عدم وضوح المصطلحات في نظام تقدير الصحائف Scoring Rubrics، أو تعدد هذا النظام، أو عدم كفاية تدريب القائمين بالتقدير، أو تباين المهام في الصنوف المختلفة، أو اختلاف الظروف التي يتم العمل في ضوئها.

وفيما يلى نقدم مثلاً لبعض التساؤلات في مقابلة لمناقشة الصحيفة الوثائقية للطالب التي تتضمّن مجموعة من أعماله وأنشطته:

- (١) ما العمل الذي تعتز به أكثر من غيره؟
- (٢) ما العمل الذي تود مراجعته بدرجة أكبر، أو تحتاج مزيداً من الوقت لتحسينه؟ ولماذا؟

(٣) ما العمل الذي لست راضيا عنه؟ ولماذا؟

(٤) إذا أتيحت لك فرصة تنفيذ أيّ من هذه الأعمال مرة أخرى، ما التغييرات التي يمكنك إجراءها عليها؟

(٥) أي هذه الأعمال كان تنفيذه أصعب؟ ولماذا؟

(٦) أي هذه الأعمال كان أيسر تنفيذاً؟ ولماذا؟

(٧) أي هذه الأعمال تعلمته منه بدرجة أكبر من غيره؟ ولماذا؟

تساؤلات تقدم موشرات عن علاقة تعليم الطالب بعمله:

(١) أي هذه الأعمال قمت بتنفيذها في جماعة؟

(٢) كيف يختلف العمل في جماعة عن العمل وحده؟

(٣) أي هذه الأعمال يتطلب منك تخطيطاً وتنظيمًا أكثر؟

(٤) كيف استفدت من هذا التخطيط والتنظيم؟

(٥) ماذا تعلمت من هذا التخطيط والتنظيم؟

(٦) هل تتضمن صحيحتك الوثائقية مشروعًا استغرق أسبوعين أو أكثر؟

(٧) ما الاستراتيجية التي اتبعتها لإدارة وقتك وتنظيمه أثناء عملك في المشروع؟

ومن الجدير بالذكر أن تنفيذ متطلبات نظام التقييم التوثيقى الشامل فى الصنف المدرسى Classroom - Based Portfolio Assessment يتطلب تطوير تقنيات الحاسوب فى المدارس، وابتكار صيغ تجديدية لهذا النظام يمكن استدعاؤها بسهولة وتكون محدودة المساحة.

وعموماً، فإن هذا النظام المستحدث يعد أداة فاعلة في توثيق النمو التحصيلي للطالب من خلال وسائل متعددة تشمل: قوائم المراجعة، والملاحظة، والسجل القصصي، وموارين التقدير، وكتابات الطلاب ودراساتهم، والمحاكاة باستخدام الحاسوب، والعروض الشفوية والسمعية والبصرية، والصور، والرسوم، والقصص، والتسجيلات، والمقابلات، والمؤتمرات، وغير ذلك. وثراء وتنوع هذه الأساليب والوسائل يمكننا من تقويم إنجازات الطالب تقويمًا شمولياً Holistic Evaluation في مضمون أكثر واقعية من مضمون التقويم السائد في وقتنا الحاضر الذي يقتصر على الاختبارات التقليدية.

خلاصة

تناولنا في هذا الفصل بعض أهم التوجهات التطبيقية للفياس والتقويم التربوي والنفسى استنادا إلى التطورات المعاصرة التي أوضحتها في الفصل السابق. فقد أسهمت نظرية الاستجابة للمفردة الاختبارية إسهاماً مثمناً في مجالات تطبيقية متعددة عجزت عن مواجهتها النظرية الكلاسيكية للختارات، وقد أوضحتنا بعض المجالات المتعلقة بتطوير أساليب قياس وتقدير التحصيل الدراسي والكافيات وهي: استخدام الاختبارات مرجعية المحك في تقدير التحصيل، وتقنيات تفرييد قياس التحصيل باستخدام الاختبارات المفصلة وبناء نظم بنوك الأسئلة. وقدمنا تصميماً مقترحاً لإنشاء نظام بنكى للأسئلة متعدد الأبعاد والإجراءات استناداً إلى نموذج راش. كما أشرنا إلى مشروعات البنك الدولى للأسئلة، وتقدير التحصيل من منظور دولى. ونظراً لأن هناك توجهاً معاصرًا نحو توسيع نطاق مفهوم التحصيل وارتباطه بالكافيات والأداء المتميز فقد تناولنا مجالاً مستحدثاً أصبح يطبق في العديد من الدول المتقدمة يتعلق بالتقدير التوثيقى الشامل للتحصيل والكافيات Portfolio Assessment ، وأوضحتنا مكونات الصياغات الوثائقية Portfolios وصيغها، وبعض المشكلات المتعلقة بتصميمها وتنفيذها وتقديرها، والإفادة من تقنيات الحاسوب في التغلب على هذه المشكلات.

ومن الجدير بالذكر أن تعدد وتنوع التطورات المعاصرة في مختلف مجالات الفياس والتقويم التربوي والنفسى التي عرضناها في هذا الكتاب يدل على ثراء هذا العلم وارتباطه بالاحتياجات المتزايدة للمجتمعات المتقدمة في هذا القرن، وتأثيره تأثيراً ملحوظاً بالتطورات التقنية التي أسهمت إسهاماً كبيراً في تقدم حركة التنظير السيكومترى والإديومترى، وتطبيقاتها المتعددة في المجالات التربوية والنفسية والاجتماعية.

المراجع

- 1 - Adorno, T. , Frenkel- Bruswik, E. , Levinson, D. and Sanford, R. The Authoritarian Personality. N.Y. : Harper & Row, 1950.
- 2 - Aiken, L. Readings in Educational and Psychological Testing. Boston: Allyn and Bacon, Inc. , 1973.
- 3 - Aiken, L. Psychological Testing and Assessment. (5th Ed.) , N.Y. : Allyn and Bacon, 1986.
- 4 - Airasian, P. & Madaus, G. " Linking Testing and Instruction: Policy Issues" **Journal of Educational Measurement**, 1983,20,2, 103-118.
- 5 - Allam, S. Minimum Measurement Competencies for Classroom Teachers. N.J. : Educational Testing Service , 1980.
- 6 - Allport, G. Personality: A Psychological Interpretation. N.Y. : Holt, Rinehart & Winston, 1937.
- 7 - Anastasi, A. Testing Problems in Perspective. N.J. : ETS, 1965.
- 8 - Anastasi, A. Psychological Testing (4th ed.) . N.Y. : Macmillan Publishers, 1976.
- 9 - Anastasi, A. Fields of Applied Psychology. N.Y. :McGraw - Hill, 1979.
- 10 - Anderson, S. , Ball, S. and Murphy, R. Encyclopedia of Educational Evaluation . San Francisco: Jossey-Bass Publishers, 1975.
- 11 - Apten , M . The Computer Simulation of Behavior . London: Hutchinson & Co. Publishers, 1970.

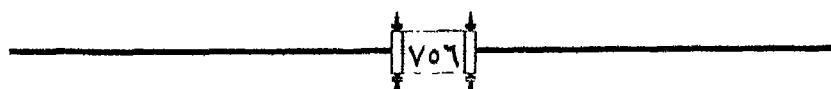
- 12 - Arndt, W. **Theories of Personality**. N.Y. : Macmillan Publishing Co. Inc. , 1974.
- ~~13 - Arter , J . The Portfolios for Assessment and Instruction . ERIC Digest, 1995.~~
- 14 - Assessment Systems Corporation. Computerized Testing Products Catalog. MN: ASC, 1992.
- 15 - Ayala, R. ; Dodd, B. and Koch, W. "A Simulation and Comparison of Flexilevel and Bayesian Computerized Adaptive Testing" **Journal of Educational Measurement**, 1995, 27, - 239.
- 16 - Babbie, E. **Survey Research Methods**. Calif. : Wadsworth Publishing Company , Inc., 1973.
- 17 - Barnette, W. **Readings in Psychological Tests and Measurements** (Revised ed.). Ill. : The Dorsey Press, 1968.
- 18 - Barry, N. & Shannon, D. "Portfolios in Teacher Education: A Matter of Perspective" .**The Educational Forum**, 1997, 61 , 320-328.
- 19 - Bauernfeind, R. **Building a School Testing Program**. Ill. :Educational Testing Service, 1963.
- 20 - Beaton, A. & Zwick, R. "Overview of the National Assessment of Educational Progress". **Journal of Educational Statistics**, 1992, 17,2,95-109.
- 21 - Beatty , W. **Improving Educational Assessment and Inventory Measures of Affective Behavior**. Washington: NEA, 1969.
- 22 - Berdie, D. & Anderson, J. **Questionnaire: Design and Use** . N.J.: Scarecrow, Metuchen, 1974.



- 23 - Berk , R . Consumers' Guide to Criterion- Referenced Test Item Statistics. NCME Publications, 1978, 9, 1.
- 24 - Bertrand, A. & Cebula, J. Tests, Measurement, and Evaluation: A Developmental Approach. Calif. : Addison-Wesley, 1980.
- 25 - Birenbaum, M. & Dochy, F. Alternativs in Assessment of Achievement, :Learning Processes and Prior Knowledge. The Netherlands: Kuwer Academic Publishers, 1995.
- 26 - Birrell, J. & Ross, S. "Standardized Testing and Portfolio Assessment: Rethinking the Debate". **Reading Research and Instruction**, 1996, 35,385-396.
- 27 - Blalock , H. Measurement in the Social Sciences: Theories and Strategies. Ill. : Aldine Publishing Company, 1974.
- 28 - Bloom, B. ; Madaus, G. and Hastings, T. Evaluation to Improve Learning. N.Y. : McGraw-Hill, 1981.
- 29 - Borgatta, E. & Lambert, W. Handbook of Personality Theory and Research. Ill. : Rand McNally & Co. , 1968.
- 30 - Bormuth, J. On the Theory of Achievement Test Items. Chicago, IL.: University of ChicagO Press, 1970.
- 31 - Boud, D. Enhancing Learning Through Self Assessment. London: Kagan Page, 1995.
- 32 - Brennan, R. Some Applications of Generalizability Theory to the Dependability of Domain - Referenced Tests. A Paper Presented at the Annual Johns Hopkins University National Symposium on CRM, 1978.
- 33 - Brennan , R . Handbook for GAPID : A Computer Program for Generalizability Analysis with Single Facet Design. ACT Technical Report. Iowa: ACT, 1979.



- 34 - Brennan , R . Elements of Generalizability Theory . Iowa: ACT Publications, 1983.
- 35 - Brown, F. Principles of Educational and Psychological Testing (2nd ed.) . N.Y. : Holt, Rinehart and Winston, 1976.
- 36 - Burns, E. The Development, Use and Abuse of Educational Tests. Ill.: Charles Tomas, 1979.
- 37 - Burns, R. The Self Concept: Theory, Measurement and Behavior. N.Y.: Longman, 1979.
- 38 - Butcher, H. & Lomax, D. Readings in Human Intelligence. N.Y. : Harper & Row Publishers, Inc. , 1972.
- 39 - Byrne, D. An Introduction to Personality. N.Y. : Prentice. Hall, Inc. , 1966.
- 40 - Carlson , R . & Locklin , R . Item Response Theory: Comparing BILOG and MicroCAT Calibration for a Mathematics Ability Test. MI. : ERIC, 1995.
- 41 - Cartwright, C. & Cartwright, P. Developing Observation Skills. N.Y.: McGraw-Hill, 1974.
- 42 - Cattell, R. Handbook of Modern Personality Theory. Chicago: Aldine, 1970.
- 43 - Cattell, R. ; Eber, H. ; and Tatsuoka, M. Handbook for the Sixteen Personality Factor Questionnaire . Ill . : Institute for Personality and Ability Testing, 1970.
- 44 - Chase, C. Measurement for Educational Evaluation (2nd ed.) . Calif.: Addison-Wesley, 1978.
- 45 - Cherryholmes, C. "Construct Validity and Discourses of Research". **American Journal of Education**, 1988, 96,3, 421-457.



