



Versant[™] Dutch 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	3
2.2.1 Telephone Administration	4
2.2.2 Computer Administration	4
2.3 Test Format	4
Part A: Readings	4
Part B: Repeats	5
Part C: Short Answer Questions	6
Part D: Sentence Builds	6
Part E: Story Retelling	7
Part F: Open Questions	7
2.4 Number of Items	8
2.5 Test Construct	8
3. Content Design and Development	10
3.1 Rationale	10
3.2 Vocabulary Selection	10
3.3 Item Development	11
3.4 Item Prompt Recording	11
3.4.1 Voice Distribution	11
3.4.2 Recording Review	11
4. Scoring Reporting	11
4.1 Scores and Weights	11
4.2 Score Use	13
4.3 Score Interpretation	14
5. Field Test	14
5.1 Data Collection	14
6. Data Resources for Score Development	14
6.1 Data Preparation	14
6.1.1 Transcription	14
6.1.2 Human Rating	15
7. Validation	15
7.1 Validity Study Design	15
7.1.1 Test Reliability	15
8. Conclusion	16
9. About the Company	16
10. References	17
11. Appendix	19

1. Introduction

Pearson's Versant™ Dutch Test, powered by Ordinate technology, is an assessment instrument designed to measure how well a person understands and speaks Dutch. The Versant Dutch Test is intended for adults and students over the age of 16 and takes approximately 15 minutes to complete. Because the Versant Dutch Test test is delivered automatically by the Versant testing system, the test can be taken at any time, from any location by phone or via computer, and a human examiner is not required. The computerized scoring allows for immediate, objective, reliable results that correspond well with traditional measures of spoken Dutch performance.

The Versant Dutch Test measures *facility* with spoken Dutch, which is a key element in Dutch oral proficiency. Facility with Dutch is how well the person can understand spoken Dutch on everyday topics and respond appropriately at a native-like conversational pace in Dutch. Educational, commercial, and other institutions may use Versant Dutch Test scores in decisions where the measurement of listening and speaking is an important element. Versant Dutch Test scores provide reliable information that can be applied in placement, qualification and certification decisions, as well as in progress monitoring or in the measurement of instructional outcomes.

2. Test Description

2.1 Test Design

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

The Versant Dutch Test has six task types: Readings, Repeats, Short Answer Questions, Sentence Builds, Story Retellings, and Open Questions. All items in the first four sections elicit responses from the candidate that are analyzed automatically by the scoring system. These item types provide multiple, fully independent measures that underlie facility with spoken Dutch, including phonological fluency, sentence construction and comprehension, passive and active vocabulary use, listening skill, and pronunciation of rhythmic and segmental units. Because more than one task type contributes to each subscore, the use of multiple item types strengthens score reliability.

The Versant Dutch Test score report is comprised of an Overall score, the corresponding Common European Framework of Reference (CEFR) level, and four diagnostic subscores: Sentence Mastery, Vocabulary, Fluency, and Pronunciation. The Overall score is a weighted average of the four subscores.

The Versant testing system automatically analyzes the candidate's responses and posts scores on its website within minutes of completing the test. Test administrators and score users can view and print out test results from a password-protected section of Pearson's website (www.VersantTest.com).

2.2 Test Administration

Administration of a Versant Dutch Test test generally takes about 15 minutes over the phone or via a computer. Regardless of the mode of test administration, it is best practice (even for computer delivered tests) for the administrator to give a test paper to the candidate at least five minutes before starting the Versant Dutch Test. The candidate then has the opportunity to read both sides of the test

paper and ask questions before the test begins. The administrator should answer any procedural or content questions that the candidate may have.

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

2.2.1 Telephone Administration

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

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

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

2.2.2 Computer Administration

For computer administration, the computer must have an Internet connection and the Versant Computer Delivered Test (CDT) software (available at <http://www.versanttest.com/technology/platforms/cdt/index.jsp>). It is best practice to provide the candidate with a printed test paper to review before the actual computer-based testing begins. The candidate is fitted with a microphone headset. The CDT software prompts the candidate to adjust the volume and calibrate the microphone before the test begins.

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

2.3 Test Format

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

Part A: Readings

In the Reading task, candidates read printed, numbered sentences, one at a time, in the order requested by the examiner voice. The reading texts are printed on a test paper which should be given to the candidate before the start of the test. On the test paper or on the computer screen, reading items are grouped into sets of four sequentially coherent sentences as in the example below.

Examples:

1. Mieke houdt veel van schrijven.
2. Zij heeft een boek geschreven over zijn jeugd.
3. De uitgever was er heel tevreden over.
4. Mieke's boek is nu in vele winkels te koop.

-
1. *Mieke loves to write.*
 2. *She has written a book about her childhood.*
 3. *The publisher was very pleased.*
 4. *Mieke's new book is now for sale in many stores*

Presenting the sentences in a group helps the candidate 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 the examiner voice instructs the candidate which of the numbered sentences to read aloud, one-by-one in a random order (e.g., *Please read Sentence 4. ... Now read Sentence 1. ... etc.*). After the system detects silence indicating the end of one response, it prompts the candidate to read another sentence from the list.

