

Testing

Test types: Contrasting types of ESL tests

Knowledge
Subjective
Productive
Language subskills
Norm-referenced
Discrete-point tests
Proficiency

Performance (or Skills) tests
Objective
Receptive
Communication skills
Criterion-referenced
Integrative
Achievement

Testing: Morrow

Keith Morrow. Communicative language testing: revolution or evolution? *The Communicative Approach to Language Teaching*, Edited by C.J. Brumfit and K. Johnson. Oxford University Press. Pp. 143-157.

I. Reliability

One of the most significant features of psychometric tests as opposed to those of 'pre-scientific' days is the development of the twin concepts of reliability and validity.

The basis of the reliability claimed by Lado is objectivity. The rather obvious point has, however, not escaped observers that Lado's tests are objective only in terms of actual assessment. In terms of the evaluation of the numerical score yielded, and perhaps more importantly, in terms of the construction of the test itself, subjective factors play a large part.

It has been equally noted by observers that an insistence on testing procedures which can be objectively assessed has a number of implications for the data yielded. Robinson (*op. cit.*) identifies three areas of difference between testing procedures designed to yield data which can be objectively assessed and those which are open to subjective assessment.

1. The amount of language produced by the student. In an objective test, students may actually produce no language at all. Their role may be limited to selecting alternatives rather than producing language.

2. Thus the type of ability which is being tested is crucially different. In a subjective test the candidate's ability to produce language is a crucial factor; in an objective test the ability to recognize appropriate forms is sufficient.
3. The norms of language use are established on different grounds. In an objective test the candidate must base his responses upon the language of the examiner; in a subjective test, the norms may be his own, deriving from his own use of the language. Thus an objective test can reveal only differences and similarities between the language norms of the examiner and candidate; it can tell us nothing of the norms which the candidate himself would apply in a use situation.

The above factors lead to what Davies has called the reliability-validity 'tension'. Attempts to increase the reliability of tests have led test designers to take an over-restrictive view of what it is that they are testing.

Other terms of importance: norm referencing, criterion referencing,

II. Validity

The idea that language test designers should concern themselves with validity-in other words that they should ask themselves whether they [147] are actually testing what they think they are testing, and whether what they are testing is what they ought to be testing-is clearly an attractive one. But unfortunately, because of the 'tension' referred to above, designers working within the tradition we are discussing seem than totally convincing.

Five types of validity which a language test may claim are traditionally identified (cf. Davies).

Face	the test looks like a good one.
Content	the test accurately reflects the syllabus on which it is based.
Predictive	the test accurately predicts performance in some subsequent situation.
Concurrent	the test gives similar results to existing tests which have already been validated.
Construct	the test reflects accurately the principles of a valid theory of foreign language learning.

Statistical techniques for assessing validity in these terms have been developed to a high, and often esoteric level of sophistication. But unfortunately, with two exceptions (face, and possibly predictive) the types of validity outline above are all ultimately circular. Starting from a certain set of assumptions about the nature of language and language learning will lead to language tests which are perfectly valid in terms of these assumptions, but whose value must inevitably be called into question if the basic assumptions themselves are challenged. Thus a test which perfectly satisfies criteria of content, construct or concurrent validity may nonetheless fail to show in any interesting way how well a candidate can perform in or use the target language. This may occur quite simply if the construct of the language learning theory, and the content of the syllabus are themselves not related to this aim, or if the test is validated against other language tests which do not concern themselves with this objective. There is clearly no such thing in testing as 'absolute' validity. Validity exists only in terms of specified criteria, and if the criteria turn out to be the wrong ones, then validity claimed in terms of them turns out to be spurious. *Caveat emptor.*

Notes:

The material discussed above was not presented for you to simply accept; in fact, if you simply accept it, then I would be somewhat disappointed in you.

In the discussion of testing, what we were discussing was the meanings of the TECHNICAL TERMS reliability and validity. We were NOT saying that if a test was reliable and valid in a technical sense that you yourself would necessarily think it was a reliable and valid test in a non-technical sense. Quite the opposite. I suspect that many tests that are RELIABLE and VALID do not really seem reliable and valid to you and me. So what we did in class was discuss what test makers mean when they use these terms. One of the main points was to see what the difference was between how they use these terms and what they mean by them and how you and I use these terms and what we mean by them. I should hope that you do not blindly accept the test makers uses of the word reliable and valid.