



Versant™ by Pearson English Speaking and Listening Test

Test Description and Validation Summary

Table of Contents

1. Introduction.....	3
2. Test Description	3
2.1 Test Design.....	3
2.2 Test Administration.....	4
2.2.1 Mobile App Administration	4
2.2.2 Computer Administration.....	5
2.3 Test Format	5
Record a speech sample	5
Part A: Give a short answer to the question.....	6
Part B: Repeat the sentence	6
Part C: Answer a question about a conversation	7
Part D: Answer questions about a passage.....	7
Part E: Retell a passage.....	8
Part F: Give your opinion.....	8
2.4 Number of Items	9
2.5 Test Construct.....	9
3. Content Design and Development.....	12
3.1 Vocabulary Selection.....	12
3.2 Item Development	13
3.3 Item Recording	13
4. Score Reporting.....	13
4.1 Scores and Weights.....	13
4.2 Scaling onto the Global Scale of English (GSE).....	15
4.3 Score Use.....	16
4.4 Score Interpretation.....	16
5. Validation.....	17
5.1 Validation Sample	17
5.2 Internal Validity.....	17
5.2.1 Standard Error of Measurement.....	17
5.2.2 Reliability.....	18
5.2.4 Correlations between the <i>Versant by Pearson English Speaking and Listening Test</i> and Human Scores.....	19
5.3 Relationship to Known Populations: First-Language English Speakers and English Language Learner Group Performance.....	20
5.4 Relationship of Scores to the Common European Framework of Reference.....	21
6. Conclusions.....	22
7. About the Company.....	23

8. References	24
9. Appendix	27

1. Introduction

The Versant™ by Pearson English Speaking and Listening Test is an assessment instrument designed to measure how well a person understands and speaks English. The *Versant by Pearson English Speaking and Listening Test* is intended for adults and students over the age of 15 and takes approximately 20 minutes to complete. Because the *Versant by Pearson English Speaking and Listening Test* is delivered automatically by the Versant by Pearson testing system, the test can be taken at any time, from any location using the phone app or a 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 by Pearson English Speaking and Listening 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 smooth conversational pace in English. Academic institutions, corporations, and government agencies throughout the world use the Versant by Pearson assessment portfolio 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 by Pearson English Speaking and Listening Test* provide reliable information that can be applied to placement and qualification decisions, as well as to monitor progress and measure instructional outcomes.

Scores are returned on the Global Scale of English (GSE) – Pearson’s comprehensive framework of capabilities and learning objectives related to proficiency in English – as well as the Common European Framework of Reference for Languages (CEFR). The GSE is a detailed, numerical scale running from 10 at the lowest level to 90 at the highest and is fully aligned both with the CEFR and with Pearson’s range of language learning resources. GSE scores offer more precision than traditional CEFR bands, meaning that employers can pinpoint a suitable cut score for hiring and more easily track incremental improvement over time.

2. Test Description

2.1 Test Design

The *Versant by Pearson English Speaking and Listening Test* may be taken at any time from any location using a mobile phone app or computer. During test administration, the Versant by Pearson 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 speakers of English from the United States, United Kingdom, and Australia, providing a range of speaking styles.

The *Versant by Pearson English Speaking and Listening Test* has six item types: 'Repeat the sentence', 'Give a short answer to the question', 'Answer a question about a conversation', 'Answer questions about a passage', 'Retell a passage', and 'Give your opinion'. All item types elicit responses that are analyzed automatically. These item types provide multiple, fully independent measures that underlie facility with spoken English, including fluency, sentence construction and comprehension, vocabulary use, listening skills, intelligibility, and pronunciation of English sounds. Because more than one item type contributes to each sub-score, the use of multiple item types strengthens score reliability.

The Versant by Pearson 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 by Pearson English Speaking and Listening 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 smooth conversational pace in intelligible English. The *Versant by Pearson English Speaking and Listening Test* score report is comprised of an Overall score and diagnostic sub-scores for Speaking and Listening. Additional insight into Speaking performance is provided via a Manner of Speaking sub-score. Together, these scores describe the candidate's facility with spoken English.

2.2 Test Administration

Administration of the *Versant by Pearson English Speaking and Listening Test* generally takes about 20 minutes via a mobile app or computer. The delivery of the recorded test questions is interactive – the system detects when the candidate has finished responding to one item and then presents the next item.

2.2.1 Mobile app administration

Test administration on a mobile phone requires the Versant by Pearson Test app and an internet connection. The testing app can be downloaded at no cost from the iOS App Store or Google Play store. The candidate can use a wired headset, earbuds with microphone, or speakerphone. The testing app prompts the candidate to enter the Test Identification Number they have received from their test administrator. 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 displayed verbatim on the device screen 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 close the testing app.

