International Journal of Academic Research in Management (IJARM)
Vol. 5, No. 3, 2016, Page: 28-36, ISSN: 2296-1747

www.elvedit.com

Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research

Authors

Hamed Taherdoost
Research and Development Department, Hamta Business Solution Sdn Bhd
Research and Development Department, Ahoora Ltd | Management Consultation Group

hamed.taherdoost@gmail.com
Kuala Lumpur, Malaysia

Abstract

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). This review article explores and describes the validity and reliability of a questionnaire/survey and also discusses various forms of validity and reliability tests.

Key Words

Research Instrument, Questionnaire, Survey, Survey Validity, Questionnaire Reliability, Content Validity, Face Validity, Construct Validity, and Criterion Validity.

I. Introduction

Validity explains how well the collected data covers the actual area of investigation (Ghauri and Gronhaug, 2005). Validity basically means "measure what is intended to be measured" (Field, 2005). In this paper, main types of validity namely; face validity, content validity, construct validity, criterion validity and reliability are discussed. Figure 1 shows the subtypes of various forms of validity tests exploring and describing in this article.

Electronic copy available at: https://ssrn.com/abstract=3205040

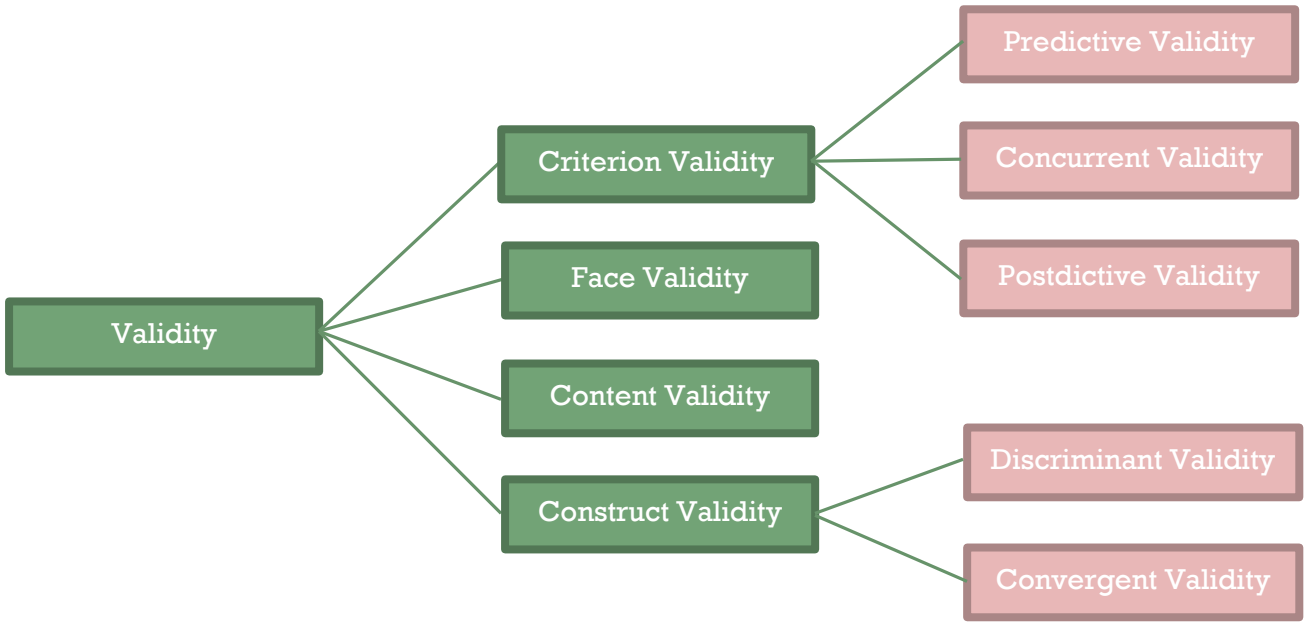

FIGURE 1: SUBTYPES OF VARIOUS FORMS OF VALIDITY TESTS

II. 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 (Oluwatayo, 2012).

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. (1975) 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.

Electronic copy available at: https://ssrn.com/abstract=3205040

III. 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. 2004). 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 (Lewis et al., 1995, Boudreau et al., 2001). 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. This poses a limitation to conduct validity on a survey instrument when experts are located in different geographical areas (Choudrie and Dwivedi, 2005). 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 (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.

