

University of Tikrit

College of Education for

Humanities English Department

Ph.D. Candidates



**Language Testing
Part 2**

Prof. Istabraq Tariq Alazzawi (Ph.D.)

2025-2026

1.2 Theories of validity

Language tests, like other tests, are deliberate samples related to an individual's language knowledge or language behavior. To get general state of that person's knowledge or ability, language tests require technical skill in their construction and application, in order to make the inferences that we draw from test results interpretable and supportable. The validation of language tests refers to the process of gathering arguments and evidence in support of the interpretations and uses we may wish to make of test scores.

Generally, there are **two basic processes** involved: **1) an articulation** of the exact character of the knowledge we seek and of the purpose for which we seek it; and **2) the gathering of empirical evidence** to support the interpretations we wish to make of candidate performance.

Messick (1989) clarifies the fundamental aspects of test validation in the form of a matrix. The column **“Test interpretation”** involves considering the evidence of test validity outside any specific context of its deployment; the column **“Test use”** looks at the actual use of the test in a specific context. The social consequences of the actual deployment of tests are the focus of the final cell in the matrix.

- **Messick's approach represents a unified model of validation including**

	TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity	Construct validity + Relevance/utility
CONSEQUENTIAL BASIS	Value implications	Social consequences

Figure 31.2 Facets of validity (from Messick, 1989, p. 20)

three various kinds of validity:

1) Content validity: concerning the representativeness of test content.

- 2) **Construct validity:** concerning the coherence and defensibility of the theory (of knowledge, skill, and so on) on which the test is based on.
- 3) **Criterion-related validity:** the extent to which measures produced by the test makes sense due to relevant measures of skill, or predict outcomes associated with the skill being measured in the test.

2) Validation Research in Language Testing:

Test validation is the process of investigating the meaningfulness and defensibility of the inferences we make about individuals based on their test performance. The need to defend our interpretations of the meanings of test scores is because of the necessarily indirect relationship between the test and what we ultimately want to know about.

Validation: is the process of investigating the relationship between the claims of the test and evidence in support of these claims, both from the test scores themselves, or from independent evidence.

Language aptitude: is defined as “the extent to which an individual possesses specific language learning ability”. It differs from language proficiency (current level of attainment in second language) in that, language aptitude can be measured before an individual begins second language study. Language aptitude tests are designed to enable us to choose those with the greatest potential to benefit from language study, or to reject those who will struggle to succeed.

As a result, the relationship between test construct and test performance is a two-way street: tests are built on constructs, but can offer evidence of the validity of the constructs on which they are built. The investigation of these relationships is an important component of language test validation research.

3) Language Testing as Institutional Practice:

Language testing functions not only as an educational tool but also as a social and political mechanism that regulates access to opportunities and resources. It plays a gatekeeping role, influencing who can progress within or across societies. Within educational systems, **language tests serve multiple purposes:**

- ✓ They ensure responsibility,

- ✓ They provide feedback for students and parents
- ✓ They are used for certification and degree assessment.
- ✓ They help determining the distribution of educational opportunities and resources.
- ✓ They often indicate both achievement and potential for further study.

The globalization of education and work has expanded the institutional role of language testing. For instance, in Europe, the Common European Framework of Reference for Languages (CEFR) has standardized language assessment across countries, linking testing with educational and employment systems in which assessment has become a key policy tool in managing labor movement and educational quality.

Moreover, the internationalization of English has increased reliance on language tests for student admissions and employment, reinforcing their institutional power. This has led scholars such as Shohamy (2001) to develop the concept of Critical Language Testing, which uses critical social theory to examine how language testing supports or challenges systems of control and inequality.

4) Language Testing Research and Language learning:

Language testing research mainly looks at big international tests like **TOEFL** and **IELTS**, which are used to measure English ability. These tests are important because they act as gatekeepers, they decide who can study, work, or immigrate to English-speaking countries. Institutions spend a lot on creating and checking these tests to make sure they are fair and accurate. **The research mostly focuses on using tests for:**

- University admission,
- Immigration requirements, and
- Employment or professional certification (like teachers or health workers).

Some curriculum-based scales and frameworks for describing students' progress over time. have dual purposes:

- ✓ Guide teachers in measuring student progress
- ✓ Provide a standard system for reporting learner achievement.

According to such scales and frameworks, **formal testing** has been introduced as means of monitoring the achievement of educational outcomes, they are used to check how well students and schools are performing. Another area of study is **self-assessment**, where learners judge their own language skills in which it helps learners become more aware of their abilities and take active responsibility for their learning progress.

5) Current and Future Development in Language Testing Research:

Language testing has become more advanced because of new statistical tools that help researchers study test validity and score patterns. Measurement theories like Generalizability Theory and Multi-Faceted Rasch Measurement help identifying differences in test scores, particularly the differences between raters in terms of their relative severity and consistency.

Item Response Theory; allows powerful generalizations about candidates and items based on the responses of samples of candidates to samples of items. It also allows adaptive testing, where the computer selects items suited to the test-taker's level. This makes testing more accurate and efficient.

Test materials can now be delivered online, allowing learners to assess themselves anytime. This approach is being used in large-scale multilingual projects in Europe, such as DIALANG, and other similar applications.

Even with all these technological and statistical advances, the core issue remains validity, whether test results truly represent a learner's language ability. Researchers must also think about:

- ❖ The impact of tests on teachers and learners (e.g., stress, motivation, teaching focus).
- ❖ The social and political values behind tests.