- 46 - Coffman, W. **Frontiers of Educational Measurement and Information System.** Boston: Houghton Mifflin Co., 1973.
- 47 - Cohen, G. **The Psychology of Cognition** . N.Y. : Academic Press, 1977.
- 48 - Collins, H. **Educational Measurement and Evaluation: A Worktext.** Ill.: Scott, Foresman, 1976.
- 49 - Cooke, D. ; Craven, H. & Clarke, G. **Basic Statistical Computing.** Scotland: Thomson Ltd., 1982.
- 50 - Coombs, C. **A Theory of Data.** N.Y. : John Wiley, Inc., 1964.
- 51 - Couch, A. & Keniston, K. "Response Sets and Social Desirability". **Journal of Abnormal and Social Psychology**, 1961, 62,165-169.
- 52 - Cox, R. & Vargas, J. **A Comparison of Item Selection Techniques for Norm-Referenced and Criterion- Referenced Tests .** Paper Presented at the Annual Meeting of the NCME . Chicago, 1966.
- 53 - Crombag, H. & Gruijter, N. **Contemporary Issues in Educational Testing.** Paris: Mouton, 1974.
- 54 - Cronbach, L. **Essentials of Psychological Testing (3 rd ed.)** N.Y. : Harper & Row, Publishers, 1970.
- 55 - Cronbach, L. **Toward Reform of Program Evaluation.** Calif.: Jossey-Bass Publishers, 1980.
- 56 - Cronbach, L. **Five Perspectives On Validity Argument.** In H. Wainer (Ed.) , **Test Validity.** N.Y.: Erlbaum, 1988.
- 57 - Cronbach, L. **Construct Validity After Thirty Years.**In R. Linn (Ed.), **Intelligence : Measurement, Theory and Public Policy.** Urbana: University of Illinois Press, 1989.



- 58 - Cronbach, L. & Drenth, P. Mental Tests and Cultural Adaptation. Netherlands: Mouton Publishers, 1972.
- 59 - Cronbach, L. & Glaser , R . Psychological Tests and Personnal Decisions (2nd ed.) Ill. : University of Illinois Press, 1965.
- 60 - Cronbach, L. & Meehl, P. "Construct Validity in Psychological Tests". **Psychological Bulletin**, 1955, 52.
- 61 - Cronbach, L. ; Glaser, R . ; Nanda, H. and Rajaratnam, N . The Dependability of Behavioral Measurement : Theory of Generalizability for Scores and Profiles. N.Y.: Wiley, 1972.
- 62 - Cropley , A . "Creativity and Intelligence" . **British Journal of Educational Psychology**, 1966, 36, 259 - 266.
- 63 - Crowne, D. The Experimental Study of Personality. N.Y. : Lawrence Erlbaum Publishers, 1979.
- 64 - Crowne , D . & Marlow , D . The Approval Motive: Studies in Evaluative Dependence. N.Y. : John Wiley, Inc., 1964.
- 65 - Dahistrom, W. ; Welsh, G. and Dahistrom, L. An MMPI Handbook: Clinical Interpretation (A Revised ed.) . Minn. : University of Minnesota Press, 1972.
- 66 - D'Angelo, E. The Teaching of Critical Thinking. Amestrdam : Gruner, 1971.
- 67 - Dann, P. ; Irvine S. and Collins, J. Advances in Computer - Based Human Assessment. MA. : Kluwer Academic Publishers, 1991.
- 68 - Davis, I. Competency Based Learning: Technology, Management, and Design . N.Y. : McGraw- Hill, 1973.



- 69 - Dockrell, W. On Intelligence. London: Methuen & Co. Ltd. , 1970.
- 70 - Doner, D. ; Karn , M. and David, A. Students React to Portfolio Assessment. **Contemporary Education**, 1998, 69,3, 159, 165.
- 71 - Dong , T. & Sarti, L. Design and Production of Multimedia and Simulation - Based Learning Material. The Netherlands: Kuwer Academic Publishers, 1994.
- 72 - Dunn. Rankin. P. Scaling Methods. N.Y. : Erlbaum, Hillsdale, 1983.
- 73 - Ebel, R. Essentials of Educational Measurement, (3 rd ed.) N.J. : Engleswood Cliffs, 1979.
- 74 - Edwards, A. Techniques of Attitude Scale Construction. N.Y.: Appleton - Century- Crofts, Inc. 1957.
- 75 - Edwards, A. The Social Desirability Assessment and Research. N.Y. : Dayden, 1957.
- 76 - Edwards, A. The Measurement of Personality Traits By Scales and Inventories. N.Y. : Holt , Rinehart and Winston, 1970.
- 77 - - Edwards, A. & Walker, J. "Social Desirability and Agreement Response Set " **Journal of Abnormal and Social Psychology**, 1961, 62, 180 , 183.
- 78 - ETS The Redesign of Testing For the 21st Century. N.J.: ETS, 1985.
- 79 - Eysenck, H. The Structure of Human Personality (3 rd ed.) , London: Methuen & Co. Ltd., 1970.
- 80 - Eysenck, H. The Measurement of Intelligence. Lancaster: MTP, 1973.
- 81 - Eysenck, H. The Measurement of Personality. Lancaster: MTP, 1976.

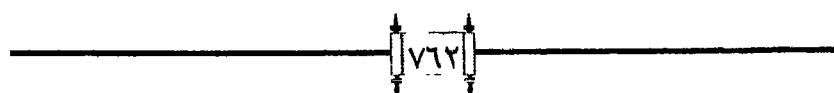


- 82 - Eysenck, H. **The Structure and Measurement of Intelligence** . N.Y.: Springer - Verlage, 1979.
- 83 - Fahmlow, H. & Woodley, K. **Objective - Based Testing: A Guide to Effective Test Development** . N . J . : Educational Technology, 1969.
- 84 - Feder, P. **The Complete Guide to Taking Tests**. N.J. : Prentice- Hall, 1979.
- 85 - Ferguson, L. **Personality Measurement (2 nd ed.)** . N.Y. : McGraw - Hill, 1971.
- 86 - Fincher, J. **Human Intelligence**. N.Y. : Putnam's Sons, 1976.
- 87 - Fine, B. **The Stranglehold of the I.Q.** N.Y. : Doubleday & Co. , Inc. , 1975.
- 88 - Fishbin, M. **Readings in Attitude Theory and Measurement**. N.Y.: John Wiley, 1967.
- 89 - Fiske, D. **Measuring the Concept of Personality**. Chicago: Aldine Publishing Company, 1971.
- 90 - Forgas, R. & Shulman. B. **Personality: A Cognitive View**. N.J. : Prentice- Hall, Inc., 1979.
- 91 - Frahm , R . & Covington , J. **What's Happening in Minimum Competency Testing**. **Phi Delta Kappa**, 1979.
- 92 - Frediksen, N; Mislevy, R. and Bejar, I. **Test Theory for a New Generation of Tests**. N.J. : Erlbaum, 1993.
- 93 - Freeman, J. ; Butcher, H., and Christie, T. **Creativity: A Selective Review of Research**. London: Society for Research into Higher Education Ltd. ; 1968.



- 94 - Fricke, B. Opinion, Attitude and Interest: Survey Handbook, A Guide to Personality and Interest Measurement. MI : OAIS Testing Program, 1965.
- 95 - Gable, R. & Wolf, M. Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporates and School Settings. The Netherlands: Kuwer Academic Publishing, 1994.
- 96 - Gekoski, N. Psychological Testing (2 nd ed.) . Ill: Charles Thomas, 1964.
- 97 - Georgi , D. & Crowe , J . "Digital Portfolios : A Confluence of Portfolio Assessment and Technology". **Teacher Education Quarterly**, 1998, 25, 73 - 84.
- 98 - Getzels, J. & Jakson, P. Creativity and Intelligence, N.Y. :Wiley, 1962.
- 99 - Goldman, L. Using Tests in Counselling (2 nd ed.). N.Y. : Meredith Corporation, 1971.
- 100 - Gorow , F. Better Classroom Testing. Penn. : Chandler Publishing Co. , 1966.
- 101 - Gorth. W. Comprehensive Achievement Monitoring: A Criterion - Referenced Evaluation Technology, 1975.
- 102 - Gredler, M. "Implications of Portfolio Assessment for Program Evaluation" **Studies in Educational Evaluation**, 1995, 21, 431 - 437.
- 103 - Green , B. Adaptive Testing by Computer: New Directions for Testing and Measurement. No. 17, Edited by R. Ekstrom. San Francisco: Jossey- Bass, 1983.

- 104 - Gregg, L. Knowledge and Cognition . N.Y. : John Wiley, 1974.
- 105 - Gronlund, R. Readings in Measurement and Evaluation. N.Y. : The Macmillan Company, 1968.
- 106 - Gronlund, N. Preparing Criterion-Referenced Tests for Classroom Instruction. N.Y. : Macmillan, 1973.
- 107 - Gronlund, N. Measurement and Evaluation in Teaching (4 th ed.). N.Y. : Macmillan , 1981.
- 108 - Gruijter , D . & Kamp . L . Advances in Psychological and Educational Measurement. N.Y.: Wiley & Sons, 1976.
- 109 - Guilford, J. Personality. N.Y. : McGraw- Hill, 1959.
- 110 - Guilford, J. The Nature of Intelligence. N.Y. : McGraw - Hill, 1967.
- 111 - Guilford, J. & Hoepfner, R. The Analysis of Intelligence. N.Y. : McGraw-Hill , 1971.
- 112 - Guion, R. Personnal Testing . N.Y. : McGraw-Hill ,1965.
- 113 - Gulliksen, H. Theory of Mental Tests. N. Y. : Wiley, 1950.
- 114 - Guslkey, T. Implementing Mastery Learning. Calif.: Wadsworth Publishing Company, 1985.
- 115 - Hambleton, R. & Gruijter, D. "Applications of Item Response Models to Criterion- Referenced Test Item Selection". **Journal of Educational Measurement**, 1983, 20 , 4 , 355-367.
- 116 - Hambleton , R . & Zaal , J. Advances in Educational and Psychological Testing: Theory and Applications. MA. : Kluwer Academic Publishers, 1991.
- 117 - Haney, W. "Testing Reasoning and Reasoning About Testing" . **Review of Educational Research**, 1984, 54, 4 , 597-654.



- 118 - Harman , H . Modern Factor Analysis , (2nd ed.). Chicago:
University of Chicago Press, 1967.
- 119 - Harris, C. & Wilcox, R. Achievement Test Items: Methods of
Study. Calif. : CSE, 1977.
- 120 - Henerson, M. ; Morris, L.; and Tayler, C. How to Measure
Attitudes. Calif. : Sage Publications, 1978.
- 121 - Hewitt, G. Shoud We Use Pupils Writing Protfolios to Grade
Schools? They Guide Improvement. **American Teacher**,
1978, 83 , 2 , P. 4.
- 122 - Heywood, J. Assessment in Higher Education. N.Y.: John Wiley,
1977.
- 123 - Hills, J. Measurement and Evaluation in the Classroom . Ohio :
A Bell & Howell Company, 1976.
- 124 - Hively, W. Domain- Referenced Testing . N.J. : Englewood Cliffs,
1974.
- 125 - Holtzman , W . Computer-Assisted Instruction , Testing and
Guidance N.Y.: Harper & Row , Publishers, 1970.
- 126 - Hopkins , K . & Stanley , J. Educational and Psychological
Measurement and Evaluation (6 th ed .) . N . J . :
Prentice-Hall Inc., 1981.
- 127 - Horrocks, J. & Schoonover, T. Measurement for Teachers. Ohio:
Charles Merrill Publishing Company, 1968.
- 128 - Houts , P . The Myth of Measurability: IQ Tests. N.Y.: Hart
Publishing Company, 1977.
- 129 - Hunt, E. & Pellegrino, J. "Using Interactive Computing to Expand
Intelligence Testing : A Critique and Prospects " .
Intelligence, 1982, 9 , 3 , 207 - 230.



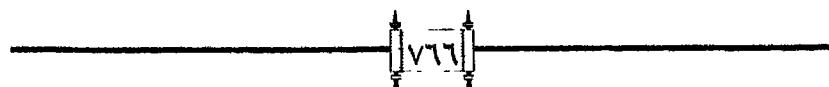
- 130 - Izar , J . Assessing Learning Achievement. Paris: UNESCO,
Document 60 , 1990.
- 131 - Jackson, D. & Messick, S. Problems in Human Assessment. N.Y. :
McGraw-Hill, 1967.
- 132 - Jensema, C. "An Application of Latent Trait Mental Test Theory".
**British Journal of Mathematical and statistical
Psychology.** 1974, 27.
- 133 - Jensen, A. Bias in Mental Testing. N.Y.: Free Press, 1980.
- 134 - Kantowitz , B . Human Information Processing : Tutorials in
Performance and Cognition. N.Y. : John Wiley, 1974.
- 135 - Kaplan, R. Psychological Testing : Principles, Applications, and
Issues. Calif. : Brooks, 1983.
- 136 - Kaufman, A. Intelligence Testing with the WISC-R. N.Y.: Wiley,
1979.
- 137 - Keats, J. An Introduction to Quantitative Psychology. N.Y. : John
Wiley, 1971.
- 138 -Keeves, J. Educational Research, Methodology and Measurement:
An International Handbook. N.Y. : Pergamon, 1997.
- 139 - Kelly, E. Assessment of Human Characteristics. Calif. : Brooks /
Cole Publishers. , 1967.
- 140 - Kelly. G. The Psychology of Personal Constructs. N.Y. : Noron &
Co. , Inc., 1955, 2 Vols.
- 141 - Kimlig, L. & Jarvik, L. Genetics and Intelligence: **A Review
of Science**, 1963 , 142, 1477-1479 .
- 142 - Klahr, D. Cognition and Instruction. N.Y.: Lawerence Erlbaum
Publishers, 1976.



