

Document published by the
Knowledge Technologies group of

PEARSON



For additional information about the
Knowledge Technologies group of Pearson
and their automated written and spoken
assessment products, visit:

Spoken assessment: www.versanttest.com

Written assessment: www.pearsonkt.com



VersantTM Aviation English Test

Test Description
&
Validation Summary

Contents

Versant™ Aviation English Test	1
1. Introduction	5
2. Test Description	6
Test Administration	6
Test Format	7
Number of Items	12
Scores and ICAO Levels	12
3. Test Construct	12
Test Construct	12
Relationship between Tasks and Subscores	13
4. Test Development	18
Item Development	18
Field-Testing	19
Scoring Development	21
5. Scoring Technology	21
6. Compliance with ICAO Language Proficiency Requirements	22
7. Reliability and Validity	24
Reliability	25
Correlations Among Subscores	25
Validity	26
Content Validity	26
Accuracy of Machine Scores	27
Construct Validity	28
8. Conclusions	30
9. Pearson: Testing and Technology	31
10. References	32
11. Appendices (A - E)	33

1. Introduction

The **Versant™ Aviation English Test (VAET)** is a spoken English test that measures the ability of test takers to understand aviation and plain English and to respond intelligibly in spoken English at a fully-functional pace. The VAET test was developed under a co-operative research and development agreement with the U.S. Federal Aviation Administration (FAA) and is intended for both pilots and air traffic controllers. The test is aligned with the mandate of the International Civil Aviation Organization (ICAO) as described in the *Manual on the Implementation of ICAO Language Proficiency Requirements* (2004), which requires that all pilots and air traffic controllers serving international traffic show at least Level 4 English language proficiency on the ICAO Language Proficiency Rating Scales.

The ICAO Manual (Section 6.6.3) states the following four characteristics as an appropriate assessment in the context of aviation language testing:

- a) The test must be a proficiency test of speaking and listening.
- b) The test must be based on the ICAO Rating Scale and holistic descriptors.
- c) The test must test speaking and listening proficiency in a context appropriate to aviation.
- d) The test must test language use in a broader context than in the use of ICAO phraseologies alone.

The VAET is a proficiency test that is designed to measure the test taker's speaking and listening abilities. The VAET is scored with reference to the ICAO Language Proficiency Rating Scale and reports all six ICAO subskills: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions. The VAET evaluates these subskills based on the test taker's performance on tasks that employ ICAO phraseology and common English. In addition, the test follows the recommended model in the ICAO Manual and it is "comprised of a mix of both aviation-specific content alongside less aviation-specific content." (Section 6.8.7) Common, concrete, and work-related topics are presented throughout the test.

Section 6.7 on the ICAO Manual states that "tests that do not evaluate speaking and/or listening skills directly, either through direct interaction or through an audio- or video-taped exchange, are not appropriate to meet the requirements of the ICAO language proficiency testing guidelines." The VAET directly measures how well the test taker can understand spoken English on common, concrete, and work-related topics, and respond to it in real time through a series of audio prompts that are recorded by various native and non-native speakers of English. The VAET is thus considered a semi-direct testing procedure and is listed in the ICAO Manual as one of the preferable testing procedures for testing English language proficiency (Sections 6.7.5 and 6.7.8).

The VAET therefore meets the ICAO guidelines and can be used for certifying a pilot's or an air traffic controller's spoken English proficiency as specified and required by the ICAO language proficiency scale.

2. Test Description

Test Administration

The VAET can be delivered over the telephone or on a computer. During the test, the Versant testing system presents a series of spoken prompts in English at a conversational pace and elicits oral responses from the test taker in English. The VAET test consists of eight sections:

- Aviation Reading
- Common English Reading
- Repeat
- Short Answer Questions
- Readback
- Corrections and Confirmations
- Story Retelling
- Open Questions

The test generally takes about 25 minutes to complete. Instructions for the test are spoken over the testing system in an examiner voice and are also presented verbatim on a printed test paper for telephone delivery and on the computer screen for computer delivery. Test items themselves are presented in various native and non-native speaker voices that are distinct from the examiner voice.

Performance from each section of the test contributes to the scoring of six ICAO subskills:

- Pronunciation
- Structure
- Vocabulary
- Fluency
- Comprehension
- Interactions

Each subskill is evaluated based on different, independent aspects of the spoken response material.