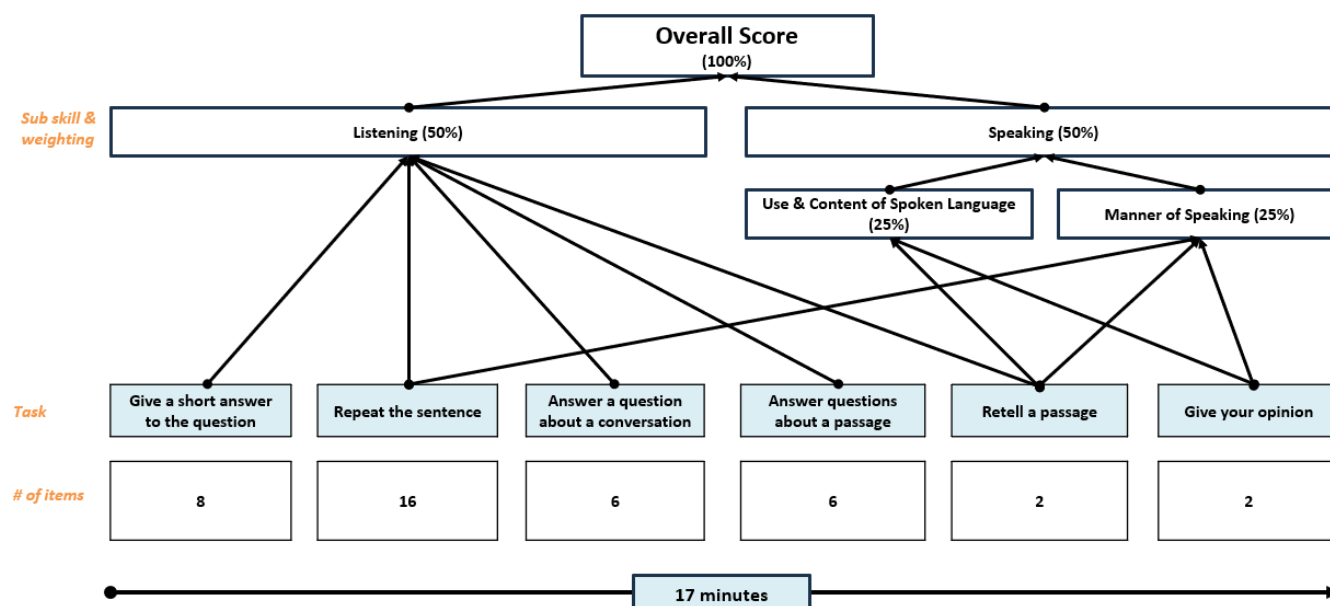
2.2.2 Computer administration

For computer administration, the candidate takes the test via the online testing site. The computer used must have an internet connection. The candidate should be wearing a wired microphone headset. The system allows 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.

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 by Pearson English Speaking and Listening Test*.



Record a speech sample

In this task, candidates listen to a spoken question that asks them to talk about an everyday task or give their opinion on a topic. Candidates have up to 30 seconds to respond to the question.

Examples:

Do you prefer speaking with someone by a voice call or a video call? Explain why.

Do you think it's important to learn English? Why or why not?

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

Part A: Give a short answer to the question

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?

Does a tree usually have more trunks or branches?

To correctly respond to the questions, a candidate must understand the question and have a good idea about what the correct answer would be within a few seconds. Give a short answer to the question measures listening comprehension and receptive and productive vocabulary within the context of spoken questions presented in a conversational style.

Part B: Repeat the sentence

In this task, candidates are asked to listen to sentences, then repeat them back verbatim. The sentences are presented to the candidate in approximate order of increasing difficulty. Sentences range in length from five to 15 words. The sentences are spoken in a conversational manner and do not contain highly specialized vocabulary.

Examples:

You're taller than he is.

He wanted to know if the price included breakfast.

They're trying to fix the problem, but so far, they've had no luck.

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 can remember five-word phrases as a chunk with meaning (e.g., "the really big apple tree"), rather than as five separate words, 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 chunks 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. Chunking is a very natural and automatic process as people are more proficient in a language, so this item type provides good insights into candidates' proficiency in understanding sentence-level English.

Because the Repeat the sentence items require candidates to chunk language, they 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, pronunciation, and intelligibility in continuous spoken English.

Part C: Answer a question about a conversation

In this task, candidates listen to a conversation between two speakers, which typically consists of three speaking turns. Immediately after the conversation, candidates are asked a comprehension question, and they answer the question by saying a word or short phrase.

Example:

Speaker 1: Congratulations on graduating!

Speaker 2: Thanks! It was a lot of work.

Speaker 1: I know. You deserve a party.

Question: Why does that man deserve a party?

This task measures candidates' listening comprehension ability. Answer a question about a conversation items are recorded at a conversational pace covering a range of topics. The task requires candidates to follow speaking turns and extract the topic and content from the interaction at a conversational pace. Quick word recognition and decoding and efficient comprehension of meaning are critical in correctly answering the questions.

Part D: Answer questions about a passage

In this task, candidates listen to a spoken passage (usually a narrative) and then they listen to three comprehension questions about the passage. The passages range from 50 to 150 words in length. Most passages are simple stories with a situation involving a character (or characters), a setting, and an ending. The body of the story typically describes an action performed by the agent of the story followed by a possible reaction or implicit sequence of events. The ending typically introduces a result, new situation, actor, thought, or emotion.

Example:

Jason woke up feeling sick. He called his boss and explained that he could not come in to work. Immediately after making the phone call, he took some medicine. A few hours later, Jason no longer felt sick. Rather than waste the afternoon at home, he decided to go to work after all.

After listening to a passage, the candidate hears and responds to three comprehension questions.

Question 1: What problem did Jason have when he woke up?
 Question 2: What did he do right after calling his boss?
 Question 3: What did Jason do that afternoon?

For each passage, candidates are asked to answer three comprehension questions. Correct answers to the questions (or information needed for simple inferences) are all included in the passage. Questions typically ask for the main idea and details of the passage. This task measures candidates' listening comprehension ability.

Part E: Retell a passage

In this task, candidates listen to a passage, either a short narrative or an expository text, and then are asked to retell the passage in their own words. Candidates have thirty seconds to respond to each passage. Candidates are encouraged to retell as much of the passage as they can. The passages consist of four to ten sentences and contain from 50 to 150 words.

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.

Retell a passage items assess candidates' ability to listen to and understand a passage, then retell it using their own choice of vocabulary and grammar. 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 Retell a passage provides a measure of fluency, pronunciation, intelligibility, vocabulary, and accuracy of retell (based on listening and speaking abilities).

Part F: Give your opinion

In this task, candidates listen to spoken questions that elicit an opinion and are asked to provide an answer with details. Candidates have 40 seconds to respond to each question. The questions ask about the advantages and disadvantages of various things, explanations on how to do various activities, personal preferences on daily topics, or similarities and differences between two items.

Examples:

What are two advantages of communicating by email? What are two disadvantages?
 Explain two things that you can do to help you learn a new language.
 Do you prefer reading a book or watching a movie? Explain why.
 Describe two ways that a doctor and a teacher are similar; and describe two ways that they are different.

This task involves fully and accurately representing an appropriate response in clear and coherent

speech using accurate grammar and vocabulary. As such, it is a measure of test-takers' mastery of language use and content, in addition to their pronunciation, fluency, and intelligibility. Good responses address the topic, include several ideas, and give reasons for the ideas expressed.

2.4 Number of Items

In the administration of the *Versant by Pearson English Speaking and Listening Test*, the testing system presents approximately 37 items in six separate sections to each candidate. The items are drawn at random from a large item pool. This means that most or all items are 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 approximate number of items presented in each section. The exact number of items in each test may change from time to time as new, unscored items are added to and removed from the test. The responses to the unscored items do not impact the candidates' scores nor do they impact the test experience. The responses are used to build scoring models for new items, which allows Pearson to add new content to the test in order to keep the item bank secure and up-to-date.

