Versant™ English Test

Test Description and Validation Summary

Table of Contents

1. Introduction

The Versant™ English Test, powered by Ordinate technology, is an assessment instrument designed to measure how well a person understands and speaks English. The Versant English Test is intended for adults and students over the age of 15 and takes approximately 15 minutes to complete. Because the Versant English Test is delivered automatically by the Versant testing system, the test can be taken at any time, from any location by phone or via computer. A human examiner is not required. The computerized scoring allows for immediate, objective, and reliable results that correspond well with traditional measures of spoken English performance.

The Versant English Test measures *facility* with spoken English, which is a key element in English oral proficiency. Facility in spoken English is how well the person can understand spoken English on everyday topics and respond appropriately at a native-like conversational pace in English. Academic institutions, corporations, and government agencies throughout the world use the Versant English Test to evaluate the ability of students, staff, and officers to understand spoken English and to express themselves clearly and appropriately in English. Scores from the Versant English Test provide reliable information that can be applied to placement, qualification and certification decisions, as well as monitor progress and measure instructional outcomes.

2. Test Description

2.1 Test Design

The Versant English Test may be taken at any time from any location using a telephone or a computer. During test administration, the Versant testing system presents a series of recorded spoken prompts in English at a conversational pace and elicits oral responses in English. The voices of the item prompts are from native speakers of English from several different regions in the U.S, providing a range of speaking styles.

The Versant English Test has six item types: Reading, Repeats, Short Answer Questions, Sentence Builds, Story Retelling, and Open Questions. All item types except for Open Questions elicit responses that can be analyzed automatically. These item types provide multiple, fully independent measures that underlie facility with spoken English, including phonological fluency, sentence construction and comprehension, passive and active vocabulary use, listening skill, and pronunciation of rhythmic and segmental units. Because more than one item type contributes to each subscore, the use of multiple item types strengthens score reliability.

The Versant testing system analyzes the candidate's responses and posts scores to a secure website usually within minutes of the completed test. Test administrators and score users can view and print out test results from a password-protected website.

The Versant English Test provides numeric scores and performance levels that describe the candidate's facility in spoken English – that is, the ability to understand spoken English on everyday topics and to respond appropriately at a native-like conversational pace in intelligible English. The Versant English Test score report is comprised of an Overall score and four diagnostic subscores: Sentence Mastery, Vocabulary, Fluency, and Pronunciation. Together, these scores describe the candidate's facility in spoken English.

2.2 Test Administration

Administration of a Versant English Test generally takes about 15 minutes over the telephone or via a computer. It is best practice (even for computer delivered tests) for the administrator to give a test paper to the candidate at least five minutes before starting the test (see Appendix). The candidate then has the opportunity to read both sides of the test paper and ask questions before the test begins. The administrator should answer any procedural or content questions that the candidate may have.

The delivery of the recorded item prompts is interactive – the system detects when the candidate has finished responding to one item and then presents the next item.

2.2.1 Telephone Administration

Telephone administration is supported by a test paper. The test paper is a single sheet of paper with material printed on both sides. The first side contains general instructions and an explanation of the test procedures. These instructions are the same for all candidates. The second side has the individual test form, which contains the phone number to call, the Test Identification Number (TIN), the spoken instructions written out verbatim, item examples, and the printed sentences for Part A: Reading. The individual test form is unique for each candidate.

When the candidate calls the Versant testing system, the system will ask the candidate to use the telephone keypad to enter the Test Identification Number that is printed on the test paper. This identification number is unique for each candidate and keeps the candidate's information secure.

A single examiner voice presents all the spoken instructions for the test. The spoken instructions for each section are also printed verbatim on the test paper to help ensure that candidates understand the directions. Candidates interact with the test system in English, going through all six parts of the test until they complete the test and hang up the telephone.

2.2.2 Computer Administration

For computer administration, the computer must have an Internet connection and Pearson's Computer Delivered Test (CDT) software (available at http://www.versanttest.com/technology/platforms/cdt/index.jsp). The candidate is fitted with a microphone headset. The CDT software prompts the candidate to adjust the volume and calibrate the microphone before the test begins.

The instructions for each section are spoken by an examiner voice and are also displayed on the computer screen. Candidates interact with the test system in English, speaking their responses into the microphone. When a test is finished, the candidate clicks a button labeled, "End Test".

2.3 Test Format

The following subsections provide brief descriptions of the item types and the abilities required to respond to the items in each of the six parts of the Versant English Test.

Part A: Reading

In this task, the candidate reads printed, numbered sentences, one at a time, as prompted. For telephone administration, the sentences are printed on the test paper. For computer administration, the sentences are displayed on the computer screen. Reading items are grouped into sets of four sequentially coherent sentences, as in the examples below.

Examples:

1. Larry's next door neighbors are awful.
2. They play loud music all night when he's trying to sleep.
3. If he tells them to stop, they just turn it up louder.
4. He wants to move out of that neighborhood.

Presenting the sentences as part of a group helps the candidate disambiguate words in context and helps suggest how each individual sentence should be read aloud. The computer screen or test paper contains three groups of four sentences (i.e., 12 items). Candidates are prompted to read eight of the twelve sentences in a random order. The system tells the candidate which of the numbered sentences to read aloud (e.g., "Now, please read sentence 7."). After the candidate has read the sentence (or has remained silent for a period of time), the system prompts him or her to read another sentence from the list.

The sentences are relatively simple in structure and vocabulary, so they can be read easily and in a fluent manner by literate speakers of English. For candidates with little facility in spoken English but with some reading skills, this task provides samples of their pronunciation and reading fluency. The readings appear first in the test because, for many candidates, reading aloud presents a familiar task and is a comfortable introduction to the interactive mode of the test as a whole.

Part B: Repeat

In this task, candidates are asked to repeat sentences that they hear verbatim. The sentences are presented to the candidate in approximate order of increasing difficulty. Sentences range in length from three words to 15 words. The audio item prompts are spoken in a conversational manner.

Examples:

Get some water.
Let's meet again in two weeks.
Come to my office after class if you need help.

To repeat a sentence longer than about seven syllables, a person must recognize the words as spoken in a continuous stream of speech (Miller & Isard, 1963). Highly proficient speakers of English can generally repeat sentences that contain many more than seven syllables because these speakers are very familiar with English words, phrase structures, and other common syntactic forms. If a person habitually processes five-word phrases as a unit (e.g. "the really big apple tree"), then that person can usually repeat utterances of 15 or 20 words in length. Generally, the ability to repeat material is constrained by the size of the linguistic unit that a person can process in an automatic or nearly automatic fashion. As the sentences increase in length and complexity, the task becomes increasingly difficult for speakers who are not familiar with English sentence structure.

Because the Repeat items require candidates to organize speech into linguistic units, Repeat items assess the candidate's mastery of phrase and sentence structure. Given that the task requires the candidate to repeat full sentences (as opposed to just words and phrases), it also offers a sample of the candidate's fluency and pronunciation in continuous spoken English.

Part C: Short Answer Questions

In this task, candidates listen to spoken questions and answer each question with a single word or short phrase. The questions generally present at least three or four lexical items spoken in a continuous phonological form and framed in English sentence structure. Each question asks for basic information or requires simple inferences based on time, sequence, number, lexical content, or logic. The questions do not presume any knowledge of specific facts of culture, geography, history, or other subject matter; they are intended to be within the realm of familiarity of both a typical 12-year-old native speaker of English and an adult who has never lived in an English-speaking country.

Examples:

> What is frozen water called?
> How many months are in a year and a half?
> Does a tree usually have more trunks or branches?

To correctly respond to the questions, a candidate must identify the words in phonological and syntactic context, and then infer the demand proposition. Short Answer Questions measure receptive and productive vocabulary within the context of spoken questions presented in a conversational style.

Part D: Sentence Builds

For the Sentence Builds task, candidates hear three short phrases and are asked to rearrange them to make a sentence. The phrases are presented in a random order (excluding the original word order), and the candidate says a reasonable and grammatical sentence that comprises exactly the three given phrases.

