



A smarter way to test speaking skills

Versant[™] English Test

Test Design and Validation Research

Table of Contents

1. Introduction	3
2. Test Description	3
Test Format	3
Part A: Reading	3
Part B: Repeat	4
Part C: Short Answer Questions	5
Part D: Sentence Builds	5
Part E: Story Retelling	5
Part F: Open Questions	5
Number of Items	6
Test Administration	7
Telephone Administration	7
Computer Administration	7
Test Construct	7
3. Content Design and Development	9
Vocabulary Selection	10
Item Development	10
4. Scoring	11
Score Use	13
Score Interpretation	13
5. Validation	16
Native and Non-native Group Performance	16
Correlations Among Subscores	17
Scoring Precision and Reliability	18
Correlations Between the Versant English Test and Human Scores	19
Correlations with Other English Language Tests	20
Conclusions	21
6. About the Company	22
7. References	22
8. Appendix: Test Paper and Instructions	24

1. Introduction

The **Versant™ English Test** with Ordinate® technology evaluates the facility in spoken English of people whose native language is not English. Academic institutions, corporations, and government agencies throughout the world use the Versant English Test to evaluate the ability of students, staff or officers to understand spoken English and to express themselves clearly and appropriately in English. The test is intended for use with adults and with students over 15 years of age.

2. Test Description

The Versant English Test is a fifteen-minute spoken English test for adult non-native speakers of English. The test is delivered over the telephone or on a computer and is scored automatically. During the test, the system presents a series of spoken prompts in English at a conversational pace and elicits oral responses in English. The Versant English Test provides numeric scores and performance levels that describe the test taker'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 has six sections: Reading, Repeats, Short Answer Questions, Sentence Builds, Story Retelling, and Open Questions. All items in the first five sections 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 comprehension, vocabulary, and pronunciation of rhythmic and segmental units. More than one task type contributes to each subscore; thus, the use of multiple item types maximizes score reliability.

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

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 test taker's facility in spoken English.

Test Format

Instructions for the test are spoken over the testing system in an examiner voice and are also presented verbatim on a printed test paper during telephone administration and on the computer screen during computer administration. Test items themselves are presented in various native-speaker voices that are distinct from the examiner voice.

The following subsections provide brief descriptions of the task 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, test takers read printed, numbered sentences, one at a time, in the order requested. For telephone administration, the printed text is available from a test paper, which is given to the test taker before beginning the test. For computer administration, the text is displayed on the computer screen. Reading items are grouped into sets of four sequentially coherent sentences, as in the example below.

Examples:

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.

Presenting the sentences in a group helps the test taker disambiguate words in context and helps suggest how each individual sentence should be read aloud. The test paper or computer screen presents three sets of four sentences and asks the test taker to read eight of these sentences in a random order. The system tells the test taker which of the numbered sentences to read aloud. After the system hears the end of one sentence, it prompts the test taker 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 test takers with little facility in spoken English but with some reading skills, this task provides samples of their pronunciation and reading fluency. The readings start the test because, for many test takers, 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, test takers repeat sentences verbatim. The sentences are presented to the test taker in approximate order of increasing difficulty. Sentences range in length from 3 words to 15 words. The audio item prompts are read in a colloquial manner.

Examples:

Biology requires study.
It's supposed to rain tomorrow, isn't it?
There are three basic ways in which a story might be told to someone.

To repeat a sentence longer than about seven syllables, the test taker has to 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. "her 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 test takers to organize speech into linguistic units, it tests their sentence mastery. In addition, the task has them repeat back full sentences (as opposed to just words and phrases), and therefore, it also offers a sample of the test taker's pronunciation and fluency in spoken English.

Part C: Short Answer Questions

In this task, test takers listen to spoken questions in English and answer each of these questions with a single word or short phrase. The questions generally present at least three or four (sometimes more) lexical items spoken in a continuous phonological form and framed in an 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 Anglo-American 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 season comes before spring?
What is frozen water called?
Does a tree usually have fewer trunks or branches?