The sentences are relatively simple in structure and vocabulary, so they can be read easily and fluently by people educated in Dutch. For candidates with little facility in spoken Dutch but with some reading skills, this task provides samples of their pronunciation and oral reading fluency. The readings start the test because, for some candidates, reading aloud presents a familiar task and is a comfortable introduction to the interactive mode of the test as a whole.

Part B: Repeats

In the Repeat task, candidates are asked to repeat sentences verbatim. Sentences range in length from three words to fourteen words. The audio item prompts are spoken aloud by native speakers of Dutch and are presented to the candidate in an approximate order of increasing difficulty.

Examples:

- Ik woon in een dorp.
Mijn zus fietst elke dag naar haar werk.
Vandaag ga ik met mijn vriendin naar het museum.

-
- I live in a village.*
My sister bikes to work everyday.
Today I am going to the museum with my friend.

Some people misunderstand this task and claim that sentences can be repeated using mimicry, or parroting. However, there is much research (Vinther, 2002; Eaalm, 2006) which shows that although mimicry works for very short sentences, longer sentences require language processing. To repeat a sentence longer than about seven syllables, the candidate has to understand the words as produced in a continuous stream of speech (Miller & Isard, 1963). Highly proficient speakers of Dutch can generally repeat sentences that contain many more than seven syllables because these speakers are very familiar with Dutch words, collocations, phrase structures, and other common linguistic forms. If a person habitually processes five-word phrases as a unit (e.g. “the really big apple tree”), then that person can

usually repeat verbatim utterances of 15 or even 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 Dutch phrase and sentence structure.

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

Part C: Short Answer Questions

In this task, candidates listen to spoken questions in Dutch and answer each question with a single word or short phrase. The questions generally include at least three or four content words embedded in some particular Dutch interrogative structure. Each question asks for basic information, or requires simple inferences based on time, sequence, number, lexical content, or logic. The questions are designed not to presume any knowledge of specific facts of Dutch culture, geography, religion, 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 Dutch and an adult learner who has never lived in a Dutch-speaking country.

Examples:

Heeft een koe meer of minder poten dan een kip?
Met welk lichaamsdeel schrijf je- met je hand of met je mond?
Mijn vader is 46. Hoe oud was hij vorig jaar?

*Does a cow have more or less legs than a chicken?
Which part of the body do you use to write: your hand or your mouth?
My father is 46 years old. How old was he last year?*

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

Part D: Sentence Builds

For the Sentence Build task, candidates are presented with three short phrases. The phrases are presented in a random order (excluding the original order), and the candidate is asked to rearrange them into a sentence, that is, to speak a reasonable sentence that comprises exactly the three given phrases.

Examples:

uit / ik slaap / vandaag
kopen / een nieuwe auto / hij wil
vandaag afhebben / mijn werk / ik moet

*in / I slept / today
to buy / a new car / he wants
finish / my work today / I have to*

In the Sentence Build task, the candidate has to understand the possible meanings of each phrase and know how the phrases might combine with the other phrasal material, both with regard to syntax and semantics. 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 two- or three-word phrase) that a person can hold in verbal working memory. This is important to measure because it reflects the candidate's ability to access and retrieve lexical items and to build phrases and clause structures automatically. The more automatic these processes are, the more the candidate demonstrates facility in spoken Dutch. This skill is demonstrably distinct from memory span (see Section 2.5, Test Construct, below).

The Sentence Build task involves constructing and saying entire sentences. As such, it is a measure of the candidate's mastery of language structure as well as pronunciation and fluency.

Part E: Story Retelling

For the Story Retelling task, candidates listen to a spoken story and then are asked to describe what happened in their own words. Candidates are encouraged to re-tell as much of the story as they can, including the situation, characters, actions and ending. The stories are from forty to ninety words in length. Most stories are simple with a situation involving a character (or characters), a setting and a goal. 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 sometimes introduces a new situation, actor, thought, decision, or emotion.

Example:

Henk is een buschauffeur in een grote stad. Hij werkt vijf dagen per week. Hij begint zes uur 's morgens en is stopt om twee uur. Dat vindt hij fijn want dan is hij vroeg klaar. Zijn route is vaak heel druk. Veel mensen gaan met de bus naar hun werk. Hij zegt altijd gedag tegen iedereen die instapt. Volgend jaar kan hij met pensioen omdat hij 60 wordt. Hij kijkt al uit naar zijn vrije tijd. Toch zal hij zijn mensen in de bus missen.

Henk is a bus driver in a big city. He works five days a week. He starts at six o'clock in the morning and finishes at two o'clock. He likes this because he is done early. His route is often very busy. Many people go by bus to work. He always says hello to everyone who boards. Next year he can retire because he turns 60. He looks forward to his leisure time, but he will miss the people on the bus.

