IJMRR/ April 2013/ Volume 3/Issue 4/Article No-8/2752-2758 ISSN: 2249-7196

INTERNATIONAL JOURNAL OF MANAGEMENT RESEARCH AND REVIEW

ISSUES IN RELIABILITYAND VALIDITY OF RESEARCH

S. Lakshmi*[1], Dr.M.Akbar Mohideen[2]

[1]H.O.D, (Research Scholar), PG & Research Department of Commerce, Adhiparasakthi College of Arts & Sciences, Kalavai.

[2]Associate Professor & H.O.D of Corporate Secretaryship & Research Guide in Commerce, C. Abdul Hakeem College, Melvisharam.

ABSTRACT

The use of reliability and validity are common in quantitative research and now it is reconsidered in the qualitative research paradigm. The challenges of achieving reliability and validity are among the most difficult faced by researchers. This article discusses the issues in validity and reliability of research. First, the meaning of reliability and validity of research are discussed. Secondly, the factors which are affecting the reliability and validity of research are discussed. Finally this article attempts to explore the measurement related concepts as well as some of the issues pertaining thereto.

Keywords: Reliability, Validity, Consistent, estimates.

INTRODUCTION

In research, validity has two essential parts: internal and external. Internal validity encompasses whetherthe results of the study are legitimate because of the way the groups were selected, data was recorded or analysis performed. External validity, often called "generalizability", involves whether the results given by the study are transferable to other groups (i.e. populations) of interest. An important point to remember when discussing validity is that without internal validity, we cannot have external validity. A common threat to internal validity is reliability. Assuming the same initial conditions for a test assessment or process the test must provide the same result every time it is performed for it to be deemed reliable. Reliability is often at risk when assessments are taken over time, performed by different people or the assessments are highly subjective. As a researcher, we must ensure that these reliability errors are minimized so that if differences are seen in the data that can be attributed to the intervention and not to sloppy weight measurements. Threats to study a validity and reliability exist at almost every time in the research process. Across disciplines, competent researchers often not only fail to report the reliability of their measures, but also fall short of grasping the inextricable link between scale validity and effective research. Instrument validity and reliability lie at the heart of competent and effective study. However, these phenomena have often been somewhat misunderstood or under emphasized. How productive can any research is if the instrument used does not actually measure what it purports to? How legitimate or justifiable is research that is based on an inconsistent instrument? What constitutes a valid instrument? What are the implications of proper andimproper testing? This paper attempts to explore these measurement related concepts as well as some of the issues pertaining thereto.

MEANING OF RELIABILITY

Reliability is the degree to which measures are free from error and therefore yield consistent results (i.e.the consistency of a measurement procedure). If a measurement device or procedure consistently assigns the same score to individuals or objects with equal values, the instrument is considered reliable. Reliability involves the consistency, or reproducibility, of test scores i.e., the degree to which one can expect relatively constant deviation scores of individuals across testing situations on the same, or parallel, testing instruments.

RELIABILITY ISSUES IN RESEARCH

Reliability Estimation

Repeatability, or stability-over-time reliability, may be measured with the test-retest method, whereby the same scale or measure is administered to the same respondents at two separate points in time (Zikmund, 2003 p 300), i.e. comparing the scores from repeated testing of the same participants with the same test. Reliable measures should produce very similar scores, e.g. IQ tests typically show high test-retest reliability. However, test-retest procedures may not be useful when participants may be able to recall their previous responses and simply repeat them upon retesting.

Internal consistency, or homogeneity, may be measured by using either the split-half method, alternate-form method, or Cronbach's alpha method. The split-half method is one that measures the degree of internal consistency by checking one half of the results of a set of scaled items against the other half, i.e. comparing scores from different parts of the test. The method demands equal item representation across the two halves of the instrument. Clearly the comparison of dissimilar sample items will not yield an accurate reliability estimate. One can ensure equal item representation through the use of random item selection, matching items from one half to the next or assigning items to halves based on an even/odd distribution.

The alternate-form method is one that measures the correlation between alternative instruments, designed to be as equivalent as possible, administered to the same group of subjects i.e. by comparing scores from alternate forms of the test. In cases where administering the exact same test will not necessarily be a good test of reliability, we may use equivalent/alternate forms reliability. As the name implies, two or more versions of the test are constructed that are equivalent in content and level of difficulty, e.g. professors use this technique to create makeup or replacement exams because students may already know the questions from the earlier exam.

