

Versant™ English Test

Validation Summary

A common question about any test is “how do we know this is a good, accurate, consistent test?” To answer that question, one needs to look at two factors that are critical to high quality tests:

1. **Validity and Accuracy:** Does the test really measure what it is intended to measure? Is it a valid assessment of the construct it claims to test? Can a machine score like a human expert can?
2. **Reliability:** Can you expect consistent results from the test? Would a person score the same if they took the test over and over again?

For a test to be high quality, it must be valid, accurate, and reliable, as demonstrated by empirical validation investigations. A series of studies have been conducted on the Versant English Test as part of ongoing research, and the data show that the Versant English Test has strong validity, impressive accuracy, and high reliability.

Validity and Accuracy

Over the years, the Versant Test Development team has collected data on parallel administrations of the Versant English Test and other well-established language examinations. These data allow for a measure of concurrent validity for the Versant English Test. Listed in the table below are the correlation coefficients (indicating the degree of correspondence between each other language assessment and the Versant English Test), as well as the number of test takers whose data were utilized in calculating each correlation coefficient.

Instrument	Correlation (<i>r</i>)	Num of test-takers
TOEFL Overall	.75	392
TOEFL Reading	.64	321
TOEIC Listening	.71	171
TOEFL Listening	.79	321
New TOEFL Listening	.78	321
TSE	.88	58
New TOEFL Speaking	.84	321
Common European Framework of Reference	.88	303
ILR Speaking	.75	51
TOEFL iBT Speaking	.75	130
IELTS Speaking	.76	130

Note that the table is divided into three sections: the upper section presents data from tests of written language, which should be expected to have only a moderate correlation with a speaking test such as the Versant English Test. The middle section contains data on tests of listening comprehension, which, because they are presented orally, should have a somewhat higher correlation with the Versant English Test. The bottom section presents correlations with instruments

for assessing oral skills. These tests are expected to show the highest correlation with the Versant English Test.

The above data strongly indicate that the Versant English Test measures the same constructs as other examinations of English language, and that it overlaps substantially with other instruments designed to assess spoken skills. In short, the Versant English Test is a highly valid instrument for assessing fluency in spoken English.

Another important aspect of any machine-graded test is to assure that automatically generated scores are as accurate as those of trained human evaluators. In one study, 50 Versant English test takers were evaluated by both the standard Versant English Test machine scoring and by expert human raters. The correlation coefficient for the two sets of scores (human vs. machine) was an impressive 0.97 (correlations run on a scale from 0 to 1). This means that when the Versant English Test automated scoring system provides scores they accurately predict scores given by human experts.

Reliability

Another estimate of a test's reliability is the test-retest reliability, in which scores from the same test taker are compared between two administrations of the test. In a study to investigate consistency of scoring, 140 test takers were administered the Versant English Test three times in succession. On each test occasion they were given a standard Versant English Test (each with unique randomly generated items). The correlation between the three test scores was 0.97. This is strong evidence that the Versant English Test provides extremely consistent scores each time a test taker takes it, and that familiarity with the test format does not artificially boost test scores. These data indicate that the Versant English Test is a very reliable test.

A common estimate of a test's reliability uses the split-half method, and the split-half reliability of the Versant English Test was calculated during a study reported in the Test Description and Validation Summary document. The split-half reliability of the Overall score is 0.97 (on a scale from 0 to 1). The high reliability score is a good indication that the computerized assessment will be consistent for the same test taker whenever he or she takes the test. Reliabilities for the subscores are similarly high: 0.97 for Pronunciation, 0.95 for Fluency, 0.88 for Vocabulary, and 0.93 for Sentence Mastery. This high degree of reliability is comparable to that found for a human rater in the same validation study.