

Psychometrics

Psychometrics is a field of study concerned with the theory and technique of psychological measurement. As defined by the US National Council on Measurement in Education (NCME), psychometrics refers to psychological measurement. Generally, it refers to the field in psychology and education that is devoted to testing, measurement, assessment, and related activities.^[1]

The field is concerned with the objective measurement of skills and knowledge, abilities, attitudes, personality traits, and educational achievement. Some psychometric researchers focus on the construction and validation of assessment instruments such as questionnaires, tests, raters' judgments, and personality tests. Others focus on research relating to measurement theory (e.g., item response theory; intraclass correlation).

Practitioners are described as psychometricians. Psychometricians usually possess a specific qualification, and most are psychologists with advanced graduate training. In addition to traditional academic institutions, many psychometricians work for the government or in human resources departments. Others specialize as learning and development professionals.

Contents

Historical foundation

- Victorian stream
- German stream
- 20th century

Definition of measurement in the social sciences

Instruments and procedures

Theoretical approaches

- Key concepts

Standards of quality

- Testing standards
- Evaluation standards

Non-human: animals and machines

See also

References

- Bibliography
- Notes

Further reading

External links

Historical foundation

Psychological testing has come from two streams of thought: the first, from Darwin, Galton, and Cattell on the measurement of individual differences, and the second, from Herbart, Weber, Fechner, and Wundt and their psychophysical measurements of a similar construct. The second set of individuals and their research is what has led to

the development of experimental psychology, and standardized testing.^[2]

Victorian stream

Charles Darwin was the inspiration behind Sir Francis Galton who led to the creation of psychometrics. In 1859, Darwin published his book "The Origin of Species", which pertained to individual differences in animals. This book discussed how individual members in a species differ and how they possess characteristics that are more adaptive and successful or less adaptive and less successful. Those who are adaptive and successful are the ones that survive and give way to the next generation, who would be just as or more adaptive and successful. This idea, studied previously in animals, led to Galton's interest and study of human beings and how they differ one from another, and more importantly, how to measure those differences.

Galton wrote a book entitled "Hereditary Genius" about different characteristics that people possess and how those characteristics make them more "fit" than others. Today these differences, such as sensory and motor functioning (reaction time, visual acuity, and physical strength) are important domains of scientific psychology. Much of the early theoretical and applied work in psychometrics was undertaken in an attempt to measure intelligence. Galton, often referred to as "the father of psychometrics," devised and included mental tests among his anthropometric measures. James McKeen Cattell, who is considered a pioneer of psychometrics went on to extend Galton's work. Cattell also coined the term *mental test*, and is responsible for the research and knowledge which ultimately led to the development of modern tests. (Kaplan & Saccuzzo, 2010)

German stream

The origin of psychometrics also has connections to the related field of psychophysics. Around the same time that Darwin, Galton, and Cattell were making their discoveries, Herbart was also interested in "unlocking the mysteries of human consciousness" through the scientific method. (Kaplan & Saccuzzo, 2010) Herbart was responsible for creating mathematical models of the mind, which were influential in educational practices in years to come.

E.H. Weber built upon Herbart's work and tried to prove the existence of a psychological threshold, saying that a minimum stimulus was necessary to activate a sensory system. After Weber, G.T. Fechner expanded upon the knowledge he gleaned from Herbart and Weber, to devise the law that the strength of a sensation grows as the logarithm of the stimulus intensity. A follower of Weber and Fechner, Wilhelm Wundt is credited with founding the science of psychology. It is Wundt's influence that paved the way for others to develop psychological testing.^[2]

20th century

The psychometrician L. L. Thurstone, founder and first president of the Psychometric Society in 1936, developed and applied a theoretical approach to measurement referred to as the law of comparative judgment, an approach that has close connections to the psychophysical theory of Ernst Heinrich Weber and Gustav Fechner. In addition, Spearman and Thurstone both made important contributions to the theory and application of factor analysis, a statistical method developed and used extensively in psychometrics. In the late 1950s, Leopold Szondi made an historical and epistemological assessment of the impact of statistical thinking onto psychology during previous few decades: "in the last decades, the specifically psychological thinking has been almost completely suppressed and removed, and replaced by a statistical thinking. Precisely here we see the cancer of testology and testomania of today."^[3]

More recently, psychometric theory has been applied in the measurement of personality, attitudes, and beliefs, and academic achievement. Measurement of these unobservable phenomena is difficult, and much of the research and accumulated science in this discipline has been developed in an attempt to properly define and quantify such phenomena.

