

Aproximación histórica y conceptos básicos de la psicometría

Julio Meneses

PID_00198627

Índice

Introducción.....	5
1. La psicometría en el contexto de la psicología.....	7
1.1. Una aproximación histórica a la psicometría	7
1.2. La psicometría hoy	12
1.3. La psicometría en el contexto de la evaluación psicológica	15
2. Fundamentos de la psicometría.....	18
2.1. Definición y clasificación de los tests	18
2.2. Modelos de medida psicométrica	23
2.3. Teoría clásica de los tests	27
2.4. El proceso de inferencia psicométrica	30
3. Construcción y administración de tests.....	34
3.1. El proceso de construcción de tests	34
3.2. Criterios para la valoración de tests	37
3.3. Aspectos éticos y deontológicos en el uso de tests	40
Bibliografía.....	47

Introducción

Esta Introducción a la psicometría se propone dar al estudiante algunas claves importantes para abordar la complejidad de los conocimientos, procedimientos y valores vinculados con el desarrollo y la administración de tests. Partiendo del encuadre de la psicometría en el contexto general de la psicología, empezaremos desarrollando una aproximación histórica a su nacimiento, desde los antecedentes remotos hasta los primeros desarrollos, describiremos el estatus actual de la psicometría como disciplina científica y, finalmente, construiremos una definición formal teniendo en cuenta el papel que desempeñan los tests como instrumentos de evaluación psicológica. A continuación abordaremos los fundamentos de la psicometría, donde tendremos la oportunidad de ofrecer una definición y clasificación de los tests, abordar los modelos de medida psicométrica, introducir la teoría clásica de los tests y recapitular el proceso de inferencia psicométrica en el que se basan los tests. Finalmente, discutiremos las cuestiones relativas a la construcción y administración de tests, tratando las diferentes fases implicadas en el desarrollo y ofreciendo algunos criterios importantes para la valoración de los tests disponibles en la literatura. Concluiremos esta introducción con una discusión de los aspectos éticos y deontológicos vinculados al uso de tests en la práctica profesional de la psicología.

1. La psicometría en el contexto de la psicología

Entender un campo de estudio determinado implica, en la mayoría de las ocasiones, conocer sus raíces fundacionales. Es decir, conocer los problemas a los que se enfrentaron los pioneros, la manera como los entendían y las soluciones que les dieron. A pesar de que este no es el lugar para hacer un viaje en profundidad¹, una aproximación histórica a la psicometría debe empezar reconociendo lo que se consideran sus orígenes y antecedentes. A pesar de que el desarrollo y la administración de tests psicológicos es una práctica desarrollada fundamentalmente a partir del siglo XX, es posible encontrar algunos antecedentes remotos en culturas tan antiguas como la china. Partiendo de estos antecedentes, y repasando algunas de las contribuciones más importantes que han contribuido a su desarrollo como disciplina científica, estaremos en disposición de ofrecer una definición formal de la psicometría que nos permita situarla en el contexto general de la psicología.

⁽¹⁾Entre otros, podéis consultar los trabajos clásicos de Goodenough (1949), DuBois (1970), Hilgard (1987), y más recientemente, Buchanan y Finch (2005), o Jones y Thissen (2007).

1.1. Una aproximación histórica a la psicometría

Tal y como se suele señalar, el desarrollo de las primeras dinastías del **antiguo imperio chino** generó los primeros sistemas de evaluación de los individuos en función de su habilidad. A pesar de que algunos cronistas han apuntado a referencias tan antiguas como el año 2000 a. de C.², el estudio de las evidencias arqueológicas ha puesto en entredicho esta antigüedad. En cualquier caso, tal y como ha apuntado Bowman (1989), las pruebas documentales que se conservan permiten situar estos orígenes en un periodo relativamente más reciente, durante la dinastía Tang (618-907). Durante estos años se desarrolló un sistema de evaluación imperial que permitió la selección y promoción de los funcionarios de los diferentes departamentos de la Administración. Este sistema tuvo un importante impulso durante la dinastía Ming (1368-1644), cuando se estableció un examen institucional según el mérito para todos los funcionarios de los diferentes niveles territoriales –desde el nivel municipal al nacional–, poniendo en marcha uno de los primeros sistemas de clasificación oficial mediante la expedición de los primeros títulos formales que acreditaban el acceso a los diferentes niveles de responsabilidad.

⁽²⁾Por ejemplo, podéis ver DuBois (1970).

Estos sistemas de evaluación son fácilmente homologables a los que la **educación formal europea**, especialmente la universitaria, desarrolló a partir de la introducción de los exámenes orales a sus estudiantes a partir del siglo XIII. Tal y como señala Rogers (1995), la invención y la incorporación del papel a la vida cotidiana a partir del siglo XVI facilitó el tránsito hacia las pruebas escritas que durante el siglo XIX se convirtieron en los primeros mecanismos de selección competitiva de los estudiantes universitarios. Como antecedentes remotos de la psicometría, tanto la evaluación en la educación formal como el

establecimiento del sistema de evaluación imperial supusieron un importante cambio en la concepción sobre el juicio de las capacidades de las personas. De este modo, se fue trasladando la confianza en el juicio personal basado en impresiones, hacia la administración de pruebas institucionales basadas en una autoridad imparcial, que objetivaba las destrezas y los conocimientos requeridos en los ámbitos educativo y administrativo.

A pesar de que en algunas ocasiones se ha pasado por alto, otro antecedente remoto se encontraría en los inicios de la **evaluación psiquiátrica** a mediados del siglo XIX. De acuerdo con el trabajo de Bondy (1974), es conveniente tener presente el esfuerzo de los primeros profesionales orientados al estudio de los problemas mentales y las lesiones cerebrales en el establecimiento de lo que se podrían considerar las primeras pruebas de evaluación psicológica. Así, por ejemplo, se desarrollaron los primeros tests para evaluar las consecuencias del daño cerebral, alguno de ellos tan elaborado que exigía una administración durante periodos de 100 horas. A pesar de sus limitaciones, entre otras la ausencia de procedimientos estándares en su uso, estas primeras pruebas incorporaron muchos de los elementos que, con la evolución del estudio de los síntomas psicológicos, todavía son utilizados en las pruebas de diagnóstico actuales.

Desarrolladas las bases para el examen individual, tanto en contextos de actividad cotidiana como en situaciones de trastorno psicológico o accidente, los antecedentes recientes de la psicometría se encontrarían en el desarrollo del **estudio sistemático de las diferencias humanas** durante el siglo XIX. Primero gracias a los trabajos de Friedrich W. Bessel (1784-1846) y Carl F. Gauss (1777-1855) –que fueron pioneros en el estudio de las diferencias individuales en la percepción en el campo de la astronomía– y después en las contribuciones de Gustav T. Fechner (1801-1887) y Hermann von Helmholtz (1821-1894) en el desarrollo de la psicofísica –que supuso el inicio de la psicología como disciplina académica–, ambas aproximaciones constituyeron un avance importante en la sistematización de la medida de las sensaciones psicológicas producidas por la estimulación física (Boring, 1978, para una revisión histórica). Así, mediante el desarrollo de las primeras leyes que relacionan estímulos físicos y sensaciones, se fueron asentando las bases para la medida psicológica. De este modo se perfeccionaron los métodos de presentación de los estímulos y el registro de las respuestas, se trabajó en la mejora de la precisión de las medidas y se adoptaron condiciones controladas para su consecución. Todos estos elementos son indispensables para el planteamiento posterior de los problemas más importantes que la psicometría debería afrontar en su desarrollo inicial.

Así, una vez apuntados los antecedentes remotos y más recientes, debemos citar los trabajos de Sir Francis Galton (1822-1911), James McKeen Cattell (1860-1944) y Alfred Binet (1857-1911) como los verdaderos **pioneros de la psicometría moderna**. Como continuadores de las aproximaciones anteriores al estudio sistemático de las diferencias humanas, Galton y Cattell contri-

buyeron necesariamente al establecimiento de la psicología experimental en Europa y Estados Unidos, respectivamente, mediante la creación de los primeros laboratorios antropométricos para el estudio de las características humanas (Valentine, 1999; Sokal, 1982) y el desarrollo de los primeros tests para la evaluación de las diferencias sensoriales, perceptivas y de comportamiento. Abriendo un nuevo camino para el estudio científico de la psicología, los dos autores se aventuraron a postular relaciones entre sus medidas y el intelecto, llegando a conclusiones a veces controvertidas y otras ampliamente criticadas con posterioridad por su simplismo. Por su parte, Binet adoptó un enfoque innovador y fue responsable de lo que se considera el primer test de aplicación general para la medida de las habilidades cognitivas.

Respondiendo a la necesidad de identificar a los estudiantes con problemas para lograr los objetivos de la instrucción ordinaria en las escuelas de París a principios del siglo xx, Binet y su colega Théodore Simon recogieron los incipientes movimientos en la evaluación de la discapacidad cognitiva mediante baterías de tests (Nicolas y Ferrand, 2002). A partir de ellos, y a petición del Ministerio de Educación francés, desarrollaron en 1905 la primera prueba que permitió clasificar a los niños según su inteligencia (Wolf, 1969). Este trabajo permitió el nacimiento del interés de la psicología hacia la atención a las necesidades educativas especiales, pero no fue hasta 1908 cuando estos autores publicaron una revisión de su escala, que permitió medir lo que denominaron el *nivel mental*. Gracias a la adopción de un grupo de referencia compuesto por niños de entre 3 y 13 años, y una vez ordenados los ítems en función de la edad en la que eran típicamente resueltos, desarrollaron el primer test para cuantificar la inteligencia. Esta cuantificación, tomando como referencia la edad biológica del niño, condujo a la definición de lo que se conoce, hasta la actualidad, como *edad mental*.

Adoptando el mismo enfoque orientado a la práctica, la incipiente psicometría se vio impulsada en los circuitos académicos en torno al debate sobre la **medida de la inteligencia** mediante el desarrollo de tests (Martin, 1997). A partir de los trabajos de Binet en el campo de la educación, el interés por la evaluación de la cognición dio el salto al circuito internacional de la mano de Lewis M. Terman (1877-1956), quien en 1916 publicó la revisión Stanford-Binet. De acuerdo con el relato de Goodenough (1949), esta contribución estableció un importante hito en el desarrollo de los tests tal y como los conocemos en la actualidad. Revisando los ítems originales, incorporando pautas claras de administración y destacando la importancia de la representatividad de las muestras para la correcta interpretación de sus puntuaciones, se convirtió en la escala de referencia para la medida de la inteligencia durante las siguientes décadas. De hecho, fue otro paso importante para el desarrollo de la psicometría, en la medida en que se extendió el uso de esta escala al ejército norteamericano durante la Primera Guerra Mundial. En este contexto, y gracias al apoyo del Gobierno de Estados Unidos, la psicometría creció en el centro de las políticas militares³.

⁽³⁾Entre otros, podéis ver Zeidner y Drucker (1988).

Recuperando el espíritu clasificador, los tests deberían servir como instrumentos de evaluación psicológica para el reclutamiento de los nuevos soldados. Así, a partir de la generalización del uso de la revisión Stanford-Binet, se desarrollaron nuevos tests y la administración en grupo para hacer más eficiente el procedimiento de medida. De estos esfuerzos, entre otros, hay que destacar la impronta de Robert Yerkes (1876-1956) en el desarrollo de los tests Army Alpha y Beta (Carson, 1993), que también abordaron una controversia importante sobre los **sesgos culturales de los tests**, que se fue construyendo durante estas primeras décadas del siglo xx. Tal y como se ha discutido ampliamente (Jensen, 1980), las nuevas pruebas para medir la inteligencia podían no estar libres de influencias culturales, de manera que infravalorarían a aquellos que no hablaran la lengua inglesa, a los analfabetos y a quienes sufrieran alguna discapacidad visual o auditiva. Recogiendo los incipientes movimientos en torno a las pruebas no verbales, la versión Beta del test de inteligencia del ejército norteamericano supuso el reconocimiento de la importancia de estas diferencias, así como la necesidad de minimizarlas en cualquier contexto en el que se utilicen los tests como instrumentos de evaluación psicológica.

Una cuestión interesante, por las consecuencias que tuvo más adelante en el desarrollo de la psicometría, fue la concepción que se fue construyendo sobre el propio concepto de inteligencia. Arraigada en las nociones biologicistas y hereditarias de los pioneros, y sustentada en los desarrollos estadísticos de Charles Spearman (1863-1945), una importante corriente concibió la inteligencia como un único factor –el factor g, según Spearman–, que explicaría las puntuaciones en los diferentes tests de inteligencia desarrollados hasta el momento. En cambio, desde la posición “alternativa”, que encabezó Louis Leon Thurstone (1887-1955), la inteligencia estaría compuesta en realidad por varios factores específicos. Sus avances en el desarrollo estadístico de las **técnicas de análisis factorial** que introdujo Spearman permitieron comprender mejor esta aparente contradicción⁴. De hecho, se podía entender el factor general de inteligencia como un tipo de factor subyacente que recogería la variabilidad de las aptitudes específicas de Thurstone, ya fueran relativas a la comprensión verbal, la competencia numérica o la habilidad espacial, entre otras. Del mismo modo, una vez controlado el efecto del factor general, el planteamiento de unos factores grupales por parte de Spearman permitiría recoger la variabilidad común observada en tests que comparten demandas verbales, numéricas o espaciales similares. Más allá de las consecuencias teóricas sobre la medida de la inteligencia, el perfeccionamiento del análisis factorial permitió pavimentar el camino estadístico para el desarrollo de la psicometría hasta finales del siglo xx.

⁽⁴⁾Por ejemplo, podéis ver Cattell (1943).

Además, esta visión de la inteligencia como un conjunto determinado de aptitudes cognitivas específicas contribuyó necesariamente al desarrollo de los **primeros estudios sistemáticos de validez**. Tal y como sucedió durante la Primera Guerra Mundial, el Gobierno norteamericano volvió a incorporar estas prácticas a sus programas de reclutamiento unos años antes de participar en la Segunda Guerra Mundial. A pesar de que los esfuerzos anteriores habían

dado sus frutos, el ejército norteamericano encontró en este planteamiento la manera de resolver algunas limitaciones del enfoque unitario de la inteligencia, principalmente por su escasa capacidad para seleccionar candidatos con perfiles muy especializados. Fue durante este periodo cuando, bajo la dirección de John C. Flanagan (1906-1996), las fuerzas de aviación administraron un conjunto de baterías de tests para seleccionar y clasificar a los pilotos, ingenieros de vuelo y otros técnicos encargados de los instrumentos de navegación. El estudio de la relación entre las puntuaciones en las diferentes aptitudes y el éxito en la formación posterior de los reclutas fueron clave en este proceso (Goslin, 1963), permitiendo la mejora de los procedimientos de selección y clasificación, y conformando uno de los primeros estudios empíricos de la validez de las medidas psicológicas.