The Versant testing system analyzes the test taker's responses and posts scores usually within a few minutes after a test is completed. Test administrators and score users can then view and print out test results from a password-protected section of the www.VersantTest.com website.

Telephone Delivery

For telephone delivery, a corded, land-line telephone is required for accurate test results. Telephone administration is supported by a test paper. The test paper presents the telephone number to call and a Test Identification Number (TIN) that is unique for each test administration. When the telephone number is dialed, the Versant testing system prompts the user to enter a TIN on the telephone keypad. During the test, the instructions are spoken by an examiner voice, and the test items are spoken by a range of speakers, distinct from the examiner voice. Test takers interact with the test system in English until they complete the test and hang up the telephone.

Computer Delivery

For computer delivery, the test taker is fitted with a microphone headset connected to a computer. The computer must have an internet connection and the Versant Computer-Delivered Test (CDT) software. Pearson can provide more information about the software and system requirements upon request. As in telephone delivery, the test taker is provided with a unique TIN. Before the test begins, the system guides the test taker through adjusting the volume and calibrating the microphone. The spoken instructions for each section are displayed at the appropriate time on the computer screen. Test takers interact with the test system in English, speaking their responses into the microphone headset. When the test is finished, the test taker clicks a button labeled, “END TEST”.

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 eight tasks of the VAET. Responses from Parts A to G are automatically scored and contribute to generation of each ICAO subscore. The last task, Open Questions, is not scored automatically but is provided as an additional feature. Responses on this task are recorded and made available to authorized administrators who wish to hear samples of a test taker’s spontaneous speech in English.

Part A: Aviation Reading

In this task, the test taker reads aloud sentences that are printed on the test paper. The sentences consist of ICAO phraseology. The test taker is prompted by the examiner’s voice and reads the sentences one at a time, in the order requested.

Examples:

On the test paper you see:

1. World Air 891, request descent.
2. World Air 891, maintain flight level 280 expect descent shortly.
3. Maintaining flight level 280, World Air 891.

When you hear: “Please read sentence number three.”

You read: “Maintaining flight level 280, World Air 891.”

The sentences can be read easily and fluently by native, English-speaking aviation professionals. This task provides samples of the test taker’s pronunciation and reading fluency on ICAO phraseology. The Aviation Reading task starts the test because, for many test takers, reading aloud presents a familiar task, and provides a comfortable introduction to the interactive mode of the test as a whole.

Part B: Common English Reading

In this task, the test taker reads aloud sentences that are printed on the test paper. The sentences consist of common English. The test taker is prompted by the examiner’s voice and reads the sentences one at a time, in the order requested.

The sentences are relatively simple in structure and vocabulary, and they can be read easily and fluently by literate, native English speakers. This task provides more samples of the test taker's pronunciation and reading fluency.

Examples:

On the test paper you see:

1. Our flight didn't start very well.
2. First there were some minor mechanical issues that needed to be resolved.
3. Then all outgoing flights were delayed because of the weather conditions.

When you hear: "Please read sentence number one."

You read: "Our flight didn't start very well."

Part C: Repeat

In the Repeat task, the test taker listens to a sentence in common English and then attempts to repeat the sentence aloud verbatim. The sentences are presented in order of increasing difficulty. Topics used are common, concrete, and work-related topics. Priority lexical domains (ICAO Manual, page B-10) and the top 250 four-word clusters in spoken English (ICAO Manual, page B-18) were referenced to create some of the items.

Examples:

When you hear: "Visibility is very poor today."

You say: "Visibility is very poor today."

When you hear: "Do you happen to know what time it is?"

You say: "Do you happen to know what time it is?"

When you hear: "I try to get there a couple of hours before the scheduled departure."

You say: "I try to get there a couple of hours before the scheduled departure."

To repeat a sentence longer than about seven syllables, the test taker has to recognize the words as produced 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 sentence structures. If a person habitually processes five-word phrases as a unit (e.g., "the very large descending aircraft"), then that person can usually repeat utterances of 15 or 20 words in length. Generally, the ability to repeat material is constrained by the size of the linguistic unit that a person can process in an automatic or nearly automatic fashion. As the sentences increase in length and complexity, the task becomes increasingly difficult for speakers who are not familiar with English sentence structures.

