ISSN: 2454 - 2016

Validating Measures in Business Research: Practical Implications

By Fredrick Aila[1] and Benjamin Ombok[2],

Abstract

Instrument validity, though at the heart of quantitative business research, has not been adequately treated by novice researchers and advisors alike, potentially invalidating otherwise good research. This paper reviews validity in business research and presents practical implications for its assessment. The paper finds immediate use among novice researchers and advisors in assessing instrument validity and by extension entire research validity. As a result, it is hoped quality of researches will improve warranting scrutiny by researchers in the fields of business studies and beyond.

Key words: Postgraduate Studies; School of Business and Economics; Reliability; Measurement; Business Studies.

Introduction

The ethos of validity measurement are well documented in business research methods texts (see Zikmund *et al*., 2010; Sekaran and Bougie, 2010; Saunders, Lewis and Thornhill, 2012; Hair *et al*., 2010; De Vellis, 2012; de Vaus, 2013; Creswell, 2014). The question of instrument validity however poses practical problems for novice researchers especially in the fields of business studies[3]. Yet validity is at the heart of study believability (Creswell, 2014), and competent and effective study (Thanasegaran, n.d.).

---

[1] Dr. Fredrick Aila is Chair, School Postgraduate Studies Committee (SPGSC), and Lecturer in the School of Business and Economics, Maseno University.

[2] Dr. Benjamin Ombok[2] is Dean, School of Business and Economics and Lecturer in the Department of Accounting and Finance, School of Business and Economics, Maseno University.

[3] Majority of postgraduate business students completing master's program have an observed difficulty in assessing instrument validity at the researcher's institution. This has also been observed in some business PhD students and therefore warrants scrutiny.

ISSN: 2454 - 2016

Novice researchers, especially in the field of business studies seldom collect sufficient data[4] in order to validate instruments[5]. Alternatively, once some data are collected, no attempt is made to ensure the validity of instruments[6]. Therefore, results of such studies rarely percolate to mainstream research and policy circles. Moreover, published data tend to be in obscure journals whose editors are preoccupied with article acceptance for "monetary gain" or "publicity" rather than the requisite scientific rigour. As a result, such studies do not greatly influence the research community.

Where attempts at validity are made the procedures and results are not explicitly reported for replicability[7] (Thanasegaran, n.d). Moreover, Kimberlin and Winterstein (2008) show that most data sources involve a greater degree of subjectivity in judgment or other potential sources of error in measurement. Two important questions for this study are: "How can validity measurement be more practical for ease of uptake by novice researchers?" and "In what ways can student advisors guide their students to execute and report instrument validation?" This paper therefore explores the practicality of validity measurement for uptake by both novice researchers and advisors.

Meaning of and Types of Validity.

Validity refers to the degree that an instrument actually measures what it is designed or intended to measure (Netemeyer, Bearden and Sharma, 2003; Burton and Mazerolle, 2011; Bolliger and Inam, 2012). It is the accuracy of a measure or the extent to which a score truthfully measures a concept (Zikmund *et al.,* 2010). Or simply put, it is as the extent to which an

[4] Factor analysis methods of construct validation (exploratory factor analysis (EFA) and confirmatory factor analysis (CFA)) are large sample techniques (n>50), yet pretesting, if any is usually done on small samples (n<30).

[5] The business school has several junior postgraduate faculty serving as student advisors. Having transited to postgraduate faculty status, they might overlook the criticality of instrument validation. Moreover, most postgraduate students rarely assess validity beyond translation validity (Trochim, 2006).

[6] Most novice researchers rely on face validity assessed by student advisors. Where the phenomena is not clearly understood by the advisor, then chances of wrong translation are likely.

[7] Several reported studies published in journals fail to report instrument validity. It may be that no validity was assessed at all.

ISSN: 2454 - 2016

instrument measures what it purports to measure (Kimberlin and Winterstein, 2008).

Validity is concerned with the meaningfulness of research components (Drost, 2011). Drost (2011) suggests that somewhat confusing to the novice researcher is the notion that a reliable measure is not necessarily a valid measure. In other words, measures must be valid as well as being reliable. Bollen (1990) explains that reliability is that part of a measure that is free of purely random error and that nothing in the description of reliability requires that the measure be valid. It is possible to have a very reliable measure that is not valid. Therefore, reliability is a necessary but not a sufficient condition for validity (Kimberlin and Winterstein, 2008).

