

# Validity and Reliability of the Research Instrument

Subjects: Management | Education & Educational Research | Information Science & Library Science

Contributor: Hamed Taherdoost

Questionnaire is one of the most widely used tools to collect data in especially social science research. The main objective of questionnaire in research is to obtain relevant information in most reliable and valid manner. Thus the accuracy and consistency of survey/questionnaire forms a significant aspect of research methodology which are known as validity and reliability. Often new researchers are confused with selection and conducting of proper validity type to test their research instrument (questionnaire/survey).

Keywords: Research Instrument ; Questionnaire ; Survey Validity ; Questionnaire Reliability ; questionnaire validity ; discriminant validity ; construct validity ; survey reliability ; convergent validity ; research methods ; research methodology

---

## 1. Introduction

Validity explains how well the collected data covers the actual area of investigation<sup>[1]</sup>. Validity basically means “measure what is intended to be measured”<sup>[2]</sup>.

## 2. Face Validity

Face validity is a subjective judgment on the operationalization of a construct. Face validity is the degree to which a measure appears to be related to a specific construct, in the judgment of non-experts such as test takers and representatives of the legal system. That is, a test has face validity if its content simply looks relevant to the person taking the test. It evaluates the appearance of the questionnaire in terms of feasibility, readability, consistency of style and formatting, and the clarity of the language used.

In other words, face validity refers to researchers' subjective assessments of the presentation and relevance of the measuring instrument as to whether the items in the instrument appear to be relevant, reasonable, unambiguous and clear<sup>[3]</sup>.

In order to examine the face validity, the dichotomous scale can be used with categorical option of “Yes” and “No” which indicate a favourable and unfavourable item respectively. Where favourable item means that the item is objectively structured and can be positively classified under the thematic category. Then the collected data is analysed using Cohen's Kappa Index (CKI) in determining the face validity of the instrument. DM. et al.<sup>[4]</sup> recommended a minimally acceptable Kappa of 0.60 for inter-rater agreement. Unfortunately, face validity is arguably the weakest form of validity and many would suggest that it is not a form of validity in the strictest sense of the word.

## 3. Content Validity

Content validity is defined as “the degree to which items in an instrument reflect the content universe to which the instrument will be generalized” (Straub, Boudreau et al.<sup>[5]</sup>). In the field of IS, it is highly recommended to apply content validity while the new instrument is developed. In general, content validity involves evaluation of a new survey instrument in order to ensure that it includes all the items that are essential and eliminates undesirable items to a particular construct domain<sup>[6]</sup>. The judgemental approach to establish content validity involves literature reviews and then follow-ups with the evaluation by expert judges or panels. The procedure of judgemental approach of content validity requires researchers to be present with experts in order to facilitate validation. However it is not always possible to have many experts of a particular research topic at one location. It poses a limitation to conduct validity on a survey instrument when experts are located in different geographical areas (Choudrie and Dwivedi<sup>[7]</sup>). Contrastingly, a quantitative approach may allow researchers to send content validity questionnaires to experts working at different locations, whereby distance is not a limitation. In order to apply content validity following steps are followed:

1. An exhaustive literature reviews to extract the related items.

2.A content validity survey is generated (each item is assessed using three point scale (not necessary, useful but not essential and essential)).

3.The survey should sent to the experts in the same field of the research.

4.The content validity ratio (CVR) is then calculated for each item by employing Lawshe<sup>[8]</sup> (1975) 's method.

5.Items that are not significant at the critical level are eliminated. In following the critical level of Lawshe method is explained.

## **4. Construct Validity**

If a relationship is causal, what are the particular cause and effect behaviours or constructs involved in the relationship? Construct validity refers to how well you translated or transformed a concept, idea, or behaviour that is a construct into a functioning and operating reality, the operationalization. Construct validity has two components: convergent and discriminant validity.