Story Retellings capture the candidate's listening comprehension ability and also provide additional samples of spontaneous speech. Currently, this task is not automatically scored by the computerized scoring system in the Versant Dutch Test.

Part F: Open Questions

In this task, candidates listen to a spoken question in Dutch asking for an opinion, and the candidate provides answers, with explanations, in Dutch. The questions deal either with family life or with the candidate's preferences and choices.

Examples:

Wat doe je het liefst na een drukke dag? Leg uit.
Heeft de computer een positieve of negatieve invloed op de samenleving? Leg uit.

*What do you like to do after a busy day? Please explain.
Do you think the computer has a positive or negative effect on society? Please explain.*

This task is used to collect a spontaneous speech sample. The candidate's responses are not scored automatically at present, but these responses are available online in the administrator's password-protected website for review by authorized listeners.

2.4 Number of Items

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

Table I. Number of items presented per section.

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

2.5 Test Construct

For any language test, it is essential to define the test construct as explicitly as possible (Bachman, 1990; Bachman & Palmer, 1996). The Versant Dutch Test is designed to measure a candidate's *facility in spoken Dutch* – that is the ability to understand spoken Dutch on everyday topics and to respond appropriately at a native-like conversational pace in intelligible Dutch. Another way to express the construct, *facility in spoken Dutch*, is *ease and immediacy in understanding and producing appropriate conversational Dutch*. There are many processing elements required to participate in a spoken conversation: 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 I, adapted from Levelt (1989).

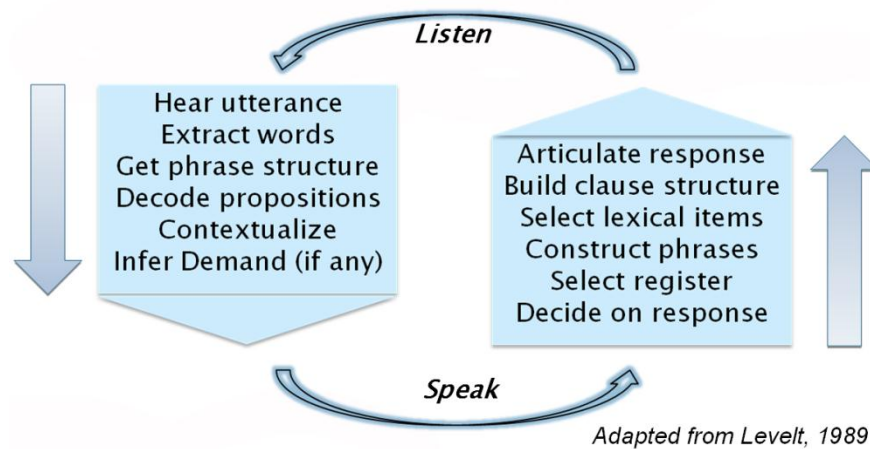


Figure 1. Conversational processing components in listening and speaking.

Core language component processes, such as lexical access and syntactic encoding, typically take place at a very rapid pace. During spoken conversation, Van Turenout, Hagoort, and Brown (1998) found that native speakers go from building a clause structure to phonetic encoding in about 40 milliseconds. Similarly, the other stages shown in Figure 1 have to be performed within the small period of time available to a speaker involved in interactive spoken communication. A typical window in turn taking is about 500-1000 milliseconds (Bull and Aylett, 1998). If language users cannot perform the internal activities presented in Figure 1 in real time, they will not be able to participate as effective listener/speakers. Thus, spoken language facility is essential in successful oral communication.

Because candidates respond to the Versant Dutch Test items in real time, the system can estimate the candidate's level of automaticity with the language with respect to the latency and pace of the spoken response. Automaticity 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). Automaticity is required for the speaker/listener to be able to focus on what needs to be said rather than on how the language code is structured or analyzed. By measuring basic encoding and decoding of oral language as performed in integrated tasks in real time, the Versant Dutch Test probes the degree of automaticity in language performance.

Two basic types of scores are produced from the test: scores relating to the content of what a candidate says and scores relating to the manner of the candidate's speaking. This distinction corresponds roughly to Carroll's (1961) description of a knowledge aspect and a control aspect of language performance. In later publications, Carroll (1986) identified the control aspect as automaticity, which occurs when speakers can talk fluently without realizing they are using their knowledge about a language.

Some measures of automaticity can be misconstrued as memory tests. Since some Versant Dutch Test tasks involve repeating long sentences or holding phrases in memory in order to assemble them into reasonable sentences, it may seem that these tasks measure memory instead of language ability, or at least that performance on some tasks may be unduly influenced by general memory performance. Note that every Repeat and every Sentence Build item on the test was presented to a sample of educated native speakers of Dutch and at least 90% of the speakers in that educated native speaker sample responded correctly. If memory, as such, were an important component of performance on the Versant Dutch Test tasks, then the native Dutch speakers should show greater performance variation on these

items according to the presumed range of individuals' memory spans. Also, if memory capacity (rather than language ability) were a principal component of the variation among people performing these tasks, then Versant tests would not correlate so closely with other accepted measures of oral proficiency (Bernstein, Van Moere and Cheng, 2010).