- 143 - Kleinmuntz, B. Personality Measurement : An Introduction. Ill. :
the Dorsey Press, 1967.
- 144 - Kline, P. New Approaches in Psychological Measurement. N.Y. :
John Wiley, 1973.
- 145 - Kline, P. Psychological Testing : The Measurement of Intelligence,
Ability and Personality. N.Y. : Crane Russak, 1976.
- 146 - Kluchnikoff, B. Education and Learning for the 21st Century: A
Priority Agenda. Paris: UNESCO, 1992.
- 147 - Koch, W. & Dodd. B. Operational Characteristics of Adaptive Test-
ing Procedure using Partial Credit Scoring. Paper Presented
at the Annual Meeting of the AERA. San Francisco, 1986.
- 148 - Labaw, P. Advanced Questionnaire Design. Mass. : Abt, 1980.
- 149 - Lachar, D. The MMPI: Clinical Assessment and Automated Inter-
pretation. Calif. : WPS, 1974.
- 150 - Lancaster, O. Effective Teaching and Learning . N.Y. : Gordon and
Breach, 1974.
- 151 - Lazarsfeld, P. The Logical and Mathematical Foundation of Latent
Structure Analysis. In S.A. Stouffer et al., Measurement
and Prediction. Princeton: Princeton University Press,
1950.
- 152 - Lawler , J . , IQ, Heritability and Racism. N.Y. : International
Publishers, 1978.
- 153 - Lesgold, A. ; Pellegrino, J. ; Fokkema, S. and Glaser, R. Cognitive
Psychology and Instruction. N.Y. : Plenum Press, 1977.
- 154 - Lewis, G. Assessment in Education. N.Y.: Wiley, 1975.



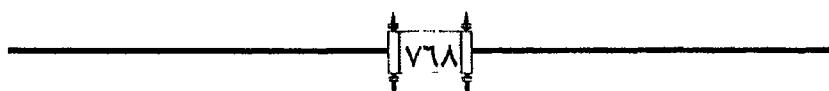
- 155 - Lien, A. Measurement and Evaluation of Learning . Iowa: Brown Company Publishers, 1961.
- 156 - Lindeman, P. & Merenda, P. Educational Measurement, (2 nd ed.). Ill. : Scott Foresman, 1979.
- 157 - Linden, K. & Linden, J. Modern Mental Measurement: A Historical Perspective. MA. : Houghton Mifflin, 1968.
- 158 - Linn, R. Educational Measurement (3 rd ed.). Washington, D. C. : NCME, 1989.
- 159 - Lord, F. & Novick, M. Statistical Theories of Mental Test Scores. Mass. : Addison Wisley, 1968.
- 160 - Lord, F. Applications of Item Response Theory to Practical Testing Problems. N.J. : Hillsdale, 1980.
- 161 - Macintosh, H. Techniques and Problems of Assessment. London: Edward Arnold, 1974.
- 162 - Madaus G. & Stufflebeam, D. Educational Evaluation: Classic Works of Ralph Tylor . MA . : Kluwer Academic Publishers, 1991.
- 163 - Magnusson, D. Test Theory. Pal Alto: Addison-Wesley, 1966.
- 164 - Matarazzo , J. Wechsler's Measurement and Appraisal of Adult Intelligence. Baltimore: The Williams & Wilkins Co. , 1972.
- 165 - McArthur, D. Alternative Approaches to the Assessment of Achievement . MA. : Klwuer Academic Publishers, 1991.
- 166 - McCaughey, S. Using Process-folios with High Ability Students in Visual Arts. **Gifted Child Today Magazine**, 1997 , 20, 40-42.



- 167 - McKinney , M . Preservice Teachers' Electronic Portfolios : Investigating Technology, Self-Assessment, and Reflection. **Teacher Education Quarterly**, 1998, 25, 85-103.
- 168 - McReynolds, P. Advances in Psychological Assessment (Volume1) Calif. : Science and Behavior Books, Inc., 1968.
- 169 - Meehl , P . Clinical Versus Statistical Prediction. Minneapolis: University of Minnesota Press, 1954.
- 170 - Meeker, M. The Structure of Intellect: Its Interpretations and Uses. Ohio: Merrill Publishing Company, 1964.
- 171 - Mehrens , W . & Ebel , R . Principles of Educational and Psychological Measurement. Ill. : Rand McNally& Co. , 1967.
- 172 - Mehrens, W. & Lehman, I. Standaradized Tests in Education, (2 nd ed.). N.Y. : Holt, Rinehart and Winston, 1975.
- 173 - Mehrens , W . & Lehmann , L . Measurement and Evaluation in Education and Psychology (4 th ed.). Fort Worth: Holt, Rinehart & Winston, 1991.
- 174 - Micklo , S . Math Portfolios in the Primary Grades. **Childhood Education**, 1997, 73, 194-199.
- 175 - Miller, K. Psychological Testing in Personnel Assessment . N.Y. : John Wiley, 1975.
- 176 - Miller, D. Handbook of Research Design and Social Measurement (3 rd ed.). N.Y. : David Mckay Co. , Inc. , 1977.
- 177 - Millman , J . & Arter , J . Issues in Item Banking. **Journal of Educational Measurement**, 1984, 21, 4, 315-335.
- 178 - Mills, E. Augmented Title: Submission of Portfolios on a Computer Disk. **International Journal of Instructional Media**, 1997, 24, 23-29.



- 179 - Morris, L. & Tayler, C. How to Measure Achievement. Calif. : Sage Publications, 1978.
- 180 - Moss , P . Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. **Review of Educational Research**, 1992, 62, 3, 229-258.
- 181 - National Computer Sytems (NCS). MICRO TEST SCORE II, User's Guide for the Sentre Plus System. MN.:NCS, 1986.
- 182 - Nettles , M. & Nettles, A. Equity and Excellence in Educational Testing and Assessment . The Netherlands: Kuwer Academic Publishers, 1995.
- 183 - Nitko, A. Educational Tests and Measurement: An Introduction. N.Y. : Harcourt Brace Jovanovich, Inc. , 1983.
- 184 - Noll, V. & Scannell, D. Introduction to Educational Measurement (3 rd ed.). N.Y. : Houghton Mifflin Co. , 1972.
- 185 - Nunnally, J. Psychometric Theory . N.Y. : McGraw-Hill, 1978.
- 186 - Oakland, T. & Hambleton, R. International Perspectives on Academic Assessment.The Netherlands: Kuwer Academic Publishers, 1995.
- 187 - O' Neil, H. Learning Strategies. N.Y. : Academic Press, 1978.
- 188 - Osgood, C. ; Suci, G.; and Tannenbaum, P. The Measurement of Meaning. Ill. : University of Illinois Press, 1957.
- 189 - Osterlind, S. Constructing Test Items . MA. : Kluwer Academic Publishers, 1991.
- 190 - Passow, A. ; Noah , H. ; Eckstein, M. The National Case Study: An Empirical Comparative Study of Twenty- One Educational Systems. N.Y. : John Wiley, 1976.



- 191 - Passow, H . The Gifted and the Talented : Their Education and Development. Ill. : The University of Chicago Press, 1979.
- 192 - Patterson, C . The Wechsler- Bellevue Scales : A Guide for Counselors. Ill. : Charles Thomas Publishing, 1953.
- 193 - Payne, D. The Assessment of Learning: Cognitive and Affective. MA. : D. C. Heath, 1974.
- 194 - Payne, D. Designing Educational Project and Program Evaluation. The Netherlands: Kuwer Academic publishers, 1994.
- 195 - Payne, D . & McMorris , R . Evaluation and Measurement Contributions to Theory and Practice. N.J. : General Learning Press, 1975.
- 196 - Pervin , L . Personality : Theory , Assessment , and Research (2 nd ed.). N.Y. : John Wiley, 1975.
- 197 - Popham, J. Evaluation in Education. Calif. : McCutchan Publishing Corporation, 1974.
- 198 - Popham, J. Criterion-Referenced Measurement . N.J. : Prentice-Hall, 1978.
- 199 - Popham. J. Modern Educational Evaluation. N.J. : Prentice- Hall, 1978.
- 200 - Rasch, G. Probabilistic Models for Some intelligence and Attainment Tests. University of Chicago, 1980.
- 201 - Resnick, L. The Nature of Intelligence. N.J. : Lawrence Erlbaum, 1976.
- 202 - Richard, R. & Andrey , H. Fundamentals of Behavioral Statistics (5th ed.). N.Y. : Addison - Wesley Publishing Company, 1984.
- 203 - Riley. M ; Riley, J. and Toby, J. Scale Analysis: Applications, Theory and Procedures.N.J.:Rutgers University Press, 1954.



- 204 - Rohwer , W . ; Ammon , P. and Cramer, P. Understanding Intellectual Development. Ill. : the Dryden Press, 1974.
- 205 - Roid, G. & Haladyna, T. Technology of Test Item Writing. N.Y. : Academic Press, 1982.
- 206 - Rorrer, L. The Great Response Style Myth. **Psychological Bulletin**, 1965, 63, 129-156.
- 207 - Roskam, E. & Vandenwollenberg M. Recent Developments in Item Response Theory and Applications in the Netherland . Paper presented at the Annual Meeting of the AERA, 1985.
- 208 - Rotter, L. "Generalized expectancies for Internal Versus External Control of Reinforcement". **Psychological Monographs**, 1966, 80, 1.
- 209 - Rushby , N. Computers in the Teaching Process. N.Y. : John Wiley, 1979.
- 210 - Safrit , M. & Wood T. Introduction to Measurement in Physical Education. N.Y.: Mosby, 1995.
- 211 - Salvia, J. & Yasseldy Key, T. : Assessment in Special Remedial Education. Boston: Houghton Mifflin Co. , 1978.
- 212 - Sandven, J. Projectometry. Oslo: Universitesforlaget, 1975.
- 213 - Sax, G. Principles of Educational and Psychological Measurement and Evaluation. Belmont, C.A. : Wadsworth, 1989.
- 214 - Schader , W . New Directions for Testing and Measurement: Measuring Achievement, Progress Over a Decade, San Francisco: Jossey-Bass, 1980.
- 215 - Schroder, H. & Scuedfeld, P. Personality Theory and Information Processing . N.Y. : The Ronald Press Co. , 1971.



- 216 - Semoenoff, B. Projective Techniques N.Y. : John Wiley, 1976.
- 217 - Shavelson, R. & Webb, N. Generalizability Theory: 1973-1980 .
British Journal of Mathematical and Statistical Psychology, 1981.
- 218 - Shaw, M. & Wright, J. Scales for the Measurement of Attitudes. N.Y. : McGraw-Hill, 1967.
- 219 - Shontz , F. Research Methods in Personality. N.Y. : Appleton - Century - Crofts, 1965.
- 220 - Sleeman, D. & Brown, B. Intelligent Tutoring Systems. N.Y. : Academic Press, 1982.
- 221 - Smith, L. & Klass, C. Assessing the Whole child. In : H. Simons and J . Elliott , Rethinking Appraisal and Assessment. Milton Keynes: Open University Press, 1989.
- 222 - Smith , P . ; Kandall, L. and Hulin , C . The Measurement of Satisfaction in Work and Retirement: A Strategy for the Study of Attitudes. Ill. : Rand McNally, 1969.
- 223 - Snow , R . "Attitude and Achievement" In New Directions for Testing . N.J. : ETS, 1980.
- 224 - Sonquist, J. & Dunkelbery, W. Survey and Opinion Research. N.J. : Prentice-Hall, Inc. , 1977.
- 225 - Srivastava , M . & Carter , E . An Introduction to Applied Multivariate statistics (3 rd ed.). N.Y. : North Holland, 1983.
- 226 - Stake, R. Advances in Program Evaluation . Greenwich: Jai Press Inc. , 1994.
- 227 - Stanley, J. & George , W. The Gifted and the Creative: A Fifty - Year Perspective . Baltimore : The Johns Hopkins University Press, 1977.



