

## Computer-based versus Phone delivery of Versant English Test spoken language tests

*This document describes an experimental study conducted by the Versant Test Development team to examine the possible grading differences between two different Versant administration modalities: computer-delivered (“CDT”) and phone. The results demonstrate that performance on the Versant English Test is consistent across test administration modalities.*

### **Design Overview**

In this within-subjects design, each participant took three Versant English Tests: a practice test (taken prior to the participants’ arrival at the study), one computer-delivered test (“CDT”), and one telephone test (“Phone”). Participants always took the practice test first in order to familiarize themselves with the test format. The order of presentation of the CDT versus phone modality of the test was randomized and counterbalanced across participants.

### **Participants**

Sixty-eight non-native English speakers (48 females and 20 males) served as participants. The participants ranged in age from 19-80 years (mean age = 39.0,  $sd = 13$ ). Participants represented a wide range of English proficiency levels and spoke a wide variety of native languages including Arabic, Cantonese, French, German, Gujarat, Hindi, Igbo, Japanese, Korean, Mandarin, Portuguese, Russian, Slovak, Spanish, Swedish, Tamil, Telugu, Thai, and Urdu.

### **Materials**

The computer-delivered version of the test was administered on one of four computers: A Fujitsu laptop with 12” screen, a Compaq laptop with 12” screen, a Dell laptop with 12” screen, or a Dell Desktop with 14” screen. Each computer was equipped with either an Altec Lansing headset and microphone or Logitech headset and microphone. Sixty-eight Test Identification Numbers (TINs) were used for CDT administration (on a staging system), and 68 test papers were used for the phone version of the Versant English Test.

### **Procedure**

Prior to the testing session, each participant was provided with a practice test and asked to take it from their home phone. When participants arrived at the testing session, they were given the opportunity to complete the practice test if they had not already done so. Each participant was tested individually. Participants took the two tests, beginning with either the computer based or the phone delivery version of the test.

For the CDT version of the test, the proctor assisted each participant in putting on the headset and in adjusting the microphone. When taking the phone version of the test, participants were given the test paper and allowed to read through it. They were then shown the telephone and instructed on proper dialing procedure. After making sure the participant had no questions, the proctor left the room. Each version of the test lasted approximately 20 minutes. At the conclusion of each version of the test, the participant was asked to complete a short questionnaire relating to their perception of the exam. At the conclusion of the entire session, the participant was asked the forced-choice question of which of the two modalities he or she preferred.

## Results

Paired, two-tailed t-tests for significance ( $\alpha = .05$ ) were conducted for the overall score and for each subscore. No significant difference was found in the overall scores of the CDT version ( $m = 52.3, sd = 13.9$ ) and the phone delivered version ( $m = 52.7, sd = 14.5$ ) ( $t(67) = -0.66, n.s.$ ). There were, however, two significant differences found between the subscores. The CDT-based subscore for *sentence mastery* ( $m = 55.9, sd = 14.1$ ) was higher than the phone-based subscore ( $m = 51.9, sd = 14.8$ ) ( $t(67) = 4.13, p < .001$ ). The CDT subscore for *pronunciation* ( $m = 48.6, sd = 13.9$ ) was lower than that of the phone version ( $m = 52.1, sd = 15.6$ ) ( $t(67) = 5.4, p < .001$ ). There were no significant differences between the *vocabulary* and *fluency* subscores. Table 1 summarizes these results.

Table 1. Mean overall score and subscores on Versant English Test across CDT and phone

Overall		Sentence Mastery		Vocabulary		Fluency		Pronunciation	
Phone	CDT	Phone	CDT	Phone	CDT	Phone	CDT	Phone	CDT
52.3	52.7	51.9	55.9	56.5	53.8	49.5	51.6	52.1	48.6
p = .51		p < .001*		p = .13		p = .13		p < .001**	

\* CDT Higher,  $p < .001$

\*\* CDT Lower,  $p < .001$

Differences across presentation modality for sentence mastery (4.0) and pronunciation (3.5) fell within the Standard Error of the Mean (SEM) for these subscores (5.9 and 3.8, respectively). Therefore, post-hoc analyses were conducted using Cohen's d to determine effect size, which reflects whether differences between two means are consistent and large enough to be of "practical concern". The Cohen's d values obtained for sentence mastery ( $d = -0.23$ ) and pronunciation ( $d = 0.20$ ) subscores constitute what is traditionally considered a small effect (i.e., 0.2 is a small effect, 0.5 is a medium effect, and 0.8 is a large effect).