Sin embargo, no debemos olvidar el trabajo que paralelamente se llevó a cabo en relación con la **evaluación de la personalidad**. Todavía en el contexto de las políticas militares de la Primera Guerra Mundial, el Gobierno de Estados Unidos se enfrentó a otro tipo de problema práctico. Más allá de la selección y clasificación de los reclutas, un importante esfuerzo se dirigió a la identificación de candidatos susceptibles de sufrir trastornos psicológicos. Para minimizar su presencia en sus filas, Robert S. Woodworth (1869-1962) fue el encargado de desarrollar un nuevo tipo de prueba que permitiera evaluar la estabilidad emocional de los soldados. Este test, su *Personal Data Sheet*, introdujo un conjunto de preguntas con respuesta positiva o negativa que, a diferencia de la evaluación de la inteligencia, no contenía respuestas necesariamente correctas o incorrectas. De este modo, Woodworth desarrolló una prueba que permitía detectar los casos problemáticos no ya mediante la comparación de la respuesta individual con una muestra representativa de la población, sino a partir de las respuestas dadas por individuos con trastornos psicológicos ya diagnosticados. Este trabajo asentó las bases para el desarrollo de los tests de evaluación de la personalidad posteriores (Gibby y Zickar, 2008), así como empezó a definir las bases para tratar adecuadamente la posibilidad de fraude en la respuesta, tanto para ocultar como para simular que se sufre un trastorno.

El prolífico Thurstone, que desarrolló las primeras pruebas de análisis de consistencia interna de los tests en el ámbito de la evaluación de la personalidad, hizo otra contribución importante al desarrollo de la psicometría. Partiendo de los experimentos desarrollados por los pioneros (Gulliksen, 1968), especialmente en la rama de la psicofísica, propuso en 1927 la ley del juicio comparativo como el método para la **medida de las actitudes, las preferencias y los valores**. Esta innovación consistió en trasladar el juicio perceptivo sobre parejas de estímulos físicos –por ejemplo, la elección entre dos sonidos según su intensidad– a la valoración de características psicológicas no estrictamente relacionadas con las propiedades físicas –por ejemplo, la elección entre dos comportamientos según su aceptabilidad. Después de proponer una segunda ley, la ley del juicio categórico, la aproximación de Thurstone a la medida de las actitudes estaba preparada para ser aplicada al juicio general, es decir, sin requerir la comparación entre parejas de fenómenos. De este modo, tal y como

pretendían Thurstone y sus continuadores (Bock y Jones, 1968), la psicometría dispondría de un método de escalamiento para medir y tratar numéricamente valoraciones individuales subjetivas en la búsqueda de una representación objetiva de los fenómenos psicológicos. Este método dio un soporte metodológico importante en las décadas siguientes de investigación en psicología social.

1.2. La psicometría hoy

Pese a la dificultad de llevar a cabo una aproximación histórica a partir de la selección de algunas de las aportaciones más importantes, este ejercicio nos permite ilustrar el nacimiento de una disciplina a partir de los problemas y las diferentes aproximaciones de los pioneros. La psicometría, un nuevo espacio de trabajo metodológico en torno al desarrollo y a la administración de tests, se fue conformando con un componente aplicado, orientada a las demandas en diferentes contextos, y formó parte de algunos de los debates teóricos más importantes de la propia psicología. Sin embargo, hasta la década de los años treinta del siglo XX no podemos situar los **inicios de su constitución como disciplina científica** tal y como la conocemos en la actualidad. Es importante empezar señalando de nuevo el papel decisivo de Thurstone, quien, con el objetivo de establecer y promover la psicología como una ciencia cuantitativa (Samejima, 2000), en 1935 fundó y fue el primer presidente de la Psychometric Society. Además, fue el impulsor de la primera revista especializada que todavía es de referencia obligada, la revista *Psychometrika*, que publicó algunos de los trabajos más importantes sobre los que se formalizó la psicometría. Pocos años después, en 1946, fue también el primer presidente de la División de Evaluación y Medida de la American Psychological Association.

En este sentido, podríamos definir el periodo entre los años treinta y sesenta como la **época dorada de la psicometría**. En este periodo es cuando se publican, además, los libros y manuales más importantes que la vertebrarían. Entre los manuales hay que destacar *The reliability and validity of tests*, de Thurstone (1931), que sistematizaba lo que se había desarrollado hasta el momento en relación con la teoría de los tests y sugería el papel central de la fiabilidad como requisito para la validez de las medidas en la psicometría. Volveremos más adelante a estas cuestiones, elementos básicos de la medida indirecta de los fenómenos psicológicos mediante tests. De manera contemporánea, fue también importante la primera edición del manual *Psychometric methods*, de Guilford (1936), un intento de organizar el campo propio de la psicometría en torno a la teoría de los tests, el escalamiento psicológico y el psicofísico. La teoría clásica de los tests, como fue ampliamente conocida a partir de los trabajos de Spearman sobre la estimación de los errores de medida, empezaba su camino en los circuitos docentes universitarios encargados de formar a los futuros psicólogos.

Por otro lado, durante los años treinta, y especialmente en los cuarenta, uno de los debates más apasionantes abiertos por la psicometría tuvo lugar sobre el propio concepto de *medida*. Inspirados por la British Association for the Ad-

Para saber más

El estudiante interesado en profundizar sobre la constitución de la psicometría como disciplina científica desde una perspectiva histórica puede consultar las revisiones de Hambleton (1994), Brennan (1997), Traub (1997) o Bock (1997).

vancement of Science, psicólogos y físicos se propusieron el reto de decidir si, con los avances en el desarrollo de los tests, la medida psicológica podía ser encuadrada dentro del modelo general de medida de los atributos físicos. En síntesis, medir consistía en cuantificar, es decir, determinar la magnitud con la que un atributo está presente en un objeto. Y para hacerlo, la medida de los atributos físicos dependía de la capacidad de observación de relaciones entre objetos como consecuencia de una operación o manipulación empírica. Por cuanto que esta posibilidad solo podía ser satisfecha con los objetos físicos, la controversia tuvo una respuesta inicial contundente. Tal y como concluyeron Ferguson y sus colaboradores (1940), la medida psicológica que se proponía la psicometría simplemente no era posible, puesto que este modelo no se podía extender para incluir los atributos psicológicos que, por definición, no eran observables ni manipulables empíricamente. El debate, sin embargo, no quedó cerrado y fue Stevens (1946), con su trabajo *On the theory of scales of measurement*, quien dio un paso fundamental para el desarrollo de lo que posteriormente fue denominado el modelo representacional de medida psicométrica.

Stevens definió la medida como el proceso de asignación de números a objetos o acontecimientos de acuerdo con unas reglas, producto de las cuales se obtendrían los diferentes tipos de escalas propuestas: nominal, ordinal, de intervalo y de razón. Medir, en este sentido, no consistiría únicamente en cuantificar, sino que sería el producto de la utilización de diferentes reglas que, en último término, determinarían el tipo de operaciones –o técnicas– estadísticas permitidas en cada escala. No exenta de críticas, es importante resaltar la importancia de su contribución como un primer intento para superar las restricciones impuestas por la cuantificación de los atributos físicos con el objetivo de resolver la controversia en torno a la medida de los fenómenos psicológicos. Así, el debate sobre los modelos o paradigmas de medida formó parte importante de la agenda psicométrica y se pusieron las bases para el desarrollo posterior de otros modelos, como el operacional y el clásico, sobre los que volveremos más adelante. Por otro lado, cerrando este breve repaso a las contribuciones importantes durante los años cuarenta, no podemos dejar de hacer mención a la publicación por parte de Thurstone (1947) del influyente *Multiple factor analysis*, que a partir de los trabajos de Spearman, Kelley y Burt proporciona el soporte estadístico necesario para la construcción y validación de los tests durante las siguientes décadas.

En la década de los cincuenta asistimos a la publicación de las otras dos obras de referencia para la psicometría moderna. Instalada en el circuito académico, el trabajo llevado a cabo por varios autores converge en lo que se ha denominado teoría clásica de los tests (TCT). Y lo hace a manos de Gulliksen (1950) en su *Theory of mental tests*, donde formaliza el modelo lineal clásico por primera vez y define sus asunciones principales. Como teoría de los tests, la TCT propone un nuevo enfoque basado en el concepto de puntuación verdadera. Partiendo de la puntuación empírica obtenida mediante los tests, y siguiendo un conjunto determinado de supuestos, el objetivo es descomponerla en dos partes fundamentales para valorar el error asociado al proceso de medida y,

así, inferir el valor real que se pretende medir. No sin dificultades se convirtió en poco tiempo en la teoría de los tests de referencia y estimuló tanto el debate sobre la propia medida psicológica como el proceso de desarrollo de los tests durante las siguientes décadas. De manera análoga, Torgerson (1958) publica *Theory and methods of scaling*, y establece el canon para el escalamiento psicofísico y psicológico, es decir, la ordenación de los estímulos de manera paralela al de las personas.

Finalmente, los años sesenta son una década de **revisión crítica y apertura de nuevos caminos** para la psicometría. A partir de la TCT, el debate sobre la estimación de los errores cristalizó en dos nuevas corrientes en el desarrollo de tests: la teoría de la generalizabilidad y la teoría de respuesta al ítem. Para incrementar la precisión de las medidas, Cronbach y sus colaboradores⁵ propusieron la teoría de la generalizabilidad, que, mediante la aplicación del análisis de varianza, permitiría descomponer el error genérico propuesto por la TCT en la búsqueda de sus diferentes elementos. Así, de acuerdo con esta teoría, el análisis de la fiabilidad se basa en el diseño de investigaciones que permiten analizar las diferentes fuentes de error –facetas, según sus términos– que afectarían al proceso de medida. Entre estas, por ejemplo, la forma del test, los ítems que lo componen, las ocasiones en las que se administra o los participantes. A pesar de que fue una importante innovación, la complejidad de esta teoría limitó su difusión a la práctica y en poco tiempo sus avances fueron básicamente reformulados por un segundo enfoque, la teoría de respuesta al ítem (TRI).

⁽⁵⁾Cronbach, Rajaratnam y Gleser (1963); Gleser, Cronbach y Rajaratnam (1965).

Así, la TRI se presenta como la respuesta a las críticas principales que había recibido la TCT. En este sentido, es conveniente hacer mención de la publicación por parte de Lord y Novick (1968) de su *Statistical theories of mental test scores*. En este trabajo, también con espíritu de síntesis y análisis crítico del trabajo desarrollado en el contexto de la TCT, se plantean las dificultades que este modelo no estaba siendo capaz de resolver. Básicamente, los problemas principales podrían ser organizados en torno a la dependencia de los instrumentos y las muestras utilizadas en el proceso de construcción y administración de los tests. Durante las décadas posteriores la TRI ocupará, junto a la TCT, un lugar privilegiado entre las teorías de los tests desarrolladas por la psicometría. Volveremos más adelante sobre estas cuestiones, pero vale la pena cerrar esta etapa dorada de la psicometría moderna señalando un último desarrollo que también tuvo sus inicios en la década de los sesenta. Se trata de los tests referidos a criterio, que, en el contexto de la educación, tienen como objetivo evaluar la destreza de las personas en un campo de conocimiento muy bien delimitado.

Evaluar en función de criterios y no de la norma –es decir, escalar u ordenar a los individuos comparando sus puntuaciones– no es un enfoque nuevo. Al contrario, como hemos visto en el recorrido histórico del nacimiento y establecimiento de la psicometría como disciplina científica, este modo de evaluación ha sido objeto de interés con anterioridad, pero gracias a las contribu-

ciones de Glaser (1963) y Popham y Husek (1969) el trabajo en este contexto queda formalizado. Lo que es innovador, desde el punto de vista del desarrollo de los tests, es el enfoque en el procedimiento para establecer los estándares de evaluación, así como en la consistencia y precisión con la que se clasifica a los individuos de acuerdo con estos estándares. A pesar de que nació en el marco de la TCT, aplicaciones posteriores de la TRI han permitido aprovechar los avances en el desarrollo de los tests referidos a criterio como instrumentos de medida de los fenómenos psicológicos.

1.3. La psicometría en el contexto de la evaluación psicológica

Una vez recorrido el camino de los antecedentes y los orígenes de la psicometría moderna, estamos en disposición de abordar formalmente su definición. Y podemos hacerlo a partir de las definiciones que se han ido proponiendo en los textos de referencia. Sin embargo, esta tarea no está exenta de riesgo. Los autores elaboran definiciones concretas que, probablemente, no reflejan todos los matices con los que después las contextualizan en sus textos. A pesar de esto, podemos aprovechar estas propuestas para valorar los diferentes aspectos en los que se pone un énfasis especial. Al fin y al cabo, no cabe todo en una definición.

Más allá del tópico que dice que hay tantas definiciones como autores, podemos distinguir tres grandes aproximaciones. Por un lado, un primer grupo de definiciones aborda la psicometría a partir de los **instrumentos que utiliza**. Desde esta corriente se circunscribe la psicometría como la disciplina encargada de desarrollar los fundamentos para la construcción y administración de tests (por ejemplo, ver Martínez Arias, Hernández Lloreda y Hernández Lloreda, 2006). Esta definición, a pesar de que corta y concisa, presenta algunas dificultades. En primer lugar, la palabra *test* es polisémica, por lo que según el contexto en el que sea empleada puede ser sinónimo de otros términos con significados diferentes, como *prueba*, *examen* o incluso *ensayo*. En segundo lugar, más allá de la precisión semántica que se puede sobreentender en el contexto de la psicología, no deja de ser en cierto modo restrictiva. Si bien es cierto que los tests psicológicos son los instrumentos específicos que desarrollan y administran los psicómetras, también lo es que son el resultado aplicado de un proceso más amplio caracterizado por el desarrollo de teorías y métodos orientados a la medida indirecta de los fenómenos psicológicos.

Una segunda corriente define la psicometría a partir del **objeto al que somete o aplica su interés**. Así, la psicometría se define de manera más o menos genérica como la disciplina científica orientada a la evaluación o medida de los fenómenos psicológicos (por ejemplo, ver Rust y Golombok, 2009). Esta definición se ajusta de hecho al significado etimológico de la palabra *psicometría*, que en sus orígenes griegos podemos encontrar en la yuxtaposición de las palabras *psique* –que significa ‘alma’, ‘aliento’ o ‘intelecto’– y *metron* –en referencia al proceso de medida. En esta línea se puede situar la definición de Kline (1998), que aborda específicamente la psicometría como la tarea de

desarrollar medidas científicas fundamentales en las áreas de la personalidad y las capacidades. En este contexto, la científicidad se convierte en sinónimo de *estandarización* y, a pesar de no hacer referencia explícita, alude indirectamente al uso de tests en el contexto del método científico. Otro implícito importante es, con relación al estatus métrico de las puntuaciones obtenidas como resultado del proceso de medida psicológica, el carácter cuantitativo que debería tener la psicometría como disciplina científica. Como ya hemos avanzado en la aproximación histórica, este es un tema controvertido al que deberemos volver más adelante.