To respond to the questions, the test taker needs to identify the words in phonological and syntactic context, and then infer the demand proposition. Short Answer Questions manifest a test of receptive and productive vocabulary within the context of spoken questions presented in a conversational style.

Part D: Sentence Builds

For the Sentence Build task, test takers are presented with three short phrases. The phrases are presented in a random order (excluding the original word order), and the test taker is asked to rearrange them into a sentence.

Examples:

in / bed / stay
Ralph / this photograph / could convince
we wondered / would fit in here / whether the new piano

For this task, the test taker has to 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 test taker demonstrates facility in spoken English.

The Sentence Build task involves constructing and saying entire sentences. As such, it is a measure of test takers' mastery of sentences in addition to their pronunciation and fluency.

Part E: Story Retelling

In this task, test takers listen to a story and are then asked to describe what happened in their own words. The test taker is encouraged to tell as much of the story as they can, including the situation, characters, actions and ending. The stories consist of two 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 from the agent of the story followed by a possible reaction or implicit sequence of events. The ending 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 to help it, but the small bird flew away, and the girl ended up with her feet covered with mud.

The Story Retelling items assess the test takers' ability to listen and understand a passage, reformulate the passage using their own vocabulary and grammar, and then retell it in their own words. This elicits longer, more open-ended speech samples than other items in the test, and will allow for the assessment of a wider range of spoken interactions. Performance on Story Retelling will feed into Fluency and Vocabulary scores.

Part F: Open Questions

In this task, test takers listen to a spoken question in English asking for an opinion, and the test taker provides an answer, with an explanation, in English. The questions deal either with family life or with the test taker'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 a spontaneous speech sample. The test taker's responses are not scored automatically at present, but these responses are available for human review by authorized listeners.

Number of Items

In the administration of the Versant English Test, the Ordinate testing system presents a series of discrete items. In total, 63 items are presented to each test taker in six separate sections. The 63 items are drawn at random from a larger item pool. For example, each test taker is presented with ten Sentence Builds from among those items available in the pool, but most items will be different from one test administration to the next. Proprietary algorithms are used by the Ordinate testing system to select from the item pool – the algorithms take into consideration, among other things, the item's level of difficulty and the order of presentation. Table 1 shows the number of items presented in each section.

Table 1. Number of items presented per task.

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

Test Administration

The Versant English Test generally takes 17-18 minutes to complete. Tests can be administered over the telephone or on a computer.

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 introduction to the test procedures (see Appendix). These instructions are the same for all test takers. On the second side is the individual test form, which contains the phone number to call, the Test Identification Number, the spoken instructions written verbatim, item examples, and the printed sentences for Part A: Reading. The individual test form is unique for each test taker.

It is best practice for the administrator to give the test paper to the test taker at least five minutes before starting the Versant English Test. The test taker 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 test taker may have.

When the test taker calls into the Ordinate testing system, the system will ask the test taker to use the telephone keypad to enter the Test Identification Number on the test paper. This identification number keeps the test taker's information secure.

An examiner voice speaks all the instructions for the test. The spoken instructions for each section are also printed verbatim on the test paper to help ensure that test takers understand the directions. Test takers 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.

Computer Administration

For computer administration, the computer must have an Internet connection and the Versant Computer Delivered Test (CDT) software, which is available at www.ordinate.com/idt. During the test administration, the test taker is fitted with a microphone headset. The system allows the test taker 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. Test takers interact with the test system in English, speaking their responses into the microphone. When the test is finished, the test taker clicks a button labeled, "END TEST".

Test Construct

The Versant English Test measures 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 express the construct facility in spoken English is “ease and immediacy in understanding and producing appropriate conversational English.” 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, on occasion, formulate and produce a relevant and intelligible response. These component processes of listening and speaking are schematized in Figure 1, adapted from Levelt (1989).

