

CONSTRUCT VALIDITY IN  
PSYCHOLOGY  
**AND**  
EDUCATIONAL MEASUREMENT

**BY**  
**ABDULLAH SALEM AL-MANNAI**  
**Ph . D.**

# CONSTRUCT VALIDITY

## INTRODUCTION :

The single most important aspect of instrumentation is validity . Messick (1980) stated that "validity is the overall degree of justification of test interpretation and use " (P.1014). It is quite possible for a measuring instrument to be relatively valid for measuring one kind of phenomenon but entirely invalid for assessing other phenomena . Therefore, the American Psychological Association has published guidelines for determining test validity . In general, validity can be best defined " as the extent to which any measuring instrument measures what it is intended to measure" ( Carmines and Zeller, 1979 ). Cronbach, 1971 , stated that " one validates not a test , but an interpretation of data arising from a specified procedure" ( P.447 ).

There are controversial viewpoints regarding the validity concept . The National Committee on Test Standards (1974 ) has taken the point of view that there are three kinds of validity : content , criterion - related , and construct validity . Each type of validity is capable of achieving a different goal of testing . This viewpoint is adopted and repeated by some writers , such as (Hopkins & Stanley , 1981 ; Nunnally , 1978 ; Thorndike & Hagen, 1969 ) . In respect to this viewpoint , I am going to define the three types of validity . Content validity is defined as " the extent to which a subject's responses to the items of a test may be considered to be a representative sample of his responses to a real or hypothetical universe of situations which together constitute the area of concern to the person interpreting the test " ( Lenore 1956 ) . Nunnally (1978 ) has given a useful definition of criterion - related validity . He notes, that it is at issue when the purpose is to use an instrument to estimate some important form of behavior that is external to the measuring instrument itself , the latter being referred to as the criterion " . Finally , there is construct validity , which will be defined later.

The second viewpoint is that all validity evidence is construct - validity evidence and that all types of validity can be subsumed under construct validity, because construct validity is considered to be the essence of validity ( Loevinger , 1957 ; Cronbach, 1980 ; Messick, 1981 ) .

I take the position of the second viewpoint , which is that construct validity is the essence of validity and that all other concepts of validity can be subsumed under it . It is important to start with the definition of the concept of construct validity . Mehrens and Lehmann (1984 ) define construct validity as follows : "It is the degree to which one can infer certain constructs in a

psychological theory from the test scores " ( P. 294 ) . Construct validity is a comprehensive concept , which includes the other types of validity , such as content , criterion- related , face , internal , etc . Anastasi (1969) stated that all information provided by any validation procedure is relevant to construct validity . Several measurement specialists claimed that all types of validity can be subsumed under the concept of construct validity (Anastasi, 1969 Cronbach. 1984 ; Tenopyr, 1977 ) .

Lemke and Wiersma, ( 1976 ) said that to investigate the construct validity for any test is not essentially different from the general scientific procedures for developing and confirming a theory . From the previous viewpoints, we can infer that the techniques for establishing construct validity is not less important or the same as the general scientific procedures for developing and confirming a theory . Thus , construct validity is very important and reliable ; therefore , other types of validity are subsumed under construct validity .

Cronbach and Meehl (1955 ) have described the logic of construct validation , which could be summarized in the following steps : First, to define something is to specify the interlocking system of construct ( syntax ) . Relationships can be defined between observables and observables ( that is, two things observed ), observables and constructs ( one thing observed, the second a trait ), or constructs and constructs ( both are traits ) . Finally , Cronbach and Meehl suggest that to understand "why" or to " learn about something " is to clarify both the constructs and the relationships between the constructs . Also , they suggest that construct validity is evaluated by "determining what psychological construct account for test performance" . The inference to be made from Cronbach and Meehl's description of construct validity is that these steps confirm that the idea of investigating a test's construct validity is not essentially different from the general scientific procedures for developing and testing a theory .

Cronbach, (1980) stated that "All validation is one, and in a sense all is construct validation". My point of view is that the partition of validity into the previously described three types will result in weakening, the concept of validation and does not serve all the purposes which specialists are trying to accomplish. Integrating all three of them into a single process in order to validate tests is necessary most of the time, as wolf, (1982) stated that, one type of validity cannot be freely substituted for others.