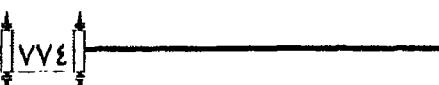
- 228 - Stephenson , W . The Study of Behavior : Q Technique and its Methodology. Ill. : University of Chicago Press, 1953.
- 229 - Sternberg, R. Intelligence, Information Processing, and Analogical Reasoning . N.J. : Lawrence Erlbaum Publishers, 1977.
- 230 - Sternberg, R. & Detterman, D. Human Intelligence: Perspectives on its Theory and Measurement. N.J. : ABLEX Publishing Corporation, 1979.
- 231 - Steven, R. "A Comparison of Item - and Person Fit Methods of Assessing Model - Data Fit in IRT " . **Applied Psychological Measurement** 1990, 14 , 2 , 127 - 137.
- 232 - Stocking, M. & Swanson, L. "A Method for Severely Constrained Item Selection in Adaptive Testing " . **Applied Psychological Measurement**, 1993, 17 , 3 , 277-292.
- 233 - Stone , B. " Problems, Pitfalls, and benefits of Portfolios ". **Treacher Education Quarterly**, 1998 , 25 , 105-114.
- 234 - Storms , B . ; Nuez, A. and Thomas, W. "Using Portfolios to Demonstrate Student Skills". **Thurst for Educational Leadership**, 1996, 25, 6-9.
- 235 - Sudman, S. & Bradburn , N. Asking Questions : A Practical Guide to Questionnaire Design. Calif. : Jossey-Bass, Publishers, 1982.
- 236 - Summers , G . Attitude Measurement. Ill. : Rand McNally & Company, 1971.
- 237 - Sundberg, N. Assessment of Persons. N.J. Prentice-Hall, Inc. , 1977.
- 238 - Tatsuoka, K. Rule Space: An Approach for Dealing With Misconceptions Based on Item Response Theory". **Journal of Educational Measurement**, 1983, 20, 4, 345-353.



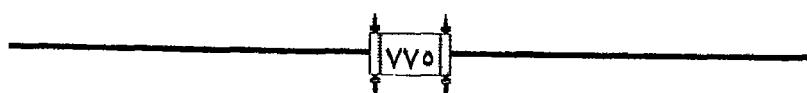
- 239 - Tayler, C. Widening Horizons in Creativity. N.Y. : John Wiley, 1964.
- 240 - Tayler, I. & Getzels, J. Perspectives in Creativity. Ill. : Aldine Publishing Co. , 1975.
- 241 - Taylor, R. Assessment of Exceptional Students (3 rd ed.). Borton: Allyn and Bacon, 1993.
- 242 - TenBrink, T. Evaluation: A Practical Guide for Teachers. N.Y. : McGraw-Hill, 1974.
- 243 - Thorndike , R . Personnel Selection: Test and Measurement Techniques. N.Y. : John Wiley, 1961.
- 244 - Thorndike, R. Educational Measurement (2 nd ed.). Washington, D.C. : American Council on Education, 1971.
- 245 - Thorndike, R. Educational Measurement: Theory and Practice. In: D. Spearritt (Ed.). The Improvement of Measurement in Education and Psychology: Contributions of Latent Trait Theories. Hawthorn: ACER, 1982.
- 246 - Thorndike , R . & Hagen, E. Measurement and Evaluation in Psychology and Education . N.Y. : Wiley & Sons , Inc. 1969.
- 247 - Thurstone , L . The Measurement of Values. Ill. : University of Chicago Press, 1959.
- 248 - Thurstone , L. & Chave, E. The Measurement of Attitude . Ill. : The University of Chicago Press, 1929.
- 249 - Tiedman , H. Fundamentals of Psychological and Educational Measurement . Ill. : Charles Thomas Publisher, 1972.
- 250 - Tierney, R. ; Carter, M. and Desai, L. Portfolio Assessement in Reading - Writing Classroom. M.A. : Gordon, 1991.



- 251 - Toh, P. ; Swaminathan, G. "Application of Item Response Theory in Standard Setting". **Studies in Educational Evaluation**, 1989, 15, 285-293.
- 252 - Torrance, E. Guiding Creative Talent. N. J. : Prentice-Hall, 1962.
- 253 - Travers, R. Second Handbook of Research on Teaching . Ill. : Rand McNally & Company, 1973.
- 254 - Trow , C . & Haddan , E. Psychological Foundations of N.J.: Educational Technology Publications, Inc. , 1976.
- 255 - Tukey , J . Exploratory Data Analysis. Mass . : Addison-Wesley Publishing Company, 1977.
- 256 - Tyler, R. & Wolf, R. Crucial Issues in Testing. Calif. : McCutchan Corporation, 1974.
- 257 - UNESCO, International Conference on Education and Informatics. Paris: UNESCO, 1989.
- 258 - United States Information Agency . "America 2000. The President Educational Strategy". U.S.A. , 1991.
- 259 - Urry, V . A Monte Carlo Investigation of Logistic Test Models . Unpublished Doctoral Dissertation, Princeton Universty, 1970.
- 260 - Vernon, P. Personslity Assessment. London: Methuen & Co. , Ltd. , 1965.
- 261 - Vernon , P. ; Adamson, G. and Vernon , D. The Psychology and Education of Gifted Children. London: Methuen & Co. Ltd. , 1977.
- 262 - Wallerg , H . & Haertel , G. The International Encyclopedia of Educational Evaluation. N.Y. : Pergaman Press, 1990.



- 263 - Walsh, W. & Tyler, L. Tests and Measurements, (3 rd ed.). N. J. : Prentice-Hall, 1979.
- 264 - Wechsler, D. WISC Manual . N.Y. : The Psychological Corporation, 1979.
- 265 - Weiner, E. & Stewart, B. Assessing Individuals: Psychological and Educational Tests and Measurements. Boston: Little, Brown and Company, 1984.
- 266 - Wenzel, L. ; Briggs, K.; and Puryear, B. "Portfolio: Authentic Assessment in the Age of the Curriculum Revolution". **Journal of Nursing Education**, 1998, 37, 5, 208-212.
- 267 - Wenzlaff, T. "Dispositions and Protfolio Development: is there a Connection?" **Education Journal**, 1998, 118, 4, 564-572.
- 268 - Wenzlaff, T. & Cummings, K. " The Portfolio as a Metaphor for Teacher Reflection". **Contemporary Education**, 1996, 67, 109-112.
- 269 - Williams, P. Item Drift in Competency Based Programs. A Paper Presented at the Annual Meeting of the AERA , New York, 1982.
- 270 - Withers, G. & Batten , M. Defining Types of Assessment. In: B. Low and G. Withers (Eds.). Developments in School and Public Assessment. Hawthorn: ACER, 1990.
- 271 - Wolf, D. ; Bixby, J. ; Glenm, J. ; and Gardner, H. "To Use Their Minds Well: Investigating New Forms of Student Assessment". **Review of Research in Education**, 1991, 16, 31-74.



- 272 - Wood, R. Trait Measurement and Item Banks . In Gruijeter, D. & Kamp, L. Advances in Psychological and Educational Measurement . N. J. : Wiley, 1976.
- 273 - Wright, B. & Bell, S. Fair and Useful Testing With Item Banks. Chicago: MESA, Research Memo, 1980.
- 274 - Wright, B. & Douglas, G.. Better Procedures for Sample- Free Item Analysis. Research Memo No. 20 , University of Chicago, 1975.
- 275 - Wright, B. & Stone, M. Best Test Design : A Handbook for Rasch Measurement. Palo-Alto: Scientific Press, 1978.
- 276 - Zytowski, D. Contemporary Approaches to Interest Measurement. Minn. : University of Minnesota Press, 1973



معجم مصطلحات القياس والتقويم (إنجليزي - عربي)

A

Ability	قدرة
Ability Scale	ميزان القدرة
Ability Test	اختبار القدرة
Abscissa	المحور الأفقي / محور السينات
Absolute Zero	الصفر المطلق
Academic Aptitude	استعداد أكاديمي
Accommodation	مواهبة
Accountability	المسؤولية التربوية
Achievement Test	اختبار تحصيل
Acquiescence	القبول / الإذعان
Actuarial Prediction	التنبؤ الإحصائي
Adaptive Tests	اختبارات مواهمة للفرد
Adjustment	تكييف
Aesthetic	متعلق بالذوق
Affect	وجдан
Affective Domain	المجال الوجداني
Age Norms	معايير العمر
Age Scale	ميزان العمر (مقاييس الذكاء)
Alpha Coefficient	معامل ألفا (ثبات)
Alternate Forms	صيغ بديلة / متكافئة (اختبارات)
Ambiguity	غموض
Analogy Test	اختبار متشابهات
Analysis of Variance	تحليل التباين

Analysis of Covariance	تحليل التغاير
Anecdotal Record	سجل قصصي (لحياة الفرد)
Aptitude Test	اختبار استعداد
Arithmetic Mean	متوسط حسابي
Assembly Test	اختبار تجميع أشياء
Assessment	تقدير
Association	اقتران
Asymptote	خط تقاربي
Attenuation	انخفاض قيمة الارتباط بين متغيرين نتيجة عدم دقة عملية القياس
Attitude	اتجاه
Attitude Scale	مقاييس (ميزان) اتجاه
Authentic Assessment	تقدير واقعي
Automated Scoring	تصحيح آلي

B

Basal Age	العمر القاعدي (مقاييس الذكاء)
Basic Skills	مهارات أساسية
Battery of Tests	بطارية اختبارات
Bayes Statistics	إحصاء بيز
Behavioral Domain	نطاق سلوكي
Behavioral Objective	هدف سلوكي
Bias	تحيز
Bimodal Distribution	توزيع ثنائى المtnوال
Binomial Distribution	توزيع ذى الحدين
Bipolar	ثنائى القطب

Biserial Correlation	معامل الارتباط الثنائي المتسلسل
Bivariate Distribution	توزيع ثنائي للمتغير
Branching Model	نموذج تفريعي (اختبارات)

C

Causality	علية / سبية
Checklist	قائمة مراجعة
Chi- Square test	اختبار مربع كاي (إحصاء)
Chronological Age	العمر الزمني
Classification	تصنيف
Clerical Aptitude	الاستعداد الكتابي
Clinical Prediction	التنبؤ الكلินيكي
Cognition	معرفة
Cognitive Processes	عمليات معرفية
Cognitive Psychology	علم النفس المعرفي
Cognitive Styles	أساليب معرفية
Common Scale	میزان مشترك (اختبارات)
Competency - Based Evaluation	تقويم يستند إلى كفايات
Competency Test	اختبار كفاية
Completion Item	مفردة تكملة
Composite Score	درجة مرکبة
Computer Simulation	المحاكاة باستخدام الحاسوب
Computerized Spreadsheet	مصفوفة بيانات الكترونية
Concurrent Validity	الصدق التلازمي
Conditional Probability	الاحتمال المشروط
Conjoint Measurement	القياس المتجدد

Construct Validity	صدق التكوين الفرضي
Content Validity	صدق المحتوى
Convergent Thinking	التفكير التقاريبي (النمطي)
Correction for Guessing	التصحيح من أثر التخمين
Correlation	معامل الارتباط
Cost / Benefit Ratio	نسبة الكلفة إلى العائد
Counseling	إرشاد
Covariance	تغاير
Criterion Groups	مجموعات محاكمة
Criterion - Referenced Test	اختبار مرجعى المحك
Criterion - Related Validity	الصدق المتعلق بمحك
Critical Incidents	أحداث حرجة
Critivity Test	اختبار ابتكار
Cross - Validation	الصدق المستعرض
Crystallized Intelligence	الذكاء المتبلور
Culture - Fair Test	اختبار غير متحيز لثقافة
Culture - Free Test	اختبار متحرر من الثقافة
Cumulative Frequency	تكرار مجتمع
Cumulative Model	نموذج تراكمي (نموذج جتمان)
Cut-off Score	درجة القطع

D

Decile	إعشاري
Decision Studies	دراسات القرارات (اختبارات)
Decision Theory	نظرية القرارات
Dependent Variable	متغير تابع

Derived Score	درجة مشتقة (محولة)
Determination Coefficient	معامل التحديد (إحصاء)
Deterministic Models	نماذج حتمية
Deviation IQ	نسبة الذكاء الانحرافية
Diagnostic Test	اختبار تشخيصي
Dichotomous	ثنائي
Difference Score	درجة الفرق
Differential Psychology	علم النفس الفارق
Differentiation	تمايز
Discrete Variable	متغير منفصل
Discrimination Index	معامل التمييز (اختبارات)
Dispersion	تشتت
Disposition	تهيؤ / استعداد
Distractors	مُشتّتات (اختبارات)
Divergent Thinking	تفكير تباعدي (منطلق)
Domain - Referenced Test	اختبار مرجعي النطاق
Domain - Selected Validity	صدق انتقاء النطاق السلوكي