The Repeat task requires test takers to organize speech into linguistic units and tests their ability to control both basic and complex grammatical structures in real time. In addition, the task has them repeat back full sentences (as opposed to just words and phrases), and therefore, it also offers a sample of the test taker's pronunciation and fluency in spoken common English.

The test taker's response latency for each Repeat item is also recorded and measured. The response time extracted from each item contributes to the immediacy aspect of the Interactions subscore.

Part D: Short Answer Questions

In the Short Answer Questions task, the test taker listens to spoken questions about common, concrete, and work-related topics in common English and answers each of these questions with a single word or short phrase. The questions generally include at least three or four content words embedded in an English interrogative 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 cultural, geographic, or historic facts 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:

When you hear: "I have a pair of suitcases. How many suitcases do I have?"

You say: "Two" or "You have two"

When you hear: "Is land that's almost entirely surrounded by water a peninsula or a pond?"

You say: "Peninsula" or "A peninsula"

When you hear: "A plane has just taken off. Is the plane on the ground or in the air now?"

You say: "In the air"

Short Answer Questions measure the test taker's ability to understand common, concrete and work-related words in sentence context and to produce an appropriate word(s) in response to the questions. To successfully perform on this task, the test taker must be able to identify the words in phonological and syntactic context, infer the demand proposition, and then say an appropriate vocabulary word(s) as a response. Lexical items are based on the ICAO list of priority lexical domains (ICAO Manual, page B-8) including topics such as animals, numbers, movement, time, transportation, and weather. Since items are recorded in different native and non-native voices, the test taker must be able to comprehend a range of speech varieties. Thus, performance on this task contributes to the following three subscores: Vocabulary, Comprehension, and Interactions (for the appropriate response aspect of the Interactions subscore).

Part E: Readback

For the Readback task, the test taker hears a spoken radiotelephony message and is asked to give an appropriate readback to confirm his or her understanding of the message. The call sign in the message is printed on the test paper for support. The test taker is expected to produce a readback using aviation phraseology as recognized by ICAO.

This task requires the ability to comprehend routine phraseology messages to include the relevant information in readback responses, and thus contributes to the Comprehension subscore as manifest in predictable, routine, work-related language.

Examples:

On the test paper you see:

1. Coastal Air 315
2. World Air 395
3. Coastal Air 405

When you hear: "Coastal Air 315, maintain flight level 070."

One possible answer is: "Maintain flight level 070, Coastal Air 315."

Another possible answer is: "Maintaining flight level 070, Coastal Air 315."

When you hear: "World Air 395, exit taxiway Hotel."

One possible answer is: "Exit taxiway Hotel, World Air 395."

Another possible answer is: "Exiting taxiway Hotel, World Air 395."

Since the test taker is providing samples of continuous speech, the Readback task also contributes to the Pronunciation and Fluency scores in the routine, aviation domain. As in the Repeat responses, the information about response latency is also measured in each Readback response to derive the Interactions subscore.

Part F: Corrections and Confirmations

In the Corrections and Confirmations, the test taker hears a radiotelephony message, either from the air traffic controller's perspective or the pilot's perspective. The test taker also sees the call sign in the message printed on the test paper. Then the test taker hears a readback, which might contain the correct information, wrong information, or a request for more information. The test taker is expected to respond to the message appropriately using ICAO phraseology when possible. For example, if the response contains wrong information, the test taker is expected to provide the correct information.

Examples:

On the test paper you see:

1. Charlie Romeo 4013
2. Coastal Airline 445
3. World Air 2043

When you hear:

(Speaker 1) "Charlie Romeo 4013, continue descent to flight level 110, report passing 150."

(Speaker 2) "Descending to flight level 10 thousand, report passing 15 thousand, Charlie Romeo 4013."

You could say: "Charlie Romeo 4013, negative, continue descent to flight level 110, report passing 150."

Another possible answer is: "Charlie Romeo 4013, I say again, continue descent to flight level 110, report passing 150."