En una posición intermedia, un tercer grupo de autores sitúa la psicometría en la **intersección de las dos corrientes anteriores**. En esta línea podemos señalar por ejemplo la definición de Buchanan y Finch (2005), que especifica los dominios de medida estandarizada mediante tests en relación con las habilidades, los atributos y las características psicológicas. Es importante señalar que, a pesar de que en algunos casos se hace referencia explícita al uso de tests, entre estas definiciones podemos encontrar también un reconocimiento explícito del sentido más amplio del proceso de medida. Así, Holden (2000) define la psicometría como la teoría y la técnica de medida que, en el contexto de la psicología, se encarga de los factores que son medibles. En este sentido, Muñiz (2003) lo hace definiendo la psicometría como el conjunto de métodos, técnicas y teorías implicadas en la medida de las variables psicológicas, teniendo en cuenta su especialización en las propiedades métricas exigibles a este tipo de medida. Finalmente, Jones y Thissen (2007) describen la psicometría como la disciplina –cuantitativa, especifican en este caso– encargada de desarrollar modelos y métodos orientados principalmente al resumen, la descripción y el establecimiento de inferencias a partir de los datos recogidos en la investigación psicológica.

Como hemos podido observar, la definición de la psicometría se ha desarrollado fundamentalmente en torno a los medios o instrumentos que utiliza –los tests– y el objeto que persigue con su uso –la medida de los fenómenos psicológicos. En cambio, una aproximación comprensiva a esta definición debería considerar un tercer elemento importante vinculado a la **finalidad a la que sirve**. Tal y como hemos denominado este apartado, la psicometría adquiere todo su significado en relación con el área de la psicología en la que desempeña su papel fundamental: la evaluación psicológica. De este modo, hemos de tener presente que la medida de los fenómenos psicológicos mediante el desarrollo y la administración de tests no es más que una parte del proceso general de evaluación, y no siempre es la más importante⁶.

⁽⁶⁾Podéis ver Fernández-Ballesteros (1997) para una discusión más amplia.

La psicometría, en este sentido, contribuye al desarrollo de la evaluación psicológica proporcionando teorías, métodos y técnicas que, en última instancia, permiten describir, clasificar, diagnosticar, explicar o predecir los fenómenos psicológicos objeto de medida. De hecho, contextualizada así la psicometría podemos ir un paso más allá y señalar que, a su vez, la evaluación psicológica no se encuentra en ningún otro lugar que al servicio de la intervención psico-

lógica. Es decir, la finalidad última que guía el desarrollo y la administración de tests es la de contribuir a la recogida de las evidencias necesarias que permitan a los psicólogos tomar una decisión u orientar alguna acción. Evidentemente, esta intervención dependerá del contexto en el que la evaluación haya sido desarrollada, pero es importante tenerlo presente para entender no solo qué es la psicometría, sino también su importancia para el ejercicio profesional de los psicólogos en los diferentes contextos en los que intervienen.

Partiendo de este enfoque comprensivo a la psicometría, podemos sintetizar las diferentes aproximaciones y formular finalmente una definición que incorpore su papel en el marco de la evaluación y la intervención psicológicas.

Así, la psicometría es una rama de la psicología que, mediante teorías, métodos y técnicas vinculados al desarrollo y la administración de tests, se ocupa de la medida indirecta de los fenómenos psicológicos con el objetivo de hacer descripciones, clasificaciones, diagnósticos, explicaciones o predicciones que permitan orientar una acción o tomar decisiones sobre el comportamiento de las personas en el ejercicio profesional de la psicología.

Para concluir este ejercicio de definición formal, es importante discutir brevemente las diferencias que existen entre la psicometría y otra disciplina también interesada por la medida psicológica: la **psicología matemática**⁷. En este sentido, podemos reconocer la psicología matemática como la rama de la psicología interesada en el desarrollo de modelos de los procesos perceptivos, cognitivos y motores con el objetivo de establecer leyes que relacionen estímulos físicos con los comportamientos a partir de estudios casi exclusivamente experimentales. Son muchos los autores comunes a las dos orientaciones desde los orígenes del estudio sistemático de las diferencias humanas en el siglo XIX, pero, a diferencia de la psicometría, la psicología matemática no está tan interesada por las diferencias individuales como en la definición de leyes generales que modelen el comportamiento medio de las personas.

⁽⁷⁾Podéis ver, entre otros, Jáñez (1989) y Padilla, Merino, Rodríguez-Miñón y Moreno (1996).

2. Fundamentos de la psicometría

Una vez abordada la psicometría desde una perspectiva histórica, tratando sus raíces y su construcción como disciplina científica en el contexto de la psicología, en este segundo apartado revisaremos sus fundamentos. Empezaremos discutiendo una definición y clasificación de los tests, progresaremos por los diferentes modelos de medida psicométrica desarrollados en las diferentes teorías de los tests y realizaremos una breve introducción a la teoría más extendida en la práctica actual: la teoría clásica de los tests. Por último, recapitularemos estos fundamentos atendiendo al proceso que la psicometría sigue para establecer sus inferencias sobre los fenómenos psicológicos no observables a partir de las puntuaciones obtenidas mediante tests. Empecemos, pues, con una definición y clasificación de los tests.

2.1. Definición y clasificación de los tests

De acuerdo con la definición del manual clásico de Anastasi (1988), un test psicológico⁸ es un procedimiento de medida objetiva y estandarizada de una muestra de comportamientos. Otras definiciones son posibles⁹ pero esta contiene tres elementos fundamentales que nos permiten abordar sistemáticamente las características más importantes que cumplen de manera genérica los tests. De un modo u otro, ya las hemos ido introduciendo en este texto.

⁽⁸⁾En adelante, diremos un test.

⁽⁹⁾Podéis ver, entre otros, Kaplan y Saccuzzo (2001), Murphy y Davidshofer (2005) o Urbina (2004).

En primer lugar, la medida mediante el desarrollo y la administración de tests es, o al menos pretende ser, **objetiva**. En este sentido, la objetividad hace referencia a la sustitución del juicio personal basado en criterios subjetivos por un conjunto de normas determinadas y conocidas que permiten obtener e interpretar las puntuaciones de los individuos en igualdad de condiciones. Asimismo, la medida que pretenden los tests es **estandarizada**, en cuanto que las puntuaciones obtenidas dependen de un procedimiento establecido de administración, corrección e interpretación que las hacen, o las deberían hacer, invariantes del profesional que administra los tests, las condiciones específicas en las que lo hace y el modo como obtiene e interpreta las puntuaciones resultantes de la media. Finalmente, los tests se enfrentan a los fenómenos psicológicos no observables mediante una **muestra de comportamientos**. En la medida en que esta muestra sea representativa del conjunto, las puntuaciones obtenidas permitirán a los profesionales establecer adecuadamente sus inferencias sobre el comportamiento general de las personas más allá de los elementos específicos evaluados mediante tests.

Ya hemos discutido la naturaleza no observable de los fenómenos psicológicos y la dificultad intrínseca que representa para la **medida indirecta mediante tests** que se propone la psicometría. Pero es importante resaltar esta cuestión

cuando la comparamos con otros tipos de medida científica. En este sentido, la medida de los fenómenos psicológicos no resulta tan sencilla como la de las magnitudes físicas –por ejemplo, el peso o la longitud– de los objetos que las ciencias naturales pueden observar y manipular directamente. Además, tal y como señala García Cueto (1993), existen otras **diferencias importantes** entre este tipo de medida y la medida indirecta de los fenómenos psicológicos que deben ser tenidos en cuenta.

Por un lado, el objetivo habitual de la medida de las magnitudes físicas es obtener información sobre un único objeto. La psicometría, en cambio, se propone desarrollar instrumentos que permitan obtenerla sobre un grupo de individuos con el objetivo de extraer conclusiones sobre cada uno de ellos, sobre el grupo entero, e incluso extrapolar sus resultados a las poblaciones de referencia de donde provienen estos individuos. Por otro lado, la medida de magnitudes físicas parte de la posibilidad de repetir su procedimiento tantas veces como sea deseado sin variar las condiciones en las que se lleva a cabo la medida. Este procedimiento es incompatible con la medida mediante tests, en cuanto que la repetición en la aplicación de una misma prueba sobre los mismos individuos produce variaciones en las puntuaciones, que pueden ser explicadas por el cansancio o por el efecto de la práctica en el aprendizaje y no por variaciones sustanciales en los fenómenos de interés. A pesar de esto, podemos encontrar algunas prácticas desarrolladas en los contextos aplicados de las ciencias naturales que tienen mucho en común con la administración de tests.

Así, cuando se pretende analizar la contaminación atmosférica de una ciudad o el nivel de alcohol en sangre de una persona, los científicos desarrollan un conjunto de pruebas objetivas –llamados *reactivos*– para aplicarlas de manera estandarizada sobre una muestra representativa del objeto que pretenden medir –ya sea el aire de la ciudad o la sangre del individuo, siguiendo los mismos ejemplos. En este sentido, el proceso de desarrollo y administración de tests es comparable con estas prácticas, y pone de relieve el sentido que la psicometría atribuye a los tests como sus instrumentos de medida. Un test, lejos del dominio de la psicología, no es más que una prueba, un examen o un ensayo. En definitiva, es un **reactivo** que, en el caso de la medida indirecta de los fenómenos psicológicos mediante tests, no tiene otra finalidad que la de producir una reacción en el comportamiento de los individuos para registrarlo y obtener una puntuación como resultado del proceso de medida. Este sentido etimológico de la palabra *test* lo podemos encontrar en la raíz de su definición en el contexto de la psicometría (por ejemplo, Yela, 1984), donde se asimila directamente con el reactivo de las ciencias naturales y se espera de él que, aplicado a un individuo, revele y dé testimonio fiel de los fenómenos psicológicos no observables que son objeto de medida.

Desde este punto de vista, el valor de un test se encuentra en su capacidad para **susitar y medir comportamientos** que resulten un buen indicador –es decir, una buena representación– del conjunto global de comportamientos

implicados en los fenómenos de interés. En este sentido, debemos hacer una precisión importante con relación a los fenómenos que son objeto de medida mediante tests. En este texto hablamos de los fenómenos psicológicos de una manera general y lo hacemos en cuanto que el objetivo de un test es la medida objetiva y estandarizada de una muestra de comportamientos. Esto no significa, sin embargo, que los tests deban dirigirse exclusivamente a elementos con una larga tradición en la teoría psicológica, como son, entre otros, la inteligencia o la personalidad. Por ejemplo, algunos tests se dirigen hacia la medida de las respuestas fisiológicas, como pueden ser la conductividad de la piel o la frecuencia cardíaca. Otros se fijan en cuestiones como las actitudes racistas, el consumo de sustancias o la sociabilidad de las personas. Finalmente, en el campo de la educación, no son pocos los tests cuyo objetivo es evaluar el dominio de los estudiantes en diferentes áreas de conocimiento. En todos los casos hablamos de fenómenos psicológicos en cuanto que responden a objetos de interés para las ciencias sociales y del comportamiento y, además, recaen en los dominios de la medida indirecta mediante tests siempre que se reúnan las condiciones que establece la psicometría.

Así, la medida indirecta que se proponen los tests empieza con una definición precisa de los fenómenos objeto de interés y se sustenta fundamentalmente en dos tipos de teorías. Por un lado, en una **teoría sustantiva** sólida y bien establecida sobre el comportamiento de las personas –por ejemplo, una teoría de la inteligencia–, que dará el soporte teórico necesario para definir los elementos críticos que conforman estos fenómenos y determinará los comportamientos implicados que serán empleados como evidencias observables en el proceso de medida. Por otro lado, en una **teoría de los tests**, que permite establecer las inferencias sobre los fenómenos psicológicos no observables a partir de las puntuaciones obtenidas en el proceso de medida. En su centro, un **modelo de medida** determinado que sirve al propósito de relacionar las puntuaciones obtenidas y los fenómenos objeto de medida, a partir del cual se articula el **proceso de inferencia psicométrica**. Tanto los modelos de medida como el propio proceso de inferencia psicométrica serán tratados en detalle más adelante. Pero es importante hacer mención a ello en este punto porque resultan clave para entender la medida indirecta que se proponen los tests y que los define en relación con otros tipos de medida científica.

Finalmente, una última consideración sobre los tests y los diferentes términos con los que se hace referencia a ellos nos lleva a fijarnos en dos grandes maneras de tratar las respuestas obtenidas para medir los fenómenos psicológicos de interés. Por un lado, los denominados **tests de habilidad o de potencia** tienen como objetivo evaluar la competencia, la aptitud o los conocimientos de los individuos a partir del acierto o la calidad de sus respuestas. Son pruebas que discriminan respuestas correctas e incorrectas y esta es la base para puntuar las ejecuciones individuales. Por otro lado, los denominados **tests de personalidad**¹⁰ tienen un objetivo diferente y pretenden conocer de manera general las motivaciones, preferencias, opiniones o actitudes de los individuos frente a un determinado estímulo. Este segundo tipo de tests no

⁽¹⁰⁾A pesar de que el término *personalidad* puede hacer referencia a la rama de la psicología interesada por las características personales que influyen en sus cogniciones, emociones, motivaciones y comportamientos, en este caso se utiliza en un sentido más amplio para hacer referencia a los tests que, en oposición a los de habilidad o de potencia, no pretenden medir las capacidades de las personas.

tienen respuestas correctas y por lo tanto no sirven para evaluar el acierto o el error de los individuos. Esta distinción es relevante desde un punto de vista terminológico, dado que los tests de personalidad son muchas veces llamados **cuestionarios**, **inventarios** o **escalas**, a pesar de que este uso no es siempre consistente. De hecho, el término *escala* puede resultar sinónimo de un test cuando está conformado por diferentes partes, o incluso de cada una de estas partes para reflejar las dimensiones o características específicas que miden en el contexto del test global. Por otro lado, otro término también extendido es el de *batería*, que tampoco tiene un significado unívoco y puede hacer referencia tanto a un test compuesto por varias partes, como a una selección de diferentes tests administrados conjuntamente por un profesional en una evaluación psicológica determinada.

Después de esta discusión terminológica, y a pesar de la enorme variedad de tests existente en la actualidad, estamos en disposición de clasificarlos en función de algunas de sus características más importantes. Así:

- **Según el propósito.** Los tests pueden tener diferentes finalidades, y entre ellas podemos destacar dos fundamentales: la **diagnos**is, orientada a la evaluación de las condiciones actuales de los individuos, y la **predicción**, que se propone relacionar la medida actual con el comportamiento de las personas en situaciones futuras. En este sentido, por ejemplo, una cosa es obtener información sobre la destreza de una persona en la resolución de problemas y otra es utilizar esta información para predecir su desempeño en un trabajo determinado. Otros propósitos más específicos son también posibles, como ya hemos comentado en la discusión de la definición formal de la psicometría.
- **Según el contenido.** De manera general, podemos clasificar los tests según el área de la psicología a la que pertenecen los fenómenos psicológicos que pretenden medir. Esta clasificación, sin embargo, varía según los autores que las realizan, en cuanto que pueden fijarse en diferentes niveles de complejidad al jerarquizar estas áreas. En el nivel más general podemos distinguir básicamente tres grandes grupos, que incluirían los tests orientados a la evaluación de las **habilidades cognitivas**, la **personalidad** y las **actitudes**.
- **Según el formato.** Teniendo en cuenta los materiales utilizados es frecuente distinguir los tests de lápiz y papel, de manipulación y de medidas fisiológicas. Nos referimos a los **tests de lápiz y papel** cuando presentan los ítems o preguntas en papel y requieren que el individuo dé algún tipo de respuesta escrita. Los **tests computerizados**, a pesar de que no usan el papel, podrían ser incluidos en esta categoría en cuanto que exigen algún tipo de respuesta escrita mediante dispositivos electrónicos. Por otro lado, son **tests de manipulación** aquellos que presentan una serie de objetos, imágenes o rompecabezas que los individuos deben resolver para demostrar su habilidad. Finalmente, los **tests de medidas fisiológicas** utilizan

sensores de distinto tipo para registrar las reacciones de los individuos ante los estímulos físicos presentados.