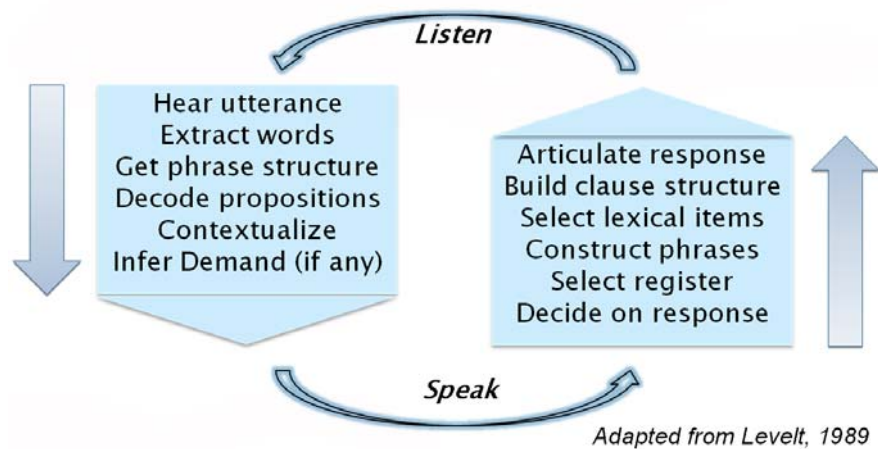


Figure 1. Conversational processing components in listening and speaking.

In the Versant English Test, the Ordinate testing system presents a series of discrete prompts to the test taker at a native conversational pace as recorded by several different native speakers, producing 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, and the items are designed to be relatively independent of social nuance and high-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 test taker’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 Turenhout, Hagoort, and Brown, 1998). Similarly, the other stages shown in Figure 1 have to be performed within the small period of time available to a speaker involved in everyday communication. The typical time window in turn taking is about 500-1000 milliseconds (Bull and Aylett, 1998). If language users involved in communication cannot perform the whole 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 such communication.

In this process, 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, and Schriefers, 2003; Levelt, 2001).

Some measures of automaticity can be misconstrued as memory tests. Since some of the 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 measure memory instead of language ability. However, psycholinguists have shown that short-term or verbal working memory for such things as remembering a string of digits is distinct from 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). Empirical findings show that virtually all

native speakers achieve high scores on the test whereas non-native speakers obtain scores distributed across the scale. If memory, as such, were an important component of performance on the Versant English Test, then the native speakers would show greater variation according to the range of memory spans. Also, the test would not correlate so highly with other accepted measures of oral proficiency as it does, since it would be measuring memory capacity and not 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 test taker's facility with the language itself and is not confounded with other factors. The goal is to tease apart familiarity with spoken language from cultural knowledge, understanding of social relations and behavior, and the test taker'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.

The Versant English Test is a measurement of 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.

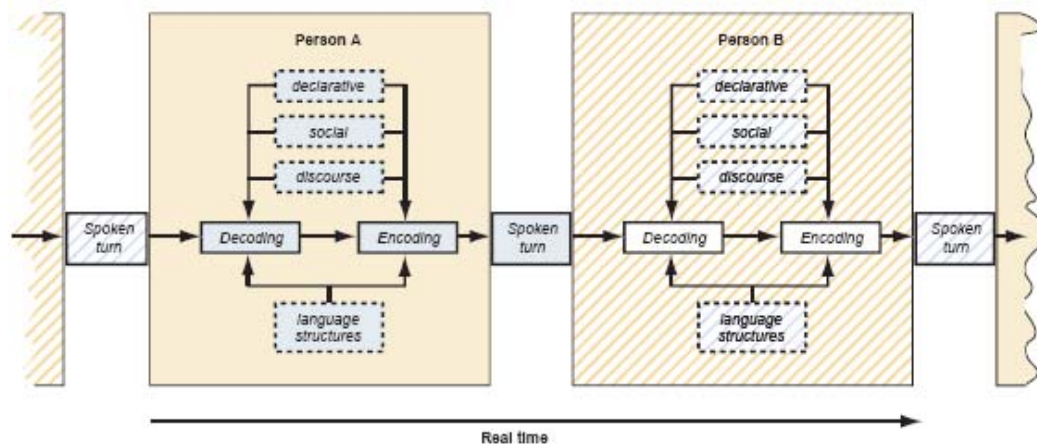


Figure 2. Message decoding and message encoding as a real-time chain-process in oral interaction.