The task requires the test taker to track an exchange between an air traffic controller and a pilot and to deal with misunderstandings by correcting, confirming, or clarifying information. Some

items reflect routine communications/situations, while others cover less routine communications/situations. A small proportion explores unexpected communications/situations as well. The immediacy and appropriateness of the responses as well as the information conveyed in them are important factors in estimating the test taker's ability to manage speaker/listener interactions. These three aspects of the Interaction subskill are extracted from the responses in this task.

Part G: Story Retelling

In the Story Retelling task, the test taker listens to an orally presented aviation scenario and is then asked to describe what happened in his or her own words. The scenarios are several sentences long and describe one or more characters, a situation, and an outcome.

Example:

You hear:

"Most of the flight between Tokyo and Hawaii was calm. An hour into the flight, the pilot encountered a line of storms which she had to fly over. The flight became bumpy and the passengers were nervous but the flight attendants kept everyone in their seats. After a few minutes, the plane cleared the storms and the rest of the flight was smooth. Several hours later, the plane arrived safely in Hawaii."

Now, you should retell this story in your own words.

Some topics are designed to be common, concrete, and familiar while others are less familiar or unexpected. The test taker must identify words in phonological and syntactic context, extract key information, and then paraphrase the relevant information in an informative, extended response. Scoring of the Story Retelling responses focuses on the test taker's vocabulary range as well as on the test taker's ability to relate the relevant information in an informative manner. The Story Retelling task contributes to Vocabulary and Interactions (for the appropriate and informative aspect of the Interactions subscore).

Part H: Open Questions

In this task, the test taker listens to spoken questions on work-related, aviation topics in common English and then describes his or her opinion, experience, or thinking for each question.

This task is used to collect spontaneous speech samples from the test taker. The test taker's responses are not scored at present, but these responses are made available for review by authorized test administrators.

Example:

You hear a question twice:

"In your experience, what kinds of weather are the most difficult for air traffic management and why? Please explain...In your experience, what kinds of weather are the most difficult for air traffic management and why? Please explain."

Now, you should give your answer.

Number of Items

The VAET presents a total of 78 items to each test taker over the eight separate sections. The 78 items are drawn at random from a larger pool of items and most items will be different from one test administration to the next.

VAET Scores and ICAO Levels

The VAET test reports both an ICAO level and a numeric score for each subskill. The final ICAO level and score is also reported. As required in the ICAO Manual (Doc. 9835), the final ICAO level is determined by the lowest subscore(s) among the six ICAO subskills. VAET numeric scores are reported on a scale of 10 to 70. Table 1 displays the ICAO levels and the corresponding score ranges on the VAET.

Table 1. Correspondence between VAET numeric scores and ICAO levels

ICAO Level	VAET Score Range
Level 1: Pre-elementary	10 - 19
Level 2: Elementary	20 - 29
Level 3: Pre-operational	30 - 39
Level 4: Operational	40 - 49
Level 5: Extended	50 - 59
Level 6: Expert	60 - 70

3. Test Construct

Test Construct

The *construct* of a test is the concept or characteristics that a test is designed to measure (*Standards for Educational and Psychological Testing*, 1999). For example, in an I.Q. test the construct is normally considered to be intelligence. The Versant Aviation English Test is designed to measure *facility in spoken aviation English and common English in the aviation domain*; that is, the ability to understand spoken English both with regard to aviation radiotelephony phraseology and topics related to aviation (such as movement, position, time, duration, weather, etc.), and the ability to respond appropriately in intelligible English at a fully-functional pace.

In order for a speaker to successfully participate in a spoken conversational exchange, numerous cognitive tasks must be completed in real-time: the speaker must 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, adapted from Levelt (1989).

Facility in spoken English is essential to successful and effective radiotelephony communications in both routine and unexpected contexts. Core language component processes normally occur at

a very rapid pace – aviation professionals must be able to track what is being said, extract meaning as speech continues, and then formulate and produce a relevant and intelligible response within a very small period of time. If test takers are unable to perform these internal processes in real time, they will most likely not be able to interact effectively in communication as either listeners or speakers, especially when emergencies occur and when attention needs to be focused on aviation-related tasks at hand to handle emergency situations.