E

Educational Profile	صفحة تربوية / بروفيل تربوي
Elaboration	تحسين / إحكام
Empirical Validity	الصدق الإيميريقي
Equivalent Form	صيغة مكافئة (اختبارات)
Errors of Estimate	أخطاء التقدير
Errors of Measurement	أخطاء القياس
Error Variance	تباین الخطأ

Essay Test	اختبار مقال
Evaluation	تقويم
Expectancy Table	جدول توقع
Experimental Design	تصميم تجربى
Expressed Interests	میول مُعبَّر عنها
External Criterion	محك خارجي
Extrapolation	استكمال (إحصاء)
Extravert	منبسط / اجتماعي

F

Face Validity	الصدق الظاهري
Facet	وجه / بُعد (اختبارات)
Factors	عوامل
Factor Analysis	التحليل العاملى
Factor Loadings	تشبعات عاملية
Fairness	عدالة
False Negatives	أخطاء سالبة
False Positives	أخطاء موجبة
Feedback	نقدية مرتجدة
Field- Dependent	معتمد على المجال الإدراكي
Figural	متعلق بالأشكال / شكلي
Flexilevel Test	اختبار ذو مستوى مرن
Fluency	طلقة
Fluid Intelligence	الذكاء المائع
Forced- Choice Items	مفردات جبرية الاختيار
Formative Evaluation	تقويم بنائي

Frequency Distribution

توزيع تكراري

Frequency Polygon

مضلع تكراري

Function

دالة / وظيفة

Functional Validity

الصدق الوظيفي

G

General Factor

عامل عام

Generalizability Coefficient

معامل إمكانية التعميم (اختبارات)

Generalizability Studies

دراسات إمكانية التعميم (اختبارات)

Generalizability Theory

نظرية إمكانية التعميم (اختبارات)

Genetic

وراثي

Goal-Free Evaluation

تقويم لا يستند إلى الأهداف

Goodness of Fit

جودة المطابقة (إحصاء)

Grade Norms

معايير الفرق الدراسية

Grade Point Average (GPA)

المعدل العام

Graded Response Model

نموذج الاستجابة المتدرجية

Graphic Rating Scale

ميزان تقييم بياني

Group Factor

عامل طائفي

Group Test

اختبار جماعي

Guessing

تخمين (اختبارات)

Guidance

توجيه

H

Halo Effect

تأثير الظاهرة

Hawthorn Effect

تأثير هاوثورن (تصميمات تجريبية)

Heuristic Models

نمذاج كشفية

Histogram	مدرج تكراري
Homogeneity	تجانس
Horizontal Equating	التكافؤ الأفقي (اختبارات)
Hypothesis	فرض

I

Imagery	تصور (عقلی)
Implicit Objectives	أهداف ضمنية
In- Basket Items	مفردات مواقف تحاکى المهام الإدارية
Incremental Validity	الصدق المتزايد
Independent Variable	متغير مستقل
Individual Test	اختبار فردي
Information Function	دالة المعلومات (اختبارات)
Information Processing	تجهيز المعلومات ومعالجتها
Intelligence	الذكاء
Intelligence Quotient	نسبة الذكاء
Intelligence Test	اختبار ذكاء
Interest Inventory	استبيان ميول
Interlocking Items	مفردات متداخلة (اختبارات)
Internal Consistency	اتساق داخلي (اختبارات)
Interpolation	استكمال من الداخل
Interval Scale	میزان فتری
Interview	مقابلة
Introvert	منطرو
Invariance	عدم التغير
Inventory	استبيان

Inventorized Interests	میول مقاسة بالاستبيان
Ipsative Measurement	قياس مرجعى الفرد
Item	مفردة (اختبارات)
Item Analysis	تحليل المفردات
Item Bank	بنك أسئلة
Item Bias	تحيز المفردة
Items Calibration	تعبير المفردات
Item Characteristic	المنحنى المميز للمفردة
Item Difficulty	صعوبة المفردة
Item Discrimination	تمييز المفردة
Items File	ملف المفردات
Item Forms	صيغ المفردات
Item Generation Rules	قواعد تكوين المفردات (اختبارات)
Item Links	وصلات المفردات
Item Logit	الترجيح اللوغاريتمي للمفردة (اختبارات)
Item Networks	شبكات المفردات
Items Pool	مجموعة من المفردات
Item Response Theory (IRT)	نظرية الاستجابة للمفردة الاختبارية

J

Job Analysis تحليل العمل

K

Kappa Coefficient معامل كابا (إحصاء)

Kuder - Richardson Formula

صيغة كيدر- ريتشاردسون (اختبارات)

L

Latent Class Model

نموذج الفئات الكامنة (نموذج لاررسفلد)

Latent Trait Theory

نظرية السمات الكامنة

Latent Variable

متغير كامن (سمة أو قدرة)

Liniency Error

خطأ متعلق بالتساهل (في التقديرات)

Likelihood Function

دالة الارجحية (إحصاء)

Likert Method

أسلوب ليكرت في قياس الاتجاهات

Linear Regression

انحدار خطى

Local Independence

الاستقلال المركزي (إحصاء)

Local Norms

معايير محلية

Logistic Models

نمذج ترجيح لوغاريتmic

Logit

وحدة ترجيح لوغاريتmic (اختبارات)

M

Machine - Scored Test

اختبار يصحح آلياً

Main Effect

تأثير أساسي (إحصاء)

Mastery Test

اختبار تمكُن

Matching Item

مفردة مزاوجة (اختبارات)

Matrix Sampling	معايير المصفوفة (اختبارات)
Maximum Likelihood	الأرجحية القصوى (إحصاء)
Mean Squares	مجموع المربعات (إحصاء)
Measurement	قياس
Median	الوسيط
Mental Age	العمر العقلى
Meta- Evaluation	التقويم الفوقي
Mode	المنوال
Moderater Variable	متغير وسيط
Monotonic Relationship	علاقة اطرادية
Multidimensional Models	نماذج متعددة الأبعاد
Multidimensional Scaling	موازين متعددة الأبعاد
Multifacet	متعدد الأوجه
Multiple- Choice item	مفردة اختيار من متعدد (اختبارات)
Multiple Correlation Coefficient	معامل الارتباط المتعدد (إحصاء)
Multiple - Factor Theory	نظرية العوامل المتعددة
Multiple - Level Test	اختبار متعدد المستويات
Multitrait - Multimethod Matrix	مصفوفة السمات المتعددة والطرق المتعددة

N

National Norms	معايير قومية
Needs Assessment	تقدير الاحتياجات
Nested Design	تصميم تجربى متشارب (متداخل)
Nominal - Response Model	نموذج الاستجابة الاسمية (التصنifieة)

Nominal Scale	میزان قیاس اسمی (تصنیفی)
Nonlinear Equating	التكافو غیر الخطی (اختبارات)
Nonparametric Statistics	إحصاء لاپارامتری (معلمی)
Nonverbal Test	اختبار غير لفظی
Norm Group	جماعة مرجعیة (معیاریة)
Normal Distribution	التوزیع الاعتدالی
Normalized Scores	درجات مُعدلة اعتدالیًّا
Norm - Referenced Test	اختبار مرجعی الجماعة أو المعيار
Norms	معايير

O

Objectives	أهداف
Objective Test	اختبار موضوعی
Observation Method	أسلوب الملاحظة
Obtained Scores	درجات ملاحظة
Odd - Even Reliability	التجزئة النصفیة (ثبات)
Ogive	منحنی تكراری متجمع
Omnibus Test	اختبار شامل
Oral Test	اختبار شفوی
Ordinal Scale	میزان قیاس رتبی
Ordinate	المحور الرأسی / محور الصادات
Originality	الأصلية

P

Paired Comparison Method	طريقة المزاوجة الثنائيّة
Paper - and Pencil Test	اختبار ورقة وقلم

Parallel Forms	صيغ متوازية (الختبارات)
Paramater	بارامتر (معلم)
Peer Ratings	تقديرات الأقران
Percentiles	مئينيات
Percentile Band	مدى مئيني
Percentile Norms	معايير مئينية
Percentile Rank	رتب مئينية
Perception	إدراك
Performance Contract	عقد أداء
Performance Standard	مستوى أداء
Performance Test	اختبار أداء
Personality	الشخصية
Personality Inventory	استبيان الشخصية
Personality Test	اختبار الشخصية
Phi Coefficient	معامل فاي (إحصاء)
Placement Test	اختبار تسكين
Point - Biserial Corelation	معامل الارتباط الثنائي المتسلسل ال حقيقي (إحصاء)
Point Scale	ميزان قياس يعتمد على النقط
Population	مجتمع
Portfolio	صفحة وثائقية
Portfolio Assessment	التقييم التوثيقى الشامل
Power Test	اختبار قوة
Practice Effect	تأثير التدرب أو الممارسة
Prediction	تنبؤ
Predictive Validity	الصدق التنبؤى
Prior Information	معلومات قبلية

Probabilistic Models	نماذج احتمالية
Profile (Psychograph)	صفحة نفسية (بروفيل نفسي)
Program Evaluation	تقويم البرنامج
Prognosis	تبؤ بنتائج
Progressive Matrices	مصفوفات متتابعة
Projective Techniques	أساليب إسقاطية
Projective Test	اختبار إسقاطي
Projectometry	القياس الإسقاطي
Proximity Data	بيانات متماثلة (مترنة)
Psychomotor	نفسحركى

Q

Q - Sort	التصنيف الترتيبى (Q)
Qualitative	كيفى / نوعى
Quantitative	كمى
Quartile	إرباعى
Questionnaire	استبيان

R

Random Effects Model	نموذج التأثيرات العشوائية (إحصاء)
Random Sample	عينة عشوائية
Range	مدى
Rapport	صلة وثام / أو ألفة
Rasch Model	نموذج راش (النموذج أحادى البارامترا)

Rating Scale	میزان تقدیر
Ratio Scale	میزان قیاس نسبی
Raw Score	درجة حام
Reactive	تفاعلی (راد لتأثير الفعل)
Readiness Test	اختبار تهیؤ
Recall Item	مفردة تتطلب استدعاء (اختبارات)
Recognition Item	مفردة تتطلب تعرف (اختبارات)
Rectangular Distribution	توزيع مستطيل (إحصاء)
Regression Equation	معادلة انحدار (إحصاء)
Regression Line	خط انحدار (إحصاء)
Reliability	الثبات
Reliability Coefficient	معامل الثبات
Representative Sample	عينة مُمثلة
Residuals	باقي
Residual Error	باقي الخطأ (إحصاء)
Response Bias	تحيز الاستجابة
Response Sets	اتجاهات عقلية
Response Styles	أساليب الاستجابات (اختبارات)

S

Sample	عينة
Sampling Theory	نظرية المعاينات
Scaling	عملية تحويل الدرجات إلى موازين
Scatter Diagram	شكل انتشاری (بين متغيرین)
Schema	مخطط عقلي
Scholastic Aptitude Test	اختبار استعداد دراسى

Screening Test	اختبار فحص
Selection	انتقاء
Selection Ratio	نسبة الانتقاء
Self - Concept	مفهوم الذات
Self - Report Inventory	استبيان تقرير ذاتي
Semantic Differential	التمايز اللغوى
Semi - Interquartile Range	نصف المدى الرئيسي (إحصاء)
Sentence Completion Test	اختبار تكميلة الجمل
Sequential Test	اختبارات تتابعية
Situational Test	اختبار مواقف
Skewness	التواء
Social Desirability	مراعاة القبول الاجتماعي
Sociogram	شكل بياني للعلاقات الاجتماعية
Sociometric Matrix	مصفوفة علاقات اجتماعية
Sociometry	القياس الاجتماعي
Software	برمجيات حاسوب
Source Traits	سمات مصدر
Spatial Ability	قدرة مكانية
Spearman-Brown Prophecy Formula	صيغة سبيرمان ويراون (ثبات)
Specific Determineres	محدد نوعى (المفردة اختيار من متعدد)
Speed Test	اختبار سرعة
Spiral Omnibus Test	اختبار حلزوني شامل
Split-Half Reliability	الثبات بطريقة التجزئة النصفية (اختبارات)
Stability Coefficient	معامل الاستقرار (اختبارات)