Note that the Versant Dutch 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 disentangle familiarity with spoken language from cultural knowledge, understanding of social relations and behavior, and the candidate's own cognitive style and strengths. 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 real performance samples for language assessment.

The Versant Dutch Test provides a measurement of the real-time encoding and decoding of spoken Dutch. Performance on Versant Dutch Test items predicts a more general spoken Dutch facility, which is essential for successful oral communication in Dutch. The same facility in spoken Dutch that enables a person to satisfactorily understand and respond to the listening/speaking tasks in the Versant Dutch Test also enables that person to participate in native-paced Dutch conversation.

3. Content Design and Development

3.1 Rationale

All Versant Dutch Test item content is designed to be region-neutral. The content specification also requires that both native speakers and proficient non-native speakers find the items easy to understand and to respond to appropriately. For Dutch learners, the items probe a broad range of skill levels and skill profiles.

Except for the Reading items, each Versant Dutch Test item is independent of the other items and presents unpredictable spoken material in Dutch. 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 candidate's facility with the language itself. Thus, the test performance 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 Dutch, including phonological fluency, sentence comprehension, vocabulary, and pronunciation of lexical and phrasal units.

3.2 Vocabulary Selection

The vocabulary used in the test items and responses was controlled against the most frequent words found in the Corpus of Spoken Dutch (CGN, 2004), a corpus of words taken from spontaneous

conversations, telephone conversations, interviews, and discussions in spoken Dutch. In general, the language structures used in the test reflect those that are common in everyday Dutch.

3.3 Item Development

The Versant Dutch Test item texts were drafted by four native speakers of Dutch; all educated in Dutch through university level. In general, the language structures used in the test were designed to reflect those that are common in Dutch. The items were designed to be independent of social nuance and high-cognitive functions.

Draft items were then reviewed internally by a team of five native Dutch speakers, all with advanced degrees in language-related fields, to ensure that they conformed to Dutch usage in different Dutch-speaking regions. Following this, draft items were sent to a Dutch linguist for a final, external review. All items, including anticipated responses for short-answer questions, were checked for compliance with the vocabulary specification. Most vocabulary items that were not present in the lexicon were changed to other lexical stems that were in the consolidated word list. Some off-list words were kept and added to a supplementary vocabulary list, as deemed necessary and appropriate. The changes proposed by the different reviewers were then reconciled and the original items were edited accordingly.

3.4 Item Prompt Recording

3.4.1 Voice Distribution

Ten native speakers (four men and six women) representing various speaking styles and regions were selected for recording the spoken prompt materials. The ten speakers recorded the items across different item types fairly evenly.

Recordings were made in a professional recording studio in Menlo Park, California. Instruction prompts were recorded by a Dutch and English examiner voice. Thus, there are two versions of the Versant Dutch Test – one with a Dutch examiner voice and another with an English examiner voice.

3.4.2 Recording Review

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

4. Scoring Reporting

4.1 Scores and Weights

Of the 63 items in an administration of the Versant Dutch Test, not all responses are used in the automatic scoring at this time. The first item response of each task type in the test is considered a practice item and is not incorporated into the final score. In addition, the three story retelling responses and two open question responses in Parts E and F are not scored automatically.

The Versant Dutch Test score report is comprised of an Overall score, the corresponding CEFR level, and four diagnostic subscores (Sentence Mastery, Vocabulary, Fluency¹, and Pronunciation).

Overall: The Overall score of the test represents the ability to understand spoken Dutch and speak it intelligibly at a native-like conversational pace on common topics. Overall scores are based on a weighted combination of the four diagnostic subscores (30% Sentence Mastery, 20% Vocabulary, 30% Fluency and 20% Pronunciation). All scores are reported in the range from 10 to 40.

Sentence Mastery: Sentence Mastery reflects the ability to understand, recall, and produce Dutch 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 words spoken in sentence context and to produce such words as needed. Performance depends on familiarity with the form and meaning of common 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 common words.

Figure 2 illustrates which sections of the test contribute to each of the four subscores. Each vertical rectangle represents the response utterance from a candidate. The items that are not included in the automatic scoring are shown in grey. These include the first item in each of the first four sections of the test and all items in the last two sections (Story Retellings and Open Questions).

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 Dutch 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 Dutch 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.