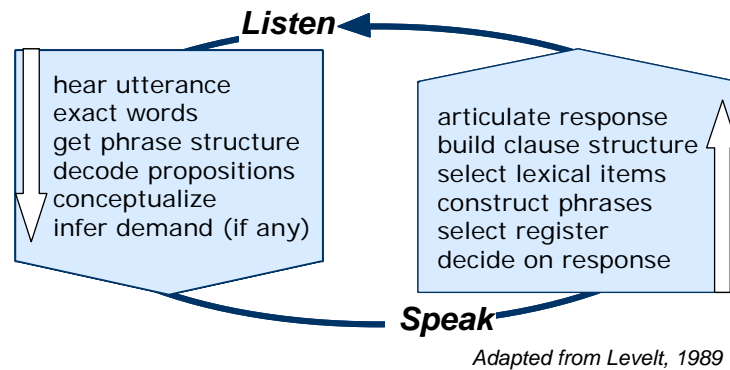


Figure 1. Processing components in listening and speaking.

The VAET is a measurement of real-time language processing and effective real-time language processing requires *automaticity* of the encoding and decoding of the core elements of oral language. An example of automaticity is when an experienced driver is driving a car: the driver performs numerous operations simultaneously without having to think about them too much. In VAET, automaticity refers to 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). For instance, because cognitive capacity is limited, automaticity is required in order for the speaker/listener to be able to pay attention to the unfolding situation rather than paying attention to what needs to be said or how the encoded message should be structured. Performance on VAET items reflects the test taker’s level of automaticity in real-time spoken language. During an emergency situation, pilots and air traffic controllers must be able to direct their attention to solving the problem rather than devoting their attention to think about how to say what needs to be said. The same facility in spoken English that enables an aviation professional to satisfactorily perform the VAET tasks in real time also enables that individual to successfully perform in predictable, less predictable, and unexpected situations with fully-functional English.

Relationships between Task and Subscore

Of the 78 items in an administration of the VAET, 71 responses are used in the automatic scoring. The first item response in the Repeat, Short Answer Questions, Readback, and Corrections and Confirmations sections of the test is considered a practice item and is not incorporated into the final score. In addition, responses to the two Open Questions in Part H are not scored automatically.

Each task in the test is specifically designed to provide information about the test taker’s ability on one or more of the six ICAO subskills. Multiple parameters are extracted from each test taker

response in each task in order to generate scores that meet the ICAO language proficiency criteria. Figure 2 illustrates which sections of the test contribute to which subscores.