Standard Deviation	الانحراف المعياري (إحصاء)
Standard Error of Measurement	الخطأ المعياري للقياس
Standard Scores	درجات معيارية
Standardization	تقنين
Standardization Group	عينة التقنين
Standardized Test	اختبار مقنن
Stanine	التساعي المعياري
Static Models	النمذج الاستاتيكية (السكنوية)
Statistic	إحصاء
Stem (of an Item)	العبارة الأساسية (مفردة اختيار من متعدد)
Stem - and - Leaf	شكل الأغصان والأوراق (إحصاء)
Stratified Sample	عينة طبقية
Structure of Intellect	بنية العقل
Subset	اختبار فرعى
Success Ratio	نسبة النجاح
Summated Ratings	تقديرات تجميعية (ميزان ليكرت)
Summative Evaluation	تقدير ختامي
Supply Item	مفردة تتطلب إجابة حرة (اختبارات)
Surface Traits	سمات سطحية
Survey Test	اختبار مسحى

T

T Scores	الدرجات التائية
Tab Test	اختبار مهام متابعة لتشخيص مشكلة واقعية وحلها

Table of Specifications	جدول مواصفات
Tailored Tests	الاختبارات المفصلة
Task Analysis	تحليل المهام
Taxonomy	تصنيف
Teacher - Made Test	اختبار يكتبه المعلم (اختبار صفي)
Test	اختبار
Test Administration	تطبيق (إجراء) الاختبار
Test Anxiety	قلق الاختبار
Tests Networks	شبكات الاختبارات
Test - Retest Reliability	الثبات بطريقة إعادة تطبيق الاختبار
Test Security	سرية الاختبار
Tetrachoric Correlation	معامل الارتباط الرباعي (إحصاء)
Threshold Score	درجة فارقة (اختبارات)
Time Sampling	معاييرة زمنية
Trait	سمة
True - False Item	مفردة صواب أو خطأ (اختبارات)
True Score	درجة حقيقة
True Variance	تبالين حقيقي

U

Unidimensional	أحادي البُعد
Unconditional Probability	احتمال غير مشروط (إحصاء)
Universe	نطاق شامل
Universe Score	درجة شاملة (اختبارات)

V

Validity	الصدق
Variability	تغير
Variance	تباین
Venn Diagram	شكل فن (لتمثيل المجموعات)
Verbal Test	اختبار لفظي
Vertical Equating	التكافؤ الرأسى (اختبارات)
Vocational Interest Test	اختبار ميول مهنية

W

WAIS	مقياس ويكسنر لذكاء الراشدين
WISC	مقياس ويكسنر لذكاء الأطفال
WITS	وحدة ترجيح لوغاريتmic مُعدلة (اختبارات)
Word Association Test	اختبار تداعى الكلمات
WPIS	مقياس ويكسنر لذكاء أطفال ما قبل المدرسة

Z

Z Score	درجة معيارية
---------	--------------

190

محتويات الكتاب

الصفحة	الموضوع
٧-٣	مقدمة
٢٩٨-٩	الباب الأول : أساسيات القياس والتقويم التربوي والنفسى
٤٩-١١	الفصل الأول : المفاهيم والمبادئ الأساسية
١٢	* مقدمة
١٣	* أهمية القياس العلمي
١٣	* المفهوم العلمي للقياس
١٥	* القياس النفسي والتربوي
١٥	* طبيعة القياس النفسي والتربوي
١٨	* أنواع القياس ومستوياته
٢١	* مدخل السمات في القياس التربوي النفسي
٢٢	* بعض الخصائص الأساسية للسمات
٢٥	* أنواع السمات النفسية والتربوية
٢٨	* عمليات القياس وأدوات القياس
٢٨	* الاختبار التربوي النفسي
٢٩	- التقني
٢٩	- الموضوعية
٣٠	- عينة السلوك
٣٠	- معيار أو محك الأداء
٣١	* التقويم والتقييم
٣٣	* استخدامات أدوات القياس والتقويم
٣٧	* تقويم تحصيل الطلاب
٣٨	* أدوار التقويم وأهدافه

الصفحة	الموضوع
٤١	* تقويم المناهج والبرامج والمشروعات التربوية المستحدثة
٤٢	* تقويم المعلم
٤٣	* تقويم المدرسة
٤٤	* تصنيف الاختبارات والمقاييس التربوية والنفسية
٤٩	* خلاصة
٦٩-٥١	الفصل الثاني: مشكلات القياس التربوي والنفسى وشروطه العلمية وأخلاقياته
٥٢	* مقدمة
٥٣	* بعض المشكلات الأساسية في القياس والتقويم في الوطن العربي
٥٦	* بعض الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربوية
٥٩	* الشروط العلمية لانتقاء الاختبارات والمقاييس النفسية والتربوية
٦١	* الشروط العلمية لتطبيق الاختبارات وتصحيحها
٦٣	* الشروط العلمية لتفسيير درجات الاختبارات وتقديم نتائجها
٦٤	* بعض ضوابط وأخلاقيات استخدام الاختبارات والمقاييس النفسية والتربوية
٦٩	* خلاصة
١٢٨-٧١	الفصل الثالث: أساسيات الإحصاء وتوظيفه في القياس والتقويم التربوي والنفسى
٧٢	* مقدمة
٧٣	* أساليب تنظيم وعرض مجموعات من الدرجات

الصفحة	الموضوع
٧٦	* التمثيل البياني لتوزيعات الدرجات
٧٩	* شكل الأغصان والأوراق
٨٤	* استخدام شكل الأغصان والأوراق في المقارنة بين مجموعتين من الدرجات
٨٦	* وصف توزيعات درجات الاختبارات والمقاييس
٨٧	* مقاييس التزعة المركزية أو الموضع
٨٧	- المتوسط الحسابي
٨٨	- الوسيط
٩٠	- المنوال
٩١	* العلاقة بين مقاييس التزعة المركزية أو الموضع
٩٣	* مقاييس التشتت
٩٤	- المدى
٩٤	- الانحراف المعياري
٩٦	- التباين
٩٦	* طريقة تقريبية لإيجاد قيمة الانحراف المعياري
٩٧	* استخدامات الانحراف المعياري
٩٩	- الدرجات المُحولة
٩٩	- الدرجات المعيارية
١٠١	* بعض عيوب الدرجات المعيارية
١٠٢	- الدرجات التائية
١٠٤	* تحديد المركز النسبي للأفراد على المنهج الاعتدالى
١٠٦	* المئنيات
١٠٧	* طرق حساب المئنيات
١١٠	* إيجاد الدرجات المقابلة للمئنيات بيانياً

الصفحة	الموضوع
١١١	* الرتب المئنية
١١٢	* طرق حساب الرتب المئنية
١١٣	* بعض أوجه قصور المئنيات والرتب المئنية
١١٤	* استخدام المنحنى الاعتدالى فى إيجاد المئنيات
١١٥	* مقاييس العلاقة بين مجموعتين من الدرجات
١١٨	* تحديد مقدار العلاقة بين مجموعتين من الدرجات
١١٨	* معامل ارتباط بيرسون
١٢٠	* برنامج الحاسوب الشخصى لإيجاد قيمة معامل ارتباط بيرسون
١٢١	* تفسير قيم معامل ارتباط بيرسون
١٢٢	* اعتبارات أخرى عند تفسير قيم معامل الارتباط
١٢٣	* استخدام معامل الارتباط فى التنبؤ
١٢٤	* معادلة خط انحدار المتغير (ص) على المتغير (س)
١٢٦	* الخطأ المعياري للتنبؤ
١٢٨	* خلاصة
١٨٤-١٢٩	الفصل الرابع : الخصائص السيكومترية للاختبارات أولا: ثبات درجات الاختبارات
١٢٩	* مقدمة
١٣٠	* ثبات درجات الاختبارات
١٣١	* الثبات والدرجات الحقيقية
١٣١	* المفهوم الإحصائى لمعامل الثبات
١٣٣	* العلاقة بين الدرجات الحقيقية والدرجات الملاحظة (مؤشر الثبات)
١٣٤	* مصادر أخطاء القياس
١٣٨	

الصفحة	الموضوع
١٣٩	أولاً : مصادر تتعلق بأداة القياس
١٤١	ثانياً : مصادر تتعلق بإجراءات تطبيق الاختبار وتصحيحه
١٤٢	ثالثاً : مصادر تتعلق بالأفراد المختبرين
١٤٣	* أنواع معاملات ثبات درجات الاختبارات
١٤٥	* كيفية تقدير قيم أنواع معاملات الثبات
١٤٥	- معامل التكافؤ
١٤٨	- معامل الاستقرار
١٥٠	- معامل الاستقرار والتكافؤ
١٥٤	- معامل الاتساق الداخلي
١٦٠	- معامل التجانس لكيودر، وريتشاردسون
١٦٢	* الصيغة (٢٠) لكيودر، وريتشاردسون
١٦٤	* الصيغة (٢١) لكيودر، وريتشاردسون
١٦٥	* معامل (α) لكرولنباك
١٦٧	* ثبات تقديرات المحكمين
١٦٨	* بعض الطرق الأخرى لتقدير الثبات
١٦٩	* ثبات الفرق بين درجتي اختبارين
١٧١	* معامل إمكانية التعميم
١٧٢	* تعقيب عام على طرق تقدير الثبات وتفسير نتائجها
١٧٤	* الخطأ المعياري للقياس
١٧٧	* تفسير الخطأ المعياري للقياس
١٧٩	* عوامل أخرى تؤثر في قيم معامل الثبات
١٨٤	* خلاصة

الصفحة	الموضوع
٢٣٢-١٨٥	الفصل الخامس : الخصائص السيكومترية للاختبارات ثانياً : صدق الاختبارات
١٨٥	* مقدمة
١٨٦	* جوانب صدق الاختبارات
١٨٨	* صدق المحتوى
١٩٠	* طرق تقدير صدق المحتوى
١٩١	* الصدق المرتبط بمحك
١٩٣	* أنواع الصدق المرتبط بمحك
١٩٤	* الصدق التبؤى
١٩٥	* طرق تقدير الصدق التبؤى
١٩٧	- طريقة تعتمد على الارتباط بين الاختبار التبؤى
١٩٧	واختبار المحك
١٩٨	- طريقة تعتمد على استخدام معادلة الانحدار
٢٠٠	* بعض المشكلات المتعلقة بالمحك
٢٠٢	- طريقة تعتمد على جداول التوقع
٢٠٧	- طريقة تعتمد على نظرية المنفعة
٢٠٨	* الصدق التلازمي وطرق تقديره
٢١٢	* بعض العوامل التي تؤثر في الصدق المرتبط بمحك
٢١٤	* صدق التكوين الفرضي (صدق المفهوم)
٢١٧	* طبيعة التكوينات الفرضية
٢١٩	* منهجية التحقق من صدق التكوين الفرضي
٢٢١	* أساليب جمع أدلة تتعلق بصدق التكوين الفرضي
٢٢٢	- أساليب تعتمد على الارتباطات
٢٢٥	- أساليب تعتمد على التجريب

الصفحة	الموضوع
٢٢٧	* أساليب تعتمد على التحليل المنطقي
٢٢٨	* منظور تكاملى لطرق تقدير الصدق والثبات
٢٣١	* خلاصة
٢٦٤-٢٣٣	الفصل السادس : معايير الاختبارات ومحكمات الأداء
٢٣٤	* مقدمة
٢٣٥	* معايير الاختبارات مرجعية الجماعة أو المعيار
٢٣٥	* أنواع الجماعات المرجعية
٢٣٦	- المعايير الوطنية
٢٣٧	- المعايير المحلية
٢٣٨	- معايير مؤسسة أو مدرسة معينة
٢٣٨	- معايير فئات خاصة
٢٣٨	* خصائص عامة للجماعة المرجعية
٢٤٠	* أنواع معايير الاختبارات
٢٤١	أولاً : الدرجات المُحوَّلة
٢٤٢	* الدرجات الثانية الخطية والمعيارية
٢٤٣	* معايير اختبار الاستعداد الأكاديمى
٢٤٤	* التساعيات المعيارية
٢٤٩	ثانياً : معايير المئينيات والرُّتب المئينية
٢٥٢	ثالثاً : معايير الارتفاع
٢٥٢	- معايير العمر الزمنى
٢٥٤	- معايير الفرق الدراسية
٢٥٧	رابعاً: الصفحات النفسية أو التربية (البروفيل)
٢٦١	* محكمات الأداء في الاختبارات مرجعية المحك
٢٦٤	* خلاصة