### **4.1 Discriminant Validity**

Discriminant validity is the extent to which latent variable A discriminates from other latent variables (e.g., B, C, D). Discriminant validity means that a latent variable is able to account for more variance in the observed variables associated with it than a) measurement error or similar external, unmeasured influences; or b) other constructs within the conceptual framework. If this is not the case, then the validity of the individual indicators and of the construct is questionable (Fornell and Larcker<sup>[9]</sup>). In brief, Discriminant validity (or divergent validity) tests that constructs that should have no relationship do, in fact, not have any relationship.

### **4.2 Convergent Validity**

Convergent validity, a parameter often used in sociology, psychology, and other behavioural sciences, refers to the degree to which two measures of constructs that theoretically should be related, are in fact related. In brief, Convergent validity tests that constructs that are expected to be related are, in fact, related.

With the purpose of verifying the construct validity (discriminant and convergent validity), a factor analysis can be conducted utilizing principal component analysis (PCA) with varimax rotation method (Koh and Nam<sup>[10]</sup>, Wee and Quazi, <sup>[10]</sup>). Items loaded above 0.40, which is the minimum recommended value in research are considered for further analysis. Also, items cross loading above 0.40 should be deleted. Therefore, the factor analysis results will satisfy the criteria of construct validity including both the discriminant validity (loading of at least 0.40, no cross-loading of items above 0.40) and convergent validity (eigenvalues of 1, loading of at least 0.40, items that load on posited constructs) (Straub et al., <sup>[11]</sup>). There are also other methods to test the convergent and discriminant validity.

## **5. Criterion Validity**

Criterion or concrete validity is the extent to which a measure is related to an outcome. It measures how well one measure predicts an outcome for another measure. A test has this type of validity if it is useful for predicting performance or behavior in another situation (past, present, or future).

Criterion validity is an alternative perspective that de-emphasizes the conceptual meaning or interpretation of test scores. Test users might simply wish to use a test to differentiate between groups of people or to make predictions about future outcomes. For example, a human resources director might need to use a test to help predict which applicants are most likely to perform well as employees. From a very practical standpoint, she focuses on the test's ability to differentiate good employees from poor employees. If the test does this well, then the test is "valid" enough for her purposes. From the traditional three-faceted view of validity, criterion validity refers to the degree to which test scores can predict specific criterion variables. The key to validity is the empirical association between test scores and scores on the relevant criterion variable, such as "job performance."

Messick<sup>[12]</sup> suggests that "even for purposes of applied decision making, reliance on criterion validity or content coverage is not enough. The meaning of the measure, and hence its construct validity, must always be pursued – not only to support test interpretation but also to justify test use". There are two types of criterion validity namely; concurrent validity, predictive and postdictive validity.

## 6. Reliability

Reliability concerns the extent to which a measurement of a phenomenon provides stable and consistent result (Carmines and Zeller<sup>[13]</sup>). Reliability is also concerned with repeatability. For example, a scale or test is said to be reliable if repeat measurement made by it under constant conditions will give the same result (Moser and Kalton<sup>[14]</sup>).

Testing for reliability is important as it refers to the consistency across the parts of a measuring instrument (Huck<sup>[15]</sup>). A scale is said to have high internal consistency reliability if the items of a scale “hang together” and measure the same construct (Huck<sup>[16]</sup> Robinson<sup>[17]</sup>). The most commonly used internal consistency measure is the Cronbach Alpha coefficient. It is viewed as the most appropriate measure of reliability when making use of Likert scales (Whitley<sup>[18]</sup>, Robinson<sup>[19]</sup>). No absolute rules exist for internal consistencies, however most agree on a minimum internal consistency coefficient of .70 (Whitley<sup>[20]</sup>, Robinson<sup>[21]</sup>).

