



Changing phase of validity: The past and now
Abraham Gyamfi

abrahamgyamfi84@gmail.com

Wesley College of Education, Kumasi, Ghana, Department of Education

Abstract

Validity is the bedding rock of all assessment theories and principles, in that, it underpins all assessment theories. It is the focus or the object of concern of every assessment process. The concept validity is and is still receiving diverging interpretation and explanation. This paper attempts to shed more light on the concept of validity as used in assessment. The changes that have emerged in the concept of validity are explained in detail. It also presents the current phase of validity; evidence of validity. All the evidences and how they are evaluated are discussed.

Keywords: validity, face validity, divergent validity

Introduction

Due to the importance of validity and it, being central to assessment, many studies have done and there are still on-going studies of validity. This is to ascertain the best evidences to support the use and interpretation of assessment results. Messick (1989) stated that new findings have changed the phases of validity over time for better understanding of the phenomena. This means that the concept validity keeps undergoing metamorphosis. It can therefore be stated that what is a valid results today, may not be a valid results tomorrow. Messick (1989) again stated that validity is a phenomenon which keeps on changing and validation is a continuing process. This is because, evidence is always not complete and it is essential to make the most current reasonable use of the assessment results which is guaranteed in advanced research.

Validity in the Pasts

Theoretically and gradually, the concept of validity has changed over the years (Anastasi, 1986 & Angoff, 1988). One or another of these forms of evidence, or combination of them gave birth to the status of types of validity in the past (Messick, 1989). Scholars based on the sources

of evidences as considered as types of validity. However, because all the sources of evidence depend on the valid interpretation and use of assessment scores, there cannot be types of validity. According to Nitko (2004), validity is a unitary concept. This means that it has been established that there cannot be types of validity. All the evidences support that unitary concept, validity.

One major evident of validity that was neglected in early views of validity is the consequential use and interpretation of assessment scores. Messick (1989) noted consequential basis of assessment validity has received little attention since the 1950s because validity has been conceptualized in terms of the functional worthiness of the assessment, that is, in terms of how well the assessment does the proposed purpose. Guilford (1946) claimed that an assessment results is valid for anything with which it correlates. Recent studies have underscored the continuing need for validation practice to address the realities of potential and actual assessment consequences on society. Emphasis is being placed on social values implied by the interpretation and use of the assessment results. The social consequences of assessment results are also seen to be subsumed as aspect of construct validity.

The 1954 technical recommendations (APA, 1954) listed four types of validity-namely, content, predictive, and construct validities. However, the 1966 standards (1966) reduced the types to three, namely, content, criterion –related and construct validities. These validity types were based on a particular aim of assessment. These aims include 1) determining how an individual is currently performing in a collection of content, 2) forecasting an individual's future standing or to estimating the individuals present standing on some important trait other than the assessment, and 3) inferring the degree to which an individual possesses some construct acclaimed to be reflected in performance of the assessment task (Messick, 1989). The American Psychological Association (1966) further pointed out that the three types of validity are by concept, independent, and seldom, one is important than other in a particular situation. All the types of validity are needed for a thorough study of assessment. The study is incomplete without the others.

Further clarification on the concept of validity was detailed in the APA (1974). Behaviour was replaced with content. That is content validity was described as how well the behaviours demonstrated in assessment constitutes a representative sample of domain of behaviours. The shift from content to behaviour means content validity cannot be evaluated by a mere professional judgement of content relevance and representativeness. Thus content validity requires evidence of reliable response which are consistent on the assessment and that the

assessment and the domain of assessment are similar or from same response (Messick, 1989). This has placed the evaluation of content validity beyond mere professional judgement.

The 1985 standards APA (1985) also showed more light on the conceptualization of validity. The standards stressed on the unitary nature of validity, referring to the appropriateness, usefulness and meaningfulness of the specific inferences made from the assessment scores. This notion nullifies the “types concept of validity” to “categories of validity evidences” as content-related, criterion-related and construct –related evidence of validity. Evidence from the related areas should be provided to support the interpretation, use and social consequences of the assessment results before it is deemed valid.

Works of Anastasi and Cronbach as cited in Messick (1989) portray some evolution of validity. Anastasi, in his work in 1954 organised validity in terms of face validity, content validity, factorial validity, and empirical validity. Face validity has been phased out in recent validity analysis because face validity which refers to what an assessment appears to measure to the layperson. Validity has come to be understood as not about the assessment itself but the results. Empirical validity has been established to an aspect of construct validity and therefore no more in operation. Empirical validity is about the procedures used to check content validity, which, construct validity measures by evaluating how well the content measures the behaviour. Factorial validity also in the work of Anastasi has been phased out. Factorial validity refers to the correlation between the assessment scores and a factor common to a group of assessment or other measures of behaviour. Contemporary construct validity is established by finding the correlation of the assessment results with other measures (Amedahe, 2000). This suggests that the Anastasi’s factorial validity is an aspect of contemporary construct-related evidence of validity.

