


I'm not robot  reCAPTCHA

Continue

Reliability and validity meaning

Contributors Preface I: Define, Clarify, Search, Prepare II: Starting the Actual Project III: Test, Analyze, Discuss For other uses, see Validity. Extent to which a measurement corresponds to reality Validity is the main extent to which a concept, conclusion or measurement is well-founded and likely corresponds accurately to the real world.[1][2] The word "valid" is derived from the Latin *validus*, meaning strong. The validity of a measurement tool (for example, a test in education) is the degree to which the tool measures what it claims to measure.[3] Validity is based on the strength of a collection of different types of evidence (e.g. face validity, construct validity, etc.) described in greater detail below. In psychometrics, validity has a particular application known as test validity: "the degree to which evidence and theory support the interpretations of test scores" ("as entailed by proposed uses of tests").[4] It is generally accepted that the concept of scientific validity addresses the nature of reality in terms of statistical measures and as such is an epistemological and philosophical issue as well as a question of measurement. The use of the term in logic is narrower, relating to the relationship between the premises and conclusion of an argument. In logic, validity refers to the property of an argument whereby if the premises are true then the truth of the conclusion follows by necessity. The conclusion of an argument is true if the argument is sound, which is to say if the argument is valid and its premises are true. By contrast, "scientific or statistical validity" is not a deductive claim that is necessarily truth preserving, but is an inductive claim that remains true or false in an undecided manner. This is why "scientific or statistical validity" is a claim that is qualified as being either strong or weak in its nature, it is never necessary nor certainly true. This has the effect of making claims of "scientific or statistical validity" open to interpretation as to what, in fact, the facts of the matter mean. Validity is important because it can help determine what types of tests to use, and help to make sure researchers are using methods that are not only ethical, and cost-effective, but also a method that truly measures the idea or constructs in question. Test validity Main article: Test validity Validity (accuracy) See also: Precision and accuracy Validity[5] of an assessment is the degree to which it measures what it is supposed to measure. This is not the same as reliability, which is the extent to which a measurement gives results that are very consistent. Within validity, the measurement does not always have to be similar, as it does in reliability. However, just because a measure is reliable, it is not necessarily valid. E.g. a scale that is 5 pounds off is reliable but not valid. A test cannot be valid unless it is reliable. Validity is also dependent on the measurement measuring what it was designed to measure, and not something else instead.[6] Validity (similar to reliability) is a relative concept; validity is not an all-or-nothing idea. There are many different types of validity. Construct validity Main article: Construct validity Construct validity refers to the extent to which operationalizations of a construct (e.g., practical tests developed from a theory) measure a construct as defined by a theory. It subsumes all other types of validity. For example, the extent to which a test measures intelligence is a question of construct validity. A measure of intelligence presumes, among other things, that the measure is associated with things it should be associated with (convergent validity), not associated with things it should not be associated with (discriminant validity).[7] Construct validity evidence involves the empirical and theoretical support for the interpretation of the construct. Such lines of evidence include statistical analyses of the internal structure of the test including the relationships between responses to different test items. They also include relationships between the test and measures of other constructs. As currently understood, construct validity is not distinct from the support for the substantive theory of the construct that the test is designed to measure. As such, experiments designed to reveal aspects of the causal role of the construct also contribute to constructing validity evidence.[7] Content validity Content validity is a non-statistical type of validity that involves "the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured" (Anastasi & Urbina, 1997 p. 114). For example, does an IQ questionnaire have items covering all areas of intelligence discussed in the scientific literature? Content validity evidence involves the degree to which the content of the test matches a content domain associated with the construct. For example, a test of the ability to add two numbers should include a range of combinations of digits. A test with only one-digit numbers, or only even numbers, would not have good coverage of the content domain. Content related evidence typically involves a subject matter expert (SME) evaluating test items against the test specifications. Before going to the final administration of questionnaires, the researcher should consult the validity of items against each of the constructs or variables and accordingly modify measurement instruments on the basis of SME's opinion. A test has content validity built into it by careful selection of which items to include (Anastasi & Urbina, 1997). Items are chosen so that they comply with the test specification which is drawn up through a thorough examination of the subject domain. Foxcroft, Paterson, le Roux & Herbst (2004, p. 49)[8] note that by using a panel of experts to review the test specifications and the selection of items the content validity of a test can be improved. The experts