Drost (2011) suggests that there are four types of validity that researchers should consider: statistical conclusion validity, internal validity, construct validity, and external validity. Statistical conclusion validity refers to inferences about whether it is reasonable to presume covariation given a specified alpha level and the obtained variances (Cook & Campbell, 1979). Statistical conclusion validity might be threatened by low statistical power, violation of assumptions, reliability of measures, reliability of treatment, random irrelevancies in the experimental setting, and random heterogeneity of respondents.

Internal validity communicates the validity of the research itself. The question is how valid is the research. Internal validity of a research design might be threatened by history, maturation, testing, instrumentation, selection, mortality, diffusion of treatment and compensatory equalization, rivalry and demoralization (Drost, 2011).

Zikmund *et al*. (2010) say construct validity exists when a measure reliably measures and truthfully represents a unique concept. It refers to how well a concept, idea, or behaviour – that is a construct – has been translated or transformed into a functioning and operating reality, the operationalization (Trochim, 2006). Finally, external validity of a study implies generalizing it

ISSN: 2454 - 2016

to other persons, settings, and times and not necessarily to the target population.

The Special Case of Construct Validity

To substantiate construct validity involves accumulating evidence in six validity types: face validity, content validity, concurrent and predictive validity, and convergent and discriminant validity (Turocy, 2002; Trochim, 2006; Zikmund *et al*., 2010; Drost, 2011; Aila, 2014). Both Trochim (2006) and Drost (2011) suggest two major approaches to construct validity: translation validity and criterion-related validity.

Translation validity centres on whether the operationalization reflects the true meaning of the construct. Therefore translation validity attempts to assess the degree to which constructs are accurately translated into the operationalization, using subjective judgment or face validity and examining content domain or content validity. Face validity is a subjective judgment on the operationalization of a construct and therefore a weak form of construct validity (Drost, 2011).

Content validity is a qualitative type of validity where the domain of the concept is made clear and the analyst judges whether the measures fully represent the domain (Bollen, 1990). Therefore, content validity is a qualitative means of ensuring that indicators tap the meaning of a concept as defined by the researcher (Kimberlin and Winterstein, 2008; Drost, 2011).

Criterion-related validity on the other hand is the degree of correspondence between a test measure and one or more external referents (criteria), usually measured by their correlation (Trochim, 2006; Drost, 2011). Concurrent validity refers to the ability of a test to predict an event in the present while predictive validity refers to the ability of a test to measure some event or outcome in the future. Convergence validity tests for convergence across different measures or manipulations of the same "thing" while discriminant validity tests for divergence between measures and manipulations of related but conceptually distinct "things" (Cook & Campbell, 1979). The scheme of construct validity types can be depicted as Figure 1 (Drost, 2011:117).