Examples:

> in / bed / stay
> she didn't notice / the book / who took
> we wondered / would fit in here / whether the new piano

To correctly complete this task, a candidate must understand the possible meanings of the phrases and know how they might combine with other phrasal material, both with regard to syntax and pragmatics. The length and complexity of the sentence that can be built is constrained by the size of the linguistic unit (e.g., one word versus a three-word phrase) that a person can hold in verbal working memory. This is important to measure because it reflects the candidate's ability to access and retrieve lexical items and to build phrases and clause structures automatically. The more automatic these processes are, the more the candidate's facility in spoken English. This skill is demonstrably distinct from memory span (see Section 2.5, Test Construct, below).

The Sentence Builds task involves constructing and articulating entire sentences. As such, it is a measure of candidates' mastery of sentences in addition to their pronunciation and fluency.

Part E: Story Retelling

In this task, candidates listen to a brief story and are then asked to describe what happened in their own words. Candidates have thirty seconds to respond to each story. Candidates are encouraged to tell as much of the story as they can, including the situation, characters, actions and ending. The stories

consist of three to six sentences and contain from 30 to 90 words. The situation involves a character (or characters), setting, and goal. The body of the story describes an action by the agent of the story followed by a possible reaction or implicit sequence of events. The ending typically introduces a new situation, actor, patient, thought, or emotion.

Example:

> Three girls were walking along the edge of a stream when they saw a small bird with its feet buried in the mud. One of the girls approached it, but the small bird flew away. The girl ended up with her own feet covered with mud.

The Story Retelling items assess a candidate's ability to listen and understand a passage, reformulate the passage using his or her own vocabulary and grammar, and then retell it in detail. This section elicits longer, more open-ended speech samples than earlier sections in the test, and allows for the assessment of a wider range of spoken abilities. Performance on Story Retelling provides a measure of fluency, pronunciation, vocabulary, and sentence mastery.

Part F: Open Questions

In this task, candidates listen to spoken questions that elicit an opinion, and are asked to provide an answer with an explanation. Candidates have 40 seconds to respond to each question. The questions relate to day-to-day issues or the candidate's preferences and choices.

Examples:

> Do you think television has had a positive or negative effect on family life? Please explain.
> Do you like playing more in individual or in team sports? Please explain.

This task is used to collect longer spontaneous speech samples. Candidates' responses to items in this section are not scored, but are available for review by authorized listeners.

2.4 Number of Items

In the administration of the Versant English Test, the testing system serially presents a total of 63 items in six separate sections to each candidate. The 63 items are drawn at random from a large item pool. For example, each candidate is presented with 10 Sentence Builds from among those items available in the pool, so most or all items will be different from one test administration to the next. Proprietary algorithms are used by the testing system to select from the item pool – the algorithms take into consideration, among other things, an item's difficulty level and similarity to other presented items. Table 1 shows the number of items presented in each section.

Table 1. Number of Items Presented per Section

Task	Presented
A. Reading	8
B. Repeat	16
C. Short Answer Questions	24
D. Sentence Builds	10
E. Story Retelling	3
F. Open Questions	2
Total	63

2.5 Test Construct

For any language test, it is essential to define the test construct as explicitly as possible (Bachman, 1990; Bachman & Palmer, 1996). The Versant English Test is designed to measure a candidate's facility in spoken English – that is, the ability to understand spoken English on everyday topics and to respond appropriately at a native-like conversational pace in intelligible English. Another way to describe the construct *facility in spoken English* is "the ease and immediacy in understanding and producing appropriate conversational English" (Levelt, 1989). This definition relates to what occurs during the course of a spoken conversation. While keeping up with the conversational pace, a person has to track what is being said, extract meaning as speech continues, and then formulate and produce a relevant and intelligible response. These component processes of listening and speaking are schematized in Figure 1.

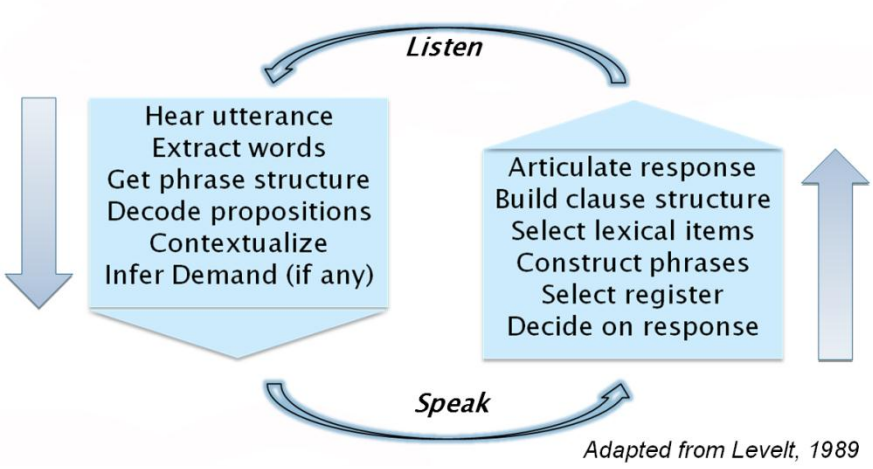

Figure 1. Conversational processing components in listening and speaking.

During a test, the testing system presents a series of discrete prompts to the candidate at a conversational pace as recorded by several different native speakers who represent a range of native accents and speaking styles. These integrated "listen-then-speak" items require real-time receptive and productive processing of spoken language forms. The items are designed to be relatively independent of social nuance and higher cognitive functions. The same facility in spoken English that enables a person to participate in everyday native-paced English conversation also enables that person to satisfactorily understand and respond to the listening/speaking tasks in the Versant English Test.

The Versant English Test measures the candidate's control of core language processing components, such as lexical access and syntactic encoding. For example, in normal everyday conversation, native

speakers go from building a clause structure to phonetic encoding (the last two stages in the right-hand column of Figure 1) in about 40 milliseconds (Van Turennout, Hagoort, & Brown, 1998). Similarly, the other stages shown in Figure 1 must be performed within the short period of time available to a speaker during a conversational turn in everyday communication. The typical time window in turn taking is about 500-1000 milliseconds (Bull & Aylett, 1998). If language users involved in communication cannot successfully perform the complete series of mental activities presented in Figure 1 in real-time, both as listeners and as speakers, they will not be able to participate actively in conversations and other types of communication.

Automaticity in language processing is required in order for the speaker/listener to be able to pay attention to what needs to be said/understood rather than to how the encoded message is to be structured/analyzed. Automaticity in language processing is the ability to access and retrieve lexical items, to build phrases and clause structures, and to articulate responses without conscious attention to the linguistic code (Cutler, 2003; Jescheniak, Hahne, & Schriefers, 2003; Levelt, 2001). Some measures of automaticity in the Versant English Test may be misconstrued as memory tests. Because some tasks involve repeating long sentences or holding phrases in memory in order to piece them together into reasonable sentences, it may seem that these tasks are measuring memory capacity rather than language ability. However, psycholinguistic research has shown that verbal working memory for such things as remembering a string of digits is distinct from the cognitive resources used to process and comprehend sentences (Caplan & Waters, 1999).

The fact that syntactic processing resources are generally separate from short-term memory stores is also evident in the empirical results of the Versant English Test validation experiments (see Section 5: Validation). Virtually all native English speakers achieve high scores on the Versant English Test, whereas non-native speakers obtain scores distributed across the scale. If memory, as such, were being measured as an important component of performance on the Versant English Test, then native speakers would show greater variation in scores as a function of their range of memory capacities. The Versant English test would not correlate as highly as it does with other accepted measures of oral proficiency, since it would be measuring something other than language ability.

The Versant English Test probes the psycholinguistic elements of spoken language performance rather than the social, rhetorical, and cognitive elements of communication. The reason for this focus is to ensure that test performance relates most closely to the candidate's facility with the language itself and is not confounded with other factors. The goal is to separate familiarity with spoken language from other types of knowledge including cultural familiarity, understanding of social relations and behavior, and the candidate's own cognitive style. Also, by focusing on context-independent material, less time is spent developing a background cognitive schema for the tasks, and more time is spent collecting data for language assessment (Downey et al., 2008).

The Versant English Test measures the real-time encoding and decoding of spoken English. Performance on Versant English Test items predicts a more general spoken language facility, which is essential in successful oral communication. The reason for the predictive relation between spoken language facility and oral communication skills is schematized in Figure 2. This figure puts Figure 1 into a larger context, as one might find in a social-situated dialog.