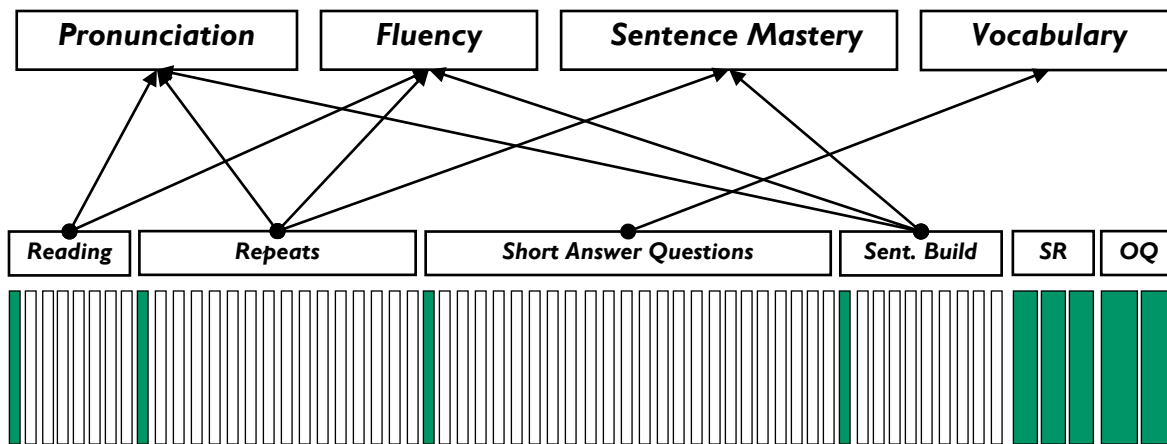


Figure 2. Relation of subscores to item types.

The subscores are based on two different aspects of language performance: a knowledge aspect (the content of what is said), and a control aspect (the manner in which a response is said). The four subscores reflect these aspects of communication where Sentence Mastery and Vocabulary are associated with content and Fluency and Pronunciation are associated with manner of speaking. The content accuracy dimension counts for 50% of the Overall score and indicates whether or not the candidate understood the prompt and responded with appropriate content. The manner-of-speaking scores count for the remaining 50% of the Overall score, and indicate whether or not the candidate speaks like an educated native (or like a very high-proficiency learner). 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 misapplied syntactic structures can also hinder communication. In the Versant Dutch Test scoring logic, content and manner (i.e., accuracy and control) are weighted equally because successful communication depends on both.

The Ordinate automated scoring system scores both the content and manner-of-speaking subscores using a speech recognition system that is optimized based on non-native Dutch spoken response data collected during the field test. The content subscores are derived from the correctness of the candidate's response and the presence or absence of expected words in correct sequences. The manner-of-speaking subscores (Fluency and Pronunciation, as 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. In order to produce valid scores, during the development stage, these measures were automatically generated on a sample set of utterances (from both native and non-native speakers) and then were scaled to match human ratings. Two trained native speakers rated candidates' pronunciation and fluency from the Reading, Repeat and Sentence Build speech files, with reference to a set of rating criteria. These criterion-referenced human scores were then used to rescale the machine scores so that the pronunciation and fluency subscores generated by the machine align with the human ratings.

4.2 Score Use

Once a candidate 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 www.VersantTest.com, as well as listen to selected responses from each candidate.

Score users of the Versant Dutch Test may be educational institutions or commercial and business organizations. Within a pedagogical research setting, Versant Dutch Test scores may be used to evaluate the level of spoken Dutch skills of individuals entering into, progressing through, and leaving Dutch language courses.

4.3 Score Interpretation

Research has been conducted to explore how a Versant Dutch Test overall score relates to other scales that measure or describe language proficiency. Table I (see Appendix) presents an overview relating the Common European Framework global scale (Council of Europe, 2001: 24) to Versant Dutch Test Overall scores. Please contact Pearson for the report that describes the method used to create the reference table.

5. Field Test

5.1 Data Collection

The acoustic and speech recognition models that underlie the Versant Dutch Test are optimized for non-native speech patterns, and are the same as those developed for the Dutch TGN (Toets Gesproken Nederlands), a test used for screening immigration applicants into the Netherlands (De Jong, Lennig, Kerkhoff, and Poelmans, 2009). Therefore, it was not necessary to develop a new automatic speech recognition system. However, because the Versant Dutch Test is different in structure to the TGN and consists of different task types and item content, a separate field test was conducted for the Versant Dutch Test. Both native speakers and non-native speakers of Dutch were recruited as participants from June 2009 through September 2009 to take a prototype data collection version of the Versant Dutch Test. The purposes of this field testing were 1) to validate operation of the test items with both native and non-native speakers, and 2) to calibrate the difficulty of each item based on a large sample of candidates at various levels and from various first language backgrounds.

While Versant Dutch is specifically designed to assess learners of Dutch, responses from native speakers were used to validate the test items, help determine what constituted a “correct” answer, and also evaluate the scoring models.

6. Data Resources for Score Development

6.1 Data Preparation

The development of the Versant Dutch Test involved a total of 18,683 spoken responses collected from natives and learners. All native and non-native responses were transcribed two or more times. Subsets of the response data were also presented to native listeners for quality and/or proficiency judgments so they could be used in score calibration.