In the work of Cronbach in 1949, Cronbach, organised his work on validity in terms of logical validity and empirical validity, as in the work of Anastasi. Cronbach’s logical validity was based on judgement of precisely what the assessment results measures. It was evaluated by making a careful study of the assessment itself. On the bases that validity is about the assessment result and not the assessment itself, this logical validity of Cronbach has been phased out. The empirical validity of Cronbach has been phased out on the justification for the phasing out of Anastasi’s empirical validity.

The works of Mehrens and Lehmann, (1991); Plake, Impara, & Buckendahl, (2004); and Smisko, Twing, & Denny, (2000) on validation process gave birth to a so-called type of validity known as curricular validity. Curricular validity is evaluated by comparing the assessment

instrument to the curriculum that was dictated for the assessment. The so-called curricular validity has been phased out from contemporary validation process on the bases that curriculum is reflected in the content of the assessment. Therefore, curriculum validity perfectly subsumes under content validity.

APA (2014) again highlighted some concerns with regard to validity of assessment results 1) construct underrepresentation or constructs deficiency and 2) construct irrelevant variance or constructs contamination. “Construct underrepresentation refers to the degree to which assessment results fails to capture the important aspect of the construct” (APA, 2014, pg 27). Construct validity is said to be underrepresented when there are no evidence to support a complete representation of essential constructs to be measured by the assessment instrument. This results in a narrow meaning ascribed to the assessment results. Construct –irrelevant variance also refers to the degree to which the assessment results are affect by extraneous variable. Constructs are possibly influenced by factors that are not intended, for example measuring students’ mathematical ability may influenced by vocabulary, or anxiety. The ability of the construct to be devoid of such extraneous variables ensures construct validity. According to APA (1999, pg 19) “sources of validity evidence are classified under content, response processes, internal structure, consequences of assessment and relation to other variables”. Each source perfectly comes under one of the three related evidence of validity.

Other concepts that have emerged in validity analysis are convergent and discriminant/divergent evidence of validity. The “convergent and discriminant validity” have not been popular in previous studies (Bollen, 2011). For convergent evidence measures that in reality correlates should perfectly correlate in evaluating construct validity and for discriminant/divergent validity, those that in reality do not correlate should not in any way correlate (APA, 1999). The APA (2014) however separated all the sources of validity evidences under three main themes, 1) “establishing intended uses and interpretations, 2) issues regarding samples and 3) settings used in validation and specific forms of validity” (pg 21). All the sources of validity suggested in APA (1999) are clustered under specific forms of validity in APA (2014).

Concept of validity today

Messick (1989) stated that new findings have the existing evidence of validity evidence. This means that the concept validity keeps undergoing metamorphosis. It can therefore be stated that what is a valid results today, may not be a valid results tomorrow. Messick (1989) again stated that validity is an always-changing property and validation is a continuing process. This is

because, one source of evidence is always not complete and it is essential to make the most current reasonable use of the assessment results which is guaranteed in advanced research. The theoretical conception of validity has gradually changed over the years (Anastasi, 1986 & Angoff, 1988).

According to Messick (1989), “since the early 1950s, validity has been broken into three or four different types. Specifically, validity has been divided into three types, of which one comprises two subtypes” (pg 232). These are content validity, criterion-related validity comprising predictive and concurrent validity, and construct validity. These are what APA (1954 & 1966) ascribed as traditional validity types.

Categories of validity evidence

Research has proven that these perceived types are rather sources of evidence that support the unitary concept, validity.

Content related evidence

This evidence is about the content representativeness and relevance of the assessment results. Content-related evidence of validity is assessed by showing the degree to which the content of assessment results represent the content about which conclusions are to be drawn. The judgement on content relevance focuses on whether tasks included in the assessment are in the test domain definition. The relevance of the assessment results is the extent to which the assessment matches the school's curriculum target. There should an overlap between the assessment domain and the curriculum. The weight given to each content area should be appropriate to the local curriculum (Nitko, 2004). According to Nitko, to ensure content validity, the items should have the following characteristics: (1) reflect current thinking of the subject matter of what is essential to teach and assess (2) accurately represent the subject matter (3) keyed correctly and (d) contain meaningful and relevant content.