- **Según el tipo de administración.** Otra manera de clasificar los tests puede tener en cuenta el modo como son administrados y nos permite distinguir los **tests individuales y grupales**. Esta clasificación se solapa con la que hemos hecho según el formato, pero nos permite distinguir tests que requieren una administración a un único individuo o permiten hacerlo con un grupo de individuos al mismo tiempo. También podemos distinguir los **tests verbales y no verbales**, en función de si se hace o no una presentación oral o escrita del test y sus instrucciones.
- **Según el tratamiento de las respuestas.** Siguiendo la diferencia que establecíamos entre los **tests de habilidad o de potencia** y los **tests de personalidad**, podemos clasificar también los tests en dos grandes grupos, en función de si evalúan el acierto en las respuestas para determinar la competencia, la aptitud o los conocimientos de los individuos, o si tratan de evaluar motivaciones, preferencias, opiniones o actitudes. Estos dos tipos de pruebas son también conocidas como **pruebas de ejecución máxima** y **pruebas de ejecución típica**, respectivamente.
- **Según la interpretación de las puntuaciones.** Tal y como ya hemos comentado, las puntuaciones obtenidas mediante tests pueden ser interpretadas según la norma o en referencia a un criterio. Así, los **tests normativos** permiten comparar la puntuación del individuo con la ejecución observada en un grupo de referencia que previamente ha respondido al mismo test. Este grupo puede estar compuesto por muestreo probabilístico, cuando representa a la población de referencia, o no probabilístico, cuando estamos interesados en comparar las puntuaciones con un grupo de personas que cumple unas características determinadas. Por otro lado, los **tests referidos a criterio** toman como referencia la definición de un dominio de conocimientos o habilidades específicos y permiten medir la ejecución del individuo no ya en comparación con un grupo de referencia, sino en función de su grado de adecuación o consecución de este criterio.
- **Según el estatus comercial.** Finalmente, los tests se pueden clasificar en función de si son propietarios o no. Los **tests propietarios o comerciales** son tests que requieren el pago para su uso e incluyen los ejemplares del propio test y un manual que contiene información sobre su desarrollo y sus propiedades psicométricas, las hojas de corrección de las respuestas y las tablas para interpretar las puntuaciones en relación con los grupos de referencia. En otros casos hablamos de **tests abiertos o no comerciales**, y generalmente se pueden obtener contactando con el autor, que normalmente también ha desarrollado algún tipo de manual que contextualiza el test. En algunas ocasiones esta información queda reducida a la publi-

cación de un artículo científico en el que el autor lo introduce, realiza una primera administración y presenta sus propiedades psicométricas.

2.2. Modelos de medida psicométrica

Todo tipo de medida científica, y la medida indirecta de los fenómenos psicológicos mediante tests no es una excepción, se fundamenta en una **definición del propio concepto de medida**. En este sentido, tal y como ya hemos introducido, en el núcleo de las teorías de los tests podemos encontrar un conjunto de modelos que establecen qué es exactamente medir y, como consecuencia, especifican la relación existente entre las puntuaciones empíricas obtenidas y los fenómenos psicológicos no observables objeto de interés. Son los **modelos o paradigmas de medida** que la psicometría ha desarrollado de manera formal a partir del debate iniciado en los años cuarenta sobre la viabilidad de la medida de los fenómenos psicológicos. No obstante, antes de caracterizar estos modelos retomaremos la controversia a partir de las conclusiones de Ferguson y sus colaboradores (1940) para la British Association for the Advancement of Science. Como veremos a continuación, suponen un importante punto de partida a partir del cual la psicometría ha desarrollado sus propios modelos de medida para abordar específicamente los fenómenos psicológicos no observables.

De acuerdo con el trabajo de Campbell (1920), físico y miembro de la comisión que elaboró este informe, el **modelo general de medida desarrollado en la física** se basaría en la existencia de una equivalencia –un isomorfismo, en sus palabras– entre algunos atributos físicos y las propiedades aditivas de los números. Medir consistiría en asignar números a estos atributos para representar la aditividad de los objetos físicos y esta representación solo sería posible en cuanto que se satisficieran dos condiciones: que el atributo se pudiera ordenar, es decir, que fuera posible determinar que el atributo presente en un objeto es menor, igual o mayor que otro; y que se pudiera demostrar empíricamente la existencia de la aditividad mediante la manipulación empírica. Es lo que denominó *medida fundamental*, definiendo la medida de los atributos físicos por analogía con los números a partir de las operaciones llevadas a cabo sobre los objetos para cuantificar sus atributos.

En un ejemplo sencillo, el peso de dos o más objetos puede ser observado empíricamente mediante una balanza para ordenarlos de menor a mayor, así como se puede demostrar empíricamente que el peso obtenido después de poner estos objetos juntos en la balanza da como resultado la suma de los pesos parciales.

Campbell (1928) extendió este modelo general a lo que denominó como *medidas derivadas*, donde la cuantificación de ciertos atributos físicos no dependería de la manipulación empírica sino del descubrimiento de relaciones matemáticas entre dos medidas fundamentales. Por ejemplo, sería el caso de la densidad de los objetos, medida derivada que respondería a una razón entre su masa y su volumen. En cualquier caso, la clave de la medida para el modelo

de Campbell se encontraría en la necesidad de observar y manipular empíricamente los objetos, bien para cuantificar un atributo físico mediante su observación directa, bien para establecer funciones matemáticas que relacionen otros atributos observables. En la medida en que los fenómenos psicológicos no pueden ser observados ni manipulados empíricamente, ni son producto de una relación entre otros atributos observables, la medida indirecta que persigue la psicometría no sería equiparable a la medida fundamental o derivada de la física y, por lo tanto, resultaría inviable. Esta fue la conclusión de la comisión organizada por la British Association for the Advancement of Science, que, como ya hemos comentado, lejos de cerrar la controversia se convirtió en uno de los debates más importantes para la medida indirecta de los fenómenos psicológicos mediante tests.

Como respuesta a las restricciones impuestas por el modelo general de medida desarrollado en la física, la psicometría avanzó en el debate sobre la aproximación indirecta a los fenómenos psicológicos desarrollando sus propias teorías y modelos de medida. A pesar de que otros autores utilizan otros términos¹¹, en este texto seguiremos a Michell (1986) y los denominaremos *modelo representacional*, *operacional* y *clásico*. A continuación, caracterizaremos estos modelos y analizaremos brevemente las diferencias en su planteamiento del proceso de medida para hacerlo extensivo a los fenómenos psicológicos. Siendo un debate todavía abierto, y susceptible a diferentes interpretaciones según los diferentes autores, optamos por la exposición de Michell en cuanto que plantea, desde una visión crítica, las incógnitas más importantes que la psicometría debe resolver en su construcción como disciplina científica.

En primer lugar, el **modelo representacional** fue desarrollado a partir de los trabajos de Stevens (1946) y Suppes (1951), y define la medida como el proceso de asignación de números a objetos a partir de unas reglas, de modo que reflejen relaciones empíricas entre los objetos. Estas relaciones, de manera comparable al modelo de Campbell, quedarían representadas por las propiedades de los números, pero no se centrarían exclusivamente en la aditividad. De hecho, Stevens abrió la puerta a otros tipos de relaciones para llevar a cabo la medida, aceptando por ejemplo la equivalencia o el orden. Medir no sería únicamente cuantificar, sino la representación numérica de los hechos empíricos en sentido amplio, y daría como resultado cuatro escalas clásicas para la psicología según el tipo de relación representada: escala nominal, ordinal, de intervalo y de razón.

Tal y como hace el modelo desarrollado en la física, el modelo representacional asume la correspondencia entre las propiedades de los números –las puntuaciones obtenidas mediante tests, en este caso– y las relaciones de los objetos que representan. En cambio, traslada el foco desde la demostración empírica de estas relaciones al establecimiento de un conjunto de restricciones sobre el tipo de operaciones estadísticas permitidas en cada escala que garantice la in-

⁽¹¹⁾ Por ejemplo, podéis ver Fraser (1980), Swistak (1990) o Hand (1996).

Para saber más

El estudiante interesado puede encontrar una discusión mucho más amplia de la que podemos plantear en este texto en Michell (1999).

Para saber más

En relación con el debate en torno a la estadística permitida, podéis ver la interesante discusión de Gaito (1980) y Velleman y Wilkinson (1993).

variancia de las relaciones empíricas representadas. Es el debate sobre la estadística permitida, que ha generado una gran controversia entre los partidarios del modelo representacional y el operacional.

En segundo lugar, el **modelo operacional** se basa en las contribuciones de Bridgman (1927) y Dingle (1950), a partir de las cuales se propuso la definición de cualquier concepto mediante las operaciones necesarias para medirlo. Así, de acuerdo con este modelo, la medida no sería otra cosa que el conjunto de operaciones necesarias para definir un concepto que, en última instancia, acaban produciendo números. Esta definición de la medida tiene puntos en común con la del modelo representacional, dado que asignar números – las puntuaciones obtenidas mediante tests– de acuerdo con unas reglas es un tipo particular de las operaciones posibles, y los números resultantes no son otra cosa que el producto de las operaciones llevadas a cabo. En cambio, una importante diferencia entre estos dos modelos de medida se encuentra en el requisito según el cual los números representan o no un sistema de relaciones empíricas.

Para el modelo representacional, estas relaciones empíricas son previas a la medida y tienen una existencia independiente de las operaciones llevadas a cabo para producir los números. El modelo operacional, en cambio, limita el dominio de las operaciones a aquellas que de una manera consistente producen números y en ningún caso está interesado en la existencia de una realidad empírica que apoye estos números. Esta diferencia en el enfoque de la medida tiene además importantes implicaciones en cuanto al debate sobre la estadística permitida, dado que los partidarios del modelo operacional no ven la necesidad de establecer ninguna restricción a las operaciones estadísticas. No siendo el objetivo la búsqueda de relaciones empíricas comparables a las relaciones entre los números, sino los números en sí mismos, ninguna restricción puede ser aplicable a unas puntuaciones que, en palabras de Lord (1953), no tienen conciencia de dónde provienen.

Finalmente, el **modelo clásico** se propondría resolver esta discusión planteando la naturaleza cuantitativa de los fenómenos psicológicos como condición para la medida. Partiendo de los trabajos de Rozeboom (1966) y Jones (1971), y arraigado en el desarrollo de las teorías cuantitativas de la psicología de las décadas anteriores, este modelo define la medida como la determinación de la cantidad –del cuánto– en la que un atributo se presente en el objeto medido. Es decir, medir consiste en determinar cuántas unidades están presentes en el atributo observado, pero a diferencia de los modelos de Campbell y Stevens, no exige la existencia de una relación empírica entre los objetos. De hecho, el modelo clásico mide atributos, no objetos, y por lo tanto no contempla en su núcleo la demostración de la aditividad mediante la observación y la manipulación empírica de los objetos físicos. Es más, a diferencia de los modelos representacional y operacional, la medida no se considera la asignación de números a objetos de acuerdo con unas reglas ni el resultado de un tipo

particular de operaciones que produce números, respectivamente, sino que se define como el proceso de descubrimiento de relaciones numéricas entre los valores observados en un atributo cuantitativo.

Siguiendo el ejemplo del peso de dos o más objetos físicos, sus medidas en una unidad determinada (por ejemplo, en kilos) proporcionan un sistema para relacionarlos matemáticamente en términos de sus magnitudes relativas, y permiten afirmar que uno de ellos pesa n veces más o menos que otro. Estas relaciones numéricas son, para el modelo clásico, tan válidas como las relaciones observables entre los objetos y por lo tanto comparten el mismo estatus como evidencias empíricas necesarias para el desarrollo científico. En este sentido, es muy importante señalar el papel de la teoría sustantiva, puesto que es la base sobre la que se fundamentan tanto la hipótesis sobre la naturaleza cuantitativa de los fenómenos psicológicos objeto de medida como el proceso de determinación de las relaciones numéricas a partir de las medidas obtenidas. Finalmente, el modelo clásico no contempla las diferentes escalas de medida aceptadas por los modelos representacional y operacional. De esta manera, las escalas nominales y ordinales quedan fuera del dominio de la medida –que, recordemos, es cuantitativa por definición– y por lo tanto no es necesaria ninguna prescripción en torno al debate sobre la estadística permitida.

Como ya hemos señalado, los modelos de medida psicométrica se encuentran en la base de las diferentes teorías de los tests y durante las últimas décadas han proporcionado varias alternativas para la aproximación indirecta a la medida de los fenómenos psicológicos que se propone la psicometría. Así, estos tres modelos de medida –decíamos, representacional, operacional y clásico– se corresponden con la **teoría representacional**, la **teoría clásica de los tests** y los **modelos de variable latente**, respectivamente. Abordar todas estas teorías alternativas está fuera de los objetivos de un texto introductorio como este, aunque el estudiante interesado puede encontrar un análisis muy detallado, por ejemplo, en Borsboom (2005). En cambio, optaremos por restringir el tratamiento de las teorías de los tests a aquella que goza del apoyo mayoritario en la práctica actual de la psicometría. Hablamos de la teoría clásica de los tests (TCT), que es además la base para el desarrollo de los módulos posteriores de este texto. Es importante resaltar, para evitar confusiones, que la TCT no se denomina así por el modelo de medida psicométrica que utiliza. De hecho, su definición de medida encaja con el modelo operacional y, en cambio, adopta este nombre por oposición a otros enfoques más modernos, como el de la teoría de respuesta al ítem (TRI). Finalmente, introduciremos muy brevemente algunas de las innovaciones que, de acuerdo con el modelo clásico de medida propio de los modelos de variable latente, la TRI ha desarrollado para tratar de resolver algunas de las limitaciones de la TCT.

2.3. Teoría clásica de los tests

La TCT es la teoría de los tests más extendida actualmente en la práctica de la psicometría y se basa en el **modelo lineal clásico** propuesto por Spearman¹², sistematizado por Gulliksen (1950) y reformulado posteriormente por Lord y Novick (1968), que articula el proceso de medida definiendo tres conceptos fundamentales: la puntuación verdadera, la puntuación empírica y el error de medida. Para introducirlo, en este texto seguiremos la exposición de Muñiz (1996 y 2003).