For an exploratory or pilot study, it is suggested that reliability should be equal to or above 0.60 (Straub et al.<sup>[22]</sup>). Hinton et al.<sup>[23]</sup> have suggested four cut-off points for reliability, which includes excellent reliability (0.90 and above), high reliability (0.70-0.90), moderate reliability (0.50-0.70) and low reliability (0.50 and below)(Hinton et al.,<sup>[24]</sup>). Although reliability is important for study, it is not sufficient unless combined with validity. In other words, for a test to be reliable, it also needs to be valid<sup>[25]</sup>.

---

## References

1. ACKOFF, R. L. 1953. *The Design of Social Research*, Chicago, University of Chicago Press.
2. BARTLETT, J. E., KOTRLIK, J. W. & HIGGINS, C. C. 2001. Organizational research: determining appropriate sample size in survey research. *Learning and Performance Journal*, 19, 43-50.
3. BOUDREAU, M., GEFEN, D. & STRAUB, D. 2001. Validation in IS research: A state-of-the-art assessment. *MIS Quarterly*, 25, 1-24.
4. BREWETON, P. & MILLWARD, L. 2001. *Organizational Research Methods*, London, SAGE.
5. BROWN, G. H. 1947. A comparison of sampling methods. *Journal of Marketing*, 6, 331-337.
6. BRYMAN, A. & BELL, E. 2003. *Business research methods*, Oxford, Oxford University Press.
7. CARMINES, E. G. & ZELLER, R. A. 1979. *Reliability and Validity Assessment*, Newbury Park, CA, SAGE.
8. CHOUDRIE, J. & DWIVEDI, Y. K. Investigating Broadband Diffusion in the Household: Towards Content Validity and Pre-Test of the Survey Instrument. *Proceedings of the 13th European Conference on Information Systems (ECIS 2005)*, May 26-28, 2005 Regensburg, Germany.
9. DAVIS, D. 2005. *Business Research for Decision Making*, Australia, Thomson South-Western.
10. DM., G., DP., H., CC., C., CL., S. & ., P. B. 1975. The effects of instructional prompts and praise on children's donation rates. *Child Development* 46, 980-983.
11. ENGELLANT, K., HOLLAND, D. & PIPER, R. 2016. Assessing Convergent and Discriminant Validity of the Motivation Construct for the Technology Integration Education (TIE) Model. *Journal of Higher Education Theory and Practice* 16, 37-50.
12. FIELD, A. P. 2005. *Discovering Statistics Using SPSS*, Sage Publications Inc.
13. FORNELL, C. & LARCKER, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
14. FOWLER, F. J. 2002. *Survey research methods*, Newbury Park, CA, SAGE.
15. GHAURI, P. & GRONHAUG, K. 2005. *Research Methods in Business Studies*, Harlow, FT/Prentice Hall.
16. GILL, J., JOHNSON, P. & CLARK, M. 2010. *Research Methods for Managers*, SAGE Publications.
17. HINTON, P. R., BROWNLOW, C., MCMURRAY, I. & COZENS, B. 2004. *SPSS explained*, East Sussex, England, Routledge Inc.
18. HUCK, S. W. 2007. *Reading Statistics and Research*, United States of America, Allyn & Bacon.
19. KOH, C. E. & NAM, K. T. 2005. Business use of the internet: a longitudinal study from a value chain perspective. *Industrial Management & Data Systems*, 105 85-95.
20. LAWSHE, C. H. 1975. A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.

21. LEWIS, B. R., SNYDER, C. A. & RAINER, K. R. 1995. An empirical assessment of the Information Resources Management construct. *Journal of Management Information Systems*, 12, 199-223.
22. MALHOTRA, N. K. & BIRKS, D. F. 2006. *Marketing Research: An Applied Approach*, Harlow, FT/Prentice Hall.
23. MAXWELL, J. A. 1996. *Qualitative Research Design: An Intractive Approach* London, Applied Social Research Methods Series.
24. MESSICK, S. 1989. Validity. In: LINN, R. L. (ed.) *Educational measurement*. New York: Macmillan.
25. MOSER, C. A. & KALTON, G. 1989. *Survey methods in social investigation*, Aldershot, Gower.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/49364>