The most common method of assessing internal consistency reliability estimates is through the use of coefficient alpha. Though there are three different measures of coefficient alpha, the most widely used measure is Cronbach's coefficient alpha. Cronbach's alpha is actually an average of all the possible split-half reliability estimates of an instrument Cronbach's alpha is a reliability coefficient that measures inter-item reliability or the degree of internal consistency between variables measuring oneconcept i.e. the degree to which different items measuring the same variable attain consistent results. This coefficient varies from 0 to 1 and a value of 0.6 or less generally indicates unsatisfactory internal consistency reliability. In the social sciences, acceptable reliability estimates range from .70 to .80

FACTORS AFFECTING RELIABILITY

Low internal consistency estimates are often the result of poorly written items or an excessively broad content area of measure. However, other factors can equally reduce the reliability coefficient, namely, the homogeneity of the testing sample, imposed time limits in the testing situation, item difficulty and the length of the testing instrument.Group homogeneity is particularly influential when one is trying to apply a norm-referenced test to a homogenous test sample. In such circumstances, the restriction of range of the test group (i.e. low variability) translates into a smaller proportion of variance explained by the test instrument, ultimately deflating the reliability coefficient. It is essential to bear in mind the intended use of the instrument when considering these circumstances and deciding how to use an instrument Imposed time constraints in a test situation pose a different type of problem, i.e. time limits ultimately affect a test taker's ability to fully answer questions or to complete an instrument. Lastly, test length also factors into the reliability estimate. Simply, longer tests yield higher estimates of reliability. However, one must consider the reliability gains earned in such situations, as infinitely long tests are not necessarily desirable.

DIFFICULTIES OF ACHIEVING RELIABILITY

It is important to understand some of the problems concerning reliability which might arise. It would be ideal to reliably measure, every time, exactly those things which we intend to measure. However, researchers can go to great lengths and make every attempt to ensure accuracy in their studies, and still deal with the inherent difficulties of measuring particular events or behaviors. Sometimes, and particularly in studies of natural settings, the only measuring device available is the researcher's own observations of human interaction or human reaction to varying stimuli. As these methods are ultimately subjective in nature, results may be unreliable and multiple interpretations are possible. Three of these inherent difficulties are quixotic reliability, diachronic reliability and synchronic reliability.

Quixotic reliability refers to the situation where a single manner of observation consistently, yet erroneously, yields the same result. It is often a problem when research appears to be going well. This consistency might seem to suggest that the experiment was demonstrating perfect stability reliability. This, however, would not be the case.

For example, if a measuring device used in an Olympic competition always read 100 meters for every discus throw, this would be an example of an instrument consistently, yet erroneously, yielding the same result. However, quixotic reliability is often more subtle in its occurrences than this.

Diachronic reliability refers to the stability of observations over time. It is similar to stability reliability in that it deals with time. While this type of reliability is appropriate to assess features that remain relatively unchanged over time, such as landscape benchmarks or buildings, the same level of reliability is more difficult to achieve with socio-cultural phenomena.

For example, in a follow-up study one year later of reading comprehension in a specific group of school children, diachronic reliability would be hard to achieve. If the test were given to the same subjects a year later, many confounding variables would have impacted the researchers' ability to reproduce the same circumstances present at the first test. The final results would almost assuredly not reflect the degree of stability sought by the researchers.

Synchronic reliability refers to the similarity of observations within the same time frame; it is not about the similarity of things observed. Synchronic reliability, unlike diachronic reliability, rarely involves observations of identical things. Rather, it concerns itself with particularities of interest to the research.

MEANING OF VALIDITY

Validity has been defined by "the extent to which [a test] measures what it claims to measure" (Gregory, 1992, p.117). A measure is valid if it measures what it is supposed to measure, and does so cleanly – without accidentally including other factors. The focus here is not necessarily on scores or items, but rather inferences made from the instrument i.e. the behavioral inferences that one can extrapolate from test scores is of immediate focus. In order to be valid, the inferences made from scores need to be "appropriate, meaningful, and useful. These distinctions illuminate the inextricable link between validity and reliability. A valid instrument must be reliable, but a reliable instrument may not necessarily be valid.

Violations of instrument validity severely impact the function and functioning of a testing instrument. In some ways, validity inadequacies impart even more serious consequences on an instrument than its reliability counterpart.

VALIDITY ISSUES IN RESEARCH

Effective validity studies not only demand the integration of multiple sources of evidence, but also must continually take place over time, i.e. a measure cannot be deemedvalid in a simple instance of study. Rather, multiple studies must be implemented over different samples, and the collection of validity evidence must cover specified areas. Moreover, in recent years researchers have expanded the understanding of validity to comprise more dimensionality than previously recognized.

The unified concept of validity is best understood and examined within the context of its four discrete facets: content validity, construct validity, criterion validity and consequential validity.

CONTENT VALIDITY

Content validity considers whether or not the items on a given test accurately reflect the theoretical domain of the latent construct it claims to measure. Items need to effectively act as a representative sample of all the possible questions that could have been derived from the construct In the social sciences where theories and constructs involved are innately intangible (e.g. anxiety, intelligence, depression, etc.), their measurement depends on the operationalization of variables deemed to be representative of the domain. Experts suggest that employing the following four steps to effectively evaluate content validity: 1) identify and outline the domain of interest, 2) gather resident domain experts, 3) develop consistent matching methodology, and 4) analyze results from the matching task.