ISSN: 2454 - 2016

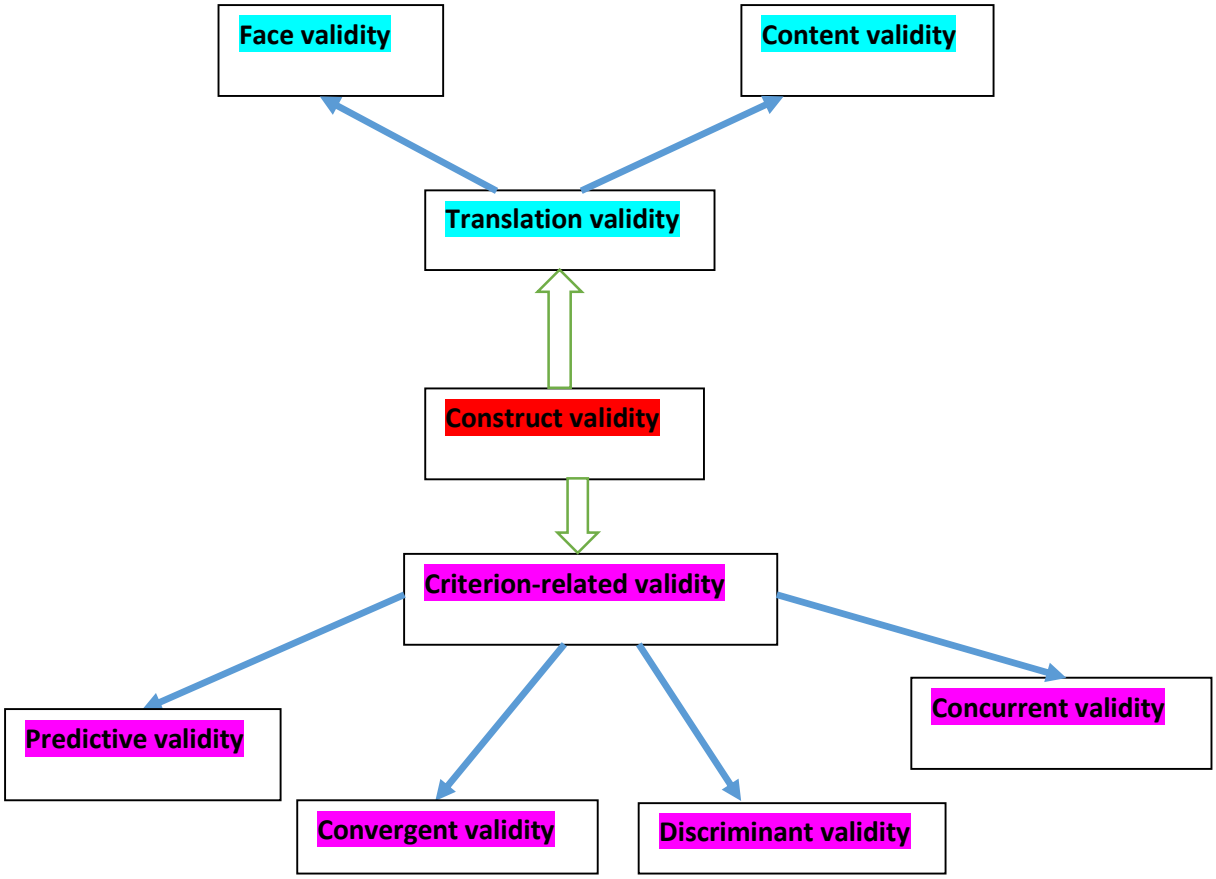

Figure 1: Types of Construct Validity. Source: Drost (2011)

Practical Issues in Assessing Construct Validity

Face validity can be established through analyst assessment (Bolliger and Inam, 2012; Aila, 2014). Aila (2014) asserts that it is important to demonstrate qualitatively the measure's relevancy, consistency and suggestions for revision. Therefore, mere mention of face validity assessment is not sufficient much as mere existence of the instrument is sufficient.

Content validity can be assessed through literature survey/searches (Zikmund *et al*., 2010) to ensure items are based on the domain of the study concepts (DeVellis, 2012) corroborated by expert/analyst judgement and review suggestions (Bolliger and Inam, 2012). In other words, use of experts/analysts alone is no panacea to content validity. Moreover, the

ISSN: 2454 - 2016

question "Who is an expert[8]?" may need to be answered by distinguishing experts from non-experts and by discriminating non-experts from assessing the instrument for validity.

Some correlation analysis is required for one to assess criterion-related validity[9]. So data with correlation ability need to be collected and analyzed for correlations. The questions to raise include: how do similar constructs correlate? High correlations among similar constructs reveals convergent validity. Low correlations indicate they do not tap on the same construct; in other words, they discriminate one another. Therefore dissimilar constructs will have low inter-correlations signifying discriminant validity.

Concurrent validity is seen when one half of the construct correlates to the other half. This means the half measure concurrently validates the other half[10]. Alternatively, the measure being assessed is viewed in light of other related constructs. This requires the researcher to subject respondents to these existing constructs/measures as well and assess how the target measure correlates with existing measures.

How well does the measure predict future events? In a single study for instance, assess the ability of the construct to measure the phenomena through the pilot study. The correlation obtained at this stage (note Cronbach's Alpha is a correlation) predicts how well the construct will measure the main event which occurs sometime in the future. A measure that has been used severally in the past and has yielded consistent results is said to have a good predictive validity, especially in different populations[11].

[8] See Slavec and Drnovsek (2012) for further arguments on experts.

[9] Item-total statistics; correlation analysis; multitrait-multiattribute method (MTMM), confirmatory factor analysis (CFA); exploratory factor analysis (EFA) (Nargundkar, 2003; Trochim, 2006; Drost, 2011; Aila, 2014).

[10] Care should be taken when splitting uneven numbered items.