Así, partiendo del modelo operacional de medida psicométrica, esta teoría de los tests no está interesada en el sistema de relaciones empíricas, sino que centra su atención en el análisis de las puntuaciones obtenidas para valorar los errores cometidos en el proceso de medida indirecta de los fenómenos psicológicos. Es la llamada **puntuación empírica (X)**, que, de acuerdo con esta teoría, respondería a una relación lineal de dos componentes fundamentales:

$$X = V + e$$

Por un lado, la **puntuación verdadera (V)**, que sería el resultado ideal o deseado, en el que el proceso de medida mediante tests habría sido llevado a cabo libre de cualquier tipo de error. Por otro lado, el **error de medida (e)**, que sería responsable de la discrepancia entre la puntuación verdadera que se pretende conseguir y la puntuación empírica obtenida como resultado de la administración del test.

Como cualquier tipo de medida científica, la medida indirecta de los fenómenos psicológicos se encuentra sujeta a estas variaciones o desviaciones no deseadas, y el modelo lineal clásico propone un conjunto de supuestos que permiten hacer una estimación de las puntuaciones verdaderas a partir de las puntuaciones empíricas obtenidas, y una definición de tests paralelos. Así:

- **Primer supuesto:** $V = E(X)$. La puntuación verdadera (V) se define matemáticamente como la esperanza matemática de la puntuación empírica (X). Es decir, el primer supuesto del modelo lineal clásico asume que, en caso de que fuera posible la administración de un test un número infinito de veces, la media de las puntuaciones empíricas obtenidas nos daría como resultado la puntuación verdadera del sujeto en el test. La puntuación empírica, por lo tanto, no es un sustituto de la puntuación verdadera, sino la mejor aproximación disponible en un proceso de medida que, de manera explícita, es reconocido como no libre de error.
- **Segundo supuesto:** $\rho(v, e) = 0$. No existe correlación entre las puntuaciones verdaderas de los sujetos (v) en un test y sus respectivos errores de medida (e). Es decir, de acuerdo con el segundo supuesto no se espera que el

⁽¹²⁾Para una discusión histórica, podéis ver Traub (1997).

Para saber más

El estudiante interesado puede encontrar explicaciones más detalladas en Borsboom (2005), en Crocker y Algina (2006), o en de Gruijter y van der Kamp (2008).

tamaño de los errores cometidos esté sistemáticamente asociado al tamaño de las puntuaciones verdaderas.

- **Tercer supuesto:** $\rho(e_j, e_k) = 0$. Si disponemos de dos tests diferentes (j y k), no existe correlación entre los errores de medida cometidos con cada uno de ellos (e_j y e_k , respectivamente). Es decir, de acuerdo con el tercer supuesto, los errores de medida de los diferentes tests son aleatorios en cada ocasión y, por lo tanto, no se espera que exista ninguna relación entre ellos.
- **Definición de tests paralelos:** $V_j = V_k$ y $\sigma^2(e_j) = \sigma^2(e_k)$. Finalmente, dos tests (j y k) son paralelos siempre que sus puntuaciones verdaderas (V_j y V_k) y sus varianzas de los errores de medida ($\sigma^2(e_j)$ y $\sigma^2(e_k)$, respectivamente) sean idénticas.

De este modelo, con sus supuestos y la definición de tests paralelos, se deriva un conjunto de deducciones inmediatas importantes que forman la base para el módulo dedicado a la fiabilidad en este texto. Así:

- $e = X - V$. De acuerdo con la formulación inicial del modelo lineal clásico, el error de medida sería la diferencia entre la puntuación empírica y la puntuación verdadera.
- $E(e) = 0$. La esperanza matemática de los errores de medida es cero, por lo que si fuera posible administrar un test un número infinito de veces, estos errores aleatorios o no sesgados se compensarían o anularían entre ellos.
- $\mu_x = \mu_v$. La media de las puntuaciones empíricas (μ_x) es igual a la media de las puntuaciones verdaderas (μ_v).
- $\text{cov}(V, e) = 0$. De acuerdo con el segundo supuesto del modelo, las puntuaciones verdaderas no covarían con los errores de medida.
- $\text{cov}(X, V) = \text{var}(V)$. La covarianza entre las puntuaciones empíricas y verdaderas es igual a la varianza de las puntuaciones verdaderas.
- $\text{cov}(X_j, X_k) = \text{cov}(V_j, V_k)$. La covarianza entre las puntuaciones empíricas de dos tests (X_j y X_k) es igual a la covarianza de sus puntuaciones verdaderas (V_j y V_k).
- $\text{var}(X) = \text{var}(V) + \text{var}(e)$. La varianza de las puntuaciones empíricas es el producto de la suma de la varianza de las puntuaciones verdaderas y la de los errores de medida.

- $\rho(X, e) = \sigma_e / \sigma_x$. La correlación entre las puntuaciones empíricas y los errores de medida es el resultado de la división de la desviación típica de los errores entre la de las puntuaciones empíricas.
- $\mu_1 = \mu_2 = \dots = \mu_k$. Para k tests paralelos, las medias son idénticas.
- $\sigma^2(X_1) = \sigma^2(X_2) = \dots = \sigma^2(X_k)$. Del mismo modo, las varianzas de k tests paralelos son idénticas.
- $\rho(X_1, X_2) = \rho(X_1, X_3) = \dots = \rho(X_j, X_k)$. Finalmente, las correlaciones entre k tests paralelos son también idénticas.

Teniendo en cuenta que ni las puntuaciones verdaderas ni los errores de medida son observables directamente a partir de las puntuaciones empíricas, ninguno de los supuestos ni las ocho primeras deducciones inmediatas son demostrables.

Como ya hemos discutido anteriormente, la TCT resulta una aproximación útil para la estimación de las puntuaciones verdaderas a partir de las puntuaciones empíricas, pero no está exenta de **algunas limitaciones**, que no son fáciles de resolver desde el modelo lineal clásico. Tal y como discutieron Lord y Novick (1968) en su reformulación, esta teoría tiene dos dependencias importantes en relación con los instrumentos y las muestras utilizadas. Por un lado, una dependencia de las puntuaciones empíricas obtenidas respecto a los instrumentos, de manera que, por ejemplo, dos tests de inteligencia independientes utilizan escalas diferentes y, por lo tanto, las puntuaciones que se obtienen no resultan comparables directamente. Por otro lado, una dependencia de las propiedades psicométricas de los tests respecto a las muestras utilizadas para desarrollarlos, de manera que, por ejemplo, la dificultad de los ítems que conforman un test depende de las características de los individuos a quienes se administra el test.

Tratando de resolver estos problemas, la TRI se propone cambiar el foco desde el tratamiento del test entero al tratamiento individual de los ítems. Para hacerlo, esta teoría de los tests plantea la existencia de una relación entre las puntuaciones en la variable latente –recordemos, de acuerdo con el modelo clásico de medida psicométrica– y la probabilidad de acertar cada ítem introduciendo diferentes funciones matemáticas para modelar adecuadamente esta relación. Así, una vez seleccionada la función más adecuada se construyen lo que se denominan las *curvas características de los ítems*, una modelización de las respuestas que puede tener en cuenta diferentes parámetros, como la capacidad de discriminación de los ítems, su dificultad o la probabilidad de que sean acertados al azar.

Para saber más

El estudiante interesado puede encontrar el desarrollo matemático que sustenta estas deducciones en Muñiz (2003).

No es este el lugar para llevar a cabo una discusión en profundidad, pero sí para hacer énfasis en las implicaciones que tienen los diferentes modelos de medida psicométrica, a partir de los cuales la TCT y la TRI se proponen la medida indirecta mediante el uso de tests. Por un lado, tal y como hace la TCT, definiendo las puntuaciones empíricas como la suma de las puntuaciones verdaderas y el error de medida y, por otro, tal y como hace la TRI, estableciendo diferentes funciones matemáticas según el modelo de relación entre la variable latente y la capacidad de discriminación de los ítems.

2.4. El proceso de inferencia psicométrica

Para concluir con sus fundamentos, recapitularemos el proceso que la psicometría sigue para establecer sus inferencias sobre los fenómenos psicológicos no observables a partir de las puntuaciones obtenidas mediante tests. Antes de abordar sus especificidades, es importante tener presente que la medida objetiva y estandarizada de una muestra de comportamientos se ajusta, de manera general, al procedimiento establecido por el **método científico**. Lo podríamos resumir así:

- 1) Formular una pregunta de investigación o una hipótesis relevante.
- 2) Especificar y definir adecuadamente todas las variables involucradas.
- 3) Desarrollar o elegir los instrumentos y procedimientos necesarios para llevar a cabo las medidas.
- 4) Evaluar el funcionamiento de los instrumentos y procedimientos para obtener las garantías suficientes sobre la calidad del proceso de medida.
- 5) Recoger las evidencias necesarias que permitan responder a los objetivos de la investigación.
- 6) Resumir y, siempre que sea posible, tratar estadísticamente los datos obtenidos para determinar hasta qué punto los resultados son significativos y, por lo tanto, no son producto del azar.

De acuerdo con este procedimiento general, la psicometría ofrece un conjunto de teorías, métodos y técnicas vinculadas al desarrollo y la administración de tests que dan el soporte necesario a los puntos 2, 3 y 4 del método científico cuando se utiliza en el contexto de la medida indirecta de los fenómenos psicológicos no observables. Y lo hace, como hemos señalado anteriormente, partiendo de una definición precisa de los fenómenos objeto de medida y de una selección de los comportamientos implicados que serán empleados como evidencias observables. Aun así, es importante hacer énfasis en algunas dificultades específicas que, como disciplina científica, la psicometría debe afrontar para garantizar la confianza en las inferencias establecidas sobre los fenómenos psicológicos no observables a partir de las puntuaciones obtenidas me-

Para saber más

El estudiante interesado puede encontrar una comparación asequible de estas dos teorías de los tests en Muñiz (2010), así como desarrollos más detallados de la TRI en Muñiz y Hambleton (1997), Maydeu-Olivares y McArdle (2005), Borsboom (2005) o Martínez Arias, Hernández Lloreda y Hernández Lloreda (2006).

Una referencia clásica sobre los fundamentos de la TRI es la de Hambleton, Swaminathan y Rogers (1991).

dian­te tests. Todas ellas, en íntima relación con el reto esencial de la psicometría, son consecuencia directa de la dificultad añadida que supone la imposibilidad de observar y manipular directamente los fenómenos psicológicos objeto de interés. En este sentido, entre los **retos específicos** podemos señalar algunos de los más importantes:

- **La medida indirecta de los fenómenos psicológicos no es unívoca.** Como hemos ido viendo, no existe una única manera de definir el proceso de medida, así como no existe tampoco una única manera de entender los propios fenómenos objeto de medida. Son muchas las decisiones que los profesionales interesados en el desarrollo y la administración de tests deben tomar durante este proceso y, por lo tanto, son muchas las posibles soluciones a los problemas que surgen en la aproximación indirecta a la medida de los fenómenos psicológicos. Entre estas decisiones destacan las relacionadas con la elección de lo que es relevante medir, la selección y el muestreo de los comportamientos observables vinculados y la definición de las características que los tests deben tener para suscitarlos y medirlos adecuadamente.
- **La teoría desempeña un papel fundamental en la medida psicométrica.** Todas las decisiones que toman los profesionales interesados en el desarrollo y la administración de tests han de estar fundamentadas, como hemos señalado anteriormente, por dos tipos de teorías. Por un lado, por una teoría sustantiva muy establecida en torno a los fenómenos psicológicos y los comportamientos de las personas, que servirá de contexto de referencia para el proceso de medida. Por otro, una teoría de los tests que permitirá adoptar el enfoque más adecuado para llevar a cabo la medida y establecer la relación entre las puntuaciones obtenidas y los fenómenos psicológicos no observables, objeto de interés. Si los tests tienen sentido y sirven como instrumentos de medida en el campo de la psicología es, precisamente, gracias al apoyo que estos dos tipos de teorías proporcionan durante todo el proceso.
- **La selección y el muestreo de los comportamientos impone importantes limitaciones.** Un tercer reto importante para la medida indirecta que se propone la psicometría es la selección y el muestreo de los comportamientos empleados como evidencias observables. En consonancia con la teoría sustantiva, el desarrollo y la administración de tests parten de una definición que orienta hacia unos comportamientos vinculados y no a otros, que, a la vez, deben ser adecuadamente muestreados dada la imposibilidad material de suscitar y medir el dominio entero de comportamientos al que pertenecen. De hecho, un comportamiento determinado puede ser empleado como evidencia de fenómenos psicológicos diferentes en función de la perspectiva teórica con la que se fundamenta el proceso de construcción de los tests. Importantes limitaciones vinculadas al tiempo requerido

y las condiciones de administración también han de ser adecuadamente resueltas para hacer operativos los tests desde un punto de vista práctico.

- **Las puntuaciones obtenidas requieren una interpretación adecuada.** Las puntuaciones obtenidas mediante el uso de tests no son, *per se*, informativas y han de ser siempre interpretadas para responder al propósito con el que los tests han sido desarrollados. Ya sea desde un punto de vista general, con relación al contenido y su propósito, ya sea con relación al sistema de referencia empleado –según la norma, es decir, respecto a la ejecución de un grupo de referencia, o referidos a un criterio, esto es, respecto al grado de adecuación o consecución de este criterio–, el uso de tests requiere unos conocimientos y unas destrezas que solo podemos encontrar en manos de profesionales cualificados y que siempre debe cumplir con los objetivos más generales que persigue la evaluación psicológica a la que sirve.
- **La medida de los fenómenos psicológicos no está libre de error.** El error es un componente inherente a cualquier proceso de medida y debe ser objeto de un tratamiento adecuado para conocerlo y minimizarlo. La medida indirecta añade una dificultad específica que, de manera sintética, encontramos en la imposibilidad de observar directamente tanto los fenómenos psicológicos objeto de interés como el error cometido durante el proceso de medida. Partiendo de las puntuaciones obtenidas, y gracias al apoyo de los modelos de medida desarrollados en las diferentes teorías de los tests, la psicometría dedica una buena parte de sus esfuerzos a conseguir la precisión necesaria que, desde un punto de vista científico, es también exigible a la medida indirecta de los fenómenos psicológicos.
- **La medida psicológica no se puede entender si no es en relación con otras medidas o acontecimientos observables.** Pese a la importancia de la teoría sustantiva en el desarrollo de los tests, no podemos obviar finalmente que la medida indirecta de los fenómenos psicológicos no tiene sentido solo en sí misma, sino que lo tiene en relación con otras medidas derivadas de la misma teoría u otros comportamientos observables que corroboren el proceso de medida mediante tests. Como hemos dicho anteriormente, la medida mediante tests debe proporcionar información relevante con el objetivo de hacer descripciones, clasificaciones, diagnósticos, explicaciones o predicciones que permitan orientar una acción o la toma de una decisión sobre el comportamiento de las personas, y por esa razón las puntuaciones obtenidas mediante tests requieren un apoyo externo a su propia formulación para demostrar su utilidad práctica.