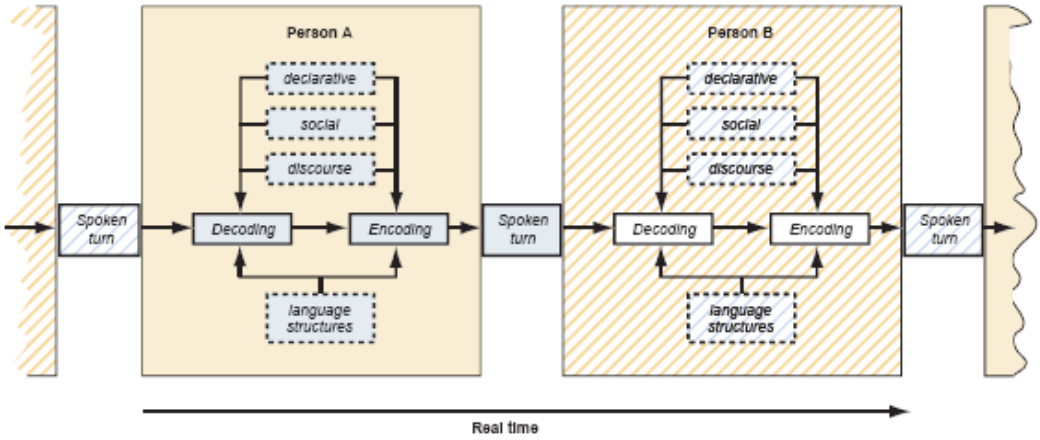

Figure 2. Message decoding and message encoding as a real-time chain-process in oral interaction.

The language structures that are largely shared among the members of a speech community are used to encode and decode various threads of meaning that are communicated in spoken turns. These threads of meaning that are encoded and decoded include declarative information, as well as social information and discourse markers. World knowledge and knowledge of social relations and behavior are also used in understanding and in formulating the content of the spoken turns. However, these social-cognitive elements of communication are not represented in this model and are not directly measured in the Versant English Test.

3. Content Design and Development

The Versant English Test measures both listening and speaking skills, emphasizing the candidate's facility (ease, fluency, immediacy) in responding aloud to common, everyday spoken English. All Versant English Test items are designed to be region neutral. The content specification also requires that both native speakers and proficient non-native speakers find the items very easy to understand and to respond to appropriately. For English learners, the items cover a broad range of skill levels and skill profiles.

Except for the Reading items, each Versant English Test item is independent of the other items and presents unpredictable spoken material in English. The test is designed to use context-independent material for three reasons. First, context-independent items exercise and measure the most basic meanings of words, phrases, and clauses on which context-dependent meanings are based (Perry, 2001). Second, when language usage is relatively context-independent, task performance depends less on factors such as world knowledge and cognitive style and more on the candidate's facility with the language itself. Thus, the test performance on the Versant English Test relates most closely to language abilities and is not confounded with other candidate characteristics. Third, context-independent tasks maximize response density; that is, within the time allotted, the candidate has more time to demonstrate performance in speaking the language. Less time is spent developing a background cognitive schema needed for successful task performance. Item types maximize reliability by providing multiple, fully independent measures. They elicit responses that can be analyzed automatically to produce measures that underlie facility with spoken English, including phonological fluency, sentence comprehension, vocabulary, and pronunciation of lexical and phrasal units.

3.1 Vocabulary Selection

The vocabulary used in all test items and responses is restricted to forms of the 8,000 most frequently

used words in the Switchboard Corpus (Godfrey & Holliman, 1997), a corpus of three million words spoken in spontaneous telephone conversations by over 500 speakers of both sexes from every major dialect of American English. In general, the language structures used in the test reflect those that are common in everyday English. This includes extensive use of pronominal expressions such as "she" or "their friend" and contracted forms such as "won't" and "I'm."

3.2 Item Development

Versant English Test items were drafted by native English-speaking item developers from different regions in the U.S. In general, the language structures used in the test reflect those that are common in everyday conversational English. The items were designed to be independent of social nuance and complex cognitive functions. Lexical and stylistic patterns found in the Switchboard Corpus guided item development.

Draft items were then reviewed internally by a team of test developers, all with advanced degrees in language-related fields, to ensure that they conformed to item specifications and English usage in different English-speaking regions and contained appropriate content. Then, draft items were sent to external linguists for expert review to ensure 1) compliance with the vocabulary specification, and 2) conformity with current colloquial English usage in different countries. Reviewers checked that items would be appropriate for candidates trained to standards other than American English.

All items, including anticipated responses for short-answer questions, were checked for compliance with the vocabulary specification. Most vocabulary items that were not present in the lexicon were changed to other lexical stems that were in the consolidated word list. Some off-list words were kept and added to a supplementary vocabulary list, as deemed necessary and appropriate. Changes proposed by the different reviewers were then reconciled and the original items were edited accordingly.

For an item to be retained in the test, it had to be understood and responded to appropriately by at least 90% of a reference sample of educated native speakers of English.

3.3 Item Prompt Recording

3.3.1 Voice Distribution

Twenty-six native speakers (13 men and 13 women) representing various speaking styles and regions were selected for recording the spoken prompt materials. The 26 speakers recorded items across different tasks fairly evenly.

Recordings were made in a professional recording studio in Menlo Park, California. In addition to the item prompt recordings, all the test instructions were recorded by a professional voice talent whose voice is distinct from the item voices.

3.3.2 Recording Review

Multiple independent reviews were performed on all the recordings for quality, clarity, and conformity to natural conversational styles. Any recording in which reviewers noted some type of error was either re-recorded or excluded from insertion in the operational test.

4. Score Reporting

4.1 Scores and Weights

The Versant English Test score report is comprised of an Overall score and four diagnostic subscores (Sentence Mastery, Vocabulary, Fluency[1] and Pronunciation).

Overall**:** The Overall score of the test represents the ability to understand spoken English and speak it intelligibly at a native-like conversational pace on everyday topics. Scores are based on a weighted combination of the four diagnostic subscores. Scores are reported in the range from 20 to 80.

Sentence Mastery**:** Sentence Mastery reflects the ability to understand, recall, and produce English phrases and clauses in complete sentences. Performance depends on accurate syntactic processing and appropriate usage of words, phrases, and clauses in meaningful sentence structures.

Vocabulary**:** Vocabulary reflects the ability to understand common everyday words spoken in sentence context and to produce such words as needed. Performance depends on familiarity with the form and meaning of everyday words and their use in connected speech.

Fluency**:** Fluency is measured from the rhythm, phrasing and timing evident in constructing, reading and repeating sentences.

Pronunciation**:** Pronunciation reflects the ability to produce consonants, vowels, and stress in a native-like manner in sentence context. Performance depends on knowledge of the phonological structure of everyday words as they occur in phrasal context.

Of the 63 items in an administration of the Versant English Test, 57 responses are currently used in the automatic scoring. The first item response in Parts A through D is considered a practice item and is not incorporated into the final score. The two Open Questions are not scored automatically. Figure 3 illustrates which sections of the test contribute to each of the four subscores. Each vertical rectangle represents a response from a candidate. The items that are not included in the automatic scoring are shown in purple.

[1] Within the context of language acquisition, the term "fluency" is sometimes used in the broader sense of general language mastery. In the narrower sense used in the Versant English Test score reporting, "fluency" is taken as a component of oral proficiency that describes certain characteristics of the observable performance. Following this usage, Lennon (1990) identified fluency as "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (p. 391). In Lennon's view, surface fluency is an indication of a fluent process of encoding. The Versant English Test fluency subscore is based on measurements of surface features such as the response latency, speaking rate, and continuity in speech flow, but as a constituent of the Overall score it is also an indication of the ease of the underlying encoding process.