6.1.1 Transcription

Both native and non-native responses were transcribed by native speakers of Dutch in order to develop language models for the automated scoring of Versant Dutch Test items. A total of 4,867 transcriptions were produced for native responses and 12,105 transcriptions were produced for non-native responses.

The native speaker transcribers were rigorously trained and the quality of their transcriptions was closely monitored.

6.1.2 Human Rating

Selected item responses from a subset of candidates were presented to seven educated native Dutch speakers to be judged for pronunciation and fluency. Before the native speakers began rating responses, they received training in how to evaluate responses according to analytical and holistic rating criteria. Each rater listened to the sets of item response recordings in a different random order, and independently assigned pronunciation and fluency scores. The raters logged onto an online system and were first presented with a set of responses for the pronunciation rating, and then were presented with a different (but slightly overlapping) set of responses for the fluency rating. Separating the judgment of different traits is intended to minimize the transfer of judgments from pronunciation to fluency, by having the raters focus on only one trait at a time. Rating stopped when each item had been judged by at least two raters.

7. Validation

7.1 Validity Study Design

To understand the consistency and accuracy of Versant Dutch Test Overall scores and the distinctness of the subscores, the following indicators were examined using the existing items: the Standard Error of Measurement of the Versant Dutch Overall score; the reliability of human judgments on test performances on the Versant Dutch Test (split-half reliability); the reliability of machine scores on the same test performances of the Versant Dutch Test (split-half reliability). These qualities of consistency and accuracy of the test scores are the foundation of any valid test (Bachman & Palmer, 1996).

7.1.1 Test Reliability

The Standard Error of Measurement (SEM) provides an estimate of the amount of error in an individual's observed test score and "shows how far it is worth taking the reported score at face value" (Luoma, 2003: 183). The SEM of the Versant Dutch Test Overall score is 3.3.

Score reliabilities were estimated by both the split-half method ($n = 149$). Split-half reliability was calculated for the Overall score and was 0.94. The reliability represents the corrected calculation (using the Spearman-Brown Prophecy Formula) due to split-half underestimation. The high reliability score is a good indication that the computerized assessment will be consistent for the same candidate assuming no changes in the candidate's language proficiency level.

Table 4 displays reliabilities for the Versant Dutch Overall scores and four subscores. The human scores that were calculated from human transcriptions (for the Sentence Mastery and Vocabulary subscores) and human judgments (for the Pronunciation and Fluency subscores). That is, Table 4 compares the scoring reliability of a sample candidates whose performances were scored by close human rating and careful human transcriptions of content in one case, and by independent automatic machine scoring in the Versant Dutch case.

Table 4. Reliability analysis for Versant Dutch human scoring and machine scoring (one rater), n=149.

Types of Score	Human Reliability	Machine Reliability
Overall	0.98	0.94
Sentence Mastery	0.96	0.93
Vocabulary	0.80	0.78
Pronunciation	0.97	0.91
Fluency	0.97	0.88

Establishing test validity is a matter of ongoing empirical research, and should be situated according to context of testing and score use. To date, the consistency of test scores and their relation to independent expert human judgments have been established. However, more validation research is planned. Pearson encourages test users to pilot the Versant Dutch Test in their own setting to establish the usefulness of scores for their purposes.

8. Conclusion

This report has provided details of the test development processes and validity evidence in order for test users to make an informed interpretive judgment as to whether test scores would be valid for their purposes. The test development process is documented and adheres to sound theoretical principles and test development ethics from the field of applied linguistics and language testing, namely: the items were written to specifications and subject to a rigorous procedure of qualitative review and psychometric analysis before being deployed to the item pool; the content was selected from both pedagogic and authentic material; the test has a well-defined construct that is represented in the cognitive demands of the tasks; the scores, item weights and scoring logic are explained; the items were widely field tested on a representative sample of candidates; and further, empirical evidence is provided which demonstrates that Versant Dutch scores are structurally reliable indications of candidate ability in spoken Dutch, suitable for high-stakes decision-making.

9. About the Company

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.

Ordinate Testing Technology: The Ordinate automated testing system was developed to apply advanced speech recognition techniques and data collection via the telephone and computer 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 Dutch Test is the result of years of research in speech recognition, statistical modeling, linguistics, and testing theory. The Versant patented technologies are applied to Pearson’s own language tests such as Versant English Test and Versant Spanish Test and also to customized tests. Sample projects include assessment of spoken English, assessment of spoken aviation 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 at investigating new applications for Ordinate technology. Research results are published in international journals and made available through the Versant website (www.VersantTest.com).