Para enfrentarse a estos retos específicos, la medida indirecta de los fenómenos psicológicos se provee de unos **principios básicos** que garantizan la confianza en las inferencias establecidas a partir de las puntuaciones obtenidas mediante

tests. Estos principios sirven para evaluar las propiedades psicométricas de los ítems y de los tests en su conjunto, y entre estos principios básicos destacan los siguientes:

- **Fiabilidad:** la precisión con la que los tests llevan a cabo la medida.
- **Validez:** la confianza en que las medidas se corresponden realmente con lo que se proponen medir.

Los diferentes métodos y procedimientos vinculados a estos principios son la clave del éxito de los tests como instrumentos de evaluación psicológica y serán, por lo tanto, objeto de un tratamiento en detalle en módulos posteriores.

3. Construcción y administración de tests

Presentados los fundamentos de la psicometría, en este último apartado nos haremos cargo del proceso de construcción y administración de tests. Para hacerlo, empezaremos desarrollando las diferentes fases implicadas en el diseño y construcción de un nuevo test. A continuación, y partiendo de las claves que nos proporciona el conocimiento de este proceso de construcción, discutiremos algunos criterios importantes para evaluar las características de los tests disponibles en la literatura y valorar su conveniencia en relación con los objetivos de la evaluación psicológica a la que deben servir. Finalmente, abordaremos los aspectos éticos y deontológicos vinculados al uso de tests en el contexto general de la práctica profesional de la psicología.

3.1. El proceso de construcción de tests

El desarrollo de instrumentos de medida es un proceso fundamental para cualquier disciplina científica. Y no lo es menos para la psicometría, en la que, como ya hemos discutido ampliamente, la imposibilidad de observar y manipular los fenómenos objeto de interés añade una importante complejidad. A continuación abordaremos **diez fases fundamentales** en las que, a modo de conclusión, podemos organizar el proceso de construcción de un nuevo test. Estas fases nos permiten dar una perspectiva de conjunto a esta introducción a la psicometría y elaborar una síntesis de los conceptos más importantes que hemos ido discutiendo.

A pesar de que no son necesariamente secuenciales, y en todo caso no tienen por qué seguir estrictamente este orden, estas diez fases representan algunas de las decisiones más importantes que los profesionales interesados en el desarrollo de nuevos tests deben tomar. Como veremos más adelante, conocer este proceso es fundamental no solo para garantizar la calidad en el diseño y la construcción de nuevos tests, sino porque nos permite desarrollar algunos criterios importantes para la evaluación de los tests disponibles en la literatura. Así:

1) Delimitación de la finalidad del test. El desarrollo de un nuevo test empieza con una determinación clara del propósito para el que se pretende recoger información relevante en el contexto de la evaluación psicológica. Tal y como hemos señalado, los tests pueden servir a multitud de finalidades, pero entre ellas podemos destacar algunas importantes, como son describir, clasificar, diagnosticar, explicar o hacer predicciones sobre el comportamiento de las personas. Un propósito bien delimitado es la primera condición que hay que cumplir para garantizar el éxito en la construcción del nuevo test y permitirá su introducción en el proceso general de evaluación psicológica.

Para saber más

El estudiante interesado puede encontrar explicaciones más detalladas en Muñiz (1996), Muñiz y Fonseca-Pedrero (2008), Murphy y Davidshofer (2005), Downing (2006), Chadha (2009) o Rust y Golombok (2009).

2) Definición de los fenómenos psicológicos objeto de medida. De acuerdo con el proceso de inferencia psicométrica, el segundo paso para la construcción de un test consiste en la delimitación precisa de los fenómenos que se pretenden medir. Para hacerlo, recordémoslo, es necesaria una teoría sustantiva sólida y bien contrastada sobre los fenómenos psicológicos y los comportamientos de las personas que servirá de referencia en el proceso de la medida mediante el nuevo test. Este marco teórico es fundamental en el momento de planificar una representación adecuada de los fenómenos psicológicos y permite conocer en detalle otros instrumentos desarrollados previamente en la investigación psicológica. Evaluar sus limitaciones y sus puntos fuertes supone una buena guía para enfocar el trabajo necesario para dar soporte teórico al nuevo test.

3) Selección y muestreo de los comportamientos observables. Una vez establecido el fenómeno objeto de medida, la teoría sustantiva proporciona también el contexto necesario para elegir los comportamientos implicados que serán empleados como evidencias observables. Su representación adecuada es fundamental para no omitir ningún comportamiento relevante, así como para evitar incluir otros no directamente relacionados con el objeto de medida. En ocasiones resulta recomendable también llevar a cabo observaciones, entrevistas a informadores clave o grupos de discusión que proporcionen información complementaria sobre los comportamientos de interés a partir de las experiencias de los participantes. Como ya hemos discutido, esta no es una fase menos importante en cuanto a que es la base para el correcto desarrollo de los ítems que conformarán el nuevo test.

4) Especificación de las características del test. Delimitado el objetivo, definido el objeto de medida y seleccionadas las evidencias observables necesarias, el siguiente paso consiste en elegir las características del nuevo test para suscitar y medir adecuadamente los comportamientos de interés. En primer lugar, decidiremos si se trata de un test de habilidad o de potencia, en el que se evaluarán el acierto y el error en las respuestas para determinar la competencia, la aptitud o los conocimientos de los individuos, o de una prueba para evaluar motivaciones, preferencias, opiniones o actitudes. Es decir, si se trata de una prueba de ejecución máxima o de ejecución típica, como también son conocidas. De acuerdo con la clasificación de los tests que hemos hecho en esta introducción a la psicometría, es el momento también para decidir el formato del nuevo test, optando por una prueba de lápiz y papel, de manipulación o de medidas fisiológicas. Asimismo, se deben preparar los materiales necesarios y se determinará el tipo de administración más adecuada, eligiendo básicamente entre una prueba individual o de administración en grupo. Finalmente, también se decidirá el método más adecuado para interpretar las puntuaciones obtenidas, ya sea en base a la norma o en referencia a un criterio.

5) Desarrollo de los ítems que conformarán el test. Una vez especificadas las características generales del nuevo test, en esta fase se llevará a cabo el desarrollo de los elementos que contendrá. Esta no es una tarea sencilla y generalmen-

te supone la colaboración de un grupo de expertos en el campo para encontrar la mejor representación de las muestras de comportamiento seleccionadas. De hecho, no es poco habitual desarrollar muchos más ítems de los estrictamente necesarios para evaluar su comportamiento en el marco del test y seleccionar los más idóneos con relación al propósito del nuevo test. El objetivo final es disponer de tantos como sean necesarios para representar adecuadamente las diferentes dimensiones de los fenómenos psicológicos objeto de interés. Este es el momento también para decidir el formato que adoptarán las respuestas, que servirán para codificar los comportamientos de manera estructurada y de acuerdo con unas reglas claras.

6) Elección de una teoría de los tests. Cerrando el círculo establecido en el proceso de inferencia psicométrica, la siguiente fase en el proceso de construcción implica la elección de una teoría de los tests que, mediante un modelo de medida psicométrica determinado, servirá para relacionar los fenómenos psicológicos objeto de interés y las puntuaciones obtenidas mediante los ítems que conforman el nuevo test. Esta elección es capital teniendo en cuenta sus consecuencias en los métodos y técnicas empleadas posteriormente para evaluar las propiedades de los ítems y del test en su conjunto con el objetivo de garantizar la confianza en las inferencias establecidas sobre el comportamiento de las personas a partir de las puntuaciones obtenidas.

7) Realización de una prueba piloto. Una vez construido el test, incluyendo los ítems potenciales y el formato de las respuestas, es necesario redactar las instrucciones que lo acompañarán y definir las condiciones en las que será administrado. Una prueba piloto servirá para evaluar el grado de comprensión de estas instrucciones, la viabilidad de las condiciones para administrar el test, detectar posibles dificultades en el momento de registrar las respuestas y llevar a cabo un primer análisis de las propiedades, tanto de los ítems como del test en su conjunto, de acuerdo con los métodos y las técnicas indicadas por la teoría de los tests utilizada. Es el momento para valorar su comportamiento en el proceso de medida y, a partir de esta información, refinar el test modificándolo, descartando o añadiendo nuevos ítems. Pruebas adicionales pueden ser requeridas para evaluar correctamente las modificaciones introducidas antes de cerrar esta fase y proceder con el desarrollo del estudio de campo final.

8) Desarrollo del estudio de campo. Una vez establecido el test definitivo, el siguiente paso consiste en su administración a la población a la que se dirige. Para hacerlo, se selecciona la muestra de participantes necesarios, que, como ya hemos señalado, puede ser probabilística o no probabilística en función de si queremos representar la población de referencia o evaluar a un grupo de personas que cumplen unas determinadas características. Con esta información se desarrollan las normas o baremos para permitir la interpretación de las puntuaciones en relación con la ejecución del grupo de referencia. En el caso de que se trate de un test referido a criterio, en lugar de normas o baremos se determinan los puntos de corte que permitirán distinguir los diferentes grados de adecuación o consecución del criterio. Asimismo, se profundizará

en el trabajo sobre las propiedades psicométricas de los ítems y del test en su conjunto, atendiendo especialmente a la fiabilidad y validez de las medidas obtenidas. Esta es una parte fundamental del proceso de construcción de tests y es la que en última instancia garantiza que la medida psicométrica cumple con todas las exigencias científicas.

9) Elaboración del manual del test. Con toda esta información se desarrolla la documentación que acompañará al nuevo test, donde se ha de incluir información relevante relativa a las diferentes fases involucradas en su construcción: fundamentación teórica, finalidad y población a la que se dirige, instrucciones para la administración, información para la interpretación de las puntuaciones obtenidas y análisis de las propiedades psicométricas. Las primeras publicaciones en revistas científicas sirven para empezar a difundir toda esta información, que como ya hemos comentado no siempre acaba constituyendo un manual propiamente dicho.

10) Revisión y mejora del test. Con la publicación del test, ya sea licenciándolo o difundiéndolo libremente, el test se pone a disposición de la comunidad científica para obtener nuevas evidencias, que, mediante el trabajo independiente de diferentes investigadores, servirán para mejorar el conocimiento sobre su funcionamiento y sus propiedades psicométricas, así como para adaptarlo a otros entornos socioculturales o a otras poblaciones diferentes para las que ha sido desarrollado el nuevo test. Cambios en los fenómenos psicológicos objeto de medida, nuevos avances en la aproximación teórica a estos fenómenos o la adaptación del test a nuevas condiciones de administración o poblaciones de interés son algunas de las razones que justifican un trabajo de revisión y actualización prácticamente indefinida para refinar su funcionamiento y valorar adecuadamente su utilidad al servicio de los objetivos de la evaluación psicológica.

3.2. Criterios para la valoración de tests

Construir un test nuevo no es la práctica más habitual en el ejercicio profesional de la evaluación psicológica. De hecho, como hemos visto, es un trabajo laborioso que requiere un importante esfuerzo por parte de varios profesionales durante un periodo de tiempo prolongado para planificar, ejecutar y analizar los datos recogidos en uno o más estudios de campo. Por ello la mayoría de las situaciones en las que se utilizan los tests no comienzan necesariamente con la creación de uno original. En cambio, parten de la búsqueda y valoración de algún test existente en la literatura que encaje con los objetivos de la evaluación psicológica. Para simplificar este proceso, organizaciones vinculadas al ejercicio profesional de la psicología como la American Psychological Association (<http://www.apa.org>) o el Consejo General de Colegios Oficiales de Psicólogos (<http://www.cop.es>), sugieren algunas **fuentes para la búsqueda de tests**:

- **Publicaciones periódicas.** Con formato libro o enciclopedia, varias publicaciones se dedican a recopilar los tests comerciales disponibles con diferente nivel de detalle. Por ejemplo, el Buros Center for Testing de la Universidad de Nebraska edita *Tests in Print* (TIP, <http://buros.org/tests-print>) y *Mental Measurements Yearbook* (MMY, <http://buros.org/mental-measurements-yearbook>). TIP es una de las recopilaciones en lengua inglesa más exhaustivas que proporciona información relevante sobre los objetivos, las poblaciones a las que se dirigen, los autores y editoriales que los publican y el precio de más de 3.000 tests. Por su parte, MMY va más allá de la estricta recopilación y dedica sus esfuerzos a proveer diferentes evaluaciones sobre la calidad de los tests. Otra referencia internacional importante es Pro-Ed (<http://www.proedinc.com>), que edita también con carácter periódico sus *Test Publisher*, con el objetivo de recopilar los tests comerciales disponibles, y *Test Critiques Publisher*, que funciona como un recurso complementario al anterior y proporciona también información sobre su calidad.
- **Bases de datos.** Por otro lado, en la Red podemos encontrar algunos directorios abiertos que proporcionan información sobre los tests indexados, tanto comerciales como no comerciales, y facilitan su acceso. Entre otros, podemos destacar la *Test Collection* (http://www.ets.org/test_link), del Educational Testing Service, que incluye más de 25.000 tests. En este sentido, otra fuente interesante es *Test Reviews Online* (<http://buros.unl.edu/buros/jsp/search.jsp>), una base de datos desarrollada por el Buros Center for Testing, que permite la búsqueda electrónica entre los tests incorporados al MMY y que ofrece acceso comercial individual a sus evaluaciones.
- **Catálogos de las editoriales.** De especial interés para la búsqueda de tests en lengua no inglesa, y en especial en castellano, diferentes editoriales españolas publican los catálogos de tests que licencian y proveen de la información necesaria para conocer sus características. Entre estos, podemos destacar los catálogos de TEA Ediciones (<http://www.teaediciones.com>), del Grupo Albor-Cohs (<http://www.psicologia365.com>), del Instituto de Orientación Psicológica EOS (<http://www.eos.es>) y de Pearson (Pearson Clinical & Talent Assessment, <http://www.pearsonpsychcorp.es>).
- **Búsqueda de tests no comerciales.** Finalmente, debemos tener presente que no siempre los tests comerciales disponibles en el mercado se ajustan a las necesidades de la evaluación. En este caso, en la búsqueda de tests no comerciales para utilizarlos en la evaluación psicológica es conveniente consultar las bases de datos que incluyen las publicaciones científicas más relevantes en el ámbito de conocimiento de la psicología. Más allá de las recomendaciones habituales, podemos destacar *PsycTESTS* (<http://www.apa.org/pubs/databases/psyc-tests>), una base de datos especializada en tests no comerciales desarrollada por la American Psychological Association y actualizada con una periodicidad mensual, en la que se recopilan

tests obtenidos de los propios autores, de publicaciones en revistas científicas, libros o tesis.