Table 1. Approximate Number of Items Presented per Section

Task	Approximate Number of Items Presented
A. Give a short answer to the question	8
B. Repeat the sentence	16
C. Answer a question about a conversation	6
D. Answer questions about a passage	6
E. Retell a passage	2-3
F. Give your opinion	2-3

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 by Pearson English Speaking and Listening 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 smooth 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” (Levitt, 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.

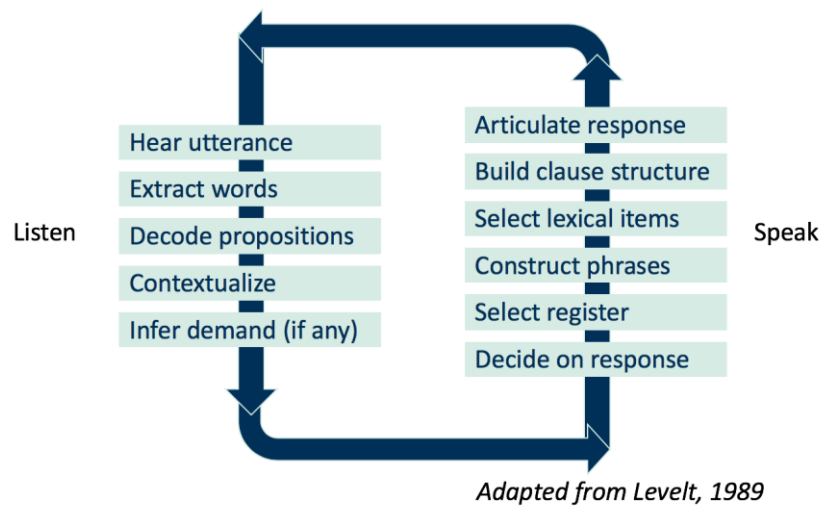


Figure 1. Conversational processing components in listening and speaking.

During a test, the testing system presents a series of discrete prompts to the candidate at a conversational pace as recorded by several different speakers who represent a range of first-language English 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 rapid English conversation also enables that person to satisfactorily understand and respond to the listening/speaking tasks in the *Versant by Pearson English Speaking and Listening Test*.

The *Versant by Pearson English Speaking and Listening Test* measures the candidate’s control of core language processing components, such as lexical access and syntactic encoding. For example, in normal everyday conversation, first-language English 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, & 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, 2002; Jescheniak, Hahne, & Schriefers, 2003; Levelt, 2001). Some measures of automaticity in the *Versant by Pearson English Speaking and Listening 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 by Pearson English Speaking and Listening Test* validation experiments (see Section 5: Validation). Virtually all first-language English speakers achieve high scores on the *Versant by Pearson English Speaking and Listening Test*, whereas English language learners obtain scores distributed across the scale. If memory, as such, were being measured as an important component of performance on the *Versant by Pearson English Speaking and Listening Test*, then first-language English speakers would show greater variation in scores as a function of their range of memory capacities. The *Versant by Pearson English Speaking and Listening 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 by Pearson English Speaking and Listening 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 by Pearson English Speaking and Listening Test* measures the real-time encoding and decoding of spoken English. Performance on *Versant by Pearson English Speaking and Listening 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.

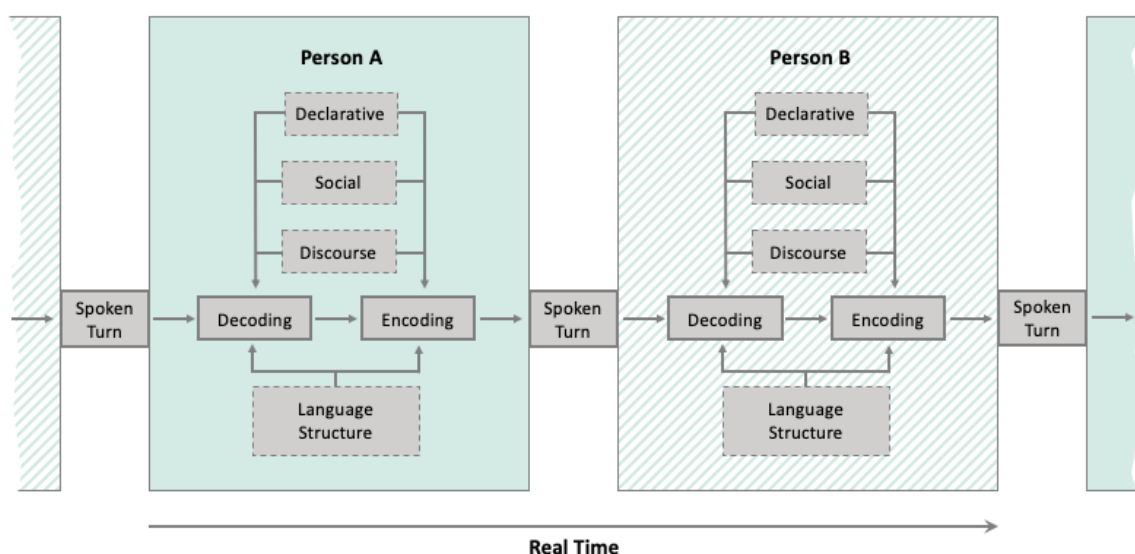


Figure 2. Message decoding and message encoding as a real-time chain-process in oral interaction.

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 by Pearson English Speaking and Listening Test*.

3. Content Design and Development

The *Versant by Pearson English Speaking and Listening Test* measures both listening and speaking skills, emphasizing the candidate's facility (ease, fluency, immediacy) in responding aloud to common, everyday spoken English. The content of all *Versant by Pearson English Speaking and Listening Test* items are designed to be region-neutral, using English that should be understood by all speakers (e.g., "tired" could appear on the test but not "knackered"). The content specification also requires that both first-language English speakers and proficient English language learners 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 by Pearson English Speaking and Listening 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 by Pearson English Speaking and Listening 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, as reflected in the subskill scores on the score report.

3.1 Vocabulary Selection

Most of the vocabulary used in the test items and responses are found among 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 pronouns, possessives, and contractions, such as “she” or “their friend” and contracted forms such as “won’t” and “I’m.”