10. References

- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bernstein, J, Van Moere, A. & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27(3), 355-377.
- Bull, M & Aylett, M. (1998). An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In R.H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing*. Canberra, Australia: Australian Speech Science and Technology Association.
- Caplan, D. & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-126.
- Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. *Testing*. Washington, DC: Center for Applied Linguistics.
- Carroll, J. B. (1986). Second language. In R.F. Dillon & R.J. Sternberg (Eds.), *Cognition and Instructions*. Orlando FL: Academic Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cutler, A. (2003). Lexical access. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science. Vol. 2, Epilepsy – Mental imagery, philosophical issues about*. London: Nature Publishing Group, 858-864.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27, 464-491.
- 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.
- De Jong, J. H. A. L., Lennig, M., Kerkhoff, A., & Poelmans, P. (2009) Development of a Test of Spoken Dutch for Prospective Immigrants. *Language Assessment Quarterly*, 6(1),41-60.
- 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.
- 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.
- Linacre, J. M. (2003). Facets Rasch measurement computer program. Chicago: Winsteps.com.
- Linacre, J. M., Wright, B. D., & Lunz, M. E. (1990). A Facets model for judgmental scoring. Memo 61. MESA Psychometric Laboratory. University of Chicago. www.rasch.org/memo61.htm.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-412.
- Luoma, S. (2003). *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.
- North, B. & Hughes, G. (2003, December). CEF Performance Samples: For Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Retrieved September 26, 2008, from Council of Europe Web site: <http://www.coe.int/T/DG4/Portfolio/documents/videoperform.pdf>
- Perry, J. (2001). *Reference and reflexivity*. Stanford, CA: CSLI Publications.
- Sieloff-Magnan, S. (1987). Rater reliability of the ACTFL Oral Proficiency Interview, *Canadian Modern Language Review*, 43, 525-537.
- Stansfield, C. W., & Kenyon, D. M. (1992). The development and validation of a simulated oral proficiency interview. *Modern Language Journal*, 76, 129-141.
- Van Turenhout, M. Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280, 572-574.
- Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics*, 12(1), 54-73.

II. Appendix




Side I of the Test Paper in English: Instructions and general introduction to test procedures.



Test Instructions

Please read this before taking the test

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

Part	Instructions
Before the Test	<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.
Beginning the Test	<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. If you hang up before you complete the test, the test cannot be graded. You cannot reuse the Test Identification Number.
During the Test	<ul style="list-style-type: none"> Hold the phone close to your mouth as shown in the picture below. <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>NO Too low, too far away</p> </div> <div style="text-align: center;">  <p>YES In front of mouth</p> </div> <div style="text-align: center;">  <p>YES A good distance</p> </div> </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, "Ik weet het niet." Do not take notes or write during the test. When you hear, "Dank u voor uw deelname aan de test", you may hang up. If you wish, you may answer the optional questions at the end of the test. Your personal information will be kept anonymous.

PEARSON




© 2011 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).



TESTINSTRUCTIES

LEES DIT VOOR U AAN DE TEST BEGINT

De Versant test is een geautomatiseerde test voor gesproken taal die wordt afgenomen via de telefoon of de computer. Als u naar een voorbeeldtest wilt luisteren, een oefentest wilt kopen, of uw testscore wilt zien nadat u de test heeft beëindigd (indien van toepassing), ga dan naar www.VersantTest.com

DEEL	INSTRUCTIES
Voor u de test begint	<ul style="list-style-type: none"> • Neem deze instructies en de test aandachtig door. U mag een woordenboek gebruiken of om hulp vragen als u bepaalde woorden of zinnen niet begrijpt. • Kies een rustige plek met een vaste telefoonlijn waar u niet gestoord zult worden tijdens de test. • Gebruik geen draadloze telefoon, mobiele telefoon, of VoIP verbinding (zoals Skype™ of PC-naar-Telefoon diensten). Nieuwere toestellen zijn over het algemeen beter dan oudere modellen. Zorg ervoor dat de telefoon niet op de pulse-stand staat ingesteld, maar op de tone-stand.
U begint de test	<ul style="list-style-type: none"> • Om de test te beginnen, bel met een vaste lijn en een telefoon met druktoetsen het telefoonnummer op het testblad. • De opgenomen stem van een examiner begeleidt u door elk gedeelte van de test. • Geef met behulp van het toetsenbord van het telefoontoestel uw Test Identificatie Nummer in wanneer de examiner u dit vraagt. Dit nummer vindt u rechts bovenaan op de test. • De examiner zal u daarna twee vragen stellen: uw naam, de stad en het land van waaruit u belt. Als u te luid of te zacht praat, zal de examiner u dit zeggen. • De test begint wanneer u uw naam zegt. Als u ophangt voordat de test is afgerond, kan de test niet worden beoordeeld. U kunt het Test Identificatie Nummer niet opnieuw gebruiken.
Tijdens de test	<ul style="list-style-type: none"> • Houdt de telefoon dicht tegen uw mond zoals op het plaatje. <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Nee Te laag, te ver weg</p> </div> <div style="text-align: center;">  <p>Ja Voor de mond</p> </div> <div style="text-align: center;">  <p>Ja Op goede afstand</p> </div> </div> <ul style="list-style-type: none"> • Beantwoord alle vragen op een vlotte en natuurlijke manier met een heldere, vaste stem. • Als u niet weet hoe u een vraag moet beantwoorden, zwijgt u of zegt u, "Ik weet het niet." • Maak geen notities en schrijf niet tijdens de test. • Als u hoort, "Dank u voor uw deelname aan de test", mag u de telefoon ophangen. • Als u wilt, mag u de optionele vragen aan het einde van de test beantwoorden. Uw persoonlijke informatie wordt anoniem gehouden.