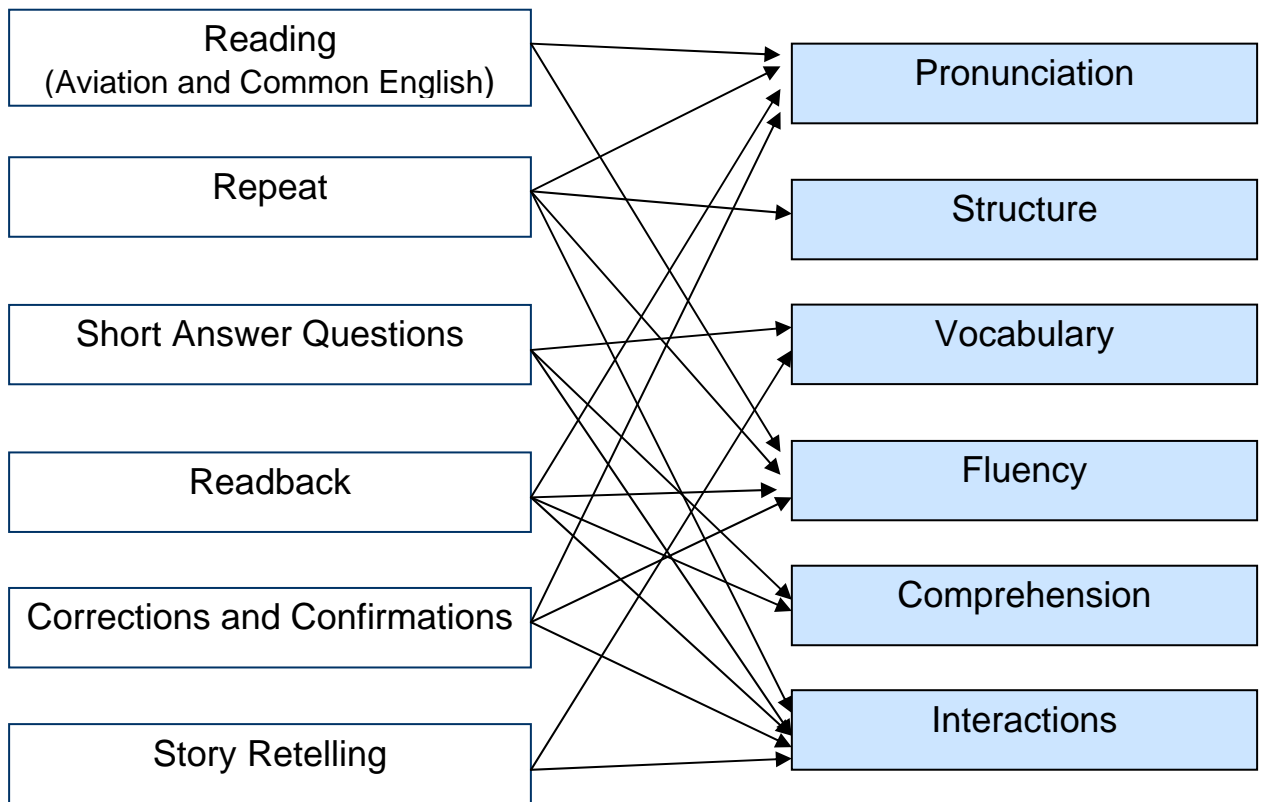


Figure 2. Relation of subscores to item types for the Versant Aviation English Test.

Pronunciation

The ICAO Language Proficiency Rating Scale defines the Pronunciation subskill levels in terms of “pronunciation, stress, rhythm, and intonation”. To derive the ICAO Pronunciation subscore, responses from the following five tasks are used: Aviation Reading, Common English Reading, Repeat, Readback, and Corrections and Confirmations. Each of these tasks elicits responses longer than a single phrase, so that all aspects of the ICAO Pronunciation construct can be observed. Aviation Reading, Readback, and Corrections and Confirmations provide the test taker’s pronunciation samples in the work-related, phraseology domain, whereas responses to Common English Reading and Repeat items display the test taker’s pronunciation ability in the common English context on more common, concrete topics. Pearson’s system extracts information about the stress and segmental forms of the words in the responses and the pronunciation of the segments in the words within a prosodic phrasal context. These aspects of pronunciation are sufficient to predict a listener’s “ease of understanding” as cited in the ICAO Language Proficiency Rating Scale.

Structure

According to the ICAO Language Proficiency Rating Scale, the Structure subskill represents the test taker's ability to control both basic and complex grammatical structures in real time. The test taker's responses from the Repeat section are used to derive scores on this subskill. In this task, the test taker listens to a sentence and then tries to repeat it verbatim. This task requires the test taker to "correctly repeat all the pertinent forms for person, tense, register" (Radloff, 1991: 9). In other words, sentence repetition requires the test taker to use grammatical knowledge in real time to repeat sentences with all phrases and clauses intact and in grammatically correct sequences. This is especially the case with longer sentences. As sentences increase in length and complexity, the task becomes increasingly difficult for less proficient speakers who are less familiar with the language's phrase structures and common syntactic forms (Radloff, 1991).

Sentence repetition should not be misconstrued as a memory test. The Versant test development teams' item analysis studies for the spoken English test and the spoken Spanish test show that virtually all native speakers achieve high scores on the sentence repetition tasks, whereas non-native speakers obtain scores distributed across the scale. If memory, as such, were an important component of performance on this task, the native speakers would also show notable variance according to their range of memory spans. Psycholinguists have also shown that short-term memory for strings of digits, for example, is distinct from cognitive resources normally used to process and comprehend sentences (Caplan & Waters, 1999). This result is consistent with the hypothesis that successful performance on the Repeat task is based on the size of the linguistic unit that the test taker can process in an automatic or nearly automatic fashion. In other words, the task assesses the test takers' syntactic processing resources and their facility with longer linguistic units (e.g. "the very large descending aircraft").