الصفحة	الموضوع
٢٩٧-٢٦٥	الفصل السابع: تحليل مفردات الاختبارات
٢٦٦	* مقدمة
٢٦٧	* أهمية تحليل مفردات الاختبارات
٢٦٨	* صعوبة مفردات الاختبارات
٢٧٠	* تصحيح قيمة معامل الصعوبة من أثر التخمين
٢٧٤	* أوجه قصور معامل الصعوبة وكيفية تلافيها
٢٧٧	* تمييز مفردات الاختبار
٢٧٨	* أساليب تقييم «تمييز» المفردات استناداً إلى محك داخلي
٢٧٩	* أسلوب يعتمد على معامل الارتباط
٢٨٠	* معامل تمييز المفردات
٢٨٠	أولاً : باستخدام معامل الارتباط الثنائي المتسلسل
٢٨٤	ثانياً : باستخدام درجات المجموعتين الطرفيتين
٢٨٧	* تقييم «تمييز» المفردات استناداً إلى محك خارجي
٢٨٨	* تفسير معامل تمييز المفردات، وعلاقته بمعامل الصعوبة
٢٩١	* تحليل فاعلية بدائل الإجابات
٢٩٢	* استخدام تحليل المفردات في انتقاء مفردات الاختبار
٢٩٣	* استخدام الحاسوب في تحليل مفردات الاختبارات
٢٩٧	* خلاصة
٤٦٤-٢٩٩	الباب الثاني: تطبيقات القياس والتقويم التربوي والنفسى
٣٤٤-٣٠١	أولاً : قياس وتقويم الجوانب المعرفية
٣٠٢	الفصل الثامن : قياس التحصيل والكافيات
٣٠٣	* مقدمة
	* قياس التحصيل الدراسي أو الأكاديمي من منظور تاريخي

الصفحة

الموضوع

٣٠٥	* مفهوم التحصيل الدراسي وعلاقته بمفهوم الاستعداد
٣٠٧	* التمييز بين اختبارات التحصيل واختبارات الاستعدادات
٣٠٩	* افتراضيات أساسية يرتكن إليها قياس التحصيل
٣١٠	* تصنيف اختبارات التحصيل واستخداماتها
٣١٦	أولاً : اختبارات التحصيل مرجعية المحك
٣١٧	* خطوات بناء الاختبارات مرجعية المحك
٣٢٩	* استخدامات الاختبارات مرجعية المحك
٣٣٢	ثانياً : اختبارات التحصيل مرجعية الجماعة أو المعيار
٣٣٤	* خطوات بناء اختبارات التحصيل المقنية
٣٤١	* بعض أوجه النقد التي توجه للاختبارات التحصيلية المقنية
٣٤٤	* خلاصة
٤١١-٣٤٥	الفصل التاسع : قياس الذكاء الإنساني
٣٤٦	* مقدمة
٣٤٦	* قياس الذكاء من منظور تاريخي
٣٤٩	* مفهوم الذكاء والمشكلات المتعلقة بتعريفه
٣٥٠	- الذكاء كمفهوم وصفى
٣٥٤	- الذكاء كمجموعة من القدرات المترابطة
٣٥٤	- الذكاء كعامل عام
٣٥٦	- الذكاء كعامل متعدد
٣٥٩	- الذكاء كبنية ثلاثة الأبعاد
٣٦٥	- الذكاء كنظام هرمي من العوامل
٣٦٦	- الذكاء ومدخل تجهيز المعلومات ومعالجتها
٣٦٩	* قياس الذكاء الإنساني

الصفحة	الموضوع
٣٦٩	* اختبارات الذكاء العام
٣٧٠	أولاً : اختبارات الذكاء الفردية
٣٧٠	* مقياس ستانفورد - بيئي للذكاء
٣٧٤	* تعقيب على مقياس ستانفورد - بيئي
٣٧٥	* مقاييس ويكسنر للذكاء
٣٧٧	* مقياس ويكسنر للذكاء الراشدين
٣٨١	* مقياس ويكسنر للذكاء الأطفال
٣٨٣	* مقياس ويكسنر للذكاء أطفال ما قبل المدرسة
٣٨٤	* مقارنة بين مقاييس ويكسنر ، ومقياس ستانفورد - بيئي
٣٨٦	* بعض مقاييس الذكاء الفردية الأخرى
٣٨٩	ثانياً: اختبارات الذكاء الجماعية
٣٩٠	* أنواع اختبارات الذكاء الجماعية
٣٩٢	* أمثلة لبعض اختبارات الذكاء الجماعية
٣٩٣	* اختبارات ذكاء جماعية لفظية متعددة المستويات
٣٩٦	* اختبارات ذكاء جماعية غير لفظية
٣٩٩	* تعقيب عام على مقاييس الذكاء
٤٠١	* بعض القضايا المتعلقة بقياس الذكاء
٤٠١	- بناء اختبارات الذكاء
٤٠٣	- بعض محددات الذكاء والعوامل المؤثرة فيه
٤٠٧	- ثبوت نسبة الذكاء
٤١٠	* خلاصة
٤٦٤-٤١٣	الفصل العاشر : قياس الاستعدادات الخاصة
٤١٤	* مقدمة
٤١٥	* قياس الاستعدادات الخاصة من منظور تاريخي

الصفحة	الموضوع
٤١٧	* مفهوم الاستعداد وعلاقته بمفهوم القدرة
٤١٨	* تأثير العوامل الوراثية والبيئية في الاستعدادات الخاصة
٤١٩	* استخدامات اختبارات الاستعدادات الخاصة
٤٢١	* أنواع اختبارات الاستعدادات الخاصة
٤٢٤	* بعض الافتراضات الأساسية في اختبارات الاستعدادات الخاصة
٤٢٦	* أساليب بناء اختبارات الاستعدادات الخاصة
٤٢٦	- أساليب تستند إلى التحليل النظري لمكونات الاستعداد
٤٢٦	- أساليب إمبريقية
٤٢٧	- أساليب تستند إلى التحليل العاملى
٤٢٧	- أساليب تستند إلى نتائج تحليل العمل
٤٢٨	* إجراءات تحليل العمل
٤٣١	* اختبارات الاستعدادات الخاصة
٤٣٢	* بطاريات الاستعداد المتعددة
٤٣٥	* اختبارات استعدادات تتعلق ب المجالات نوعية محددة
٤٣٥	* الاستعداد الميكانيكي
٤٣٩	* اختبارات الاستعدادات النفسية الحركية
٤٤٢	* الاستعداد الكتابي
٤٤٢	* اختبارات تقدير الاستعداد الكتابي العام
٤٤٤	* اختبارات تقدير استعدادات كتابية نوعية
٤٤٥	* الاستعداد الموسيقي
٤٤٨	* الاستعدادات الفنية
٤٤٩	أولاً. اختبارات الاستعداد المتعلقة بالحكم الذوقى
٤٥٠	ثانياً. اختبارات الاستعداد المتعلقة بالإنتاج الفنى

الصفحة	الموضوع
٤٥٢	* الابتكارية
٤٥٣	* الابتكارية ومفهوم التفكير التباعدي
٤٥٥	* قياس قدرات التفكير الابتكاري
٤٥٧	* اختبارات التفكير الابتكاري
٤٥٩	* بعض المشكلات المتعلقة بقياس الابتكارية
٤٦٠	- مدى عمومية سمة الابتكارية
٤٦٠	- علاقة اختبارات الابتكاريه باختبارات الذكاء
٤٦٢	- صدق اختبارات الابتكاريه
٤٦٣	- ثبات درجات اختبارات الابتكاريه
٤٦٤	* خلاصة
٦٧٤-٤٦٥	باب الثالث: تطبيقات القياس والتقويم التربوي النفسي
٥١١-٤٦٧	ثانياً : قياس وتقويم الجوانب غير المعرفية (الوجدانية والشخصية)
٤٦٨	الفصل الحادى عشر: قياس الميلو
٤٦٩	* مقدمة
٤٧١	* قياس الميلو من منظور تاريخي
٤٧٢	* الميلو وعلاقتها بالاستعدادات الخاصة
٤٧٣	* الأنماط المختلفة للميلو
٤٧٤	* أساليب بناء استبيانات الميلو
٤٧٥	- أسلوب يستند إلى أسس نظرية أو منطقية
٤٧٦	- أسلوب يستند إلى بناء فقرات متتجانسة
٤٧٧	- أسلوب يستند إلى التحقق الإمبرييفي
٤٧٨	* بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية
٤٧٩	* بعض المشكلات المتعلقة بالاستبيانات المقمنة للميلو
٤٨٠	* مشكلات تفسير درجات استبيانات الميلو

الصفحة

الموضوع

- ٤٨٨ * أساسيات بناء استبيانات الميول المهنية
- ٤٨٩ * قياس الميول باستخدام الاستبيانات المقننة
- ٤٩٠ * استبيانات سترونج - كامبل للميول
- ٤٩٤ * استبيانات كيودر للميول
- ٤٩٧ * استبيان لي - ثروب للميول المهنية
- ٤٩٧ * استبيان جاكسون لمسح الميول المهنية
- ٤٩٨ * استبيان مينيسوتا للميول المهنية
- ٤٩٩ * استبيان أوهايو لمسح الميول المهنية
- ٤٩٩ * استبيان جايست المصور للميول
- ٥٠٠ * تعقيب عام على استبيانات الميول
- ٥٠٢ * استخدامات استبيانات الميول
- ٥٠٥ * تقويم ميول الطلبة في الصف الدراسي
- ٥٠٦ * أدوات قياس ميول الطلبة في الصف الدراسي
- ٥١ * خلاصة
- ٥٧٤_٥١٣ الفصل الثاني عشر : قياس الاتجاهات
- ٥١٤ * مقدمة
- ٥١٥ * قياس الاتجاهات من منظور تاريخي
- ٥١٧ * مفهوم الاتجاه وأهميته
- ٥١٩ * علاقة مفهوم الاتجاه بغيره من المفاهيم المتعلقة به
- ٥٢١ * مكونات الاتجاهات
- ٥٢٣ * خصائص الاتجاهات
- ٥٢٦ * الأسس المنطقية لقياس الاتجاهات
- ٥٢٩ * تصنيف مقاييس الاتجاهات
- ٥٣٢ * أساليب بناء موازين الاتجاهات

الصفحة	الموضوع
٥٣٢	* أسلوب الفترات المتساوية ظاهرياً (ثيرستون)
٥٣٣	* خطوات بناء موازين الاتجاهات باستخدام أسلوب ثيرستون
٥٣٨	* بعض ميزات وعيوب أسلوب الفترات المتساوية ظاهرياً
٥٣٩	* أسلوب التقدير الجماعي (ليكرت)
٥٤١	* خطوات بناء موازين الاتجاهات باستخدام أسلوب ليكرت
٥٤٣	* مزايا وعيوب أسلوب التقدير الجماعي
٥٤٤	* الأسلوب التراكمي (جتمان)
٥٤٦	* خطوات بناء موازين الاتجاهات باستخدام أسلوب جتمان
٥٤٨	* ميزات وعيوب الأسلوب التراكمي
٥٤٩	* بعض الأساليب الخاصة
٥٤٩	* أسلوب تمييز معانى المفاهيم
٥٥	* خطوات أسلوب تمييز معانى المفاهيم
٥٥٢	* مزايا وعيوب أسلوب تمييز معانى المفاهيم
٥٥٣	* أسلوب التصنيف الترتيبى Q Sort
٥٥٤	* أسلوب البسط Unfolding
٥٥٥	* أساليب قياس العلاقات الاجتماعية
٥٥٧	* أساليب بناء موازين الاتجاهات متعددة الأبعاد
٥٥٨	* تعقیب عام على أساليب بناء موازين الاتجاهات
٥٦١	* ثبات وصدق موازين الاتجاهات
٥٦٢	* إرشادات عامة لبناء موازين الاتجاهات
٥٦٣	* أنواع أخرى من مقاييس الاتجاهات
٥٦٤	- أسلوب اختيار الخطأ
٥٦٤	- أسلوب القبول الظاهري