To determine whether this effect may have been driven by a few extreme outliers, the data were visualized in histograms. As Figure 1 shows, vocabulary and fluency subscore differences between CDT and phone are fairly evenly distributed around 0, indicating that changes in score from one modality to the other were not systematic. Sentence mastery and pronunciation subscores, however, exhibit outliers that appear to push the overall difference disproportionately in the anticipated direction. That is, for the sentence mastery subscore, 4 test takers scored 20 points lower in the computer modality than in the phone modality, while for the pronunciation subscore, 7 test takers scored >13 points higher on the phone than on the computer. These extreme values fall outside of an otherwise relatively heterogeneous distribution and appear to have driven the findings of statistical significance despite small effect size.

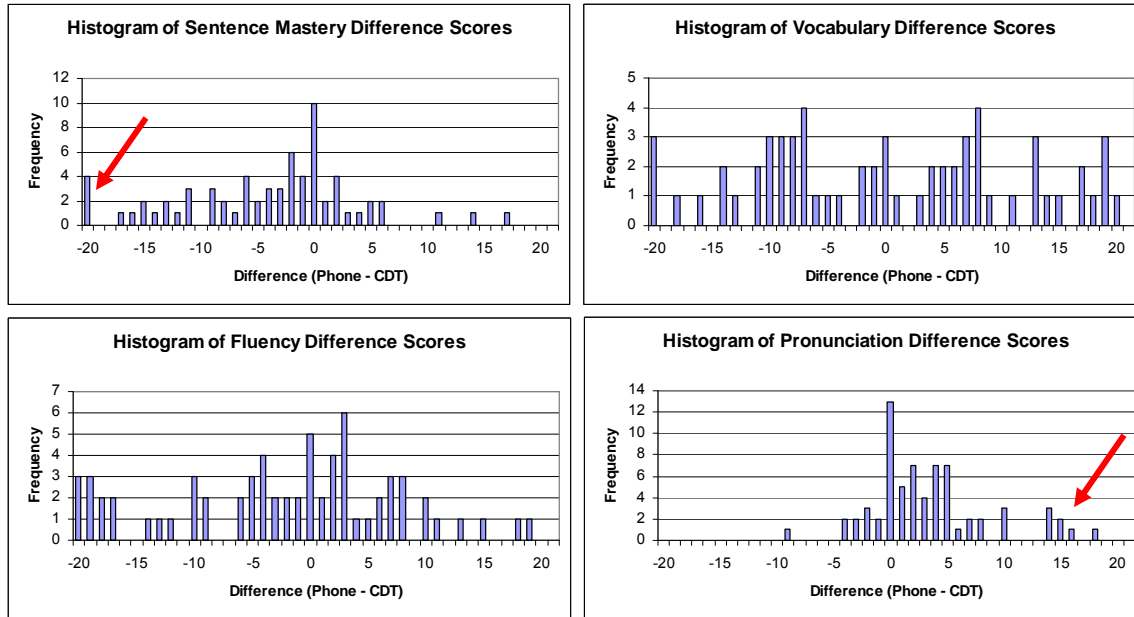


Figure 1: Histograms of difference scores (Phone – CDT) across subscores

## Discussion

In this study, the same 68 participants took the Versant English Test in both the CDT and phone modalities. Mean overall scores did not differ significantly from one modality to another. The mean *sentence mastery* subscore was 4 points higher when taken by CDT than when taken by phone, and the mean *pronunciation* subscore was 3.5 points lower by CDT than by phone. These differences were statistically significant, although they fell within the SEM of each subscore. Post-hoc analyses indicated the effect size was small for both subscores, suggesting that a handful of outliers were behind the statistical significance. This conclusion was supported by visual inspection of the data.

It is unclear why *pronunciation* and *sentence mastery* subscores should differ according to administration modality. Also unclear is why these subscores should be differentially affected (e.g., *sentence mastery* scores were, on average, higher in CDT while *pronunciation* scores were lower). One explanation for the reduced *pronunciation* subscore is that the specific headsets used in the CDT tests may have attenuated particular features of the test-takers' speech, contributing to reduced scoring precision due to the scoring system's processing of a less complete signal (i.e., compared to the signal obtained by phone). While plausible, this explanation is conjectural and does not explain the observed increase in sentence mastery subscores.

Overall, these results are strongly suggestive that there is no systematic, practically significant difference between Versant English Test scores obtained from the same test taker when the test is taken via CDT or phone.



**Contact Us**  
To try a sample test or get more information, contact us at:

US: 800.211.8378  
Int'l: +1 650.470.3505  
sales@pearsonkt.com

Or visit us online at:  
[www.VersantTest.com](http://www.VersantTest.com)