[11] Novice researchers should be encouraged to seek for validated scales and replicate them instead of attempting to develop new measures. Existing measures with known validities can be modified to suit temporal situations. This allows one to assess the measures predictive validity. Where no measure exists in literature, then the researcher will be forced to develop one according to set rules (Aila, 2014).

ISSN: 2454 - 2016

Student advisors' should convince themselves that the study they are guiding has a logical place in the body of literature. Mapping the study in literature helps the advisor to be a better expert and therefore his or her judgement is seen to be sound. The advisor should be current in his/her chosen subject as a proof of his/her expertise[12]. In other words, the advisor should be the first assessor of translation validity. Much as the study seeks to fill gaps in knowledge, the measures adopted must be valid.

Given that most studies will by and large replicate prior studies and extend them incrementally, advisors are strongly encouraged to adopt or modify validated measures to the temporal situation of the study. Does this advice hinder innovation? Certainly not. It only seeks to deter invalid innovation.

Lastly but not least, researchers need to assess how well their measures generalizes not only to the target population but to all other populations. Generalizability to the target population must be immediately demonstrated[13]. However, generalizability to populations beyond the target requires the researcher's ingenuity and is a mark of a truly novel research output[14]. All researchers should strive for this, even though it might be a mirage!

Conclusion

This paper has reviewed practical applications of instrument validity in a user friendly language for both novice researchers and advisors. It is hoped that these researchers will not only appreciate the terrain validity encompasses, but will begin assessing beyond translation validity. Advisors are equally hoped will guide their students for both instrument and research validity. Ultimately, it is hoped the quality of researches will immensely

[12] Advisors in new and interesting business fields necessarily need to read faster than their students' in order to integrate the new with existing information. The field of business is constantly changing: moreover change is inevitable.

[13] Can be assessed through the coefficient of determination, R^2.

[14] This is evidenced over time especially in terms of numbers and quality of citations the piece of research generates as assessed external parties (e.g. journal impact factors etc.)

ISSN: 2454 - 2016

improve to warrant scrutiny by researchers in fields beyond business studies by clearly impacting the research community.

References

Aila, F.O. (2014). Effect *of consumer attitudes, consumer characteristics and biosecurity principles on consumer preferences for indigenous chickens in Kisumu City, Kenya*. An unpublished PhD Thesis, Department of Business Administration, Maseno University.

Bollen K. A. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development, Spring* (1), 7-24.

Bolliger, D. U., & Inam, F. A. (2012). Development and validation of the online student connectedness survey (OSCS). *The International Review of Research in Open and Distance Learning, 13*(2), 41-65.

Burton, L. J, & Mazerolle, S. M. (2011). Survey instrument validity part I: Principles of survey instrument development and validation in athletic training education research. *Athl Train Educ J*. 6(1):27-35.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings.* Boston, MA: Houghton Mifflin Company.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* (4th ed.*).* London: Sage Publications, Inc.

de Vaus, D. A. (2013). *Surveys in social research*. (5th ed.). London: Rutledge.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. (3rd ed.). London: Sage Publications, Inc.

Drost, E.A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives, 38*(1), 105-123.

Hair, J. F. Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. (7th ed.). Upper Saddle River, NJ: Pearson.

Kimberlin, C. L. & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *Am J Health-Syst Pharm*. 65, 2276-84.

Nargundkar, R. (2003). *Marketing research: Text and cases*. (2nd ed.). New Delhi: Tata McGraw-Hill.

ISSN: 2454 - 2016

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. London: Sage.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students.* (3rd ed.) London: Prentice Hall-Financial Times.

Sekaran, U. & Bougie, R. (2010). *Research methods for business: A skill building approach*. (5th ed.). New York: John Wiley and Sons.

Slavec, A. & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review, 14*(1), 39-62.

Thanasegaran, G. (n.d.). *Reliability and validity issues in research*. Department of Management & Marketing, Faculty of Economics & Management, Universiti Putra Malaysia. Accessed from www.upm.edu.my on 15th May 2015.

Trochim, W. M. (2006). *The research methods knowledge base.* (2nd ed.). Internet www page, at URL:<http://www.socialresearchmethods.net/kb/> (version current as of October 20, 2006). Accessed on 15 May 2014.

Turocy, P. S. (2002). Survey research in athletic training: The scientific method of development and implementation. *J Athl Train*. 37(Suppl 4), S174-S179.

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2010). *Business research methods.* (8th ed.). Mason, OH: South-Western/Cengage Learning.