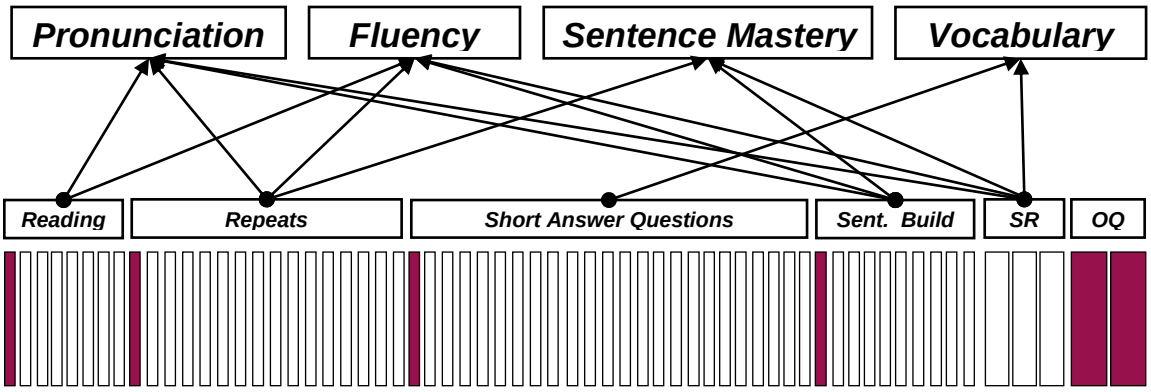

Figure 3. Relation of subscores to item types.

Among the four subscores, two basic types of scores are distinguished: scores relating to the content of what a candidate says (Sentence Mastery and Vocabulary) and scores relating to the manner (quality) of the response production (Fluency and Pronunciation). This distinction corresponds roughly to Carroll's (1961) distinction between a knowledge aspect and a control aspect of language performance. In later publications, Carroll (1986) identified the control aspect as automatization, which suggests that people speaking fluently without realizing they are using their knowledge about a language have attained the level of automatic processing as described by Schneider & Shiffrin (1977).

In all but the Open Questions section of the Versant English Test, each incoming response is recognized automatically by a speech recognizer that has been optimized for non-native speech. The words, pauses, syllables, phones, and even some subphonemic events are located in the recorded signal. The content of the responses to Reading, Repeats, SAQs, and Sentence Builds is scored according to the presence or absence of expected correct words in correct sequences. The content of responses to Story Retelling items is scored for vocabulary by scaling the weighted sum of the occurrence of a large set of expected words and word sequences that are recognized in the spoken response. Weights are assigned to the expected words and word sequences according to their semantic relation to the story prompt using a variation of latent semantic analysis (Landauer et al., 1998). Across all the items, content accuracy counts for 50% of the Overall score, and reflects whether or not the candidate understood the prompts and responded with appropriate content.

The manner-of-speaking scores (Fluency and Pronunciation, or the control dimension) are calculated by measuring the latency of the response, the rate of speaking, the position and length of pauses, the stress and segmental forms of the words, and the pronunciation of the segments in the words within their lexical and phrasal context. These measures are scaled according to the native and non-native distributions and then re-scaled and combined so that they optimally predict human judgments on manner-of-speaking. The manner-of-speaking scores count for the remaining 50% of the Overall score, and reflect whether or not the candidate speaks in a native-like manner.

In the Versant English Test scoring logic, content and manner (i.e. accuracy and control) are weighted equally because successful communication depends on both. Producing accurate lexical and structural content is important, but excessive attention to accuracy can lead to disfluent speech production and can also hinder oral communication; on the other hand, inappropriate word usage and misunderstood syntactic structures can also hinder communication.

4.2 Score Use

Once a candidate has completed a test, the Versant testing system analyzes the spoken performances and posts the scores at www.VersantTest.com. Test administrators and score users can then view and print out the test results from a password-protected section of the website.

Scores from the Versant English Test have been used by educational and government institutions as well as commercial and business organizations. Pearson endorses the use of Versant English Test scores for making valid decisions about oral English interaction skills of individuals, provided score users have reliable evidence confirming the identity of the individuals at the time of test administration. Score users may obtain such evidence either by administering *the* Versant English Test themselves or by having trusted third parties administer the test. In several countries, education and commercial institutions provide such services.

Versant English Test scores can be used to evaluate the level of spoken English skills of individuals entering into, progressing through, and exiting English language courses. Scores may also be used effectively in evaluating whether an individual's level of spoken English is sufficient to perform certain tasks or functions requiring mastery of spoken English.

The Versant English Test score scale covers a wide range of abilities in spoken English communication. In most cases, score users must decide what Versant English Test score is considered a minimum requirement in their context (i.e., a cut score). Score users may wish to base their selection of an appropriate cut score on their own localized research. Pearson can provide a Benchmarking Kit and further assistance in establishing cut scores.

4.3 Score Interpretation

Two summary tables offer a quick reference for interpreting Versant English Test scores in terms of the Common European Framework of Reference descriptors. Table 1 in the Appendix presents an overview relating the Common European Framework global scale (Council of Europe, 2001:24) to Versant English Test Overall scores. Table 2 in the Appendix provides the more specific scale of Oral Interaction Descriptors used in the studies designed to align the two scales. The method used to create the reference tables is described in the Can-Do Guide. Please contact Pearson for this report.

5. Validation

The scoring models used in the first version of the Versant English Test were validated in a series of studies to over 4,000 native and non-native English speakers. In the initial validation study, the native group comprised 376 literate adults, geographically representative of the U.S. population aged 18 to 50. It had a female:male ratio of 60:40, and was 18% African-American. The non-native group was a stratified random sample of 514 candidates sampled from a larger group of more than 3,500 non-native candidates. Stratification was aimed at obtaining an even representation for gender and for native language. Over 40 different languages were represented in the non-native norming group, including Arabic, Chinese, Spanish, Japanese, French, Korean, Italian, and Thai. Ages ranged from 17 to 79 and the female:male ratio was 50:50. More information about these initial validation studies can be found in Validation Summary for PhonePass SET-10. Please contact Pearson for this report.

The Versant English Test has undergone several modifications. The test has been previously known as *PhonePass*, *SET-10*, and *Versant for English*. Because of the introduction of several modifications, a

number of additional validation studies have been performed. With each modification, the accuracy of the test has improved but the scores are still correlated highly with previous versions. The additional validation studies used a native norming group of 775 native speakers of English from the U.S. and the U.K. and a non-native norming group of 603 speakers from a number of countries in Asia, Europe and South America. The native norming group consisted of approximately 33% of speakers from the U.K. and 66% from the U.S. and had a female:male ratio of 55:45. Ages ranged from 18 to 75. The non-native norming group had a female:male ratio of 62:38. Ages ranged from 12 to 56.

In the most recent version of the Versant English Test, Story Retelling items were introduced. Scores on Story Retelling items contribute to all four subscores. A correlation of 0.99 (n=149) was found between the current version of the Versant English Test and the version on which previous validation studies were conducted. The high correlation suggests that many of the inferences from validation studies conducted with the previous releases also apply to the new version. Some of the data presented in this section were collected in validation studies for the previous versions and are assumed to generalize to the most recent version of the test.

5.1 Validity Study Design

Validity analyses examined three aspects of the Versant English Test scores:

1. Internal quality (reliability and accuracy): whether or not the Versant English Test a) provides consistent scores that accurately reflect the scores that human listeners and raters would assign and b) provides distinct subscores that measure different aspects of the test construct.
2. Relation to known populations: whether or not the Versant English Test scores reflect expected differences and similarities among known populations (e.g., natives vs. English learners).
3. Relation to scores of tests with related constructs: how closely Versant English Test scores predict the reliable information in scores of well-established speaking tests.

5.1.1 Validation Sample

From the large body of spoken performance data collected from native and non-native speakers of English, a total of 149 subjects were set aside for a series of validation analyses. Over 20 different languages were represented in the validation sample. Ages ranged from 20 to 55 and the female:male ratio was 42:58. Care was taken to ensure that the training dataset and validation dataset did not overlap. That is, the spoken performance sample provided by the validation candidates were excluded from the datasets used for training the automatic speech processing models or for training any of the scoring models. A total of seven native speakers were included in the validation dataset, but have been excluded from the validity analyses so as not to inflate the correlations.