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 the spoken turns and in formulating the content of spoken turns. However, these social-cognitive elements of communication are not represented in this model and are not directly measures in the Versant English Test.

3. Content Design and Development

The Versant English Test measures both listening and speaking skills, emphasizing the test taker's facility (ease, fluency, immediacy) in responding aloud to common, everyday spoken English. All Versant English Test items were 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.

Each Versant English Test item is independent of the other items and presents unpredictable spoken material in English. Context-independent material is used in the test items 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 test taker's facility with the language itself. Thus, the test performance relates most closely to language abilities and is not confounded with other test-taker characteristics. Third, context-independent tasks maximize response density; that is, within the time allotted, the test taker 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 rhythmic and segmental units.

Vocabulary Selection

The vocabulary used in the test items and responses was restricted to forms of the 8,000 most frequent words found in the Switchboard Corpus (Godfrey and Holliman, 1997), a corpus of three million words taken from spontaneous telephone conversations. 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.” The 8,000 most common roots were used to create the base lexicon for the English test item development.

Item Development

Versant English Test items were drafted by item developers in the US. In general, the language structures used in the test reflect those that are common in everyday English. The items were designed to be independent of social nuance and high-cognitive functions. To ensure conversational content, conversations from 540 North Americans guided the design of test items. Lexical and stylistic patterns of these actual conversations were used in developing all item materials. Conversation samples were balanced by geography and gender and represented every major dialect of American English.

Draft items were sent for outside review to ensure that they conformed to current colloquial English usage in different countries and would be appropriate for test takers trained to standards other than U.S. English. British and Australian linguists reviewed all items to ensure conformity to colloquial usage in the United Kingdom and Australia.

All items, including anticipated responses for short-answer questions, were checked for compliance with the vocabulary specification. Vocabulary items that were not present in the lexicon were either changed to other entries that were listed or kept and added to a supplementary vocabulary list. The 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.

4. Scoring

The Versant English Test score report is comprised of an Overall score and four diagnostic subscores (Sentence Mastery, Vocabulary, Fluency¹ 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, 54 responses are currently used in the automatic scoring. The first item response in each part of the test is considered a practice item and is not incorporated into the final score. The three Story Retelling items will be incorporated into the scoring soon. 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 test taker. The items that are not included in the automatic scoring are shown in grey.

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 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.

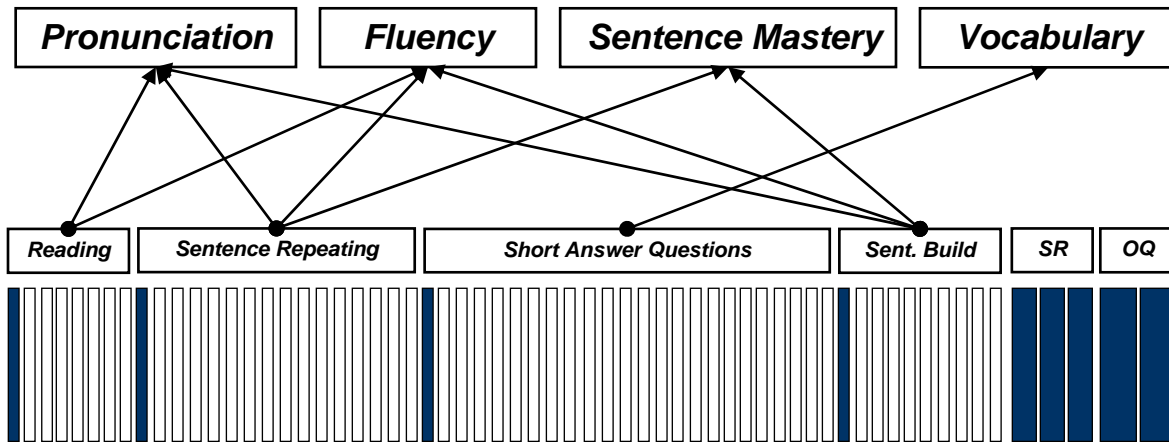


Figure 3. Relation of subscores to item types.