3.2 Item Development

Versant by Pearson English Speaking and Listening Test items were drafted by advanced English-speaking item developers residing in 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, 2) conformity with current colloquial English usage in different countries, and 3) observance of universal design and free of discrimination, prejudices or biases. Reviewers also checked that items would be appropriate for candidates trained to standards other than American English.

All items, including anticipated responses, 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, when 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 first-language English speakers.

3.3 Item Recording

More than twenty speakers, representing various speaking styles and regions such as the U.S., U.K., and Australia, were selected for recording the spoken prompt materials. Recordings were made in professional recording studios. In addition to the item prompt recordings, all the test instructions were recorded by professional voice talents whose voices were distinct from the item prompt voices.

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 irregularity was either re-recorded or excluded from insertion in the operational test.

4. Score Reporting

4.1 Scores and Weights

The *Versant by Pearson English Speaking and Listening Test* score report is comprised of an Overall score and three diagnostic subscores (Speaking, Listening, and Manner of Speaking). Scores are reported in

the range from 10 to 90 on Pearson's GSE. The corresponding Common European Framework of Reference for Languages (CEFR) level is also displayed.

Overall: The Overall score of the test represents the ability to understand spoken English and speak it intelligibly at a smooth conversational pace on everyday topics. Scores are based on a weighted combination of the three diagnostic subscores.

Speaking: Speaking reflects the ability to produce English phrases and clauses in complete sentences. The score is based on the ability to produce consonants, vowels, and stress clearly and understandably, use accurate syntactic processing and appropriate usage of words in meaningful sentence structures, as well as use appropriate rhythm, phrasing, and timing. It contributes 50% to the Overall score.

Manner of Speaking: This reflects a candidate's pronunciation, fluency, and intelligibility in their speech, including rhythm and phrasing, accuracy of consonants, vowels, and stress, and understandability of speech. It contributes 50% to the Speaking score.

Listening: Listening reflects the ability to understand specific details and main ideas from English speech. The score is based on the ability to track meaning and infer the message from English that is spoken at a conversational pace. It contributes 50% to the Overall score.

Among the three subscores, two basic types of scores are distinguished: scores relating to the content of what a candidate says and scores relating to the manner (quality) of the response production (Fluency, Pronunciation and Intelligibility). 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).

For all item types in the *Versant by Pearson English Speaking and Listening Test*, each incoming response is recognized automatically by a speech recognizer that has been optimized for English language learner accents. The words, pauses, syllables, phones, and even some subphonemic events are located in the recorded signal. The content of the responses to Repeat the sentence, Give a short answer to a question, Answer a question about a conversation, and Answer questions about a passage is scored according to the presence or absence of expected correct words in correct sequences. The content of responses to Retell a passage and Give your opinion items is scored for vocabulary, content accuracy, and grammatical accuracy 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 and language use count for 75% of the Overall score and reflect whether the candidate understood the prompts and responded with appropriate content using accurate English resources.

The manner-of-speaking scores (Fluency, Pronunciation, and Intelligibility) are calculated by measuring aspects of speech such as the latency of the response, the rate of speech, 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. The manner of speaking scores count for the remaining 25% of the Overall score and reflect test-takers' pronunciation accuracy, fluency, and intelligibility.

In the *Versant by Pearson English Speaking and Listening Test* scoring logic, Listening skills and content and accuracy from Speaking contribute to 75% of the Overall score, and Manner of Speaking (i.e. accuracy and control) contributes 25% to the Overall score. 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. So, candidates with a good balance of all these traits are most likely to get high scores.

4.2 Psychometric modeling and scaling onto the GSE

Many of the items on VESLT had been already developed, field-tested, and made machine scorable for the VET or other tests in the Versant family. Therefore, a separate process of field testing the items was not needed, because responses on the items from operational tests going back many years were already available. A very large dataset of the scored responses on nearly one million completed tests was assembled from existing data, to enable items from all Versant tests to be calibrated on the same psychometric model and scale. This large dataset of scored responses was calibrated in a Rasch psychometric model using Winsteps (Linacre, 2023). In the case of the subset of those items that are in VESLT, an average of 1,867 responses per item were included in the calibration run in Winsteps. The calibration converged easily and yielded a robust model. "Items" in this model are scores on traits for particular responses – for example, a Repeat the Sentence response is scored for accuracy, pronunciation, fluency, and intelligibility by the automated scoring models. Each of those four scores is entered into the psychometric model separately. The resulting psychometric model therefore reflects a wide array of English language skills, all placed into the same framework of difficulty.

An IRT scoring model creates scores on particular skills for a candidate, using the empirical difficulty of the items across all candidates from the calibration. In the case of VESLT, an EAP (expected a posteriori) scoring model is used. For example, in calculating the intelligibility score for a candidate, the EAP algorithm is applied to responses across all the items and item types for which intelligibility is calculated. That intelligibility score can then be combined with scores from other traits and item types to create a higher level score (in this case, Manner of Speaking). These scores are initially on a logit scale (ranging from about -3 to +3) which does not have a real-world interpretation.

In order to scale these scores onto the GSE, a standard-setting exercise was undertaken using the body of work approach. Ten applied linguistics experts from a variety of backgrounds were contracted to examine the body of work of 200 candidates. These panelists listened to candidates' complete sets of responses from all items and made judgments about their speaking and listening proficiency levels on GSE ranges (which translate directly to CEFR levels, as described in section 5.4 below). Each candidate was assessed by five expert panelists independently. Another Rasch model was used to aggregate the

scores from all panelists and arrive at the expert- derived “true” proficiency level for each candidate on each skill, adjusting for rater leniency/severity, using Facets software (Linacre, 2022). These proficiency scores were then included as the dependent variable in a linear regression model, with the candidates’ IRT scores on VESLT, weighted as described herein, as the independent variable. The resulting line of best-fit equations for speaking and listening scores yields the transformation from the internal IRT scores to the GSE scores for these skills. The correlations between GSE scores derived in this manner and expert panelist scores were .87 for Speaking and .92 for Listening. In this way, a solid link between machine-based IRT scores and GSE scores was established.

4.3 Score Use

Once a candidate has completed a test, the Versant by Pearson testing system analyzes the spoken performances and posts the scores to the password-protected test administration platform, ScoreKeeper. Test administrators can choose to make scores available to test takers. If this option is selected, test takers may be able to see them on ScoreKeeper or using the score look up function on the Pearson website.