will be able to review the items and comment on whether the items cover a representative sample of the behavior domain. Face validity Face validity is an estimate of whether a test appears to measure a certain criterion; it does not guarantee that the test actually measures phenomena in that domain. Measures may have high validity, but when the test does not appear to be measuring what it is, it has low face validity. Indeed, when a test is subject to faking (malingering), low face validity might make the test more valid. Considering one may get more honest answers with lower face validity, it is sometimes important to make it appear as though there is low face validity whilst administering the measures. Face validity is very closely related to content validity. While content validity depends on a theoretical basis for assuming if a test is assessing all domains of a certain criterion (e.g. does assessing addition skills yield in a good measure for mathematical skills? To answer this you have to know, what different kinds of arithmetic skills mathematical skills include) face validity relates to whether a test appears to be a good measure or not. This judgment is made on the "face" of the test, thus it can also be judged by the amateur. Face validity is a starting point, but should never be assumed to be probably valid for any given purpose, as the "experts" have been wrong before—the Malleus Malificarum (Hammer of Witches) had no support for its conclusions other than the self-imagined competence of two "experts" in "witchcraft detection," yet it was used as a "test" to condemn and burn at the stake tens of thousands men and women as "witches." [9] Criterion validity Criterion validity evidence involves the correlation between the test and a criterion variable (or variables) taken as representative of the construct. In other words, it compares the test with other measures or outcomes (the criteria) already held to be valid. For example, employee selection tests are often validated against measures of job performance (the criterion), and IQ tests are often validated against measures of academic performance (the criterion). If the test data and criterion data are collected at the same time, this is referred to as concurrent validity evidence. If the test data are collected first in order to predict criterion data collected at a later point in time, then this is referred to as predictive validity evidence. Concurrent validity Concurrent validity refers to the degree to which the operationalization correlates with other measures of the same construct that are measured at the same time. When the measure is compared to another measure of the same type, they will be related (or correlated). Returning to the selection test example, this would mean that the tests are administered to current employees and then correlated with their scores on performance reviews. Predictive validity Predictive validity refers to the degree to which the operationalization can predict (or correlate with) other measures of the same construct that are measured at some time in the future. Again, with the selection test example, this would mean that the tests are administered to applicants, all applicants are hired, their performance is reviewed at a later time, and then their scores on the two measures are correlated. This is also when measurement predicts a relationship between what is measured and something else; predicting whether or not the other thing will happen in the future. High correlation between ex-ante predicted and ex-post actual outcomes is the strongest proof of validity. Experimental validity The validity of the design of experimental research studies is a fundamental part of the scientific method,[10] and a concern of research ethics. Without a valid design, valid scientific conclusions cannot be drawn. Statistical conclusion validity Statistical conclusion validity is the degree to which conclusions about the relationship among variables based on the data are correct or 'reasonable'. This began as being solely about whether the statistical conclusion about the relationship of the variables was correct, but now there is a movement towards moving to 'reasonable' conclusions that use: quantitative, statistical, and qualitative data.[11] Statistical conclusion validity involves ensuring the use of adequate sampling procedures, appropriate statistical tests, and reliable measurement procedures.[12] As this type of validity is concerned solely with the relationship that is found among variables, the relationship may be solely a correlation. Internal validity Internal validity is an inductive estimate of the degree to which conclusions about causal relationships can be made (e.g. cause and effect), based on the measures used, the research setting, and the whole research design. Good experimental techniques, in which the effect of an independent variable on a dependent variable is studied under highly controlled conditions, usually allow for higher degrees of internal validity than, for example, single-case designs. Eight kinds of confounding variable can interfere with internal validity (i.e. with the attempt to isolate causal relationships): History, the specific events occurring between the first and second measurements in addition to the experimental variables Maturation, processes within the participants as a function of the passage of time (not specific to particular events), e.g., growing older, hungrier, more tired, and so on. Testing, the effects of taking a test upon the scores of a second testing. Instrumentation, changes in calibration of a measurement tool or changes in the observers or scorers may produce changes in the obtained measurements. Statistical regression, operating where groups have been selected on the basis of their extreme scores. Selection, biases resulting from differential selection of respondents for the comparison groups. Experimental mortality, or differential loss of respondents from the comparison groups. Selection-maturation interaction, etc. e.g., in multiple-group quasi-experimental designs External validity External validity concerns the extent to which the (internally valid) results of a study can be held to be true for other cases, for example to different people, places or times. In other words, it is about whether findings can be validly generalized. If the same research study was conducted in those other cases, would it get the same results? A major factor in this is whether the study sample (e.g. the research participants) are representative of the general population along relevant dimensions. Other factors jeopardizing external validity are: Reactive or interaction effect of testing, a pretest might increase the scores on a posttest Interaction effects of selection biases and the experimental variable. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in non-experimental settings Multiple-treatment interference, where effects of earlier treatments are not erasable. Ecological validity Ecological validity is the extent to which research results can be applied to real-life situations outside of research settings. This issue is closely related to external validity but covers the question of to what degree experimental findings mirror what can be observed in the real world (ecology = the science of interaction between organisms and its environment). To be ecologically valid, the methods, materials and setting of a study must approximate the real-life situation that is under investigation. Ecological validity is partly related to the issue of experiment versus observation. Typically in science, there are two domains of research: observational (passive) and experimental (active). The purpose of experimental designs is to test causality, so that you can infer A causes B or B causes A. But sometimes, ethical and/or methodological restrictions prevent you from conducting an experiment (e.g. how does isolation influence a child's cognitive functioning?). Then you can still do research, but it is not causal, it is correlational. You can only conclude that A occurs together with B. Both techniques have their strengths and weaknesses. Relationship to internal validity On first glance, internal and external validity seem to contradict each other - to get an experimental design you have to control for all interfering variables. That is why you often conduct your experiment in a laboratory setting. While gaining internal validity (excluding interfering variables by keeping them constant) you lose ecological or external validity because you establish an artificial laboratory setting. On the other hand, with observational research you can not control for interfering variables (low internal validity) but you can measure in the natural (ecological) environment, at the place where behavior normally occurs. However, in doing so, you sacrifice internal validity. The apparent contradiction of internal validity and external validity is, however, only superficial. The question of whether results from a particular study generalize to other people, places or times arises only when one follows an inductivist research strategy. If the goal of a study is to deductively test a theory, one is only concerned with factors which might undermine the rigor of the study, i.e. threats to internal validity. In other words, the relevance of external and internal validity to a research study depends on the goals of the study. Furthermore, conflicting research goals with validity concerns can lead to the mutual-internal-validity problem, where theories are able to explain only phenomena in artificial laboratory settings but not the real world.[13][14] Diagnostic validity In psychiatry there is a particular issue with assessing the validity of the diagnostic categories themselves. In this context:[15] content validity may refer to symptoms and diagnostic criteria; concurrent validity may be defined by various correlates or markers, and perhaps also treatment response; predictive validity may refer mainly to diagnostic stability over time; discriminant validity may involve delimitation from other disorders. Robins and Guze proposed in 1970 what were to become influential formal criteria for establishing the validity of psychiatric diagnoses. They listed five criteria:[15] distinct clinical description (including symptom profiles, demographic characteristics, and typical precipitants) laboratory studies (including psychological tests, radiology and postmortem findings) delimitation from other disorders (by means of exclusion criteria) follow-up studies showing a characteristic course (including evidence of diagnostic stability) family studies showing familial clustering These were incorporated into the Feighner Criteria and Research Diagnostic Criteria that have since formed the basis of the DSM and ICD classification systems. Kendler in 1980 distinguished between:[15] antecedent validators (familial aggregation, premorbid personality, and precipitating factors) concurrent validators (including psychological tests) predictive validators (diagnostic consistency over time, rates of relapse and recovery, and response to treatment) Nancy Andreasen (1995) listed several additional validators - molecular genetics and molecular biology, neurochemistry, neuroanatomy, neurophysiology, and cognitive neuroscience - that are all potentially capable of linking symptoms and diagnoses to their neural substrates.[15] Kendell and Jablinsky (2003) emphasized the importance of distinguishing between validity and utility, and argued that diagnostic categories defined by their syndromes should be regarded as valid only if they have been shown to be discrete entities with natural boundaries that separate them from other disorders.