Critics, including practitioners in the physical sciences and social activists, have argued that such definition and quantification is impossibly difficult, and that such measurements are often misused, such as with psychometric personality tests used in employment procedures:

"For example, an employer wanting someone for a role requiring consistent attention to repetitive detail will probably not want to give that job to someone who is very creative and gets bored easily."^[4]

Figures who made significant contributions to psychometrics include Karl Pearson, Henry F. Kaiser, Carl Brigham, L. L. Thurstone, Anne Anastasi, Georg Rasch, Eugene Galanter, Johnson O'Connor, Frederic M. Lord, Ledyard R Tucker, Arthur Jensen, and David Andrich.

Definition of measurement in the social sciences

The definition of measurement in the social sciences has a long history. A currently widespread definition, proposed by Stanley Smith Stevens (1946), is that measurement is "the assignment of numerals to objects or events according to some rule." This definition was introduced in the paper in which Stevens proposed four levels of measurement. Although widely adopted, this definition differs in important respects from the more classical definition of measurement adopted in the physical sciences, namely that scientific measurement entails "the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute" (p. 358)^[5]

Indeed, Stevens's definition of measurement was put forward in response to the British Ferguson Committee, whose chair, A. Ferguson, was a physicist. The committee was appointed in 1932 by the British Association for the Advancement of Science to investigate the possibility of quantitatively estimating sensory events. Although its chair and other members were physicists, the committee also included several psychologists. The committee's report highlighted the importance of the definition of measurement. While Stevens's response was to propose a new definition, which has had considerable influence in the field, this was by no means the only response to the report. Another, notably different, response was to accept the classical definition, as reflected in the following statement:

Measurement in psychology and physics are in no sense different. Physicists can measure when they can find the operations by which they may meet the necessary criteria; psychologists have but to do the same. They need not worry about the mysterious differences between the meaning of measurement in the two sciences. (Reese, 1943, p. 49)

These divergent responses are reflected in alternative approaches to measurement. For example, methods based on covariance matrices are typically employed on the premise that numbers, such as raw scores derived from assessments, are measurements. Such approaches implicitly entail Stevens's definition of measurement, which requires only that numbers are *assigned* according to some rule. The main research task, then, is generally considered to be the discovery of associations between scores, and of factors posited to underlie such associations.

On the other hand, when measurement models such as the Rasch model are employed, numbers are not assigned based on a rule. Instead, in keeping with Reese's statement above, specific criteria for measurement are stated, and the goal is to construct procedures or operations that provide data that meet the relevant criteria. Measurements are estimated based on the models, and tests are conducted to ascertain whether the relevant criteria have been met.

Instruments and procedures

The first psychometric instruments were designed to measure the concept of intelligence.^[6] One historical approach involved the Stanford-Binet IQ test, developed originally by the French psychologist Alfred Binet. Intelligence tests are useful tools for various purposes. An alternative conception of intelligence is that cognitive capacities within individuals are a manifestation of a general component, or general intelligence factor, as well as cognitive capacity specific to a given domain.

Another major focus in psychometrics has been on personality testing. There have been a range of theoretical approaches to conceptualizing and measuring personality. Some of the better known instruments include the Minnesota Multiphasic Personality Inventory, the Five-Factor Model (or "Big 5") and tools such as Personality and Preference Inventory and the Myers-Briggs Type Indicator. Attitudes have also been studied extensively using psychometric approaches. A common method in the measurement of attitudes is the use of the Likert scale. An alternative method involves the application of unfolding measurement models, the most general being the Hyperbolic Cosine Model (Andrich & Luo, 1993).^[7]

Theoretical approaches

Psychometricians have developed a number of different measurement theories. These include classical test theory (CTT) and item response theory (IRT).^{[8][9]} An approach which seems mathematically to be similar to IRT but also quite distinctive, in terms of its origins and features, is represented by the Rasch model for measurement. The development of the Rasch model, and the broader class of models to which it belongs, was explicitly founded on requirements of measurement in the physical sciences.^[10]

Psychometricians have also developed methods for working with large matrices of correlations and covariances. Techniques in this general tradition include: factor analysis,^[11] a method of determining the underlying dimensions of data; multidimensional scaling,^[12] a method for finding a simple representation for data with a large number of latent dimensions; and data clustering, an approach to finding objects that are like each other. All these multivariate descriptive methods try to distill large amounts of data into simpler structures. More recently, structural equation modeling^[13] and path analysis represent more sophisticated approaches to working with large covariance matrices. These methods allow statistically sophisticated models to be fitted to data and tested to determine if they are adequate fits.