Among the four subscores, two basic types of scores are distinguished: scores relating to the content of what a test taker 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 language performance in relation to a knowledge aspect and a control aspect. 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 each 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, the pauses, the syllables, the phones, and even some subphonemic events are located in the recorded signal. The content of the response is scored according to the presence or absence of expected correct words in correct sequences. The content accuracy dimension counts for 50% of the Overall score, and reflects whether or not the test taker understood the prompt 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 the human judgments on manner-of-speaking (when the process is run on a reference set of non-native speakers). The manner-of-speaking scores count for the remaining 50% of the Overall score, and reflect whether or not the test taker speaks like a native (or like a favorably-judged non-native).

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. In the Versant English Test scoring logic, content and manner (i.e. accuracy and control) are weighted equally because successful communication depends on both.

Score Use

Once a test taker has completed a test, the Ordinate 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.

Score users may be 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. Score users must decide what Versant English Test score can be regarded as a minimum requirement in their context. Score users may wish to base their selection of an appropriate criterion score on their own localized research. Pearson can provide a Benchmarking Kit and further assistance in establishing criterion scores.

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 2 presents an overview relating the Common European Framework global scale (Council of Europe, 2001:24) to Versant English Test Overall scores. Table 3 provides the more specific oral interaction scale of 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.

Table 2. 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

Table 3. 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 accent 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.

5. Validation

Prototype versions of the Versant English Test (previously PhonePass and SET-10) were administered in a series of validation studies to over 4,000 native and non-native speakers. The native norming 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 norming group was a stratified random sample of 514 callers sampled from a larger group of more than 3,500 non-native callers. 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 previous validation studies can be found in *Validation Summary for PhonePass SET-10*, available from the Versant website.

Because of the introduction of several modifications to the test, a number of additional validation studies were performed. These 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% speakers from the U.K. and 66% speakers from the USA 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.

The correlation between the current version of the Versant English Test and the version for which previous validation studies were conducted is 0.98 ($n=200$). This suggests that many of the inferences from validation studies conducted with the previous release remain warranted for the new version.

Native and Non-native Group Performance

Figure 4 presents the main results for the two norming groups. The figure shows the cumulative distribution 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 reported as 20 or 80.

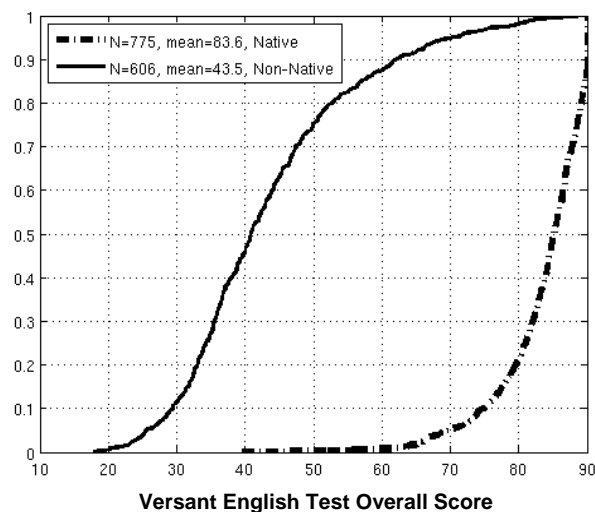


Figure 4. Cumulative density functions of Versant English Test Overall scores for the native and non-native norming groups (native n=775 and non-native n=603).

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. In sum, the Overall scores show effective separation between native and non-native test takers.

Correlations Among Subscores

Table 4 presents the correlations among the Versant English Test subscores and the Overall score for the non-native sample.

Table 4. Correlations among Versant English Test subscores for the non-native sample (n=603).

	Vocabulary	Pronunciation	Fluency	Versant Overall
Sentence Mastery	0.73	0.71	0.67	0.88
Vocabulary		0.65	0.61	0.84
Pronunciation			0.92	0.92
Fluency				0.90

Test subscores correlate with each other to some extent by virtue of presumed general covariance within the test-taker 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.