Una vez encontrado uno o más tests potencialmente interesantes, es necesario evaluar sus características para **valorar la conveniencia de su uso** en relación con los objetivos de la evaluación psicológica. Para hacerlo, podemos volver al proceso de construcción de tests expuesto en el apartado anterior y ofrecer algunos criterios a partir de las preguntas más importantes que los profesionales interesados en su uso deben formularse en relación con los objetivos del test, sus propiedades psicométricas, el proceso de administración, su estatus y, finalmente, la adecuación específica al proceso de evaluación:

- **En relación con los objetivos del test.** Entre otras cuestiones, ¿está claramente definido el objetivo del test? ¿Dispone del soporte teórico necesario? ¿Disponemos de las evidencias científicas necesarias para utilizarlo como instrumento de medida? ¿Se actualiza regularmente o al menos es objeto de investigación para valorar sus propiedades psicométricas y su utilidad según los objetivos que se propone?
- **En relación con sus propiedades psicométricas.** De acuerdo con el proceso de construcción, ¿se proporciona la información necesaria sobre el análisis de los ítems que componen el test? ¿Qué evidencias se proporcionan sobre la fiabilidad y la validez de las tamaños obtenidas mediante el test? ¿Qué tipos de muestras, y qué tamaños se han utilizado para recoger estas evidencias?
- **En relación con el proceso de administración.** ¿Dispone el test de un manual adecuado con la información y los materiales necesarios para su uso? ¿Tiene instrucciones claras y/o plantillas para la corrección? Si fuera necesario, ¿este manual dispone de normas o baremos para la interpretación de las puntuaciones obtenidas?
- **En relación con su estatus.** ¿Se trata de un test comercial? Y en caso de serlo, ¿qué coste tienen el manual, las copias del test, las hojas de respuesta y la plantilla de corrección? ¿Es necesario algún tipo de calificación o acreditación para administrarlo? ¿Ha sido evaluado por alguna entidad u organismo independiente?
- **En relación con la adecuación al proceso de evaluación.** Finalmente, ¿es pertinente de acuerdo con los objetivos de evaluación? ¿Es su marco teórico congruente con estos objetivos? ¿Se dirige específicamente a la población a la que se quiere administrar? ¿Está adaptado o dispone de una versión adaptada al contexto cultural en el que quiere ser utilizado? ¿Se ajusta a las condiciones de administración previstas en el proceso de evaluación, por ejemplo, teniendo en cuenta el medio de administración, los

materiales necesarios, el formato de respuesta a los ítems o el tiempo requerido?

Evaluar un test antes de utilizarlo es más que una práctica conveniente en el ejercicio profesional de la psicología y, de hecho, supone un importante reto al que las propias organizaciones profesionales han dedicado diferentes esfuerzos para sistematizar su abordaje. Sin detenernos en sus especificidades, y teniendo en cuenta el diferente nivel de complejidad en el que han sido desarrolladas, estas iniciativas se basan también en el análisis de las decisiones tomadas en el proceso de diseño y construcción de los tests, y proporcionan un marco de evaluación de la calidad de los tests disponibles en la literatura muy útil para la valoración de su conveniencia con relación a los objetivos de la evaluación psicológica.

Así, en el contexto internacional, debemos destacar la iniciativa de la American Educational Research Association, la American Psychological Association y el National Council on Measurement in Education (1999), que desde 1985 publican conjuntamente sus *The standards for educational and psychological testing*. Por otro lado, la European Federation of Psychologists' Associations (2012) trabaja en el desarrollo del *EFPA Review model for the description and evaluation of psychological and educational tests* con la intención de armonizar los criterios y procesos de evaluación de la calidad de los tests a escala europea. Finalmente, en el contexto estatal, debemos hacer mención a la contribución del Consejo General de Colegios Oficiales de Psicólogos para adaptar el modelo del EFPA en “**Un modelo para evaluar la calidad de los tests utilizados en España**” (Prieto y Muñiz, 2000). Este modelo ya ha sido utilizado con éxito para evaluar la calidad de diez de los tests más utilizados por los psicólogos españoles (Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez y Peña-Suárez, 2011).

3.3. Aspectos éticos y deontológicos en el uso de tests

Como hemos discutido ampliamente a lo largo de esta introducción a la psicometría, el uso de tests como instrumentos de evaluación psicológica es una de las prácticas habituales en el ejercicio de la psicología. Sea al servicio de la intervención en los diferentes campos profesionales, sea en el centro del desarrollo de los campos académicos y científicos, los tests son unos instrumentos indispensables para la medida objetiva y estandarizada de los fenómenos psicológicos. Y como tales se encuentran sujetos a una reflexión sobre las **consecuencias éticas y morales que implica su uso**. Es decir, en cuanto que instrumentos, pueden ser utilizados o no de manera adecuada y de acuerdo con unas finalidades u otras. No es este el lugar para hacer una discusión filosófica en profundidad sobre lo que significa actuar bien o hacer las cosas correctamente, pero sí para concluir esta introducción reflexionando sobre algunas cuestiones importantes que tienen que ver con un uso de los tests responsable, justo y respetuoso con los derechos y la dignidad de las personas.

Para hacerlo, repasaremos algunos de los códigos de conducta más importantes desarrollados en el campo de la psicología que, en último término, presiden el correcto desarrollo de la actividad profesional, docente e investigadora. Y analizaremos el modo como abordan los aspectos éticos y morales vinculados al uso de tests en el contexto general de la evaluación psicológica. Estos códigos son la respuesta específica que la psicología ha dado a un debate complejo que, como otras disciplinas, ha afrontado a partir de la Segunda Guerra Mundial. Partiendo del Código de Nuremberg (1947), la Declaración de Helsinki (1964), el Informe Belmont (1979) y las directrices propuestas por el Council for International Organisations of Medical Sciences (1982), la consideración de las implicaciones éticas y morales del trabajo con humanos ha formado y forma parte de la práctica cotidiana de profesionales e investigadores¹³. En el ámbito concreto de la psicología, son muchos los códigos que las organizaciones profesionales de los diferentes países han desarrollado, pero en este texto nos referiremos a los trabajos de la American Psychological Association, la European Federation of Psychologists' Associations y, en el caso español, el Consejo General de Colegios Oficiales de Psicólogos. Finalmente, comentaremos unas directrices internacionales que, en referencia al caso específico del uso de los tests, ha desarrollado la International Test Commission.

⁽¹³⁾Para una interesante discusión sobre los orígenes, podéis ver Israel y Hay (2006).

Como una de las referencias más importantes a nivel internacional, la American Psychological Association ha contribuido al debate sobre las implicaciones éticas y morales en el campo de la psicología declarando uno de los primeros códigos deontológicos. A partir de los años cincuenta, y en sucesivas revisiones hasta la actualidad, esta asociación promueve lo que denomina ***Ethical principles of psychologists and code of conduct*** (American Psychological Association, 2010), donde establece un preámbulo, cinco principios generales y diez estándares que establecen los límites y las pautas adecuadas para una práctica profesional responsable. El preámbulo y los cinco principios se presentan como un posicionamiento ético y moral, es decir, un compromiso con los ideales de la psicología y no tanto como unas normas de obligado cumplimiento. En este sentido, el código reclama la aspiración por la beneficencia y no maleficencia, la fidelidad y responsabilidad, la integridad, la justicia y el respeto por los derechos y la dignidad de las personas.

En otras palabras, los psicólogos deben hacer todo lo posible para buscar y preservar el bienestar de las personas, evitando producir cualquier mal y velando por los derechos de aquellas personas con las que interactúan en los diferentes ámbitos de actividad profesional. Asimismo, deben desarrollar relaciones de confianza, responsabilizándose de su comportamiento y tratando de resolver activamente los conflictos de interés que puedan presentarse en el ejercicio profesional. Los psicólogos deben, además, actuar honestamente y con rigor, evitando en todo momento el engaño o el equívoco. Y tienen que hacerlo también con equidad y justicia, dedicando un especial cuidado a asegurar que ningún sesgo ni límite en su competencia o en su experiencia profesional contribuya o apoye el desarrollo de prácticas injustas. Finalmente, los psicólogos han de velar por los derechos de las personas a la privacidad, con-

fidencialidad y dignidad, respetando las diferencias culturales e individuales, y evitando contribuir o dar apoyo a cualquier tipo de prejuicio en sus prácticas profesionales.

Por su parte, los estándares se presentan como un conjunto de normas de obligado cumplimiento para los miembros de la asociación y los somete a las decisiones o sanciones que su comité ético pueda decidir en caso de violación del código en el ejercicio profesional. Estos estándares están descritos de manera general y con diferente nivel de articulación, con el objetivo de que puedan ser aplicados de manera transversal a los diferentes contextos de actividad profesional, docente o investigadora. Los estándares remiten a diez aspectos fundamentales, como son la resolución de los conflictos éticos, la competencia profesional, las relaciones humanas, la privacidad y la confidencialidad, la publicidad y las contribuciones públicas, el mantenimiento de la documentación profesional y los honorarios, la formación, la investigación y la publicación de los resultados, la evaluación y la terapia.

En este sentido, y de acuerdo con los objetivos de esta discusión sobre los aspectos éticos y deontológicos vinculados al uso de tests, el estándar relativo a la evaluación describe con detalle el marco general en el que esta debe ser llevada a cabo, y plantea la necesidad de basar cualquier juicio o conclusión en las evidencias obtenidas mediante el uso apropiado de los instrumentos de evaluación. Además, hace referencia explícita al uso de tests y plantea la necesidad de utilizar tests válidos y fiables, la obligación de poner a disposición de las personas sus respuestas y su interpretación de manera confidencial, el desarrollo de nuevos tests siempre de acuerdo con los estándares de calidad de la psicometría, la desestimación de tests obsoletos o no adecuados para la población objetivo y el cumplimiento de las condiciones y licencias de uso de los tests desarrollados por terceros.

Por su parte, la European Federation of Psychologists' Associations (2005) ha propuesto un *Meta-Code of Ethics*, que sirve como marco de referencia para el contexto europeo y pretende homogeneizar el tratamiento de las cuestiones éticas y deontológicas de las diferentes organizaciones profesionales de los Estados miembros. Para hacerlo, este código contiene un preámbulo general en el que establece su obligación de velar por el desarrollo profesional de los psicólogos europeos en todas estas cuestiones y propone cuatro principios fundamentales que, articulados en un conjunto de recomendaciones, deben regir los diferentes códigos, procedimientos y comités éticos nacionales.

Así, el código europeo plantea la defensa y el respeto por los derechos y la dignidad de las personas, el reconocimiento de la importancia de la competencia de los profesionales que ejercen la psicología, la responsabilidad frente a las personas, las comunidades y la sociedad en general y la promoción de la integridad en el desarrollo de la actividad profesional, docente e investigadora. Tanto los principios como las diferentes especificaciones se consideran interdependientes, por lo que se plantea abiertamente la necesidad de estable-

Para saber más

El estudiante interesado puede encontrar en Campbell, Vasquez, Behne y Kinscherff (2010) un comentario detallado de estos estándares, así como unos dilemas a modo de ejemplo de su aplicación a la práctica.

cer un debate y diálogo profesional ante la complejidad de los conflictos éticos que se presentan en la práctica. El contenido de su articulado reclama la atención de los psicólogos hacia cuestiones importantes relativas al ejercicio profesional cotidiano comparables a las establecidas en el código de la American Psychological Association, pero a diferencia de este, no hace ninguna referencia explícita a los procesos de evaluación psicológica ni al uso de tests en este contexto.

Por otro lado, en relación con el ejercicio profesional en el campo de la psicología en España, el Consejo General de Colegios Oficiales de Psicólogos (2010) ha desarrollado su *Código deontológico* como la plasmación de unos derechos y deberes profesionales que, en última instancia, sirven de base para el juicio de la conducta de sus colegiados (Bermejo, 2009, para una discusión de las modificaciones más recientes). Dispone de un título preliminar, a modo de preámbulo, que determina su alcance y enfatiza tanto el respeto fundamental por el marco normativo y jurídico, como la obligación de rechazar cualquier tipo de impedimento o limitación en el ejercicio profesional libre e independiente. Propone también unos principios generales que, a continuación, desarrolla en un conjunto de áreas fundamentales, y un anexo final recoge el reglamento de la comisión deontológica estatal encargada de velar por la correcta interpretación y aplicación del código. Así, entre los principios generales que deben regir la actividad profesional de los psicólogos en el Estado español, plantea:

- 1) La atención primordial al bienestar, la salud, la calidad de vida y la plenitud del desarrollo de las personas.
- 2) La protección de los derechos humanos y la responsabilidad en la fundamentación objetiva y científica de las intervenciones profesionales.
- 3) El no desarrollo o contribución a prácticas que atenten contra la libertad y la integridad física y psíquica de las personas.
- 4) La obligatoriedad de informar al menos a los organismos colegiales ante el conocimiento de cualquier violación de los derechos humanos o tratamiento degradante.
- 5) El respeto por los criterios morales y religiosos de las personas.
- 6) La no discriminación por cualquier diferencia o motivo.
- 7) La denegación de cualquier beneficio o provecho que se pueda extraer como consecuencia de las relaciones de poder o superioridad establecidas.
- 8) El uso de un lenguaje prudente y crítico ante etiquetas despreciativas o discriminatorias en sus informes.

9) El respeto por la actividad de los otros profesionales y la libre competencia.

10) La denuncia de prácticas ilegítimas o intrusivas.

11) La imparcialidad ante los posibles conflictos de interés entre los psicólogos o las instituciones en las que desarrollan su actividad y los intereses de las personas.

El articulado desarrolla estos principios en siete áreas fundamentales: la competencia profesional y la relación con otros profesionales, la intervención, la investigación y la docencia, la obtención y uso de la información, la publicidad, los honorarios y la remuneración, y las garantías procesales. Articuladas con diferente nivel de detalle, estas áreas no contemplan específicamente la evaluación psicológica ni hacen ninguna referencia explícita al uso de tests. Estas cuestiones quedan relegadas a un tratamiento genérico en las áreas de la competencia profesional y de la obtención y uso de la información, de manera que se plantea la necesidad de disponer de los conocimientos necesarios para el uso de los métodos, los instrumentos, las técnicas y los procedimientos, siempre de acuerdo con las evidencias científicas necesarias que garanticen un uso adecuado. Con relación a la obtención y uso de la información, circunscribe el uso de estos métodos, instrumentos, técnicas y procedimientos de acuerdo con las condiciones de confidencialidad y secreto profesional, así como reconoce el derecho de las personas a conocer sus resultados.

Finalmente, más allá de los códigos deontológicos profesionales, debemos hacer mención al trabajo específico que la International Test Commission (2000) ha desarrollado en sus *International guidelines for test use*. Dada la disparidad en la normativa y el nivel de desarrollo de los códigos profesionales de los diferentes países, esta organización internacional plantea unas directrices específicas sobre el uso de los tests y su papel en el contexto de la evaluación psicológica que, entre otros, subcribe el Consejo General de Colegios Oficiales de Psicólogos. Tal y como señala en su introducción, la intención de estas directrices no es añadir otro conjunto de recomendaciones a las ya existentes, sino que persigue crear una estructura coherente bajo la que se puedan entender y aplicar los diferentes códigos y estándares nacionales que desarrollan los aspectos éticos y deontológicos vinculados al uso de tests. Para hacerlo, estructura su articulado a partir de un propósito general que establece un uso apropiado, profesional y ético de los tests, con respeto a los derechos de las personas, las razones por las que se utilizan y el contexto en el que se aplican.