CVR; Lawshe's Method

The CVR (content validity ratio) proposed by Lawshe (1975) is a linear transformation of a proportional level of agreement on how many "experts" within a panel rate an item "essential" calculated in the following way:

$$\text{CVR} = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}}$$

where CVR is the content validity ratio, ne is the number of panel members indicating "essential," and N is the total number of panel members. The final evaluation to retain the item based on the CVR is depends on the number of panels. Table 1 shows the guideline for the valid value of CVR for the evaluated item to be retained.

Electronic copy available at: https://ssrn.com/abstract=3205040

TABLE 1 : MINIMUM VALUE OF CVR, P = .05, SOURCE: (LAWSHE, 1975)

No. of Panellists	Minimum Value
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

IV. 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.

A. 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, 1981). In brief, Discriminant validity (or divergent validity) tests that constructs that should have no relationship do, in fact, not have any relationship.

B. 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

Electronic copy available at: https://ssrn.com/abstract=3205040

rotation method (Koh and Nam, 2005, Wee and Quazi, 2005). 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., 2004). There are also other methods to test the convergent and discriminant validity.

V. 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. From this perspective, the key to validity is the empirical association between test scores and scores on the relevant criterion variable, such as "job performance."

Messick (1989) 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.

A. Predictive Validity

The survey is predictively valid if the test accurately predicts what it is supposed to predict. It can also refer to when scores from the predictor measure are taken first and then the criterion data is collected later.in other words, the ability of one assessment tool to predict future performance either in some activity or on another assessment of the same construct. The best way to directly establish predictive validity is to perform a long-term validity study. For example, by administering employment tests to job applicants and then seeing if those test scores are correlated with the future job performance of the hired employees. Predictive validity studies take a long time to complete and require fairly large sample sizes in order to acquire meaningful aggregate data. In brief, predictive validity assesses the operationalization's ability to predict something it should theoretically be able to predict.

Electronic copy available at: https://ssrn.com/abstract=3205040

B. Concurrent Validity

Concurrent validity is a type of evidence that can be gathered to defend the use of a test for predicting other outcomes. It refers to the extent to which the results of a particular test, or measurement, correspond to those of a previously established measurement for the same construct. In brief, concurrent validity assesses the operationalization's ability to distinguish between groups that it should theoretically be able to distinguish between.

C. Postdictive Validity

For this type of validity, the criterion is in the past. That is, the criterion (e.g., another test) was administered in the past. It is a form of criterion-referenced validity that is determined by the degree to which the scores on a given test are related to the scores on another, already established test or criterion administered at a previous point in time.

VI. Reliability

Reliability concerns the extent to which a measurement of a phenomenon provides stable and consist result (Carmines and Zeller, 1979). 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, 1989).

Testing for reliability is important as it refers to the consistency across the parts of a measuring instrument (Huck, 2007). A scale is said to have high internal consistency reliability if the items of a scale "hang together" and measure the same construct (Huck, 2007, Robinson, 2009). 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, 2002, Robinson, 2009). No absolute rules exist for internal consistencies, however most agree on a minimum internal consistency coefficient of .70 (Whitley, 2002, Robinson, 2009).

For an exploratory or pilot study, it is suggested that reliability should be equal to or above 0.60 (Straub et al., 2004). Hinton et al. (2004) 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., 2004). 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 (Wilson, 2010). Table 2 compares the validity components.

Electronic copy available at: https://ssrn.com/abstract=3205040

TABLE 2: COMPARISON OF VALIDITIES THAT ARE UNDERTAKEN IN THIS RESEARCH, SOURCE: STRAUB ET AL. (2004)(NETEMEYER ET AL., 2003)(VISWANATHAN, 2005)(ENGELLANT ET AL., 2016)

Validity Component	Definition	Type	Technique Suggested
Face Validity	The extent that measurement instrument items linguistically and analytically look like what is supposed to be measured	Recommended	Post hoc theory, expert assessment of items; Cohen's Kappa Index (CKI)
Content Validity	The extent that measurement instrument items are relevant and representative of the target construct	Highly recommended	Literature review; expert panels or judges; CVRs; Q-sorting
Construct Discriminant validity	the extent that measures of different constructs diverge or minimally correlate with one another	Mandatory	MTMM; PCA; CFA; PLS AVE; Q-sorting
Construct Convergent validity	The extent that different measures of the same construct converge or strongly correlate with one another	Mandatory	MTMM; PCA; CFA; Q-sorting
Criterion Predictive Validity	the extent that a measure predicts another measure	Mandatory	Regression Analysis, Discriminant Analysis
Criterion Concurrent Validity	the extent that a measure simultaneously relates to another measure that it is supposed to relate	Mandatory	Correlation Analysis
Criterion Postdictive Validity	The extent that a measure is related to the scores on another, already established in past.	Mandatory	Correlation Analysis
Reliability Internal consistency	the extent to which a measurement of a phenomenon provides stable and consist result	Mandatory	Cronbach's a; correlations; SEM reliability coefficients