Scores from the *Versant by Pearson English Speaking and Listening Test* have been used by educational and government institutions as well as commercial and business organizations. Pearson endorses the use of *Versant by Pearson English Speaking and Listening 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 by Pearson English Speaking and Listening Test* themselves or by having trusted third parties administer the test. In several countries, education and commercial institutions provide such services.

Versant by Pearson English Speaking and Listening Test scores can be used to evaluate the level of oral 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 GSE assesses candidates of a wide range of abilities in English. In most cases, score users must decide what *Versant by Pearson English Speaking and Listening 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 tools and further assistance in establishing cut scores.

4.4 Score Interpretation

Two summary tables offer a quick reference for interpreting *Versant by Pearson English Speaking and Listening Test* scores in terms of the Common European Framework of Reference (CEFR) descriptors. Table 1 in the Appendix presents an overview relating the Common European Framework global scale (Council of Europe, 2001) to *Versant by Pearson English Speaking and Listening Test* Overall scores as reported on the GSE. Table 2 in the Appendix provides the more specific scale of the CEFR Oral Interaction Descriptors in relation to Versant by Pearson scores.

5. Validation

The scoring models used in the *Versant by Pearson English Speaking and Listening Test* were validated in a study, as described below. Two aspects of the scores were examined:

1. Internal quality (reliability and accuracy): whether or not the *Versant by Pearson English Speaking and Listening 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 the *Versant by Pearson English Speaking and Listening Test* scores reflect expected differences and similarities among known populations (e.g., L1 English speakers vs. English learners).

5.1 Validation Sample

The study included 260 literate adults spread across proficiency levels, geographically representative of the target candidates worldwide. Of the 260 participants, 123 identified as male, 135 identified as female, and 2 as non-binary. Twenty-six of the participants identified themselves as L1 English speakers, coming from the UK, North America, Australia, Southeast Asia, South Asia, and the rest of the world. The rest of the 234 participants comprised the ELL group. Thirty different languages were represented in the ELL group, including (but not limited to) Arabic, Bengali, Bulgarian, Dzongkha, Hindi, Italian, Japanese, Korean, Mandarin Chinese, Nepali, Portuguese, Spanish, Tagalog, Turkish, Vietnamese, and Ukrainian. A larger number of participants, recruited for the purposes of this study, were compensated for their time, and the 260 selected for the validation sample had completed tests with good faith responses. The 260 tests were selected because they matched demographic, test quality, and proficiency requirements.

None of the responses from the 260 participants were used to train either the automated speech recognition models or the scoring models used in the *Versant by Pearson English Speaking and Listening Test*.

5.2 Internal Validity

To understand the consistency and accuracy of the Overall scores and the distinctness of the subscores, the following indicators were examined: the standard error of measurement of the Overall score; the reliability of the *Versant by Pearson English Speaking and Listening Test* (split-half); the correlations between the Overall scores and subscores, and between pairs of subscores; comparison of machine-generated scores with listener-judged scores of the same *Versant by Pearson English Speaking and Listening Tests*. These qualities of consistency and accuracy of the test scores are the foundation of validity for a 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 a candidate'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 Overall score is 4.09 points on the GSE. SEM was calculated using the reliability estimate from the split-half method discussed below.

5.2.2 Reliability

Split-half reliability

Score reliability was estimated by the split-half method ($n=217$). For this analysis, the 260 participants were reduced to 217 so that all tests in the sample had the same number of scores. Split-half reliability was calculated for the Overall score and all subscores. The split-half reliability estimates use the Spearman-Brown Prophecy Formula to correct for underestimation. The human scores were calculated from human transcriptions and human judgments (as opposed to automated scoring). Table 2 presents split-half reliability estimates 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 test scores are highly reliable, and that the effect on reliability of using the Versant by Pearson 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 candidates who take this test receive accurate scores, and that they would likely receive similar scores on subsequent occasions taking the test (as long as their proficiency did not change).

Table 2. Split-half reliability estimates of *Versant by Pearson English Speaking and Listening Test* machine scoring versus human scoring

Score	<i>Machine Split-half reliability (n = 217)</i>	<i>Human Split-half reliability (n=217)</i>
Overall	.96	.96
Speaking	.90	.93
Listening	.92	.93
Manner of Speaking	.95	.96

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 language.

Table 5. Correlations among *Versant by Pearson English Speaking and Listening Test* Subscores for the Validation Sample (n=260)

	Speaking	Listening	Manner of Speaking	Overall
Speaking	-	.85	.92	.96
Listening		-	.82	.97
Manner of Speaking			-	.90
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, below unity, which indicates that the different scores measure different aspects of the test construct, using different measurement methods, and different sets of responses.

5.2.4 Correlations between the *Versant by Pearson English Speaking and Listening Test* and human scores

The final analysis for internal quality involved comparing scores from the *Versant by Pearson English Speaking and Listening Test* using Versant by Pearson's speech processing technologies versus careful human transcriptions and human judgments from expert raters. Table 6 presents correlations between machine-generated scores and human scores for the same subset of 260 candidates as given in section 5.2.3. The correlations presented in Table 6 suggest that the *Versant by Pearson English Speaking and Listening Test* machine-generated scores closely correspond as they should with human-based scores for the same responses. Among the subscores, the human-machine relation is closer for the content accuracy scores than for the manner of speaking scores, but the relationship is high for all four subscores. This can be taken as strong evidence of the validity of the machine scoring method used for producing scores on this test under normal circumstances.

Table 6. Correlations between the *Versant by Pearson English Speaking and Listening Test* and Human Scores (n=260)

Score Type	Correlation
Overall	.96
Speaking	.88
Listening	.98
Manner of Speaking	.85

A scatterplot of the human and machine scores for this subset is shown in Figure 4.