The following example illustrates the previous notion. Criterion validity expects that certain conditions will be met in order to achieve learning. These

conditions may include activities such as adding, typing, learning a new concept, or studying a fact, or they may include psychological traits like intelligence and anxiety. All of these elements which may form a criterion are constructs.

A relationship exists between the three types of validity which enhances the idea that all types of validity complete each other. For instance, Anastasi (1976) stated that "construct validity is a comprehensive concept, which includes the other types". This statement stresses that there is a relationship between the types of validity, or other types of validity subsumed under construct validity. Also, Anastasi, (1969) stated that "All the specific techniques for establishing content and criterion - related validity could have been listed again under construct validity ". Thus, Anastasi put a heavy emphasis on construct validity, so did Cronbach, (1971).

Constructs are important in educational measurement. For example, when we ask this question "what does the instrument really measure?" We are asking for information on construct validity . Constructs help us to interpret both measures used to appraise educational outcomes and measures used to forecast response to instruction (Cronbach , 1971 ).

In preparing a predictive ( Criterion - related ) instrument , the first step is to consider what constructs are likely to provide a basis for choosing or devising an effective test . In making predictions about new situations , the user of tests must rely on constructs of some generality . In any theoretical discussion of what is being measure by the achievement test for example , a consideration of construct validity is required . French & Michael , (1966) mentioned that to analyze construct validity , all the knowledge regarding validity will be supported .

There are different methods to evaluate the construct , such as the logical analysis method , the experimental method , predictions about group differences , and the correlational method ( Multitrait-Multimethod Matrix ) . Because I am concerned about the length of this paper , I am going to describe only the correlational method (multitrait - multimethod matrix ) .

This method is used when more than one construct or trait and more that one method are being used in the validation process. Campbell (1960) pointed out that in order to demonstrate construct validity, we must show two things. First, a test must correlate highly with other variables with which it should theoretically correlate. Second, it must not correlate significantly with variables from which it should differ. In the Multitrait-Multimethod Matrix which was proposed by Campbell and Fiske (1959) , the former process (which was

pointed out by Campbell) was described as convergent validation and the latter as discriminant validation . Their discussion emphasizes that if the measures that are believed to be related to the construct yield high correlations with the measure of the construct , this would be taken as evidence of construct validity ( convergent validation ) . When the measure of a construct does not correlate with measures that it is supposed to be unrelated to, this, too, is taken as evidence of validation (discriminant validation) . Examples of convergent validation are the correlation of a new test of scholastic aptitude with success in school and with other test of scholastic aptitude, or of the correlation between scores on a mechanical job .Examples of discriminant validation include correlation between scores on a measure of ascendancy and scores on measures of autonomy, impulsivity, and objectivity .

Both criterion- related validity and content validity have limited usefulness for assessing the validity of empirical measures of theoretical concepts employed in the social sciences. One of the limitations of criterion validation procedures is that it cannot be applied to all measurement situations in the social sciences . The most important limitation is that for some measures in the social sciences, there simply do not exist any relevant criterion variables against which the measure can be reasonable evaluated . For instance, what would be an appropriate criterion for a measure of self-esteem? As we know, there is no specific type of behavior that person(s) with high or low self-esteem display . So, there is no behavior that could be used to validate a measure of self-esteem . Moreover, it is easily understood that the more abstract the concept, the less likely one is to discover an appropriate criterion for assessing a measure of it (Carmines & Zeller, 1979) .

Content validity is limited by the adequacy of the universe specification, which is usually couched in imprecise terms and can rarely mention a pertinent aspect of the task . Bohrnstedt(1983) has argued that "while we enthusiastically endorse the procedures, we reject the concept of content validity on the grounds that there is no rigorous way to assess it " . Indeed, that is true, because the process of specifying the domain of content is considerably more complex when dealing with the abstract concepts typically found in the social sciences . In addition to that, to measure most concepts in the social sciences, it is impossible to sample content . To some degree, researchers organize or formulate a set of items that are intended to reflect the content of a given theoretical concept . However, researchers are facing the problem of the representativeness of the particular items . However, without a particular random sampling of content , they cannot insure the representativeness of the particular items (Carmines & Zeller , 1979) .