Figure 5 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 and Sentence Builds for Sentence Mastery and Repeats, Sentence Builds and Readings 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 test takers with Fluency scores in the 50-70 range have a Sentence Mastery score in the 20-40 range.

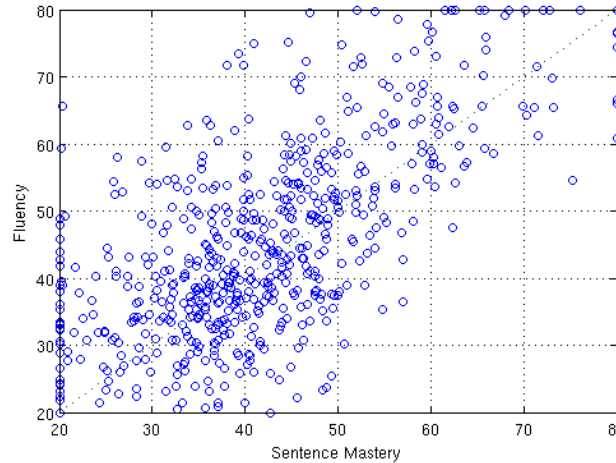


Figure 5. Machine scores of Sentence Mastery versus Fluency for the non-native norming group ($n=603$ and $r=0.67$).

Scoring Precision and Reliability

For the non-native sample ($n=603$), the Versant English Test Overall scores have a mean of 43 and a standard deviation of 13. The standard error of the Overall score is 2.9.

Table 5 displays reliabilities for a subset of 50 calls for which both machine scores and human scores were computed. The human scores were calculated from human transcriptions (for the Sentence Mastery and Vocabulary subscores) and human judgments (for the Pronunciation and Fluency subscores). That is, Table 5 compares the same individual performances, scored by close human rating in one case and by independent automatic machine scoring in the Versant English Test case. The values in Table 5 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 scoring with the Ordinate speech recognition technology, as opposed to a careful human rating, is quite small.

Table 5. Reliability analysis for human scoring (one rater) and Versant English Test machine scoring ($n=50$).

Types of Score	Human Score	Versant English Test Score
Overall	0.98	0.97
Sentence Mastery	0.96	0.93
Vocabulary	0.85	0.88
Fluency	0.98	0.95
Pronunciation	0.98	0.97

Correlations Between the Versant English Test and Human Scores

Table 6 presents correlations between machine-generated scores and human scores for the same subset of 50 test-takers. The correlations presented in Table 6 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. At the Overall score level, Versant English Test machine-generated scores are virtually indistinguishable from scoring that is done by careful human transcriptions and repeated independent human judgments.

Table 6. Correlations between the Versant English Test and human scores (n=50).

Types of Score	Human Score
Overall	0.97
Sentence Mastery	0.93
Vocabulary	0.94
Fluency	0.89
Pronunciation	0.89

The data presented in Figure 6 show human and machine scores for this subset.

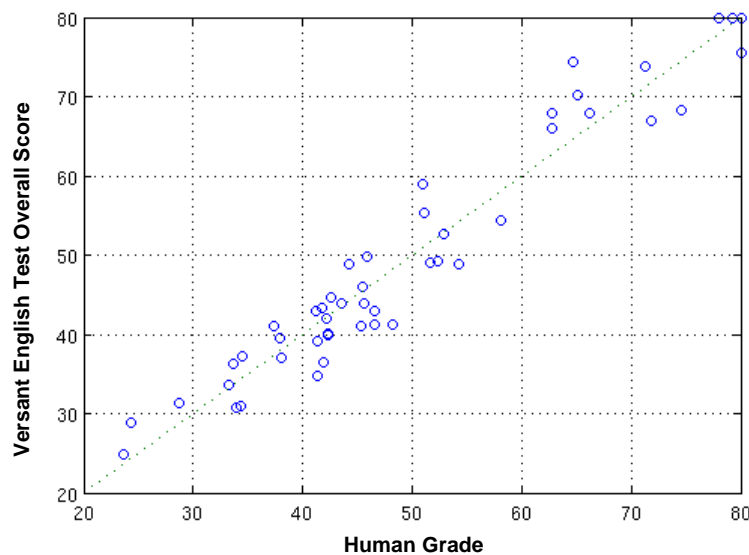


Figure 6. Versant English Test scores versus human scores (n=50).