One of the main deficiencies in various factor analyses is a lack of consensus in cutting points for determining the number of latent factors. A usual procedure is to stop factoring when eigenvalues drop below one because the original sphere shrinks. The lack of the cutting points concerns other multivariate methods, also.

Key concepts

Key concepts in classical test theory are reliability and validity. A reliable measure is one that measures a construct consistently across time, individuals, and situations. A valid measure is one that measures what it is intended to measure. Reliability is necessary, but not sufficient, for validity.

Both reliability and validity can be assessed statistically. Consistency over repeated measures of the same test can be assessed with the Pearson correlation coefficient, and is often called *test-retest reliability*.^[14] Similarly, the equivalence of different versions of the same measure can be indexed by a Pearson correlation, and is called *equivalent forms reliability* or a similar term.^[14]

Internal consistency, which addresses the homogeneity of a single test form, may be assessed by correlating performance on two halves of a test, which is termed *split-half reliability*; the value of this Pearson product-moment correlation coefficient for two half-tests is adjusted with the Spearman–Brown prediction formula to correspond to the correlation

between two full-length tests.^[14] Perhaps the most commonly used index of reliability is Cronbach's α , which is equivalent to the mean of all possible split-half coefficients. Other approaches include the intra-class correlation, which is the ratio of variance of measurements of a given target to the variance of all targets.

There are a number of different forms of validity. Criterion-related validity can be assessed by correlating a measure with a criterion measure theoretically expected to be related. When the criterion measure is collected at the same time as the measure being validated the goal is to establish *concurrent validity*; when the criterion is collected later the goal is to establish *predictive validity*. A measure has *construct validity* if it is related to measures of other constructs as required by theory. *Content validity* is a demonstration that the items of a test do an adequate job of covering the domain being measured. In a personnel selection example, test content is based on a defined statement or set of statements of knowledge, skill, ability, or other characteristics obtained from a *job analysis*.

Item response theory models the relationship between latent traits and responses to test items. Among other advantages, IRT provides a basis for obtaining an estimate of the location of a test-taker on a given latent trait as well as the standard error of measurement of that location. For example, a university student's knowledge of history can be deduced from his or her score on a university test and then be compared reliably with a high school student's knowledge deduced from a less difficult test. Scores derived by classical test theory do not have this characteristic, and assessment of actual ability (rather than ability relative to other test-takers) must be assessed by comparing scores to those of a "norm group" randomly selected from the population. In fact, all measures derived from classical test theory are dependent on the sample tested, while, in principle, those derived from item response theory are not.

Many psychometricians are also concerned with finding and eliminating test bias from their psychological tests. Test bias is a form of systematic (i.e., non-random) error which leads to examinees from one demographic group having an unwarranted advantage over examinees from another demographic group.^[15] According to leading experts, test bias may cause differences in average scores across demographic groups, but differences in group scores are not sufficient evidence that test bias is actually present because the test could be measuring real differences among groups.^{[16][15]} Psychometricians use sophisticated scientific methods to search for test bias and eliminate it. Research shows that it is usually impossible for people reading a test item to accurately determine whether it is biased or not.^[17]

Standards of quality

The considerations of validity and reliability typically are viewed as essential elements for determining the quality of any test. However, professional and practitioner associations frequently have placed these concerns within broader contexts when developing standards and making overall judgments about the quality of any test as a whole within a given context. A consideration of concern in many applied research settings is whether or not the metric of a given psychological inventory is meaningful or arbitrary.^[18]

Testing standards

In 2014, the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) published a revision of the *Standards for Educational and Psychological Testing*,^[19] which describes standards for test development, evaluation, and use. The *Standards* cover essential topics in testing including validity, reliability/errors of measurement, and fairness in testing. The book also establishes standards related to testing operations including test design and development, scores, scales, norms, score linking, cut scores, test administration, scoring, reporting, score interpretation, test documentation, and rights and responsibilities of test takers and test users. Finally, the *Standards* cover topics related to testing applications, including psychological testing and assessment, workplace testing and credentialing, educational testing and assessment, and testing in program evaluation and public policy.