5.2 Internal Validity

To understand the consistency and accuracy of the Versant English Overall scores and the distinctness of the subscores, the following indicators were examined: the standard error of measurement of the Versant English Overall score; the reliability of the Versant English Test (split-half and test-retest); the correlations between the Versant English Overall scores and subscores, and between pairs of subscores; comparison of machine-generated Versant English scores with listener-judged scores of the same Versant English tests. These qualities of consistency and accuracy of the test scores are the foundation of any valid test (Bachman & Palmer, 1996).

5.2.1 Standard Error of Measurement

The Standard Error of Measurement (SEM) provides an estimate of the amount of error in an individual's observed test scores and "shows how far it is worth taking the reported score at face value" (Luoma, 2004: 183). The SEM of the Versant English Overall score is 2.8.

5.2.2 Reliability

Split-half Reliability

Score reliabilities were estimated by the split-half method (n=143). Split-half reliability was calculated for the Overall score and all subscores. The split-half reliabilities use the Spearman-Brown Prophecy Formula to correct for underestimation and are similar to the reliabilities calculated for the uncorrected equivalent form dataset. The human scores were calculated from human transcriptions (for the Sentence Mastery and Vocabulary subscores) and human judgments (for the Pronunciation and Fluency subscores). Table 2 presents split-half reliabilities based on the same individual performances scored by careful human rating in one case and by independent automatic machine scoring in the other case. The values in Table 2 suggest that there is sufficient information in a Versant English Test item response set to extract reliable information, and that the effect on reliability of using the Ordinate speech recognition technology, as opposed to careful human rating, is quite small across all score categories. The high reliability score is a good indication that the computerized assessment will be consistent for the same candidate assuming no changes in the candidate's language proficiency level.

Table 2. Split-Half Reliabilities of Versant English Test Machine Scoring versus Human Scoring

Score	*Machine Split-half reliability (n = 143)*	*Human Split-half reliability (n=143)*
Overall	0.97	0.99
Sentence Mastery	0.92	0.95
Vocabulary	0.92	0.93
Fluency	0.97	0.99
Pronunciation	0.97	0.99

Test-Retest Reliability

Score reliabilities were also estimated by the test-retest method (n=140). Three randomly generated test forms were administered in a single session to 140 participants. Tests were administered via telephone and computer. The participants were adult learners of English studying at a community college or university and came from a wide range of native language backgrounds. The mean age was 32 years (sd = 8.75). Test administrations are referred to as Test 1, Test 2, and Test 3. Comparisons between Test 1 and Test 2 represent test-retest reliability in the *absence* of a practice test, while comparisons between Test 2 and Test 3 represent test-retest reliability in the *presence* of a practice test (i.e., Test 1). Comparisons between Test 1 and Test 3 represent "repetition effects" or "practice effects", which is the possibility that test scores naturally improve with increased experience with the task. Test-retest reliability was estimated using Pearson's correlation coefficient applied to overall

Versant English Test scores at the three different administrations. Results of the correlation analyses are summarized in Table 3.

Table 3. Correlations between Versant English Test Overall Scores (n=140)

Condition	Correlation
Without a practice test (Test 1 vs. Test 2)	0.97
With a practice test (Test 2 vs. Test 3)	0.97
Repetition effects (Test 1 vs. Test 3)	0.97

These data show that test-retest reliability is high with or without a practice test. It also suggests that increasing familiarity with the tasks does not result in any consistent change in Versant English overall scores.

To determine whether there were any statistically significant differences between scores on any of the three administrations, a separate single-factor Analysis of Variance (ANOVA) was performed with Administration Order (Test 1, 2, or 3) as a factor. Descriptive results of the scores are summarized in Table 4.

Table 4. Mean Overall Versant English Test Scores and Standard Deviations across Administration Order (n=140)

	Administration Order		
	Test 1	Test 2	Test 3
Mean (sd)	44.46 (15.30)	44.99 (14.25)	44.72 (15.17)

There were no statistically significant differences in administration order. Mean score differences are <1 point between each administration of the test, which is well within the standard error of measurement (2.8 points).

The above data were also used to examine the possible grading differences between two different Versant administration modalities: computer-delivered ("CDT") and telephone. The order of presentation of the CDT versus phone modality of the test was randomized and counterbalanced across participants. Test 1 was treated as a practice test; Tests 2 and 3 were the CDT and telephone versions of the test.

The difference of overall scores was analyzed using a paired, two-tailed t-test ($\alpha = .05$). No significant difference was found between the overall scores of the CDT version ($m = 52.3$, $sd = 13.9$) and the telephone delivered version ($m = 52.7$, $sd = 14.5$) ($t(67) = -0.66$, n.s.). These results strongly suggest that there is no systematic difference between Versant English Test scores from the same candidate when the test is taken via CDT or by telephone.

5.2.3 Dimensionality: Correlation between Subscores

Ideally, each subscore on a test provides unique information about a specific dimension of the candidate's ability. For spoken language tests, the expectation is that there will be a certain level of covariance between subscores given the nature of language learning. When language learning takes place, the candidate's skills tend to improve across multiple dimensions. However, if all the subscores

were to correlate perfectly with one another, then the subscores might not be measuring different aspects of facility with the spoken language.

Table 5 presents the correlations among the Versant English Test subscores and the Overall scores for a semi-randomly selected non-native sample.

Table 5. Correlations among Versant English Test Subscores for a Semi-randomly Selected Non-Native Sample (n=1152)

	Sentence Mastery	Vocabulary	Pronunciation	Fluency	Overall
Sentence Mastery	-	0.72	0.55	0.56	0.85
Vocabulary		-	0.51	0.53	0.78
Pronunciation			-	0.80	0.86
Fluency				-	0.88
Overall					-

As expected, test subscores correlate with each other to some extent by virtue of presumed general covariance within the candidate population between different component elements of spoken language skills. The correlations between the subscores are, however, significantly below unity, which indicates that the different scores measure different aspects of the test construct, using different measurement methods, and different sets of responses. This data set (n=1152) was semi-randomly selected from tests delivered over a six month period. A broad range of native languages is represented. A different pattern may be found when different native languages are sampled.

Figure 4 illustrates the relationship between two relatively independent machine scores (Sentence Mastery and Fluency). These machine scores are calculated from a subset of responses that are mostly overlapping (Repeats, Sentence Builds, and Story Retellings for Sentence Mastery, and Readings, Repeats, Sentence Builds, and Story Retellings for Fluency). Although these measures are derived from a data set that contains mostly the same responses, the subscores clearly extract distinct measures from these responses. For example, many candidates with Fluency scores in the 50-60 range have a Sentence Mastery score in the 30-80 range. For the non-native sample (n=1152) used in Figure 4, the Versant English Test Overall scores have a mean of 59 and a standard deviation of 11.

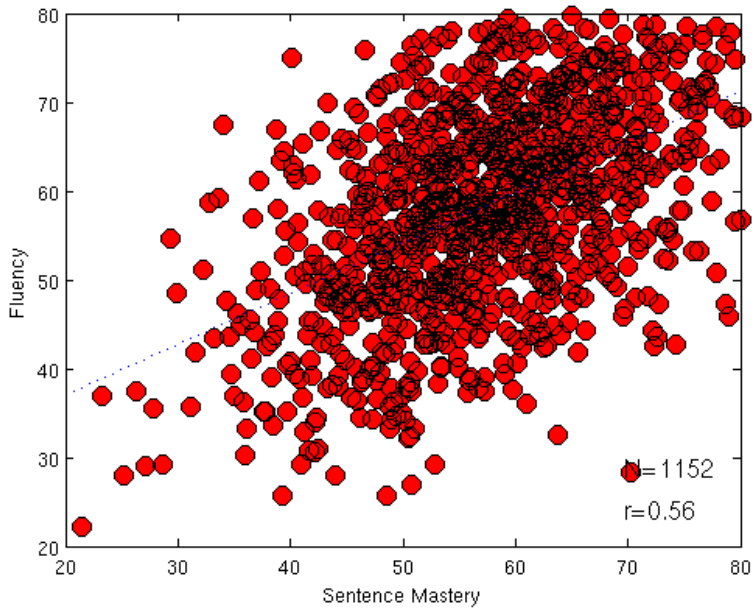

Figure 4. Machine scores of Sentence Mastery versus Fluency for a semi-randomly selected non-native sample (n=1152 and r=0.56).

