



Standard Practice for Assessing Language Proficiency¹

This standard is issued under the fixed designation F2889; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 *Purpose*—This practice describes best practices for the development and use of language tests in the modalities of speaking, listening, reading, and writing for assessing ability according to the Interagency Language Roundtable (ILR)² scale. This practice focuses on testing language proficiency in use of language for communicative purposes.

1.2 *Limitations*—This practice is not intended to address testing and test development in the following specialized areas: Translation, Interpretation, Audio Translation, Transcription, other job-specific language performance tests, or Diagnostic Assessment.

1.2.1 Tests developed under this practice should not be used to address any of the above excluded purposes (for example, diagnostics).

2. Referenced Documents

2.1 *ASTM Standards*:³

F1562 *Guide for Use-Oriented Foreign Language Instruction*

F2089 *Guide for Language Interpretation Services*

F2575 *Guide for Quality Assurance in Translation*

3. Terminology

3.1 *Definitions*:

3.1.1 *achievement test, n*—an instrument designed to measure what a person has learned within or up to a given time based on a sampling of what has been covered in the syllabus.

3.1.2 *adaptive test, n*—form of individually tailored testing in which test items are selected from an item bank where test items are stored in rank order with respect to their item difficulty and presented to test takers during the test on the basis of their responses to previous items, until it is determined

that sufficient information regarding test takers' abilities has been collected. The opposite of a *fixed-form test*.

3.1.3 *authentic texts, n*—texts not created for language learning purposes that are taken from newspapers, magazines, etc., and tapes of natural speech taken from ordinary radio or television programs, etc.

3.1.4 *calibration, n*—the process of determining the scale of a test or tests.

3.1.4.1 *Discussion*—Calibration may involve anchoring items from different tests to a common difficulty scale (the theta scale). When a test is constructed from calibrated items then scores on the test indicate the candidates' ability, i.e. their location on the theta scale.

3.1.5 *cognitive lab, n*—a method for eliciting feedback from examinees with regard to test items.

3.1.5.1 *Discussion*—Small numbers of examinees take the test, or subsets of the items on the test, and provide extensive feedback on the items by speaking their thought processes aloud as they take the test, answering questionnaires about the items, being interviewed by researchers, or other methods intended to obtain in-depth information about items. These examinees should be similar to the examinees for whom the test is intended. For tests scored by raters, similar techniques are used with raters to obtain information on rubric functioning.

3.1.6 *computer adaptive test, n*—a test administered by a computer in which the difficulty level of the next item to be presented to test takers is estimated on the basis of their responses to previous items and adapted to match their abilities.

3.1.7 *construct, n*—the knowledge, skill or ability that is being tested.

3.1.7.1 *Discussion*—The construct provides the basis for a given test or test task and for interpreting scores derived from this task.

3.1.8 *constructed response, adj*—a type of item or test task that requires test takers to respond to a series of open-ended questions by writing, speaking, or doing something rather than choose answers from a ready-made list.

3.1.8.1 *Discussion*—The most commonly used types of constructed-response items include fill-in, short-answer, and performance assessment.

¹ This practice is under the jurisdiction of ASTM Committee F43 on Language Services and Products and is the direct responsibility of Subcommittee F43.04 on Language Testing.

Current edition approved May 1, 2011. Published June 2011. DOI: 10.1520/F2889-11.

² Interagency Language Roundtable, Language Skill Level Descriptors (<http://www.govtllr.org/Skills/ILRscale1.htm>).

³ For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.