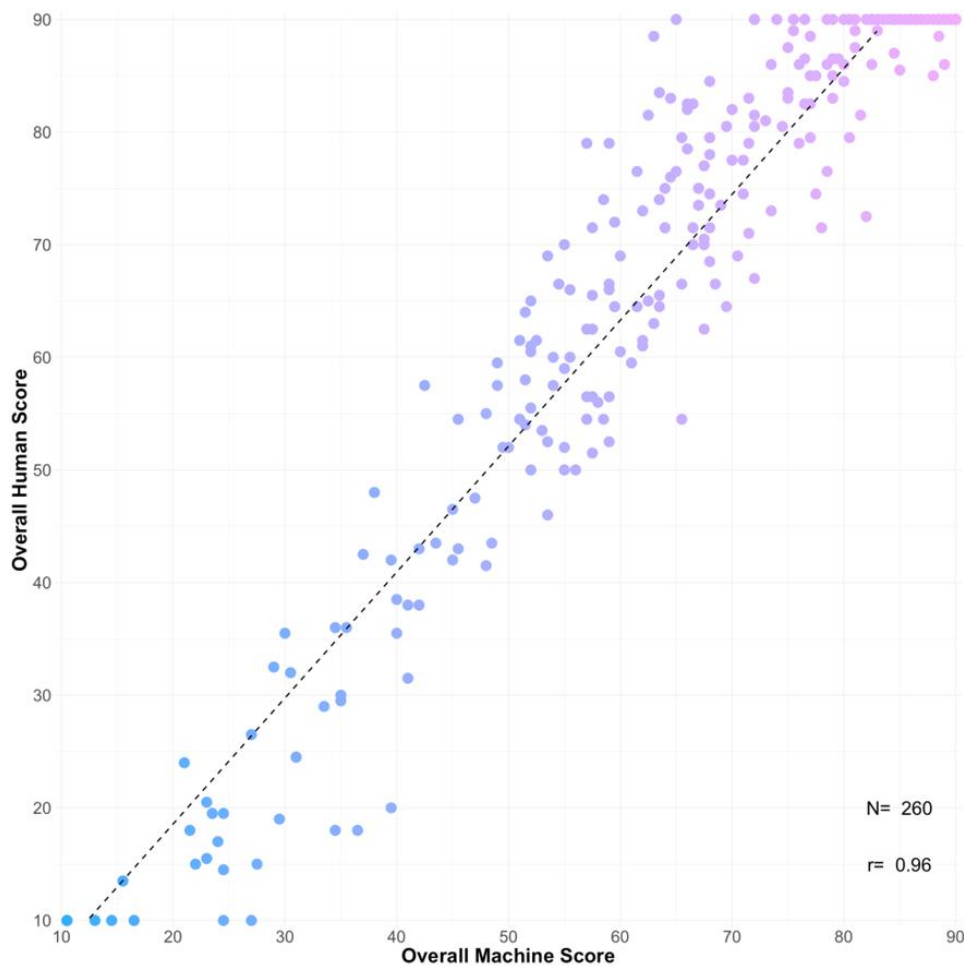


Figure 4. *Versant by Pearson English Speaking and Listening Test* Overall machine scores versus human scores ($n=260$).

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 by Pearson English Speaking and Listening 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: First-Language English Speakers and English Language Learner Group Performance

The next validity analysis examined whether the *Versant by Pearson English Speaking and Listening Test* scores reflect expected differences between first-language English speakers and English language learners (ELL). Overall scores from 26 first language English speakers and 234 English language learners representing a range of native languages were compared. Figure 5 presents cumulative distributions of Overall scores for the first-language and ELL groups.

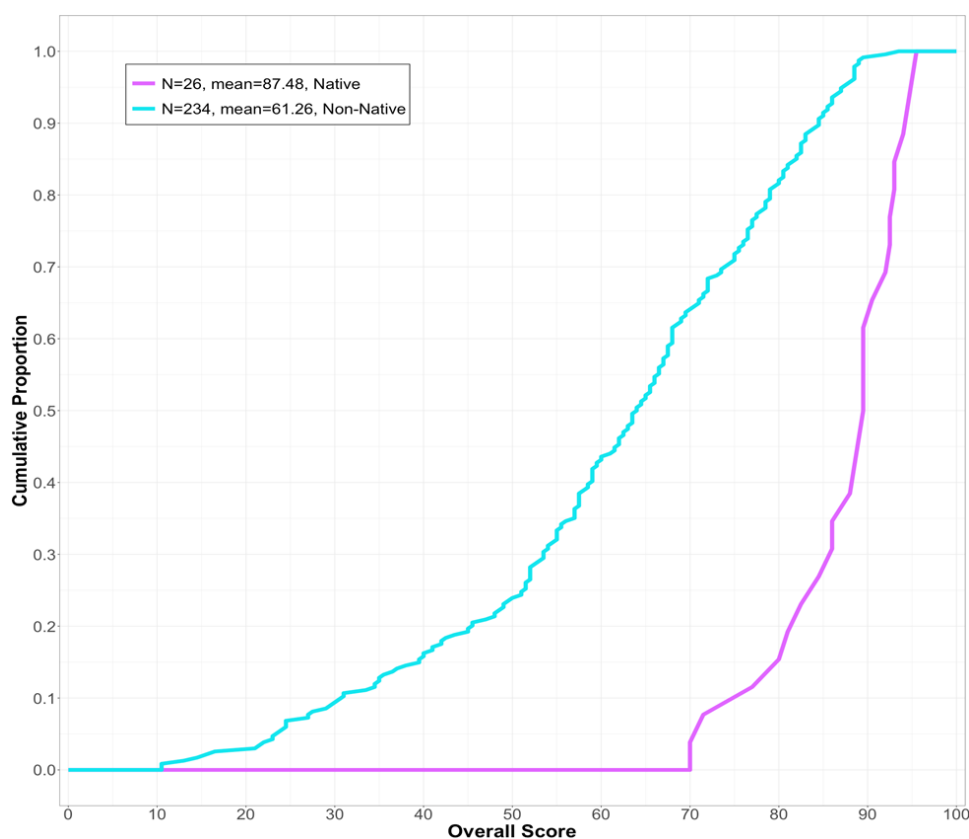


Figure 5. Cumulative density functions of *Versant by Pearson English Speaking and Listening Test* Overall scores for the L1 ($n=26$) and ELL ($n=234$) groups.

The results show that first-language English speakers consistently obtain very high scores on the *Versant by Pearson English Speaking and Listening Test* (averaging 87.48 out of 90). No one in the first language sample scored below 70. Learners of English as a second or foreign language, on the other hand, are distributed over a wide range of scores. Note that only 36% of the ELL group scored above 70. The Overall scores show effective separation between first language and ELL candidates.

