AAPPL

ACTFL ASSESSMENT OF PERFORMANCE TOWARD PROFICIENCY IN LANGUAGES

An assessment is only as valid and reliable as the research that supports it.

About the Study

The purpose of the study, **A validity argument to support the ACTFL Assessment of Performance toward Proficiency (AAPPL),** was to document the reliability and develop a validity argument for the assessment, using evidence from over 10,000 test results. The tests were administered in 2014 to students in grades 5 to 12; in French, Spanish, and Chinese. The authors, Dr. Troy Cox (Brigham Young University) and Dr. Margaret Malone (ACTFL and Georgetown University), utilized several diagnostics in Rasch measurements to provide empirical evidence on the validity of the assessment based on the three key areas listed below.

"Without evidence of the validation of a measurement instrument, test users might not know which instrument is most suited for their circumstances and how the resulting scores relate to stated learning outcomes." (p. 550)

1. Design – *the extent to which the item and test specifications function as intended*

"One evidence of validity in the design of criterion-referenced tests is the extent to which the empirical item difficulties that are calculated after the items have been administered align with the levels of difficulty described in the standards." (p. 555)

"…there were distinct differences in item difficulty, and none of the items that targeted the major proficiency levels (Novice, Intermediate, and Advanced) was misaligned." (p. 555)

"The data reported above indicate that the AAPPL has strong design evidence, especially with the productive skills, where the intended and actual item difficulties best aligned…finally, there was evidence that the proficiency guidelines formed a unitary construct." (p. 573)

2. Assessment Records – *the consistency and replicability of examinee scoring with or without raters*

"Thus, the difficulty of the items for both productive modes aligned with the underlying construct – every Novice item was easier than the Intermediate items, and those were easier than the items that targeted the Advanced level. This result provides evidence for the argument of test validity." (p. 557)

"In terms of rater reliability, the raters were self-consistent, but also reliably different from each other." (p. 571)

"The results showed strong support for the construct that underpins both the receptive and productive skills." (p. 569)

3. Interpretations – *the generalizability of scores across languages*

"For a high-stakes test, the reliability should be greater than 0.80 (…) …productive skills were most reliable, with examinee person separation reliability in speaking in Chinese of 0.95, French of 0.92, and Spanish of 0.93 and in writing ranging from Chinese = 0.97, French = 0.93, and Spanish = 0.90." (p. 567)

"…the item difficulty correlations between the three languages for the items that were used to assess productive skills (speaking and writing) were strong; in fact, they were almost perfectly correlated across all three levels…(…)…for the receptive skills, a relationship among languages exists, but it is not as strong as for the productive skills." (p. 569)

"The results of this study have already been considered and they have been, and continue to be, incorporated in new versions of the AAPPL…While ongoing improvements are being made, the AAPPL demonstrates a strong argument for test validity." (p. 573)

www.languagetesting.com/AAPPL

A Word on Proficiency and Performance

In terms of assessments, proficiency and performance are interrelated concepts that are often mistakenly used interchangeably. Proficiency tests generally elicit spontaneous, unrehearsed language in potentially unfamiliar contexts. In performance tests, test takers can practice and prepare for a familiar set of tasks and topics. The AAPPL is a marriage of both proficiency and performance tests, in that ACTFL releases the topics in advance and provides sample items from previous years. Present-year test items are regularly refreshed and are not released in advance. Students are therefore unable to rehearse content specific to the test itself.

About the AAPPL

Based on the results of a survey of 1,600 language educators, ACTFL developed a test that reflected the *World-Readiness Standards for Learning Languages* in addition to the *ACTFL Proficiency Guidelines* and *NCCSFL-ACTFL Can-Do Statements*. The AAPPL is designed to be consistent across languages, so that all test takers are assessed on similar content that aligns with proficiency-focused curricula. The AAPPL consists of four computer-administered test modules that test three ACTFL modes of communication: Interpersonal Listening and Speaking, Interpretive Listening and Reading, and Presentational Writing. There are two forms of each test module. Form A targets Novice to Intermediate levels of proficiency; Form B targets Intermediate to Advanced Low.

Reference: Cox, T.L., & Malone, M.E. (2018). A validity argument to support the ACTFL Assessment of Performance toward proficiency. *Foreign Language Annals*, 51 (3), 548-574. Retrieved from https://doi.org/10.1111/flan.12353

Proper Citation of this Document: ACTFL. (2019, January). *AAPPL Validity Argument Overview* [Brochure]. Alexandria, VA: ACTFL.