Factor analysis shows that the Sentence Mastery and Vocabulary subscores correlate well enough to be considered one aspect (content), and that Fluency and Pronunciation correlate well enough to be considered another aspect (manner). Confirmatory factor analysis was conducted on test performances from 102,077 candidates from around the world. A correlated two-factor model, with the Sentence Mastery and Vocabulary scores loaded on one common factor, and Pronunciation and Fluency loaded on the other common factor, was proposed as an *a priori* underlying factorial structure. According to selected model fit indices, this model fits the data fairly well (Chi-Square= 11.377, df =1, p=0.0007; CFI = 1; RMSEA= 0.010, SRMR = 0.001). The correlation between the two factors is 0.716. A Chi-Square difference test was conducted to compare the model fit between the proposed correlated two-factor model and a competing one-factor model. The latter model would suggest that there is only one underlying factor which accounts for test performance. The Chi-Square difference test result indicated that the two-factor model fit the data significantly better than the one-factor model. This empirically-derived correlated two-factor model confirms that the Sentence Mastery and Vocabulary subscores load on one common factor which can be conceptualized as the content aspect, and that the Pronunciation and Fluency subscores load on the other factor which can be conceptualized as the manner aspect.

In order to ensure that the Versant English Test measures the same two aspects (content and manner) of facility in spoken English across candidates with different first languages and from different learning environments, the same analysis was applied to sub-populations of this validation sample (China, Europe, India, Japan, Korea, Latin America, Philippines, and United States). The same correlated two-factor structure proved to be the best model in every sub-population. This lends support to the claim that the Versant English Test measures the same ability across different regions.

5.2.4 Correlations between the Versant English Test and Human Scores

The final analysis for internal quality involved comparing scores from the Versant English Test using Ordinate's speech processing technologies versus careful human transcriptions and human judgments from expert raters. Table 7 presents correlations between machine-generated scores and human scores for the same subset of 143 candidates as given in section 5.2.2. The correlations presented in

Table 7 suggest that the Versant English Test machine-generated scores are not only reliable, but that they generally correspond as they should with human ratings. Among the subscores, the human-machine relation is closer for the content accuracy scores than for the manner-of-speaking scores, but the relation is close for all four subscores.

Table 7. Correlations between the Versant English Test and Human Scores (n=143)

Score Type	Correlation
Overall	**0.97**
Sentence Mastery	**0.97**
Vocabulary	**0.96**
Fluency	**0.94**
Pronunciation	**0.88**

A scatterplot of human and machine scores for this subset is shown in Figure 5.

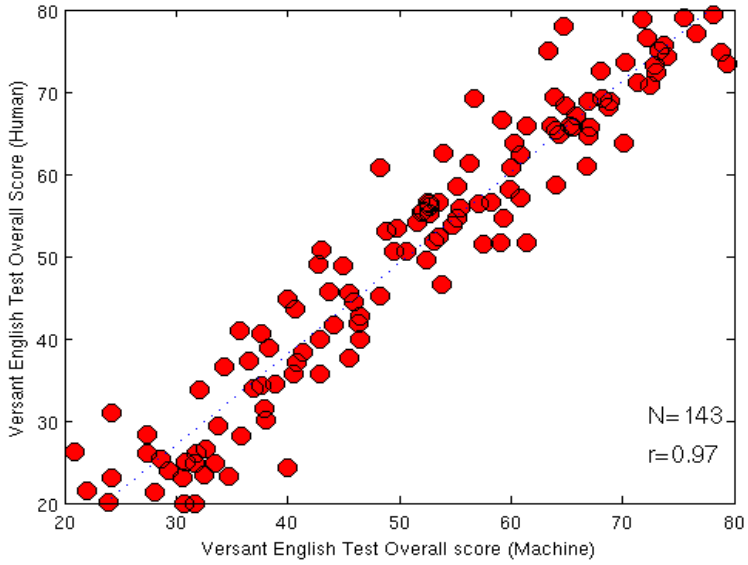

Figure 5. Versant English Test *scores versus human scores (n=143).*

In the scatterplot, all the data points fall within a tight range of the regression line with no outliers. Together the correlations and the scatterplot show that at the Overall score level, Versant English Test machine-generated scores are virtually indistinguishable from scoring based on careful human transcriptions and repeated independent human judgments.

5.3 Relationship to Known Populations: Native and Non-native Group Performance

The next validity analysis examined whether or not the Versant English Test scores reflect expected

differences between native and non-native English speakers. Overall scores from 775 native speakers and 603 non-native speakers representing a range of native languages were compared. Figure 6 presents cumulative distributions of Overall scores for the native and non-native speakers. Note that the range of scores displayed in this figure is from 10 through 90, whereas the Versant English Test scores are reported on a scale from 20 to 80. Scores outside the 20 to 80 range are deemed to have saturated the intended measurement range of the test and are therefore reported as 20 or 80.

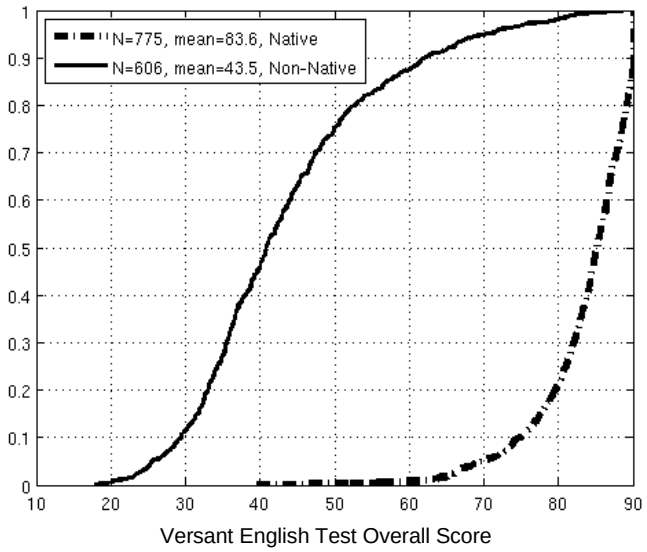

Figure 6. Cumulative density functions of Versant English Test Overall scores for the native and non-native norming groups (native n=775 and non-native n=603).

The results show that native speakers of English consistently obtain high scores on the Versant English Test. Fewer than 5% of the native sample scored below 68. Learners of English as a second or foreign language, on the other hand, are distributed over a wide range of scores. Note also that only 5% of the non-natives scored above 68. The Overall scores show effective separation between native and non-native candidates.

5.4 Relationship to Scores of Tests with Related Constructs

Over the years the Versant Test Development team and third parties have collected data on parallel administrations of *the* Versant English Test and other well-established language examinations, enabling a measure of concurrent validity of the Versant English Test.

Table 8 presents correlations of scores for these instruments with Overall scores on the Versant English Test. The table is divided into three sections: the upper section shows data from overall scores on tests that include multiple language skills (e.g., speaking, listening, writing, and reading). Scores that include skills such as writing and reading are expected to correlate only moderately with the Versant English Test, which specifically targets speaking and listening. The middle section shows tests of listening comprehension, which are expected to have a somewhat higher correlation with the Versant English Test given the oral mode of both tests. The bottom section shows correlations with instruments which focus mainly or entirely on speaking. These instruments are expected to show the highest correlation with the Versant English Test.

Table 8. Correlations of the Versant English Test with Other Measures

	Instrument	r	n
Overall	TOEFL Overall	0.75	392
	TOEFL Overall[1]	0.80	104
	TOEFL iBT Overall[2]	0.64	130
	TOEIC	0.65	494
Listening	TOEIC Listening	0.71	171
	TOEFL Listening[3]	0.79	321
	New TOEFL Listening[3]	0.78	321
Speaking	TSE	0.88	58
	New TOEFL Speaking[3]	0.84	321
	TOEFL iBT Speaking[2]	0.75	130
	Common European Framework, 1st experiment	0.84	121
	Common European Framework, 2nd experiment	0.94	150
	Common European Framework, 3rd experiment	0.88	303
	ILR Speaking[4]	0.75	51
	IELTS Speaking[2]	0.76	130
	BEST Plus, 1st experiment[5]	0.86	151
	BEST Plus, 2nd experiment[5]	0.81	151

Sources: [1]Dodigovic (2009); [2]Farhady & Hedayati (2008); [3]Enright, Bridgeman, & Cline (2002); [4]Bernstein et al. (1999); [5]Van Moere & Present-Thomas (2010); all others Versant Test Development

The data suggest that the Versant English Test measures overlap substantially with instruments designed to assess spoken language skills. For more information about how the Versant English Test relates to other instruments, see Bernstein, Van Moere, & Cheng (2010).