Loevinger, (1957) recommended that the claims of content validity should be dropped and that attention should be confined to construct validity . In sum, although researcher(s) should attempt to insure the criterion and content validity of any empirical measurement , these problems have prevented both criterion and content validation becoming fully adequate for assessing the validity of social science measures .

## **Conclusion :**

There is a tendency to move to put heavier emphasis on construct validation (Anastazi , 1969; Cronbach, 1984; Guion, 1978; Hopkins & Stanley,1981). Such a change toward this notion could be useful for psychological and educational tests, and it could guide the measurement specialists to put more effort in understanding the test validity, interpretation and use of test results . In addition , construct validation directs specialists'attention to focus on psychological theories and to help them in findig new ways in gathering data for validation .

Many measurement specialists claimed that construct validation includes the other types of validation . In other words, it subsumes the other types of validity . I believe that construct validation is a container which includes the other types of validity . We can also say that, any one has taken the point of view that there are different types of validity (content, criterion- related, and construct validity) and each type of validity is capable of achieving a different goal of testing, construct validity should be done earlier or, in other words, it is a prerequisite to other types of validity . Therefore, it is logical to conclude that construct validity is sufficient and necessary for other types of validity and measurement.

## REFERENCES

- Anastasi, A. (1969). Psychological Testing. New York : MacMillan .
- Anastasi , A. (1976). Psychological Testing (4th ed . ) . New York : MacMillan .
- Bohrnstedt , G . W . (1983) "Measurement " , in J. Wright and P. Rossi (eds . )  
Handbook of Survey Research . New york : Academic Press .
- Campbell , D . T . , & Fiske, D. W. (1959) . Convergent and Discriminate Validity by  
Multitrait-Multimethod Matrix . Psychological Bulletin , 56, 81-105 .
- Campbell, D . T. (1960) . Recommendations for APA Test Standards Regarding  
Construct, Trait, or Discriminant Validity . American Psychologist , 15,  
546-553 .
- Carmines, E. G. , & Zeller, R.A. (1979). Reliability and Validity Assessment . Sage  
Publications, Inc .
- Cronbach, L.J.(1971). Test Validation . In R . L . Thorndike (ED.), Educational  
Measurement, 2nd. ed., Washington, D.C., American Council on  
Education .
- Cronbach, L. J. (1980) . Selection Theory for a Political World . Public Personnel  
Manage , 9, 37-50 .
- Cronbach, L. J.(1984) . Essentials of Psychological Testing . New York : Harper &  
Row, Publishers.
- Cronbach, L.J., & Meehl,P.E. (1955). Construct validity in Psychological Tests .  
Psychological Bulletin , 52, 281-302 .
- French,J.W.& Mechael , W.B. (1966) . Standards for Educational and Psychological  
Tests and Manuals . American Psychological Association.
- Guion, R.(1978) . Content Validity in Moderation . Personal Psychology , 31, 205-214 .
- Hopking, K., & Stanley, J.(1981) . Educational and Psychological Measurement and  
Evaluation, Englewood Cliffs ; N.J.: Prentice-Hall .
- Lemke, E., & Wiersma, W. (1976) . Principles of Psychological Measurement .
- Lennon, R.T (1956) . "Assumptions Underlying the Use of Content Validity " ,  
Educational and Psychological Measurement , 16,294-304.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory,  
Psychological Reports , 3, 635-694 .
- Mehrens, W. A . and Lehmann, I . J . (1984). Measurement and Evaluation in  
Educational Psychology, Sounders College Publishing .
- Messick, s. (1980). " Test Validity and the Ethics of Assessment " , American  
Psychologist , 35 (11), 1012-1027 .
- Messick, S . (1981) . Evidence and Ethics in the evaluation of Tests , Educational  
Researcher , 10 (9), 9-20 .
- Nunnally, J. C. (1978) . Psychometric Testing . New York : McGraw-Hill .
- Tenopyr, M. (1977) . Content-Construct Confusion . personal Psychology , 30 ,47  
54.
- Thorndike, R., & Hangen, E. (1969) . Measurement and Evaluation in Psychology  
and Education . New York : John Wiley Sons, Inc.
- Wolf, R. (1982). Validity of Tests . In H .E . Mitzel (Ed.) Encyclopedia of  
Educational Research, (Vol. 4) . New York : MacMillan & Co.