Correlations with Other English Language Tests

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 7 presents correlations of scores for these instruments with Versant English Test Overall scores. 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), which are expected to have only a moderate correlation with the Versant English Test because it specifically targets speaking and listening. The middle section shows tests of listening comprehension, which, being in the oral mode, are expected to have a somewhat higher correlation with the Versant English Test. The bottom section shows correlations with instruments for assessing oral skills, which focus mainly or entirely on speaking. These instruments are expected to show the highest correlation with the Versant English Test. The data suggest that the Versant English Test measures overlaps substantially with that of other instruments designed to assess spoken language skills.

Table 7. Correlations of the Versant English Test with other measures.

Instrument		r	n
Overall	TOEFL Overall	0.75	392
	TOEFL iBT Overall ²	0.64	130
Listening	TOEIC Listening	0.71	171
	TOEFL Listening ¹	0.79	321
	New TOEFL Listening ¹	0.78	321
Speaking	TSE	0.88	58
	New TOEFL Speaking ¹	0.84	321
	TOEFL iBT Speaking ²	0.75	130
	Common European Framework, 1 st experiment	0.84	121
	Common European Framework, 2 nd experiment	0.94	150
	Common European Framework, 3 rd experiment	0.88	303
	ILR Speaking	0.75	51
IELTS Speaking ²	0.76	130	

Sources: ¹ Enright, Bridgeman, & Cline (2002); ² Farhady & Hedayati (2008); all others Versant Test Development

Table 7 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. 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 related to 397 responses. These raters showed perfect agreement in assigning a Common European Framework (CEF) 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 relationship between the Versant English Test score and the CEF levels that became apparent from this experiment. The correlation was 0.88. The graph also shows how both instruments (Versant English Test and the CEF) clearly separate the native and non-native norming groups.

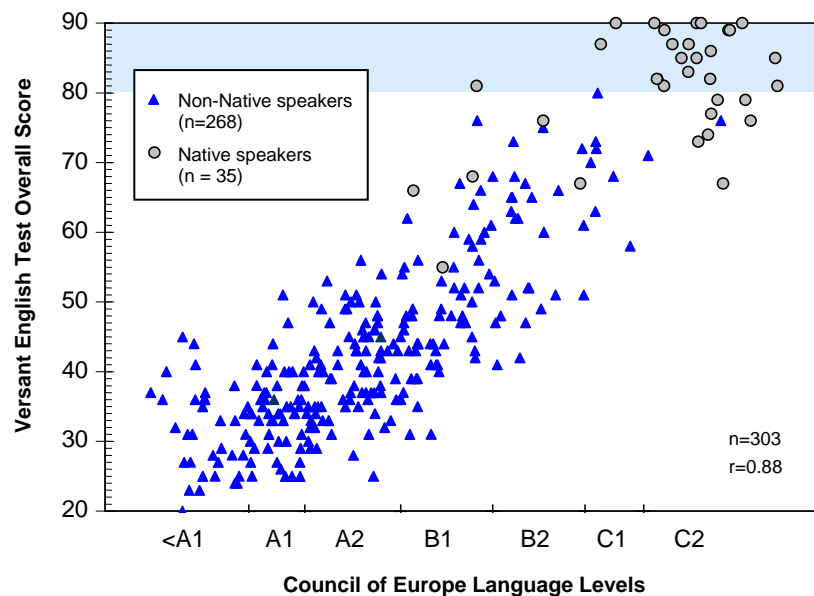


Figure 7: Correlation between Versant English Test Overall score and CEF-levels.