To judge whether as assessment the content has related evidence to support the interpretation and uses of the assessment results, table of specification is prepared and use (Nitko, 2004). The table of specification is a means of defining the domain for standardized position on achievement test. It contains the major content areas and skills to be assessed and the percentage of tasks content-skills.

Criterion related evidence

Criterion-related evidence of validity measures how well an assessment results can predict a future performance on similar content. It is established by comparing the assessment results with scores of one or more external variables (called criteria) which is considered to

provide a direct measure of the trait of interest. There are two types of criterion related evidence of validity. The predictive validity of the criterion validity gives an indication of the extent to which an individual's future performance on the criterion is predicated from a previous performance. The purpose is to predict the future performance of a criterion variable. The concurrent validity also gives an indication of the degree to which the assessment results estimates individuals' present standing on the criterion. The purpose is to substitute the assessment results for the score of a related variable. The line of difference between the predictive and the concurrent validity then becomes the time of measure of the future (criterion) and present (predictor) standing on the criterion.

To assess this evidence, the correlation coefficient of the criteria and predictor is estimated. The coefficient gives an indication as whether there is a relationship between the scores and how well the predictor predicts or relate with the criteria.

Another approach to check predictive validity is by the use of the expectancy table. It is a two-way table that allows criteria to be predicted from a score.

Construct related evidence

The construct validity is established by studying what qualities the assessment measures, that is, finding the degree to which the constructs account for the performance on the assessment. Asamoah-Gyimah and Anane (2018) stated that this evidence refers to how the assessment results can be interpreted as reflection of an individual's achievement on what is being measured. Asamoah-Gyimah and Anane (2018) mentioned the following as the methods of construct validation:

Defining the domain tasks to be measured. The construct to be measured should be well defined so that the meaning of the construct will be clear. It takes expert judgement to achieve this definition of the construct.

Analyzing the mental process required by the assessment tasks. Here students taken the assessment are made to think aloud to determine whether their thinking on the assessment is the same as intended.

Correlating the assessment scores with other measures. If two forms measures the same construct, their correlation to an assessment forms other the two should be high positive. Therefore the correlation between the assessment scores and other measure on of the same construct is estimated. If the results are positive high, then the results reflect attribute being measured.

Comparison of validity types

Loevinger (1957) pointed out that, for all the so-called validity types, construct validity is the whole and all the others are ad hoc. Thus, construct validity encompasses almost all the forms of validity evidences. Content validity, “professional judgement is the bases of the “so-called relevance and representativeness of the assessment content” of a particular domain of interest. So, the so-called content validity addresses the assessment instrument representativeness and not the scores. In that sense, the so-called validity cannot be validity. Cronbach (1980) claimed that validation is one a misconception of the unitary nature of validity.

References

- Amedahe, F. A. (2000). *Assessment in schools*. Unpublished Menograph
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: AERA, APA & NCME
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association
- American Educational Research, American Psychological Association & National Council on Measurement (1999). *Standards for educational and psychological testing*. USA: American Educational Research, American Psychological Association & National Council on Measurement
- American Psychological Association, American Educational Research & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.
- American Psychological Association.(1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Author
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual review of psychology*, 37, 1-15
- Angoff, W. H. (1988). Validity: an evolving concept. In H. Wainer & H. Braun (Eds.), *test validity*. Hillsdale, NJ: Lawrence Erlbaum.
- Asamoah-Gyimah, K. & Anane, E. (2018). *Assessment in schools*. Unpublished Menograph
- Bollen, K. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, 35(2), 359-372.

- Cronbach, L. J. (1980). *Validity on parole: how can we go straight? New directions for testing and measurement: Measuring achievement over a decade*. Proceedings of the 1979 ETS Invitational conference (pp. 99-108). San Francisco: Jossey-bass
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological report reports*, 3, 635-694.
- Mehrens, W. A., & Lehmann, J. I. (1991). *Measurement and evaluation in education and psychology* (4th ed.). New York : Harcourt Brace College Publishers.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed). New York, NY: American Council on education and Macmillan, pp. 13-104
- Nitko, A. J. (2004). *Educational measurement* (4rd Ed.). USA: American Council on Education & Praeger
- Nitko, A. J. (2004). *Educational Tests and Measurements* (3rd ed.). USA: Prentice-Hall, Inc
- Plake, B. S, Impara, J. C. & Buckendahl, C. W (2004). Technical quality criteria for evaluating district assessment portfolios used in the Nebraska STARS *Educational Measurement: Issues and Practice*, 23(2), 12-16
- Smisko, A, Twing, J. S. & Denny, P (2000). The Texas model for content and curricular validity. *Applied Measurement in Education* 1, 3(4), 333-342.