5.4 Relationship of Scores to the Common European Framework of Reference

Versant by Pearson English Speaking and Listening Test scores are reported on the GSE which gives an accurate and detailed assessment of a person's English ability on a scale of 10 to 90. Extensive research has been carried out to place the GSE on the CEFR. Two different approaches were used to assign CEFR levels to GSE scores. The first approach was the item-centered approach in which item writers and item reviewers estimated the level of ability that would be required to answer each test item. The second approach was the candidate-centered approach in which expert raters evaluated all of a test-taker's responses to give a judgement about that candidate's Speaking and Listening CEFR level. These responses were scored by Versant by Pearson's scoring engine. A transformation function was derived in order to convert the ability estimates delivered by the test's scoring engine into scores on the GSE. To establish the lower bounds of each CEFR level, the CEFR ratings for all candidates in the sample were

scaled using FACETS. A regression function was calculated for the scaled CEFR ratings obtained through FACETS analysis and item thetas. The correlation between the two measures is 0.69. In the next stage, an equipercentile equating was chosen to express the CEFR lower bounds on the item theta scale. Equipercentile equating determines the equating relationship as one where a score has an equivalent percentile on either form. Table 6 shows the resulting alignment between *Versant by Pearson English Speaking and Listening Test* scores and the CEFR. The CEFR cut-offs from the candidate-centered analysis were then compared to the CEFR cut-offs from the item-centered analysis. Both estimates, derived independently, agree to a high degree ($r=.99$).

Table 6. Concordance table between *Versant by Pearson English Speaking and Listening Test* scores and CEFR levels

CEFR Level	<i>Versant by Pearson English Speaking and Listening Test</i> Score
<A1	10-21
A1	22-29
A2	30-35
A2+	36-42
B1	43-50
B1+	51-58
B2	59-66
B2+	67-75
C1	76-84
C2	85-90

6. Conclusions

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

- The *Versant by Pearson English Speaking and Listening Test* produces precise and reliable skill estimates.
- Overall scores show effective separation between first-language English speakers and English language learners.
- Subskill scores of the *Versant by Pearson English Speaking and Listening Test* are reasonably distinct and therefore offer useful diagnostic information about the different skills.
- *Versant by Pearson English Speaking and Listening Test* machine scores show a very high correlation with human-produced ratings.

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 by Pearson English Speaking and Listening Test* in conformance with the prescriptions of the EEOC's Uniform Guidelines.

7. About the Company

Pearson: Pearson and Ordinate Corporation, the creator of the Versant by Pearson tests, were combined in January, 2008. The Versant by Pearson tests are the first to leverage a completely automated method for assessing spoken and written language.

Versant by Pearson Testing Technology: The Versant by Pearson automated testing system was developed to apply advanced speech recognition techniques and data collection to the evaluation of language skills. The system includes automatic mobile phone 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 by Pearson English Speaking and Listening Test* is the result of years of research in speech recognition, statistical modeling, linguistics, and testing theory. The Versant by Pearson patented technologies are applied to its own language tests such as the Versant by Pearson 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'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 Versant by Pearson technology. Research results are published in international journals and made available through the Versant by Pearson website:

<https://www.pearson.com/languages/hr-professionals/versant.html>

8. References

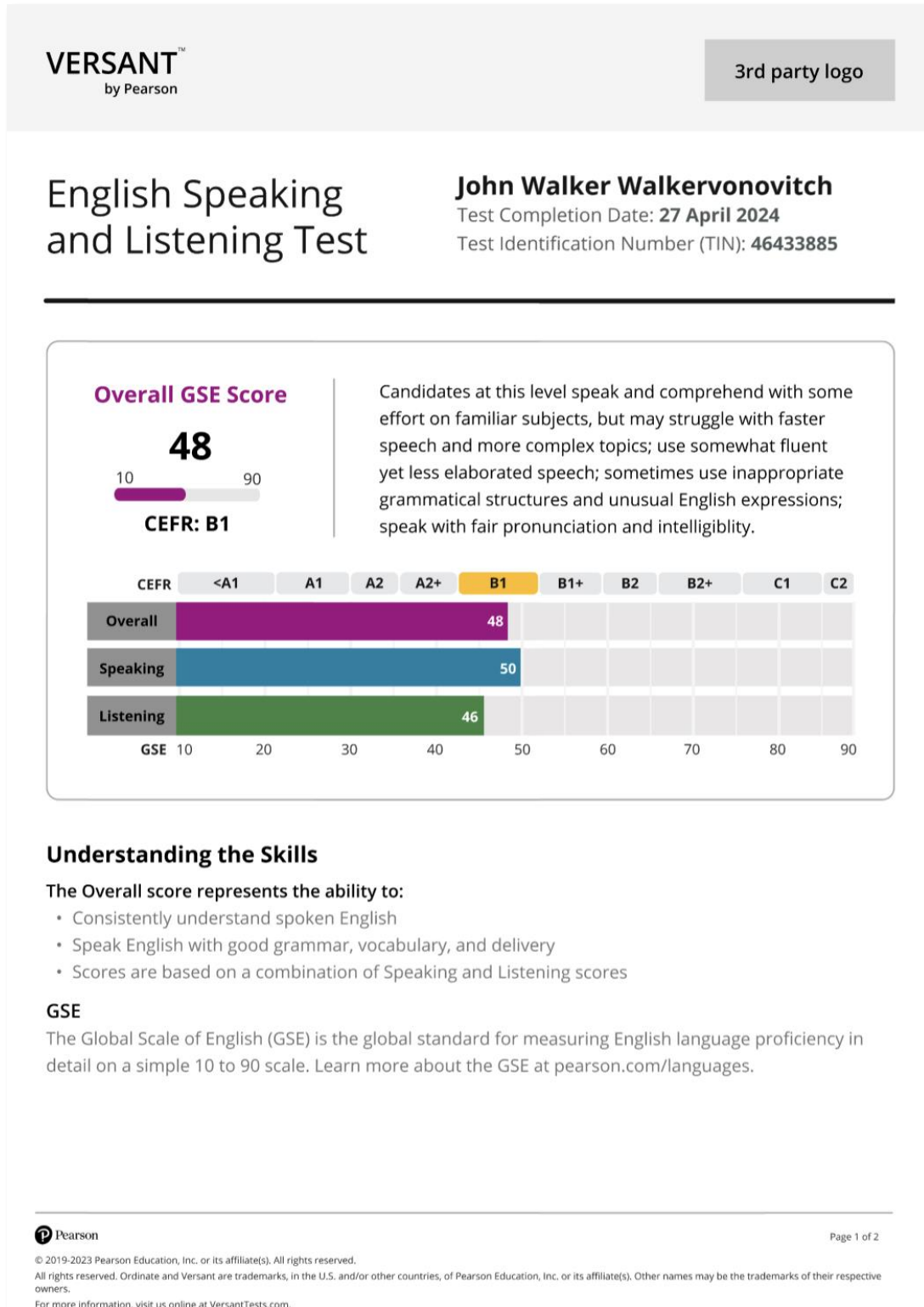
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bernstein, J., De Jong, J.H.A.L., Pisoni, D., & Townshend, B. (2000). Two experiments 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. In A. Campbell (Ed.), *Teaching English as a second language. A book of readings* (pp. 311-321) (2nd ed., 1965). New York: McGraw-Hill.
- Carroll, J. B. (1986). Second language. In R. F. Dillon & R. J. Sternberg (Eds.), *Cognition and instruction* (pp. 83-125). Academic Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cutler, A. (2002). Lexical access. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 858-864). London: Nature Publishing Group.