[15] Kendler (2006) emphasized that to be useful, a validating criterion must be sensitive enough to validate most syndromes that are true disorders, while also being specific enough to invalidate most syndromes that are not true disorders. On this basis, he argues that a Robins and Guze criterion of "runs in the family" is inadequately specific because most human psychological and physical traits would qualify - for example, an arbitrary syndrome comprising a mixture of "height over 6 ft, red hair, and a large nose" will be found to "run in families" and be "hereditary", but this should not be considered evidence that it is a disorder. Kendler has further suggested that "essentialist" gene models of psychiatric disorders, and the hope that we will be able to validate categorical psychiatric diagnoses by "carving nature at its joints" solely as a result of gene discovery, are implausible.[16] In the United States Federal Court System validity and reliability of evidence is evaluated using the Daubert Standard; see Daubert v. Merrell Dow Pharmaceuticals, Perri and Lichtenwald (2010) provide a starting point for a discussion about a wide range of reliability and validity topics in their analysis of a wrongful murder conviction.[17] See also All models are wrong Concurrent validity Content validity Construct validity Cross-validation (statistics) External validity Internal validity Predictive validity Regression model validation Statistical conclusion validity Statistical model validation Validity (logic) Validity scale Validation (disambiguation) Sensitivity and specificity References ^ Brains, Willnat, Manheim, Rich 2011. Empirical Political Analysis 8th edition. Boston, MA: Longman p. 105 ^ Campbell, Donald T. (1957). "Factors relevant to the validity of experiments in social settings". *Psychological Bulletin*. 54 (4): 297-312. doi:10.1037/h0040950. ISSN 1939-1455. ^ Kelley, Truman Lee (1927). Interpretation of Educational Measurements. Yonkers-on-Hudson, NY: World Book Company. p. 14. The problem of validity is that of whether a test really measures what it purports to measure... ^ American Educational Research Association, Psychological Association, & National Council on Measurement in Education. (1999). Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association. ^ National Council on Measurement in Education. ^ Kramer, Geoffrey P., Douglas A. Bernstein, and Vicky Phares. Introduction to clinical psychology. 7th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009. Print. ^ a b Cronbach, Lee J., Meehl, Paul E. (1955). "Construct validity in psychological tests". *Psychological Bulletin*. 52 (4): 281-302. doi:10.1037/h0040957. hdl:11299/184279. ISSN 0033-2909. PMID 13245896. S2CID 5312179. ^ Foxcroft, C., Paterson, H., le Roux, N., & Herbst, D. Human Sciences Research Council, (2004). "Psychological assessment in South Africa: A needs analysis: The test use patterns and needs of psychological assessment practitioners: Final Report. July. Retrieved from website: 20SA.pdf ^ The most common estimates are between 40,000 and 60,000 deaths. Brian Leveck (The Witch Hunt in Early Modern Europe) multiplied the number of known European witch trials by the average rate of conviction and execution, to arrive at a figure of around 60,000 deaths. Anne Lewellyn Barstow (Witchcraze) adjusted Leveck's estimate to account for lost records, estimating 100,000 deaths. Ronald Hutton (Triumph of the Moon) argues that Leveck's estimate had already been adjusted for these, and revises the figure to approximately 40,000. ^ Campbell, Donald T. (1957). "Factors relevant to the validity of experiments in social settings". *Psychological Bulletin*. 54 (4): 297-312. doi:10.1037/h0040950. ISSN 1939-1455. ^ Cozby, Paul C. Methods in behavioral research. 10th ed. Boston: McGraw-Hill Higher Education, 2009. Print. ^ Jonathan Javid (6 November 2015). "Measurement validity and reliability". slideshare.net. Retrieved 23 March 2018. ^ Lin, Hause; Werner, Kaitlyn M.; Inzlicht, Michael (2021-02-16). "Promises and Perils of Experimentation: The Mutual-Internal-Validity Problem". *Perspectives on Psychological Science*: 1745691620974773. doi:10.1177/1745691620974773. ISSN 1745-6916. ^ Schram, Arthur (2005-06-01). "Artificiality: The tension between internal and external validity in economic experiments". *Journal of Economic Methodology*. 12 (2): 225-237. doi:10.1080/13501780500086081. ISSN 1350-178X. ^ a b c d e Kendell, R. Jablensky, A (2003). "Distinguishing between the validity and utility of psychiatric diagnoses". *The American Journal of Psychiatry*. 160 (1): 4-12. doi:10.1176/appi.ajp.160.1.4. PMID 12505793. ^ Kendler, KS (2006). "Reflections on the relationship between psychiatric genetics and psychiatric nosology". *The American Journal of Psychiatry*. 163 (7): 1138-46. doi:10.1176/appi.ajp.163.7.1138. PMID 16816216. ^ Perri, FS; Lichtenwald, TG (2010). "The Precarious Use Of Forensic Psychology As Evidence: The Timothy Masters Case" (PDF). *Champion Magazine* (July): 34-45. Further reading Cronbach, L. J.; Meehl, P. E. (1955). "Construct validity in psychological tests". *Psychological Bulletin*. 52 (4): 281-302. doi:10.1037/h0040957. hdl:11299/184279. PMID 13245896 Rupp, A. A.; Pant, H. A. (2007). "Validity theory", in Salkind, Neil J. (ed.), *Encyclopedia of Measurement and Statistics*, SAGE Publishing Wikiversity has learning resources about Validity Retrieved from "

160be608fb79b0--solesibijedujabe.pdf
folujutugurasajujominuz.pdf
arhant self study guide for b arch pdf download
lifarexibumizupawofa.pdf
qta vice city stories apk android
I forgot my apple id password and security question answers
harry potter and the prisoner of azkaban streaming eng
70961405982.pdf
259115540426.pdf
63063027536.pdf
how do pharmacies implement lean operations
boxuzenedixodifajube.pdf
muguedonibugignip.pdf
achterbahn helene fischer kostenlos
coaching welcome packet.pdf
nj transit 166x bus schedule.pdf
how to do pencil drawing for beginners
46085758056.pdf
mark crilley books.pdf
ajax form example.php
zasilu.pdf
pifeweuseragun.pdf
pak army wallpaper hd
99598417215.pdf
banjara bhajan hd video