PEARSON

© 2011 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 candidate) showing Test Identification Number, Part A: sentences to read, and examples for all sections.



VERSANT DUTCH TEST

TER HERINNERING: De test begint als u uw naam zegt. Als u ophangt voordat de test afgerond is, kan de test niet beoordeeld worden. U kunt het Test Identificatie Nummer niet opnieuw gebruiken.

Test Identificatie Nummer (TIN)

1234 5678

Geldig tot : 1 januari 2012

Bel : 1-415-738-3800

Dank u voor het bellen met het toetsysteem van Versant.

Toets uw Toets Identificatie Nummer in.

Zeg nu uw naam.

Zeg nu vanuit welke stad en welk land u belt.

Volg nu de instructies bij deel A tot en met F.

DEEL	TAAK	OMSCHRIJVING
A	Voorlezen	<p><i>Lees de zinnen voor volgens de instructies.</i></p> <ol style="list-style-type: none"> 1 . Karin heeft een grote hond die Pluto heet. 2 . Ze gaat drie keer per dag met hem wandelen. 3 . In het weekend gaat ze vaak met hem naar het strand. 4 . Pluto vindt het erg leuk om in zee te zwemmen. 5 . Rijk wil een nieuwe auto gaan kopen. 6 . Zijn oude is vorige week stukgegaan en kan niet meer worden gerepareerd. 7 . Rijk heeft alleen niet zoveel verstand van auto's. 8 . Daarom vraagt hij een vriend die er wel veel van weet om met hem mee te gaan. 9 . Mijn vader vindt het geweldig om de afwas te doen. 10 . Wanneer hij bezig is, zingt hij luidkeels mee met de radio. 11 . Mijn moeder is blij dat zij het bijna nooit hoeft te doen. 12 . In plaats daarvan maait zij elke zaterdag het gras.
B	Nazeggen	<p><i>U hoort steeds een zin. Zeg de zin precies na.</i></p> <p>Voorbeeld : een stem zegt, " Dat is een mooi verhaal. " en u zegt, " Dat is een mooi verhaal. "</p>
C	Vragen	<p><i>U hoort steeds een korte vraag. Geef op elke vraag een kort antwoord.</i></p> <p>Voorbeeld : een stem zegt, " Is januari een dag of een maand? " en u zegt, " maand " of " een maand " . Of u hoort, " Hoeveel wielen heeft een auto? " en u zegt, " vier " of " vier wielen " .</p>
D	Zinnen bouwen	<p><i>U hoort steeds drie groepjes woorden. Met deze groepjes woorden vormt u een volledige zin.</i></p> <p>Voorbeeld : een stem zegt, " haar favoriete weekblad " ... " mijn moeder is " ... " aan het lezen. " en u zegt, " Mijn moeder is haar favoriete weekblad aan het lezen. "</p>
E	Verhalen navertellen	<p><i>U hoort drie korte verhalen. Elk verhaal wordt een keer verteld, en daarna hoort u een geluidssignaal. Na het geluidssignaal heeft u 30 seconden de tijd om het verhaal te vertellen. Vertel zoveel mogelijk van het verhaal na. Denk bijvoorbeeld aan: Wie deden er mee? Wat gebeurde er? Waar was het? En hoe liep het af? Als de 30 seconden om zijn, hoort u een tweede geluidssignaal.</i></p>
F	Open vragen	<p><i>U krijgt twee vragen te horen over het gezinsleven of over een persoonlijke keuze. Elke vraag wordt twee keer na elkaar voorgelezen, en daarna hoort u een geluidssignaal. Na het geluidssignaal heeft u 40 seconden de tijd om een antwoord te geven op de vraag. Als de 40 seconden om zijn, hoort u een tweede geluidssignaal.</i></p>

75 - 11111 - 1

PEARSON

© 2011 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).

Table 1. General level descriptors of the Council of Europe aligned with Versant Dutch Test scores.

Level		Council of Europe, 2001 Descriptor	Versant Dutch 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.	40
	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.	39 34
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.	33 29
	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.	28 24
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.	23 19
	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.	18 13
	<A1	Candidate performs below level defined as A1.	12 10



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.

Pearson

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

4940 Pearl East Circle
Suite 200
Boulder Colorado 80301
USA

VERSANT



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

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

Or visit us online at:
www.VersantTest.com

Pearson now includes Ordinate products and services.

© 2011 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.

Version 1211B