Evaluation standards

In the field of evaluation, and in particular educational evaluation, the [Joint Committee on Standards for Educational Evaluation](#)^[20] has published three sets of standards for evaluations. *The Personnel Evaluation Standards*^[21] was published in 1988, *The Program Evaluation Standards* (2nd edition)^[22] was published in 1994, and *The Student Evaluation Standards*^[23] was published in 2003.

Each publication presents and elaborates a set of standards for use in a variety of educational settings. The standards provide guidelines for designing, implementing, assessing and improving the identified form of evaluation.^[24] Each of the standards has been placed in one of four fundamental categories to promote educational evaluations that are proper, useful, feasible, and accurate. In these sets of standards, validity and reliability considerations are covered under the accuracy topic. For example, the student accuracy standards help ensure that student evaluations will provide sound, accurate, and credible information about student learning and performance.

Non-human: animals and machines

Psychometrics addresses *human* abilities, attitudes, traits and educational evolution. Notably, the study of behavior, mental processes and abilities of non-human *animals* is usually addressed by [comparative psychology](#), or with a continuum between non-human animals and the rest of animals by [evolutionary psychology](#). Nonetheless there are some advocates for a more gradual transition between the approach taken for humans and the approach taken for (non-human) animals.^{[25][26][27] [28]}

The evaluation of abilities, traits and learning evolution of *machines* has been mostly unrelated to the case of humans and non-human animals, with specific approaches in the area of [artificial intelligence](#). A more integrated approach, under the name of universal psychometrics, has also been proposed.^[29]

See also

- [Cattell–Horn–Carroll theory](#)
- [Classical test theory](#)
- [Computational psychometrics](#)
- [Concept inventory](#)
- [Cronbach's alpha](#)
- [Data mining](#)
- [Educational assessment](#)
- [Educational psychology](#)
- [Factor analysis](#)
- [Item response theory](#)
- [List of psychometric software](#)
- [List of schools for psychometrics](#)
- [Operationalisation](#)
- [Quantitative psychology](#)
- [Psychometric Society](#)
- [Rasch model](#)
- [Scale \(social sciences\)](#)
- [School counselor](#)
- [School psychology](#)
- [Standardized test](#)

References

Bibliography

- Andrich, D. & Luo, G. (1993). "A hyperbolic cosine model for unfolding dichotomous single-stimulus responses" (<http://apm.sagepub.com/content/19/3/269.full.pdf>) (PDF). *Applied Psychological Measurement*. **17** (3): 253–276. CiteSeerX 10.1.1.1003.8107 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1003.8107>). doi:10.1177/014662169301700307 (<https://doi.org/10.1177%2F014662169301700307>).
- Michell, J. (1999). *Measurement in Psychology*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511490040
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Reese, T.W. (1943). "The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples". *Psychological Monographs*. **55**: 1–89.
- Stevens, S. S. (1946). "On the theory of scales of measurement". *Science*. **103** (2684): 677–80. Bibcode:1946Sci...103..677S (<http://adsabs.harvard.edu/abs/1946Sci...103..677S>). doi:10.1126/science.103.2684.677 (<https://doi.org/10.1126%2Fscience.103.2684.677>). PMID 17750512 (<https://www.ncbi.nlm.nih.gov/pubmed/17750512>).
- Thurstone, L.L. (1927). "A law of comparative judgement". *Psychological Review*. **34** (4): 278–286. doi:10.1037/h0070288 (<https://doi.org/10.1037%2Fh0070288>).
- Thurstone, L.L. (1929). The Measurement of Psychological Value. In T.V. Smith and W.K. Wright (Eds.), *Essays in Philosophy by Seventeen Doctors of Philosophy of the University of Chicago*. Chicago: Open Court.
- Thurstone, L.L. (1959). *The Measurement of Values*. Chicago: The University of Chicago Press.
- S.F. Blinkhorn (1997). "Past imperfect, future conditional: fifty years of test theory". *Br. J. Math. Statist. Psychol.* **50** (2): 175–185. doi:10.1111/j.2044-8317.1997.tb01139.x (<https://doi.org/10.1111%2Fj.2044-8317.1997.tb01139.x>).

Notes

1. National Council on Measurement in Education
http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorP Archived (https://web.archive.org/web/20170722194028/http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorP) 2017-07-22 at the Wayback Machine
2. Kaplan, R.M., & Saccuzzo, D.P. (2010). *Psychological Testing: Principles, Applications, and Issues*. (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.
3. Leopold Szondi (1960) *Das zweite Buch: Lehrbuch der Experimentellen Triebdiagnostik*. Huber, Bern und Stuttgart, 2nd edition. Ch.27, From the Spanish translation, B)II *Las condiciones estadísticas*, p.396. Quotation:

el pensamiento psicologico especifico, en las ultima decadas, fue suprimido y eliminado casi totalmente, siendo sustituido por un pensamiento estadistico. Precisamente aqui vemos el cáncer de la testología y testomania de hoy.