Table 8 includes data from three independent experiments conducted by the Versant Test Development team to relate the Versant English Test reporting scale to an oral interaction scale based on the Common European Framework (Council of Europe, 2001). The first experiment was reported by

Bernstein et al. (2000); the second experiment is reported in Ordinate (2003); and the third experiment was conducted especially for the validation of the current version of the Versant English Test. For the third experiment, responses to Open Questions from a subsample of both norming groups were assigned randomly to six raters who together produced 7,266 independent ratings in an overlapping design. The ratings from the two raters with the largest amount of overlapping data were analyzed. Based on 397 responses, the raters showed perfect agreement in assigning a Common European Framework (CEFR) level to 63% of the cases and differed by only one level in a further 30% of the cases. Rater agreement overall was 0.89.

Figure 7 shows the relation between the Versant English Test score and the CEFR levels from the experiment. The correlation was 0.88. The graph shows how both instruments (Versant English Test and the CEFR) clearly separate the native and non-native norming groups.

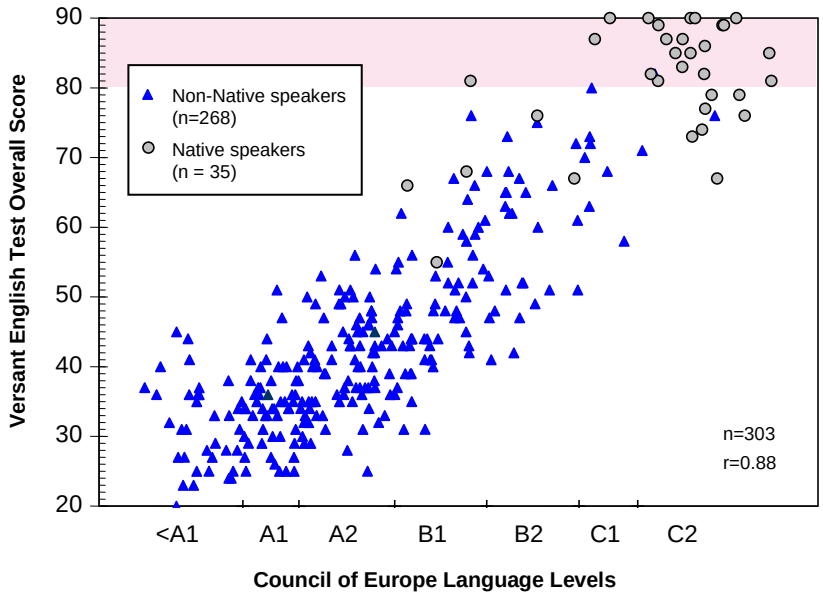

Figure 7: Correlation between Versant English Test Overall score and CEFR-levels (n=303).

6. Conclusions

Data from the validation studies provide evidence in support of the following conclusions:

- The Versant English Test produces precise and reliable skill estimates.
- Overall scores show effective separation between native and non-native candidates.
- Subscores of the Versant English Test are reasonably distinct and therefore offer useful diagnostics.
- Versant English Test scores show a high correlation with human-produced ratings.
- Versant English Test Overall scores have meaningful correlations with other related tests of English proficiency.

To assure the defensibility of employee selection procedures, employers in the U.S. follow the Equal Employment Opportunity Commission's (EEOC's) Uniform Guidelines for Employee Selection Procedures. These guidelines state that employee selection procedures must be reliable and valid. The

above information provides evidence of the reliability, validity and legal defensibility of the Versant English Test in conformance with the prescriptions of the EEOC's Uniform Guidelines.

7. About the Company

Ordinate Testing Technology**:** The Versant automated testing system was developed to apply advanced speech recognition techniques and data collection to the evaluation of language skills. The system includes automatic telephone and computer reply procedures, dedicated speech recognizers, speech analyzers, databanks for digital storage of speech samples, and score report generators linked to the Internet. The *Versant English Test* is the result of years of research in speech recognition, statistical modeling, linguistics, and testing theory. The Versant patented technologies are applied to its own language tests such as the Versant series and also to customized tests. Sample projects include assessment of spoken English, children's reading assessment, adult literacy assessment, and collections and human rating of spoken language samples.

Pearson**:** Pearson's Knowledge Technologies group and Ordinate Corporation, the creator of the Versant tests, were combined in January, 2008. The Versant tests are the first to leverage a completely automated method for assessing spoken language.

Pearson's Policy**:** Pearson is committed to the best practices in the development, use, and administration of language tests. Each Pearson employee strives to achieve the highest standards in test publishing and test practice. As applicable, Pearson follows the guidelines propounded in the Standards for Educational and Psychological Testing, and the Code of Professional Responsibilities in Educational Measurement. A copy of the Standards for Educational and Psychological Testing is available to every employee for reference.

Research at Pearson**:** In close cooperation with international experts, Pearson conducts ongoing research aimed at gathering substantial evidence for the validity, reliability, and practicality of its current products and investigating new applications for Ordinate technology. Research results are published in international journals and made available through the Versant website (www.VersantTest.com).

8. References

Bachman, L.F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A.S. (1996). Language testing in practice. Oxford: Oxford University Press.

Bernstein, J., De Jong, J.H.A.L., Pisoni, D. & Townshend, B. (2000). Two experiments on automatic scoring of spoken language proficiency. In P. Delcloque (Ed.), *Proceedings of InSTIL2000: Integrating Speech Technology in Learning*, University of Abertay Dundee, Scotland, 57-61.

Bernstein, J., Lipson, M., Halleck, G. & Martinez-Scholze, J. (1999). Comparison of Oral Proficiency Interviews and Automatic Testing of Spoken Language Facility. Paper presented at LTRC, Tsukuba, Japan.

Bernstein J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. Language Testing, *27,* 355-377.

Bull, M & Aylett, M. (1998). An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In R.H. Mannell & J. Robert-Ribes (Eds.), Proceedings of the 5th International Conference on Spoken Language Processing. Canberra, Australia: Australian Speech Science and Technology Association.

Caplan, D. & Waters, G. (1999). Verbal working memory and sentence comprehension. Behavioral and Brain Sciences*, 22,* 77-126.

Carroll, J.B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. Testing. Washington, DC: Center for Applied Linguistics.

Carroll, J.B. (1986). Second language. In R.F. Dillon & R.J. Sternberg (Eds.), Cognition and Instructions. Orlando FL: Academic Press.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Cutler, A. (2003). Lexical access. In L. Nadel (Ed.), Encyclopedia of Cognitive Science. Vol. 2, Epilepsy – Mental imagery, philosophical issues about. London: Nature Publishing Group, 858-864.

Dodigovic, M. (2009). Speech Processing Technology in Second Language Testing. In *Proceedings of the Conference on Language & Technology*, Lahore, Pakistan, 113-120.

Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M. & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English test: A response. Language Assessment Quarterly, 5*,* 160-167.

Enright, M.K., Bridgeman, B. & Cline, F. (2002). Prototyping a test design for a new TOEFL. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Equal Employment Opportunity Commission. Uniform Guidelines for Employee Selection Procedures. Retrieved from http://www.uniformguidelines.com/uniformguidelines.html

Farhady, H. & Hedayati, H. (2008). Human operated, machine mediated, and automated tests of spoken English. Paper presented at AAAL, Washington, DC.

Godfrey, J.J. & Holliman, E. (1997). Switchboard-1 Release 2. LDC Catalog No.: LCD97S62. http://www.ldc.upenn.edu.

Jescheniak, J.D., Hahne, A. & Schriefers, H.J. (2003). Information flow in the mental lexicon during speech planning: Evidence from event-related brain potentials. Cognitive Brain Research, 15*(3),* 261-276.

Landauer, T.K., Foltz, P.W. & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. Language Learning, 40*,* 387-412.