CONSTRUCT VALIDITY

The construct validity of a measure "is directly concerned with the theoretical relationship of a variable (e.g. a score on some scale) to other variables. It is the extent to which a measure 'behaves' the way that the construct it purports to measure should behave with regard to established measures of other constructs" In practice, as constructs are not readily observable, items or variables, that act as representations of the construct and serve to measure examinee scores with respect to the paradigm, must be developed.

CRITERION VALIDITY

Criterion validity refers to the ability to draw accurate inferences from test scores to a related behavioral criterion of interest. This validity measure can be pursued in one of two contexts: predictive validity or concurrent validity. In criterion-oriented validity, the investigator is primarily interested in some criterion which he wants to predict. If the criterion is obtained sometime after the test is given, predictive validity is being studied. In predictive validity, researchers are interested in assessing the predictive utility of an instrument.

CONSEQUENTIAL VALIDITY

Consequential validity refers to the notion that the social consequences of test scores and their subsequent interpretation should be considered not only with the original intention of the test, but also cultural norms (Messick, 1995). This idea points to both the intended and unintended consequences of a measure, which may be either positive or negative.

GENERAL FACTORS AFFECTING VALIDITY

An integral issue at hand in establishing validity coefficients is the actual relationship between the two variables, or constructs, that one is interested in. Beyond this, comparable measurement issues that affected the nature of reliability coefficients also affect validity coefficients, i.e. the more heterogeneous the groups are, the higher the correlations between two measures will ultimately be.

This phenomenon is most readily observable in samples with a restriction of range problem. When the data range is limited, the scores become more homogenous and the resulting correlation coefficients derived are artificially inflated. An important point to note is that the more effective an instrument is at screening individuals for a particular purpose, the less heterogeneous the resulting sample will be, which in turn results in a smaller validity coefficient.

CONCLUSION

Reliability and validity of instrumentation should be important considerations for researchers in their investigations. The goal of achieving measurement validity and reliability can be accomplished partly by a push for quality item writing, an insistence on reporting reliability data across studies, sound theoretical bases for construct measurement and the accurate operationalisation of constructs.

This objective imparts a direct responsibility on behalf of all examiners in a given field, i.e. it is essential for researchers to actively measure the reliability and validity of instrument scores over populations and time. The continual nature of both these processes should not be undermined or overlooked. Moreover, it is critical for this type of information to be easily accessible in order to facilitate the understanding and sharing of this knowledge. Without credible instrumentation that is monitored and measured over time, research results can become meaningless.

REFERENCES

Babour RS. Mixing qualitative methods: Quality assurance or qualitative quagmire? Qualitative Health Research 1998; 8(3): 352-361.

Bogdan RC, Biklen SK. Qualitative research in education: An introduction to theory and methods (3rd ed.). Needham Heights, MA: Allyn & Bacon, 1998.

Campbell T. Technology, multimedia, and qualitative research in education. Journal of Research on Computing in Education 1996; 30(9): 122-133.

Clont JG. The concept of reliability as it pertains to data from qualitative studies. Paper Presented at the annual meeting of the South West Educational Research Association. Houston, TX, 1992.

Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory into Practice 2002; 39(3): 124-131.

Davies D, Dodd J. Qualitative research and the question of rigor. Qualitative Health research 2002; 12(2): 279-289.

Denzin NK, Lincoln YS. The landscape of qualitative research: Theories and issues. Thousand Oaks: Sage Publications, 1998.

Eisner EW. The enlightened eye: Qualitative inquiry and the enhancement of educational practice. New York, NY: Macmillan Publishing Company, 1991.

Healy M, Perry C. Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. Qualitative Market Research 2000; 3(3): 118-126.

Hoepfl MC. Choosing qualitative research: A primer for technology education researchers. Journal of Technology Education 1997; 9(1): 47-63.

Hipps JA. Trustworthiness and authenticity: Alternate ways to judge authentic assessments. Paper presented at the annual meeting of the American Educational Research Association. Atlanta, GA, 1993.

Johnson BR. Examining the validity structure of qualitative research. Education 1997; 118(3): 282-292.

Joppe M. The Research Process, 2002. Retrieved February 25, 1998, from http://www.ryerson.ca/~mjoppe/rp.htm

Kirk J, Miller ML. Reliability and validity in qualitative research. Beverly Hills: Sage Publications, 1986.

Maxwell JA. Understanding and validity in qualitative research. Harvard Educational Review 1992; 62(3): 279-300.

Patton MQ. Qualitative evaluation and research methods (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc, 2002.

Seale C. Quality in qualitative research. Qualitative Inquiry 1999; 5(4): 465-478.

Stenbacka C. Qualitative research requires quality concepts of its own. Management Decision 2001; 39(7): 551-555.

Strauss A, Corbin J. Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage Publications, Inc, 1990.

Wainer H, Braun HI. Test validity. Hilldale, NJ: Lawrence Earlbaum Associates, 1998.