4. Psychometric Assessments. *Psychometric Assessments* . (<http://www.services.unimelb.edu.au/careers/student/interviews/test.html>) University of Melbourne.
5. Michell, Joel (August 1997). "Quantitative science and the definition of measurement in psychology". *British Journal of Psychology*. **88** (3): 355–383. doi:10.1111/j.2044-8295.1997.tb02641.x (<https://doi.org/10.1111%2Fj.2044-8295.1997.tb02641.x>).

6. "Los diferentes tipos de tests psicometricos - examen psicometrico" (<http://examenpsicometrico.com/los-diferentes-tipos-de-tests-psicometricos/>). *examenpsicometrico.com*.
7. Andrich, D. & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, 17, 253-276.
8. Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
9. Hambleton, R.K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff.
10. Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
11. Thompson, B.R. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. American Psychological Association.
12. Davison, M.L. (1992). *Multidimensional Scaling*. Krieger.
13. Kaplan, D. (2008). *Structural Equation Modeling: Foundations and Extensions*, 2nd ed. Sage.
14. "Home - Educational Research Basics by Del Siegle" (<http://www.gifted.uconn.edu/Siegle/research/Instrument+Reliability+and+Validity/Reliability.htm>). *www.gifted.uconn.edu*.
15. Warne, Russell T.; Yoon, Myeongsun; Price, Chris J. (2014). "Exploring the various interpretations of "test bias"". *Cultural Diversity and Ethnic Minority Psychology*. **20** (4): 570–582. doi:10.1037/a0036503 (<https://doi.org/10.1037/a0036503>). PMID 25313435 (<https://www.ncbi.nlm.nih.gov/pubmed/25313435>).
16. Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, 6, 144-150. doi:10.1037/1076-8971.6.1.144
17. Reschly, D. J. (1980) Psychological evidence in the *Larry P.* opinion: A case of right problem-wrong solution? *School Psychology Review*, 9, 123-125.
18. Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. (<http://psychology.tamu.edu/Faculty/blanton/bj.2006.arbitrary.pdf>) Archived (<https://web.archive.org/web/20060510182915/http://psychology.tamu.edu/Faculty/blanton/bj.2006.arbitrary.pdf>) 2006-05-10 at the Wayback Machine *American Psychologist*, 61(1), 27-41.
19. "The Standards for Educational and Psychological Testing" (<http://www.apa.org/science/standards.html#overview>). <http://www.apa.org>. External link in |website= (help)
20. Joint Committee on Standards for Educational Evaluation (<http://www.wmich.edu/evalctr/jc/>) Archived (<https://web.archive.org/web/20091015044732/http://www.wmich.edu/evalctr/jc/>) 2009-10-15 at the Wayback Machine
21. Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to Assess Systems for Evaluating Educators*. (<http://www.wmich.edu/evalctr/jc/PERSTNDS-SUM.htm>) Archived (<https://web.archive.org/web/20051212001638/http://www.wmich.edu/evalctr/jc/PERSTNDS-SUM.htm>) 2005-12-12 at the Wayback Machine Newbury Park, CA: Sage Publications.
22. Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards, 2nd Edition*. (<http://www.wmich.edu/evalctr/jc/PGMSTNDS-SUM.htm>) Archived (<https://web.archive.org/web/20060222025348/http://www.wmich.edu/evalctr/jc/PGMSTNDS-SUM.htm>) 2006-02-22 at the Wayback Machine Newbury Park, CA: Sage Publications.
23. Committee on Standards for Educational Evaluation. (2003). *The Student Evaluation Standards: How to Improve Evaluations of Students*. (<http://www.wmich.edu/evalctr/jc/briefing/ses/>) Archived (<https://web.archive.org/web/20060524144621/http://www.wmich.edu/evalctr/jc/briefing/ses/>) 2006-05-24 at the Wayback Machine Newbury Park, CA: Corwin Press.
24. [E. Cabrera-Nguyen. "Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research]" (<https://www.academia.edu/2395969>). *Academia.edu*. **1** (2): 99–103.
25. Humphreys, L.G. (1987). "Psychometrics considerations in the evaluation of intraspecies differences in intelligence". *Behav Brain Sci*. **10** (4): 668–669. doi:10.1017/s0140525x0005514x (<https://doi.org/10.1017/s0140525x0005514x>).
26. Eysenck, H.J. (1987). "The several meanings of intelligence". *Behav Brain Sci*. **10** (4): 663. doi:10.1017/s0140525x00055060 (<https://doi.org/10.1017/s0140525x00055060>).