- 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. Retrieved from http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=EJ876391&ERICExtSearch_SearchType_0=no&accno=EJ876391
- 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.
- Linacre, J. M. (2021) Facets (Version 3.83.6) [Rasch modeling computer software]. Winsteps.com
- Linacre, J. M. (2023) Winsteps (Version 5.6.0) [Rasch modeling computer software]. Winsteps.com
- 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.
- Ordinate (2003). *Ordinate SET-10 Can-Do Guide*. Menlo Park, CA.
- 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), 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 Turenout, M., Hagoort, P., & Brown, C. M. (1998). Brain Activity During Speaking: From Syntax to Phonology in 40 Milliseconds. *Science*, 280, 572-574.

9. Appendix

Score report example



Understanding the Skills

The Overall score represents the ability to:

- Consistently understand spoken English
- Speak English with good grammar, vocabulary, and delivery
- Scores are based on a combination of Speaking and Listening scores

GSE

The Global Scale of English (GSE) is the global standard for measuring English language proficiency in detail on a simple 10 to 90 scale. Learn more about the GSE at pearson.com/languages.

Pearson

© 2019-2023 Pearson Education, Inc. or its affiliate(s). All rights reserved.
 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.
 For more information, visit us online at VersantTests.com.

Page 1 of 2

English Speaking and Listening Test

John Walker Walkervonovitch

Test Completion Date: 27 April 2024

Test Identification Number (TIN): 46433885

Speaking:

GSE: 50/90

CEFR: B1

Candidates at this level can speak on familiar topics with somewhat fluent speech yet lacking specificity and structure; sometimes use inappropriate vocabulary, English expressions, and grammatical structures that cause misunderstanding. Speech is fairly fluent and intelligible, and candidates can interact to some extent in real time.

💡 Tips to improve:

- Describe a problem you have at work or school. Explain why it is a problem and how it could be solved.
- Practice giving detailed instructions on how to do a task, such as how to photocopy some papers.

The Speaking scores reflect your ability to:

- Use English grammar and expressions accurately and appropriately in both constrained and more open-ended responses
- Clearly and accurately produce content and confidently present a view on a topic and give reasons for it
- Speak with English pronunciation norms, conversational fluency, and in a way that proficient English speakers find understandable

Manner of speaking:

GSE: 50/90

CEFR: B1

Candidates at this level can be understood with some effort by listeners; pronounce most vowels and consonants understandably although accent is evident; use reasonable stress and rhythm in general; show frequent hesitations, repetitions, and repairs.

💡 Tips to improve:

- Find a passage you would like to practice. Listen to how a native speaker reads it and try to "shadow" that person by reading it aloud focusing on where to pause.
- Focus on pronouncing commonly used words, vocabulary in your field, and personal names rather than low-frequency words.

Manner of speaking scores reflect your ability to:

- Speak in a way that proficient English speakers find understandable without too much effort
- Pronounce English sounds in ways that are close to typical norms, without too much of an accent from the first language
- Speak at a normal conversational pace, without too much hesitation or unusual pauses between words, phrases, or sentences

Listening:

GSE: 46/90

CEFR: B1

Candidates at this level can follow some short and extended speech that is clearly articulated. Candidates' interactions demonstrate understanding of main points and some details of passages and conversations in everyday workplace situations.

💡 Tips to improve:

- Listen to interviews and identify the speakers' viewpoints and attitudes as well as the information content.
- Watch video clips that explain the steps of a complex process using non-technical language (e.g. how plants grow).

The Listening scores reflect your ability to:

- Understand specific details and main ideas from everyday and workplace speech spoken at a conversational pace

Table 1. General Level Descriptors of the Council of Europe Aligned with Versant by Pearson English Speaking and Listening Test Scores (as reported on the GSE).

Level		Council of Europe, 2001 Descriptor	Versant by Pearson English Speaking and Listening 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.	90 85
	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.	84 76
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.	75 59
	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.	58 43

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.	42
	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.	29
			22
			10-21

Table 2. Relation of *Versant by Pearson English Speaking and Listening Test* Overall scores (as reported on the GSE) to Oral Interaction Descriptors based on Council of Europe (2001) framework.

<i>Versant by Pearson English Speaking and Listening Test</i>	CEFR Level	Oral Interaction Descriptors Based on Council of Europe (2001)
90 85	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.
84 76	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.
75 59	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."
58 43	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.
42 30	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."
29 22	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.
21 10	<A1	Candidate performs below level defined as A1.

About Us

We are Pearson English Language Learning, part of the world's learning company, with expertise in educational courseware and assessment, and a range of teaching and learning services powered by technology.

With 30,000 employees in more than 70 countries, our products are used by millions of professionals, teachers and learners around the world every day. Whether you're a learner seeking swift progress towards new horizons, a teacher who's inspiring achievement in the classroom, an institution looking for measurable improvement, or a professional striving to make data-backed decisions and upskill and reskill their talent for the future, the world of language learning is evolving.

Our mission is to help people make progress in their lives through learning – because we believe that learning opens up opportunities, creating fulfilling careers and better lives.

**To try a sample test or get more information,
visit us online at:**

<https://www.pearson.com/languages/hr-professionals/versant.html>

Version 0822