VII. CONCLUSION

In this paper, validity and reliability of questionnaire/survey as a significant research instrument tool were reviewed. Various types of validity were discussed with the goal of validity improving the skills and knowledge of survey validity tests among researchers. As discussed, there are four main validity test of the questionnaire namely; face validity, content validity, construct validity and criterion validity. Depends on the types of questionnaire, some of these validity tests are mandatory to apply and some recommended (as shown in Table 2).

Electronic copy available at: https://ssrn.com/abstract=3205040

ACKNOWLEDGMENT

This research was prepared under support of Research and Development Department of Hamta Business Solution Sdn Bhd and Ahoora Ltd | Management Consultation Group.

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 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.
[26] NETEMEYER, R. G., BEARDEN, W. O. & SHARMA, S. 2003. *Scaling procedures: Issues and applications,* Thousand Oaks, CA: Sage.
[27] OLUWATAYO, J. 2012. Validity and reliability issues in educational research. *Journal of Educational and Social Research* 2**,** 391-400.
[28] ROBINSON, J. 2009. *Triandis theory of interpersonal behaviour in understanding software privace behaviour in the South African context.* Masters degree, University of the Witwatersrand.
[29] STRAUB, D., BOUDREAU, M.-C. & GEFEN, D. 2004. Validation guidelines for IS positivist research. *Communications of the Association for Information Systems,* 13**,** 380-427.
[30] VISWANATHAN, M. 2005. *Measurement error and research design,* Thousand Oaks, CA: Sage. .
[31] WEE, Y. S. & QUAZI, H. A. 2005. Development and validation of critical factors of environmental management. *Industrial Management & Data Systems,* 105**,** 96-114.

Electronic copy available at: https://ssrn.com/abstract=3205040

[32] WHITLEY, B. E. 2002. *Principals of Research and Behavioural Science,* Boston, McGraw-Hill.
[33] WILSON, J. 2010. *Essentials of business research: a guide to doing your research project,* SAGE Publication.
[34] YIN, R. K. 2003. *Case study research, design and methods,* Newbury Park, CA, SAGE.
[35] ZIKMUND 2002. *Business research methods,* Dryden, Thomson Learning.

Authors' Biography

Hamed Taherdoost is holder of Bachelor degree in the field of Science of Power Electricity, Master of Computer Science (Information Security), Doctoral of Business Administration; Management Information Systems and second PhD in the field of Computer Science.

With over 16 years of experience in the field of IT and Management, Dr Hamed has established himself as an industry leader in the field of Management and IT. Currently he is Chief Executive Officer of Hamta Business Solutions Sdn Bhd, Director and Chief Technological Officer of an IT Company, Asanware Sdn Bhd, Chief Executive Officer of Ahoora Ltd | Management Consultation Group, and Chief Executive Officer of Simurgh Pvt, an International Trade Company.

Remarkably, a part of his experience in industry background, he also has numerous experiences in academic environment. Dr.Hamed has published more than 100 scientific articles in authentic journals and conferences. Currently, he is a member of European Alliance for Innovation, Informatics Society, Society of Computer Science, American Educational Research Association, British Science Association, Sales Management Association, Institute of Electrical and Electronics Engineers (IEEE), IEEE Young Professionals, IEEE Council on Electronic Design Automation, and Association for Computing Machinery (ACM).

Particularly, he is a Certified Ethical Hacker (CEH), Associate in Project Management (CAPM), Information Systems Auditor (CISA), Information Security Manager (CISM), PMI Risk Management Professional, Project Management Professional (PMP), Computer Hacking Forensic Investigator (CHFI) and Certified Information Systems (CIS).

His research interest areas are Management of Information System, Technology Acceptance Models and Frameworks, Information Security, Information Technology Management, Cryptography, Smart Card Technology, Computer Ethics, Web Service Quality, Web Service Security, Performance Evaluation, Internet Marketing, Project Management and Leadership.

Electronic copy available at: https://ssrn.com/abstract=3205040