27. Locurto, C. & Scanlon, C (1987). "Individual differences and spatial learning factor in two strains of mice". *Behav Brain Sci.* **112**: 344–352.
28. King, James E & Figueredo, Aurelio Jose (1997). "The five-factor model plus dominance in chimpanzee personality". *Journal of Research in Personality.* **31** (2): 257–271. doi:[10.1006/jrpe.1997.2179](https://doi.org/10.1006/jrpe.1997.2179) (<https://doi.org/10.1006%2Fjrpe.1997.2179>).
29. J. Hernández-Orallo; D.L. Dowe; M.V. Hernández-Lloreda (2013). "Universal Psychometrics: Measuring Cognitive Abilities in the Machine Kingdom". *Cognitive Systems Research.* **27**: 50–74. doi:[10.1016/j.cogsys.2013.06.001](https://doi.org/10.1016/j.cogsys.2013.06.001) (<http://doi.org/10.1016%2Fj.cogsys.2013.06.001>). hdl:[10251/50244](https://hdl.handle.net/10251/50244) (<https://hdl.handle.net/10251%2F50244>).

Further reading

- Robert F. DeVellis (2016). *Scale Development: Theory and Applications* (<https://books.google.com/books?id=48ACCwAAQBAJ>). SAGE Publications. ISBN 978-1-5063-4158-3.
- Borsboom, Denny (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press. ISBN 978-0-521-84463-5. Lay summary (<http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=978-0-521-84463-5>) (28 June 2010).
- Leslie A. Miller; Robert L. Lovler (2015). *Foundations of Psychological Testing: A Practical Approach* (<https://books.google.com/books?id=8EYdCAAAQBAJ>). SAGE Publications. ISBN 978-1-4833-6927-3.
- Roderick P. McDonald (2013). *Test Theory: A Unified Treatment* (https://books.google.com/books?id=_feqA2RdyOoC). Psychology Press. ISBN 978-1-135-67530-1.
- Paul Kline (2000). *The Handbook of Psychological Testing* (<https://books.google.com/books?id=lm2RxaKaok8C>). Psychology Press. ISBN 978-0-415-21158-1.
- Rush AJ Jr; First MB; Blacker D (2008). *Handbook of Psychiatric Measures* (<https://books.google.com/books?id=ddyZUcaGaRIC>). American Psychiatric Publishing. ISBN 978-1-58562-218-4. OCLC 85885343 (<https://www.worldcat.org/oclc/85885343>).
- Ann C Silverlake (2016). *Comprehending Test Manuals: A Guide and Workbook* (<https://books.google.com/books?id=zI8PDQAAQBAJ>). Taylor & Francis. ISBN 978-1-351-97086-0.
- Fenton H (2019). "Top 10 Tips on how to prepare for a psychometric test to get that job!" (<https://boti.co.za/top-10-tips-prepare-psychometric-test-get-job>). Business Optimization Training Institute.
- Dr. Snigdha Rai (2018). "An Ultimate Guide to Psychometric Tests" (<https://mettl.com/en/psychometric-tests/psychometric-guide/>). Mercer Mettl.

External links

- APA Standards for Educational and Psychological Testing (<http://www.apa.org/science/standards.html>)
- International Personality Item Pool (<http://ipip.ori.org/>)
- Joint Committee on Standards for Educational Evaluation (<https://web.archive.org/web/20091015044732/http://www.wmich.edu/evalctr/jc/>)
- The Psychometrics Centre, University of Cambridge (<http://www.psychometrics.cam.ac.uk/>) ^[1]
- Psychometric Society and Psychometrika homepage (<http://www.psychometrika.org/>)
- London Psychometric Laboratory (<http://www.psychometriclab.com/>)

1. Sanford, David (18 November 2017). "Cambridge just told me Big Data doesn't work yet" (<https://www.linkedin.com/pulse/cambridge-just-told-me-big-data-doesnt-work-yet-david-sanford/>). *LinkedIn*.

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Psychometrics&oldid=909870155>"

This page was last edited on 8 August 2019, at 04:27 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