Asimismo, como gran eje conductor de sus directrices, propone la consecución de este propósito general mediante el desarrollo y la adquisición de las competencias necesarias para llevar a cabo la administración de tests, la interpretación y comunicación adecuadas de los resultados y la resolución de las difi-

cultades, los malentendidos y los conflictos que se puedan producir durante el proceso. De este modo, estas directrices establecen que los usuarios competentes de los tests deberían:

1) Responsabilizarse del uso ético de los tests:

- Actuando de manera profesional y ética.
- Asegurando que tienen los conocimientos y las habilidades necesarias.
- Haciéndose responsables del uso de los tests.
- Manteniendo de forma segura los materiales de los tests.
- Garantizando la confidencialidad de los resultados.

2) Comprometerse con las buenas prácticas en el uso de los tests:

- Estudiando la utilidad de los tests en los procesos de evaluación.
- Eligiendo tests muy fundamentados y apropiados para la situación.
- Atendiendo a las cuestiones relacionadas con la equidad.
- Preparando adecuadamente las condiciones de administración.
- Administrando correctamente los tests.
- Obteniendo las puntuaciones y analizándolas con exactitud.
- Interpretando los resultados correctamente.
- Comunicando los resultados de manera clara y exacta.
- Evaluando adecuadamente el funcionamiento y las propiedades de los tests.

Todas estas directrices están, a su vez, desarrolladas en un conjunto de recomendaciones concretas y representan un complejo retrato de las dificultades y los retos éticos y morales que plantea específicamente el uso de los tests como instrumentos de evaluación psicológica. Finalmente, el texto concluye con unos anexos, que desarrollan algunas indicaciones para el tratamiento específico de cuatro cuestiones importantes, como son el desarrollo de políticas y normativas sobre el uso de tests en las organizaciones y empresas, la redacción de contratos entre las diferentes partes involucradas en el uso de los tests, la administración de tests a personas con discapacidades y la traducción de las propias directrices a los diferentes idiomas por parte de las organizaciones estatales vinculadas al ejercicio profesional de la psicología.

Como hemos podido destacar, las cuestiones éticas y morales forman parte intrínseca del ejercicio profesional de la psicología. Así lo reflejan los diferentes códigos deontológicos que rigen la actividad de los psicólogos en Estados Unidos, Europa y España, mediante los cuales se reconoce el reto que supone afrontar de manera activa, responsable y comprometida la complejidad de las situaciones y los contextos en los que los profesionales intervienen de manera cotidiana. Por esta razón es obligación de los psicólogos exigirse, y exigir a los otros profesionales con quienes colaboran, un comportamiento ejemplar ante los dilemas éticos y morales que la práctica pueda conllevar. Los tests, instrumentos imprescindibles para la evaluación y la intervención psicológi-

cas, no se encuentran al margen de esta discusión a pesar de que, como hemos comentado, los códigos no siempre hacen un tratamiento específico de los procesos de evaluación y, concretamente, del papel que desempeñan los tests.

Como cualquier otro tipo de herramienta o instrumento, los tests pueden ser empleados al servicio de unas finalidades u otras, pero el compromiso debe ser la búsqueda constante e inequívoca del beneficio y no perjuicio de las personas, reaccionando ante cualquier tipo de impedimento que pueda limitar el ejercicio profesional honesto, libre e independiente. En definitiva, respetando el derecho de las personas a ser tratadas y evaluadas con justicia, equidad y responsabilidad, así como a conocer los resultados y las consecuencias que de la evaluación psicológica se puedan derivar. Como hemos podido ver, esta responsabilidad no se limita a la administración de tests con las garantías científicas necesarias, sino que implica el dominio de las teorías, los métodos y las técnicas que ha desarrollado la psicometría para comprender su funcionamiento, valorar su conveniencia con relación a los objetivos de la evaluación y, finalmente, evitar cualquier tipo de perjuicio que un uso inadecuado pudiera ocasionar. Por esa razón los psicólogos deben ser cuidadosos con su propia competencia, la clave de bóveda para un uso adecuado de los tests, con el objetivo de desarrollar los conocimientos y las habilidades necesarias que garanticen un profundo respeto por los derechos y la dignidad de las personas a quienes administran sus tests.

Bibliografía

American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: American Psychological Association. Disponible en línea en: <http://www.apa.org/ethics/code/principles.pdf>.

Anastasi, A. (1988). *Psychological testing* (6.^a ed.). New York: Macmillan.

Bermejo, V. (2009). Motivaciones para la revisión y cambios en el código deontológico de la profesión de la psicología. *Papeles del Psicólogo*, 30 (3), 195-206. Disponible en línea en: <http://www.papelesdelpsicologo.es/pdf/1748.pdf>

Bock, R. D. (1997). A brief history of Item response theory. *Educational Measurement: Issues and Practice*, 16 (4), 21-33.

Bock, R. D. y Jones, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.

Bondy, M. (1974). Psychiatric antecedents of psychological testing (before Binet). *Journal of the History of the Behavioral Sciences*, 10 (2), 180-194.

Boring, E. G. (1978). *Historia de la psicología experimental*. México: Trillas.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in modern psychometrics*. Cambridge: Cambridge University Press.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44 (3), 576-578.

Brennan, R. L. (1997). A Perspective on the history of Generalizability Theory. *Educational Measurement: Issues and Practice*, 16 (4), 14-20.

Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan.

Buchanan, R. D. y Finch, S. J. (2005). *History of psychometrics*. En B. S. Everitt y D. Howell (Ed.), *Encyclopedia of statistics in behavioral science* (pp. 875-878). Chichester: Wiley.

Campbell, L., Vasquez, M., Behne, S., y Kinscherff, R. (2010). *APA Ethics code commentary and case illustrations*. Washington, DC: American Psychological Association.

Campbell, N. R. (1920). *Physics: The elements*. Cambridge: Cambridge University Press.

Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. London: Longmans, Green, and co.

Carson, J. (1993). Army alpha, army brass, and the search for army intelligence. *Isis*, 84 (2), 278-309.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological bulletin*, 40 (3), 153-193.

Chadha, N. K. (2009). *Applied psychometry*. New Delhi: Sage.

Consejo General de Colegios Oficiales de Psicólogos (2010). *Código deontológico*. Madrid: Consejo General de Colegios Oficiales de Psicólogos. Disponible en línea en: <http://www.cop.es/pdf/codigo-deontologico-consejo-adaptacion-ley-omnibus.pdf>.

Crocker, L. y Algina, J. (2006). *Introduction to Classical and Modern test theory*. Mason: CENGAGE Learning.

Cronbach, L. J., Rajaratnam, N., y Gleser, G. C. (1963). Theory of Generalizability: A liberation of reliability theory. *British Journal of Mathematical and Statistical Psychology*, 16 (2), 137-163.

Dingle, H. (1950). A theory of measurement. *The British Journal for the Philosophy of Science*, 1 (1), 5-26.

Downing, S. M. (2006). *Twelve steps for effective test development*. En S. M. Downing y T. M. Haladyna (Ed.), *Handbook of test development* (pp. 3-25). Mahwah: Lawrence Erlbaum.

DuBois, P. H. (1970). *The history of psychological testing*. Boston: Allyn & Bacon.

European Federation of Psychologists' Associations (2005). *Meta-Code of Ethics*. Brussels: European Federation of Psychologists' Associations. Disponible en línea en: <http://www.efpa.eu/ethics/ethical-codes>.

European Federation of Psychologists' Associations (2012). *EFPA Review modelo for the description and evaluation of psychological and educational tests*. Versión 4.2.4. Brussels: European Federation of Psychologists' Associations. Disponible en línea en: <http://www.efpa.eu/professional-development>.

Ferguson, A., Myers, C. S., Bartlett, R. J., Banister, H., Bartlett, F. C., Brown, W., Campbell, N. R., Craik, K. J. W., Drever, J., Guild, J., Houstoun, R. A., Irwin, J. O., Kaye, G. W. C., Philpott, S. J. F., Richardson, L. F., Shaxby, J. H., Smith, T., Thouless, R. H., y Tucker, W. S. (1940). *Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events*. *Advancement of Science*, 1, 331-349.

Fernández-Ballesteros, R. (1997). Evaluación psicológica y tests. En A. Cordero (Ed.), *La evaluación psicológica en el año 2000* (pp. 11-26). Madrid: TEA Ediciones.

Fraser, C. O. (1980). Measurement in psychology. *British Journal of Psychology*, 71 (1), 23-34.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *History of Psychology*, 87 (3), 564-567.

García Cueto, E. (1993). *Introducción a la psicometría*. Madrid: Siglo Veintiuno de España Editoras.

Gibby, R. E. y Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11 (3), 164-184.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18 (8), 519-521.

Gleser, G. C., Cronbach, L. J., y Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika*, 30 (4), 395-418.

Goodenough, F. L. (1949). *Mental testing: Its history, principles, and applications*. New York: Rinehart.

Goslin, D. A. (1963). *The search for ability. Standardized testing in social perspective*. New York: Russell Sage Foundation.

de Gruijter, D. N. M. y van der Kamp, L. J. T. (2008). *Statistical test theory for the Behavioral Sciences*. Boca Raton: Chapman & Hall.

Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Willey.

Gulliksen, H. (1968). Louis Leon Thurstone, experimental and mathematical psychologist. *History of Psychology*, 23 (11), 786-802.

Hambleton, R. K. (1994). The rise and fall of criterion-referenced measurement? *Educational Measurement: Issues and Practice*, 13 (4), 21-26.

Hambleton, R. K., Swaminathan, H., y Rogers, H. J. (1991). *Fundamentals of Item response theory*. Newbury Park: Sage Publications.

Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159 (3), 445-492.

Hilgard, E. R. (1987). *Psychology in America: A historical survey*. Orlando: Harcourt.

Holden, R. R. (2000). *Psychometrics*. En A. E. Kazdin (Ed.), *Encyclopedia of psychology*, VI (pp. 417-419). New York: Oxford University Press.

International Test Commission (2000). *International guidelines for test use*. Disponible en línea en: <http://www.intestcom.org/upload/sitefiles/41.pdf>.

Kline, P. (1998). *The new psychometrics. Science, psychology, and measurement*. London: Routledge.

Jáñez, L. (1989). *Fundamentos de psicología matemática*. Madrid: Pirámide.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jones, L. V. (1971). *The nature of measurement*. En R. L. Thorndike (Ed.), *Educational measurement* (2.ª ed.) (pp. 335-355). Washington, DC: American Council donde Education.

Jones, L. V. y Thissen, D. (2007). *A history and overview of Psychometrics*. En C. R. Rao y S. Sinharay (Ed.), *Handbook of statistics 26. Psychometrics* (pp. 1-27). Amsterdam: Elsevier.

Kaplan, R. M. y Saccuzzo, D. P. (2001). *Psychological testing* (6.ª ed.). Belmont: Wadsworth Publishing Company.

Lord, F. M. (1953). On the statistical treatment of football numbers. *The American Psychologist*, 8 (12), 750-751.

Lord, F. M. y Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison Wesley.

Martin, O. (1997). La mesure en psychologie de Binet à Thurstone, 1900-1930. *Revue de synthèse*, 118 (4), 457-493.

Martínez Arias, M. R., Hernández Lloreda, M. J., y Hernández Lloreda, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.

Maydeu-Olivares, A. y McArdle, J. J. (2005). *Contemporary psychometrics*. Mahwah: Lawrence Erlbaum.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *History of Psychology*, 100 (3), 398-407.

Michell, J. (1999). *Measurement in psychology*. Cambridge: Cambridge University Press.

Muñiz, J. (1996). *Psicometría*. Madrid: Editorial Universitas.

Muñiz, J. (2003). *Teoría clásica de los tests*. Madrid: Pirámide.

Muñiz, J. (2010). Las teorías de los tests: Teoría clásica y Teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31 (1), 57-66. Disponible en línea en: <http://www.papelesdelpsicologo.es/pdf/1796.pdf>.

Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, A., y Peña-Suárez, E. (2011). Evaluación de tests editados en España. *Papeles del Psicólogo*, 32 (2), 113-128. Disponible en línea en: <http://www.papelesdelpsicologo.es/pdf/1947.pdf>.

Muñiz, J. y Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación*, 5, 13-25.

Muñiz, J. y Hambleton, R. K. (1992). Medio siglo de teoría de respuesta a los ítems. *Anuario de Psicología*, (52), 41-66.

Murphy, K. R. y Davidshofer, C. O. (2005). *Psychological testing* (6.ª ed.). Upper Saddle River: Pearson education.

Nicolas, S. y Ferrand, L. (2002). Alfred Binet and higher education. *History of Psychology*, 5 (3), 264-283.

Padilla, M., Merino, J. M., Rodríguez-Miñón, P., y Moreno, E. (1996). *Psicología matemática I*. Madrid: UNED.

Popham, W. J. y Husek, T. R. (1969). Implications of criterion-referenced measurement, *Journal of Educational Measurement*, 6 (1), 1-9.

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75. Disponible en línea en: <http://www.papelesdelpsicologo.es/vernumero.asp?id=1102>.

Rogers, T. B. (1995). *The psychological testing enterprise*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Rozeboom, W. W. (1966). Scaling theory and the nature of measurement. *Psychometrika*, 16 (2), 170-233.

Rust, J. y Golombok, S. (2009). *Modern psychometrics. The science of psychological assessment* (3.ª Ed.). London: Routledge.

Samejima, F. (2000). Psychometric Society. En A. E. Kazdin (Ed.), *Encyclopedia of psychology*, VI (pp. 419-420). New York: Oxford University Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 667-680.

Sokal, M. M. (1982). *James McKeen Cattell and the failure of anthropometric mental testing, 1890-1901*. En W. R. Woodward y M. G. Ash (Ed.), *The problematic science. Psychology in nineteenth-century thought* (pp. 322-345). New York: Praeger Publishers.

Suppes, P. (1951). A set of independent axioms for extensive quantities. *Portugaliae Mathematica*, 10, 163-172.

Swistak, P. (1990). Paradigms of measurement. *Theory and Decision*, 29 (1), 1-17.

Thurstone, L. L. (1931). *The reliability and validity of tests: Derivation and interpretation of fundamental formulae concerned with reliability and validity of tests and illustrative problems*. Ann Arbor: Edwards Brothers.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vector theory of the mind*. Chicago: Chicago University Press.

Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*, 16 (4), 8-14.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken: Wiley.

Valentine, E. R. (1999). The founding of the psychological laboratory, University College London: "Dear Galton...Yours truly, J Sully". *History of Psychology*, 2 (3), 204-218.

Velleman, P. F. y Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, 47 (1), 65-72.

Wolf, T. H. (1969). The emergence of Binet's conception and measurement of intelligence: A case history of the creative process. *Journal of the History of the Behavioral Sciences*, 5 (2), 113-134.

Yela, M. (1984). *Introducción a la teoría de los tests*. Madrid: Universidad Complutense.

Zeidner, J. y Drucker, A. J. (1988). *Behavioral science in the Army: A corporate history of the Army Research Institute*. Alejandría: United States Army Research Institute for the Behavioral and Social Sciences.