Levelt, W.J.M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Levelt, W.J.M. (2001). Spoken word production: A theory of lexical access. PNAS, 98*(23),* 13464-13471.

Luoma, S. (2004). Assessing Speaking. Cambridge: Cambridge University Press.

Miller, G.A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior*, 2,* 217-228.

Ordinate (2000). *Validation summary for PhonePass SET-10: Spoken English Test-10*, system revision 43. Menlo Park, CA: Author.

Ordinate (2003). Ordinate SET-10 Can-Do Guide. Menlo Park, CA: Author.

Perry, J. (2001). Reference and reflexivity. Stanford, CA: CSLI Publications.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review*, 84,* 1-66.

Van Moere, A. & Present-Thomas, R. (2010). Validation of a benchmarking study by means of repeated measures classification consistency. Paper presented at LTRC, Cambridge.

Van Turennout, M., Hagoort, P. & Brown, C. M. (1998). Brain Activity During Speaking: From Syntax to Phonology in 40 Milliseconds. Science, 280, 572-574.

9. Appendix: Test Paper

Side 1 of the Test Paper: Instructions and general introduction to test procedures. **Note:** These instructions are available in several different languages.

VERSANT™

Test Instructions

Please read this before taking the test

Versant tests are automated spoken language tests that are taken on the telephone or computer. If you would like to listen to a sample test, purchase a practice test, or view the test score after taking the test (if applicable), please visit www.VersantTest.com

Part	Instructions
Before the Test	• Carefully read this instruction page and the test paper. You may use a dictionary or ask someone for help if there are words or sentences that you don't understand. • Choose a quiet location with a landline phone where you will not be interrupted during the test. • Do not use a cordless phone, cellular phone, or VoIP phone (e.g., Skype™ or PC-to-phone services). Newer phones are generally better than older phones. Make sure that the phone is set to tone and not pulse.
Beginning the Test	• To begin the test, call the phone number on the test paper using a landline push-button telephone. • A recorded examiner's voice will guide you through each section of the test. • Enter your Test Identification Number using the telephone keypad when the examiner's voice asks you to do so. This number is printed on the top right of your test paper. • The examiner's voice will then ask you two questions: your name, and the city and the country you are calling from. If you are speaking too loudly or too quietly, the examiner's voice will tell you. • The test begins when you say your name. If you hang up before you complete the test, the test cannot be graded. You cannot reuse the Test Identification Number.
During the Test	• Hold the phone close to your mouth as shown in the picture below.  NO Too low, too far away YES In front of mouth YES A good distance • Answer all questions smoothly and naturally in a clear, steady voice. • If you don't know the proper way to respond to a test item, you can remain silent or say, "I don't know." • Do not take notes or write during the test. • When you hear, "Thank you for completing the test", you may hang up. • If you wish, you may answer the optional questions at the end of the test. Your personal information will be kept anonymous.

PEARSON

Side 2 of the Test Paper: Individualized test form (unique for each candidate) showing Test Identification Number, Part A: sentences to read, and examples for all sections.

VERSANT™

VERSANT ENGLISH TEST

REMINDER: The test begins when you say your name. If you hang up before you complete the test, the test cannot be graded. You cannot reuse the Test Identification Number.

Call: 1-415-738-3800

Test Identification Number (TIN)
1234 5678 Expires: January 1, 2012

Thank you for calling the Versant testing system.
Please enter your Test Identification Number on the telephone keypad.
Now, please say your name. Now, please say the city and country you are calling from.
Now, please follow the instructions for Parts A through F.

PART	TASK	TEST DETAILS
A	**Reading**	*Please read the sentences as you are instructed.* 1. Traffic is a huge problem in Southern California. 2. The endless city has no coherent mass transit system. 3. Sharing rides was going to be the solution to rush-hour traffic. 4. Most people still want to drive their own cars, though. 5. Larry's next door neighbors are awful. 6. They play loud music all night when he's trying to sleep. 7. If he tells them to stop, they just turn it up louder. 8. He wants to move out of that neighborhood. 9. My aunt recently rescued a dog that was sick. 10. She brought her home and named her Margaret. 11. They weren't sure she was going to live, but now she's healthy. 12. I just wish she could get along better with their cat.
B	**Repeat**	*Please repeat each sentence that you hear.* Example: a voice says, "Leave town on the next train." and you say, "Leave town on the next train."
C	**Questions**	*Now, please just give a simple answer to the questions.* Example: a voice says, "Would you get water from a bottle or a newspaper?" and you say, "a bottle" or "from a bottle" .
D	**Sentence Builds**	*Now, please rearrange the word groups into a sentence.* Example: a voice says, "was reading" ... "my mother" ... "her favorite magazine" and you say, "My mother was reading her favorite magazine."
E	**Story Retelling**	*You will hear three brief stories. Each story will be spoken once, followed by a beep. When you hear the beep, you will have 30 seconds to retell the story in English. Try to retell as much of the story as you can, including the situation, characters, actions, and ending. You will hear another beep at the end of the 30 seconds.*
F	**Open Questions**	*You will hear two questions about family life or personal choices. Each question will be spoken twice, followed by a beep. When you hear the beep, you will have 40 seconds to answer the question. You will hear another beep at the end of the 40 seconds.*

Thank you for completing the test.

70 - 11111 - 1

PEARSON

Table 1. General Level Descriptors of the Council of Europe Aligned with Versant English Test Scores.

Level		Council of Europe, 2001 Descriptor	Versant English Test Score
Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.	80 79
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibility and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.	78 69
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	68 58
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.	57 47
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.	46 36
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	35 26
			20-25

Table 2. Relation of Versant English Test Overall scores to Oral Interaction Descriptors based on Council of Europe (2001) framework.

Versant English Test		Oral Interaction Descriptors Based on Council of Europe (2001)
80 79	C2	Conveys finer shades of meaning precisely and naturally. Can express him/herself spontaneously at length with a natural colloquial flow. Consistent grammatical and phonological control of a wide range of complex language, including appropriate use of connectors and other cohesive devices.
78 69	C1	Shows fluent, spontaneous expression in clear, well-structured speech. Can express him/herself fluently and spontaneously, almost effortlessly, with a smooth flow of language. Clear, natural pronunciation. Can vary intonation and stress for emphasis. High degree of accuracy; errors are rare. Controlled use of connectors and cohesive devices.
68 58	B2	Relates information and points of view clearly and without noticeable strain. Can produce stretches of language with a fairly even tempo; few noticeably long pauses. Clear pronunciation and intonation. Does not make errors that cause misunderstanding. Clear, coherent, linked discourse, though there may be some "jumpiness."
57 47	B1	Relates comprehensibly main points he/she wants to make on familiar matters. Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair may be very evident. Pronunciation is intelligible even if a foreign access is sometimes evident and occasional mispronunciations occur. Reasonably accurate use of main repertoire associated with more predictable situations. Can link discrete, simple elements into a connected sequence.
46 36	A2	Relates basic information on, e.g., work, background, family, free time, etc. Can make him/herself understood in very short utterances, even though pauses, false starts, and reformulation are very evident. Pronunciation is generally clear enough to be understood despite a noticeable foreign accent. Uses some simple structures correctly, but still systematically makes basic mistakes. Can link groups of words with simple connectors like "and," "but," and "because."
35 26	A1	Makes simple statements on personal details and very familiar topics. Can manage very short, isolated, mainly prepackaged utterances. Much pausing to search for expressions to articulate less familiar words. Pronunciation is very foreign.
25 20	<A1	Candidate performs below level defined as A1.

About Us

The Knowledge Technologies group of Pearson creates unique technology for automated assessment of speech and text used in a variety of industry leading products and services. These include the Versant line of automated spoken language tests built on Ordinate technology, and WriteToLearnTM automated written summary and essay evaluations using the Knowledge Analysis TechnologiesTM (KAT) engine.

The Knowledge Technologies group is part of Pearson, the international media company, whose businesses also include the Financial Times Group and the Penguin Group.

Pearson

299 S. California Avenue
Suite 300
Palo Alto, California 94306
USA

VERSANT

4940 Pearl East Circle
Suite 200
Boulder Colorado 80301
USA

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

Pearson now includes Ordinate products and services.

Version 1211L