الصفحة	الموضوع
٥٦٥	- أسلوب تقدير استجابات الآخرين
٥٦٥	- أساليب تعتمد على اختبارات الإدراك والذاكرة
٥٦٦	* بعض مشكلات قياس الاتجاهات
٥٦٨	* بعض استخدامات مقاييس الاتجاهات
٥٧٢	* خلاصة
٦٧٤-٥٧٥	الفصل الثالث عشر : قياس الشخصية
٥٧٧	* مقدمة
٥٧٨	* قياس الشخصية من منظور تاريخي
٥٨٠	* مفهوم الشخصية وبعض نظرياتها
٥٨٣	* تطبيق مقاييس الشخصية
٥٨٦	* طرق وأساليب قياس الشخصية
٥٨٧	أولاً : مقاييس محددة البنية وواضحة الهدف (استبيانات الشخصية)
٥٩٠	* استراتيجية بناء استبيانات الشخصية
٥٩٠	* استراتيجية تستند إلى أحکام الخبراء
٥٩١	* استراتيجية تستند إلى الاتساق الداخلي للفقرات
٥٩٢	* استراتيجية تستند إلى أساس إمبريقي
٥٩٢	* أمثلة لبعض استبيانات الشخصية
٥٩٣	* استبيانات تستند إلى أساس نظري منطقي
٥٩٦	* استبيانات تستند إلى التحليل العاملی
٦٠٢	* استبيانات تستند إلى أساس إمبريقي
٦١٢	* استبيانات جبائية الاختيار
٦١٦	* تقييم استبيانات الشخصية
	ثانياً: مقاييس غير محددة البنية وغامضة الهدف

الصفحة	الموضوع
٦٢٢	(الأساليب الإسقاطية)
٦٢٤	* تصنیف الأساليب الإسقاطية
٦٢٧	* أمثلة لبعض الاختبارات الإسقاطية
٦٢٧	* اختبار بقع الحبر لورشاخ
٦٣١	* ثبات وصدق اختبار رورشاخ
٦٣٣	* اختبار تفهم الموضوع
٦٣٥	* ثبات وصدق اختبار تفهم الموضوع
٦٣٦	* مقارنة بين اختبار تفهم الموضوع واختبار رورشاخ
٦٣٧	* تقييم الأساليب الإسقاطية
٦٣٧	* ميزات الأساليب الإسقاطية
٦٣٨	* عيوب الأساليب الإسقاطية
	ثالثاً : مقاييس غير محددة البنية وواضحـة الهدف
٦٣٩	(الملاحظة، والمقابلة، ومقاييس مفهوم الذات)
٦٤٠	أ - أساليب ملاحظة السلوك
٦٤٠	* الملاحظة في مواقف واقعية، وموافق مضبوطة
٦٤١	* مصادر الخطأ في أساليب الملاحظة
٦٤٢	* بعض الاقتراحات للتغلب على أخطاء الملاحظة
٦٤٤	* موازین تقدیر السلوك
٦٤٥	* أنواع موازین تقدیر السلوك
٦٤٥	- موازین تقدیر رقمیة
٦٤٦	- موازین تقدیر بیانیة
٦٤٧	- موازین تقدیر جزئیة
٦٤٩	- موازین تقدیر قوائم مراجعة
٦٥١	• مصادر الخطأ في موازین التقدیر، وكيفیة التغلب عليها

الصفحة	الموضوع
٦٥٣	ب - أساليب المقابلة
٦٥٥	* أنواع أساليب المقابلة
٦٥٥	* المقابلة غير المقتنة
٦٥٦	- المقابلة المقتنة
٦٥٦	- ميزات وعيوب أساليب المقابلة
٦٥٧	* تقييم أساليب الملاحظة والم مقابلة
	رابعاً : مقاييس محددة البنية وغامضة الهدف
٦٥٩	(مقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك)
٦٦٠	* مقاييس الأنشطة الفسيولوجية
٦٦١	* مقاييس الإدراك
٦٦٥	* تقييم الأساليب الموضوعية
٦٦٦	تعمق عام على طرق وأساليب قياس الشخصية
٦٦٨	* استخدامات مقاييس الشخصية
٦٧٢	* خلاصة
	الباب الرابع : توجهات معاصرة في القياس والتقويم التربوي والنفسي
٧٥٢-٧٧٥	الفصل الرابع عشر : بعض التطورات النظرية المعاصرة
٧٢١-٧٧٧	* مقدمة
٧٧٨	.
٧٧٩	أولاً : نظرية الاستجابة للمفردة الاختبارية
٦٨٢	* نظرية الاستجابة للمفردة الاختبارية من منظور تاريخي
٦٨٤	* الأنواع الرئيسية لنماذج الاستجابة للمفردة الاختبارية
٦٨٦	* الأساس المنطقي لنماذج الاستجابة للمفردة الاختبارية
٦٨٦	* تصنيف نماذج الاستجابة للمفردة الاختبارية
٧٩.	* مفهوم المنهجي المميز للمفردة (ICC)

الصفحة	الموضوع
٦٩٣	* النماذج الاستاتيكية اللوغاريتمية أحادية البعد
٦٩٣	- نموذج راش
٦٩٤	- نموذج البارامترن
٦٩٤	- نموذج البارامترات الثلاثة
٦٩٥	* مزايا وعيوب النماذج الاستاتيكية
٦٩٨	ثانياً : نظرية الثبات متعدد الأوجه
٦٩٩	* بعض أوجه قصور المفهوم الكلاسيكي للثبات والصدق
٧٠١	* بعض المفاهيم الأساسية لنظرية إمكانية التعميم
٧٠٣	* دراسات إمكانية التعميم
٧٠٥	* دراسات القرارات
٧٠٧	* تحديد النطاقات الشاملة في دراسات إمكانية التعميم
٧١٠	* تحليل مكونات التباين في دراسات إمكانية التعميم
٧١٢	* العلاقة بين نظرية إمكانية التعميم ونظرية الصدق
٧١٣	ثالثاً : إسهام علم النفس المعرفي في تطوير أساليب القياس والتقويم
٧١٥	* علاقة النظرية المعرفية بالعملية الاختبارية
٧١٦	* تصميم الاختبارات في ضوء النظرية المعرفية
٧٢٠	* خلاصة
٧٥٢-٧٢٣	الفصل الخامس عشر: بعض التوجهات التطبيقية
٧٢٤	* مقدمة
٧٢٤	* تطوير أساليب قياس وتقويم التحصيل الدراسي
٧٢٧	أولاً : استخدام الاختبارات مرجعية المثلث في تقويم التحصيل
٧٣٢	ثانياً : تقنيات تفرييد قياس التحصيل باستخدام الاختبارات المفصلة

الصفحة	الموضوع
٧٣٥	* الاستراتيجية ذات المراحلتين
٧٣٥	* الاستراتيجية متعددة المراحل
٧٣٦	* نموذج التفريغ الثابت
٧٣٦	* نموذج التفريغ المتغير
٧٣٨	ثالثاً : بناء نظم بنوك الأسئلة
٧٤١	* مراحل إنشاء نظام بنك أسئلة مقترن
٧٤٠	* البنك الدولي للأسئلة ، وتقسيم التحصيل التربوي من منظور دولي
٧٤٦	رابعاً: التقسيم التوثيقى الشامل للتحصيل والكافيات
٧٤٧	* مكونات الصحائف الوثائقية Portfolios
٧٤٧	* صيغ الصحائف الوثائقية
٧٤٩	* بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية
٧٥٢	* خلاصة
٧٧٦-٧٥٣	* المراجع ④
٧٩٦-٧٧٧	→ معجم مصطلحات القياس والتقويم (إنجليزى - عربى)
٨١٤	* محتويات الكتاب
٨١٦	* مراجع أخرى للمؤلف
/	

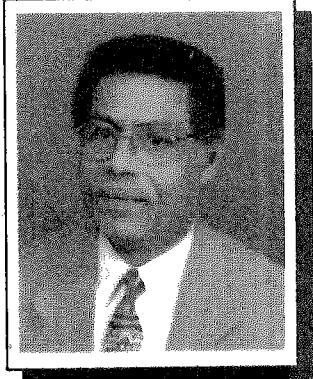
٩٩ / ٢١٠٣	رقم الإيداع
977- 10 -1210-0	I. S. B. N الترقيم الدولي

مراجع أخرى للمؤلف

- (١) اتجاهات معاصرة في تصميم وبناء نظم بنوك الأسئلة.
- (٢) الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريرية «دار الفكر العربي».
- (٣) اختبار تشخيصي مرجعى المحك لقياس مهارات المعلمين في بناء الاختبارات المدرسية.
- (٤) الأساليب الإحصائية الاستدلالية البارامترية واللابارامترية في تحليل بيانات البحوث النفسية والتربوية «دار الفكر العربي».
- (٥) تحليل البيانات في البحوث النفسية والتربوية «دار الفكر العربي».
- (٦) تحليل المحتوى وتقويمه في مرحلة التعليم الأساسي.
- (٧) تطورات معاصرة في القياس النفسي التربوي.
- (٨) تقويم البحوث التربوية في دول المشرق العربي.
- (٩) دليل المعلم في تصميم وبناء الاختبارات التحصيلية الصيفية المعاصرة.
- (١٠) دليل المعلم في تقويم الطلبة في الدراسات الاجتماعية «دار الفكر العربي».
- (١١) مفاهيم وممارسات التعليم الأساسي في بعض الدول غير العربية.
- (١٢) نظم وأساليب التقويم والامتحانات في الدول العربية: الواقع والتوجهات المستقبلية.



هذا الكتاب



- د. صلاح الدين محمود علام
- * بكالوريوس في الرياضيات والتربية من جامعة عين شمس عام ١٩٦٠م.
 - * ماجستير في التربية (علم نفس تعلمي) من جامعة عين شمس عام ١٩٧١م.
 - * دكتوراه الفلسفية في التربية تخصص قياس وتقدير وإحصاء نفسى وتروبى من جامعة بيشجان الأمريكية عام ١٩٨٠م.
 - * أستاذ القياس والتقويم والإحصاء بكلية التربية جامعة الأزهر.
 - * مستشار القياس والتقويم في عدد من المؤسسات والمنظمات التربوية العربية.
 - * صمم وأشرف على تنفيذ العديد من المشروعات التطويرية لنظم وأساليب التقويم والامتحانات في الدول العربية.
 - * أجرى العديد من البحوث والدراسات في المجالات الإحصائية والسيكومترية المعاصرة، وله عدة مؤلفات أخرى في هذه المجالات.
 - * قام بالإشراف على عدد من رسائل الماجستير والدكتوراه في مجالات القياس مرجعى المحك، والنتائج السيكومترية المعاصرة.
 - * عضو الجمعية الأمريكية للبحث التربوى، والمجلس القومى الأمريكى للقياس التربوى، وعضو مؤسس للمجلس العربى للموهوبين والمتوفين.

يعد مرجعاً أساسياً للدارسين والباحثين والممارسين في مختلف مجالات القياس والتقويم التربوى والنفسى، حيث يضم في مجلد واحد الأسس النظرية المنهجية والسيكومترية، والتطبيقات المتعددة، والتطورات والتوجهات المعاصرة المتعلقة بهذه المجالات.

ويشتمل الكتاب على أربعة أبواب رئيسية. يتناول الباب الأول الذي يشتمل على سبعة فصول المفاهيم والمبادئ الأساسية للقياس والتقويم، والخصائص الإحصائية والسيكومترية لل اختبارات والمقاييس، والأساليب المتعددة لم تقدر صدقها وثبات درجاتها، وتحليل معايرها، وتحليل مفرادتها، ومشكلات القياس التربوى والنفسى وشروطه العلمية وأخلاقياته.

ويتناول الباب الثاني الذي يشتمل على ثلاثة فصول منهجيات وأساليب قياس وتقدير الجوانب المعرفية المتممدة في التحصيل الدراسي والكمبيوتر، والذكاء الإنساني، والاستعدادات الخاصة بتنوعها المختلفة.

ويتناول الباب الثالث الذي يشتمل على ثلاثة فصول أيضاً، منهجيات وأساليب قياس وتقدير الجوانب غير المعرفية المتممدة في الميول، والاتجاهات، والشخصية الإنسانية.

ويشتمل كل من هذين البابين على أمثلة متنوعة من الاختبارات والمقاييس الشائعة الاستخدام في قياس الجوانب المعرفية، والنفسية الحركية، والوجدانية، والشخصية.

أما الباب الرابع فيشتمل على فصلين يتناول أحدهما أهم التطورات النظرية المعاصرة في القياس والتقويم التربوى والنفسى، ويتناول الآخر بعض التوجهات التطبيقية المعاصرة التي تستند إلى هذه التطورات النظرية.

يعالج الكتاب الجوانب النظرية والتطبيقية معاً بالجة متكاملة، حيث تستند التطبيقات المختلفة إلى الأطر النظرية المتعددة، ونتائج الدراسات والبحوث المتعلقة بها. كما يوضح المفاهيم وأساليب المختلفة بالعديد من الأشكال التخطيطية لكي يتيسر على القارئ استيعابها واستخدامها في مجالات اهتمامه. ويكتفى الكتاب بمراجعته مختصراً للمصطلحات الأساسية في القياس والتقويم التربوى والنفسى.

تطلب جميع منشوراتنا من وكيلنا الوحيد بالكويت دار الكتاب الحديث