Conclusions

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

- The system produces precise and reliable skill estimates.
- Overall scores show effective separation between native and non-native examinees.
- 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 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. Finally, note that Versant English Test Overall scores have highly meaningful correlations with other measures of English language proficiency.

6. About the Company

Ordinate Testing Technology: The Versant automated testing system was developed to apply advanced speech recognition techniques and data collection via the telephone to the evaluation of language skills. The system includes automatic telephone reply procedures, dedicated speech recognizers, speech analyzers, databanks for digital storage of speech samples, and scoring 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: Ordinate Corporation, creator of the Versant tests, was combined with Pearson's Knowledge Technologies group 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 at investigating new applications for Ordinate technology. Research results are published in international journals and made available through the Versant website (www.VersantTest.com).

7. References

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.


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.
- 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.
- Godfrey, J.J. & Holliman, E. (1997). *Switchboard-1 Release 2*. LDC Catalog No.: LCD97S62. <http://www ldc.upenn.edu>.
- Farhady, H. & Hedayati, H. (2008). Human operated, machine mediated, and automated tests of spoken English. *Paper presented at AAAL*, Washington, DC.
- 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.
- 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.
- 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 Turenout, M., Hagoort, P., & Brown, C. M. (1998). *Brain Activity During Speaking: From Syntax to Phonology in 40 Milliseconds*. *Science*, 280, 572-574.

8. 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.




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
<p>Before the Test</p>	<ul style="list-style-type: none"> 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.
<p>Beginning the Test</p>	<ul style="list-style-type: none"> 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. <u>If you hang up before you complete the test, the test cannot be graded. You cannot reuse the Test Identification Number.</u>
<p>During the Test</p>	<ul style="list-style-type: none"> Hold the phone close to your mouth as shown in the picture below. <div style="text-align: center; margin: 10px 0;">  </div> <ul style="list-style-type: none"> 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.



© 2008 Pearson Education, Inc. or its affiliate(s). All rights reserved. Ordinate and Versant are trademarks, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s).

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



VERSANT ENGLISH TEST (DEMO)

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.

Test Identification Number (TIN)

1111 1111

Expires: August 11, 2009

Call: +1-415-738-3800

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	<p>Please read the sentences as you are instructed.</p> <ol style="list-style-type: none"> Traffic is a huge problem in Southern California. The endless city has no coherent mass transit system. Sharing rides was going to be the solution to rush-hour traffic. Most people still want to drive their own cars, though. Larry's next door neighbors are awful. They play loud music all night when he's trying to sleep. If he tells them to stop, they just turn it up louder. He wants to move out of that neighborhood. My aunt recently rescued a dog that was sick. She brought her home and named her Margaret. They weren't sure she was going to live, but now she's healthy. I just wish she could get along better with their cat.
B	Repeat	<p>Please repeat each sentence that you hear.</p> <p>Example: a voice says, "Leave town on the next train." and you say, "Leave town on the next train."</p>
C	Questions	<p>Now, please just give a simple answer to the questions.</p> <p>Example: a voice says, "Would you get water from a bottle or a newspaper?" and you say, "a bottle" or "from a bottle".</p>
D	Sentence Builds	<p>Now, please rearrange the word groups into a sentence.</p> <p>Example: a voice says, "was reading" ... "my mother" ... "her favorite magazine" and you say, "My mother was reading her favorite magazine."</p>
E	Story Retelling	<p>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.</p>
F	Open Questions	<p>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.</p>

Thank you for completing the test.

53 - 32054 - 1



© 2008 Pearson Education, Inc. or its affiliate(s). All rights reserved. Ordinate and Versant are trademarks, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s).

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 WriteToLearn™ automated written summary and essay evaluations using the Knowledge Analysis Technologies™ (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.



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

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

4940 Pearl East Circle
Suite 200
Boulder, Colorado 80301
USA

PEARSON

((())) VERSANT™

Pearson now includes Ordinate products and services.

© 2008 Pearson Education, Inc. or its affiliate(s). All rights reserved. Ordinate and Versant are trademarks, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s). Other names may be the trademarks of their respective owners.