Vocabulary

The ICAO descriptors define the Vocabulary subskill in terms of the range and accuracy of vocabulary for effective communication and for paraphrasing. The Vocabulary subscore is generated from two tasks: Short Answer Questions and Story Retelling. In Short Answer Questions, the test taker listens to spoken questions presented in common English sentence structures and answers each question with a single word or short phrase. Responses from Short Answer Questions indicate the range of a test taker's effective reception and production of core vocabulary within common, concrete, and work-related topics. Story Retelling, on the other hand, measures deeper productive vocabulary knowledge because the test taker listens to an orally presented story or incident that deals with an aviation-specific topic, and then describes what happened in his or her own words. In this sense, the test taker's competence at retrieving lexical items for paraphrasing is assessed. Scoring of the Story Retelling responses focuses on the range of vocabulary elicited from the test taker and the accuracy of its use. The information obtained from these two tasks provides clear evidence about the test taker's ability to use vocabulary knowledge at the word, phrase, sentence, and discourse levels.

Fluency

The ICAO Fluency subskill is characterized in terms of the test taker's ability and ease in maintaining appropriate speech flow and tempo at length within a discourse, making use of discourse markers and connectors for effective communication. ICAO expert ratings were collected with reference to the ICAO Fluency descriptors to develop a criterion-referenced fluency scoring model. Fluency subscores are then generated in a way that closely matches ICAO expert

raters' assessments of fluency when rating with reference to the ICAO scales. Although the Fluency subscore is derived from the same four tasks that are used to estimate Pronunciation ability, different information is extracted from test takers' responses in order to generate ICAO Fluency levels. In other words, different features of the prompt-response discourse than are used to generate the Pronunciation subscore are extracted and analyzed for the Fluency score. Features of fluency include the rate of speaking, the position and length of pauses in each of the four tasks, and the rate and continuity of speech while accurately responding to and using discourse-level cues in the Corrections and Confirmations task. The tasks used to produce the Fluency subscore require test takers to perform in both common English structures and radiotelephony phraseology. The measures are scaled with reference to native and non-native distributions and are then rescaled such that they optimally predict expert human judgments of fluency based on the ICAO Language Proficiency Rating Scale.

Comprehension

ICAO classifies the Comprehension subskill as the degree of accuracy with which a test taker is able to comprehend common, concrete, and work-related topics. The VAET assesses the test taker's ability to understand questions and aviation messages in both phraseology and common English structures. In Short Answer Questions, the test taker is asked to listen to a series of questions posed in common English structures and infer the demand proposition about various topics including common, concrete, and work-related topics. To do this, the test taker needs to be able to identify the lexical items and process their meaning in real time for comprehension. The Readback task is designed to measure the test taker's comprehension ability in radiotelephony messages. The scoring system analyzes responses for pre-determined key content. Whether the test taker's responses contain the key content or not indicates the level of understanding of the messages in phraseology. As per the ICAO rating scale descriptors, the combination of these two tasks provides evidence about the test taker's ability to comprehend spoken English in both routine and phraseology contexts, and in more general contexts expressed in common English structures.

Interactions

In the ICAO Language Proficiency Rating Scale, the Interactions subskill is characterized in terms of *immediacy* and *appropriateness* of responses and *informativeness* of the content conveyed in responses elicited by routine and non-routine situations. In other words, the Interactions subskill is the ability to respond immediately, appropriately, and informatively using both ICAO phraseology and common English. Immediacy can be operationalized as the latency of responses. Appropriateness and informativeness of responses can be assessed based on the content of the responses. Using a combination of tasks that require ICAO phraseology and common English ensures that the test taker demonstrates performance in routine and non-routine contexts. In VAET, the Interaction subscore is derived from the following five tasks: Repeat, Short Answer Questions, Readback, Corrections and Confirmations, and Story Retelling.

Different parameters are extracted from each of these tasks. Immediacy of responses is measured from Repeat, Readback, and Corrections and Confirmations. On page 14 of Appendix A in the ICAO Doc. 9835, only one sentence is provided as a high-level description of the Interactions subskill: "Pilots and controllers should be aware that inappropriate silence may indicate a failure to understand." This illustrates the importance of modeling what is "inappropriate silence". A quantitative response latency model is implemented in VAET based on empirical spoken performance data obtained from both native and non-native aviation

professionals. The test taker's response latency is evaluated with reference to this model. The Repeat task requires the test taker to repeat common and aviation-related sentences in common English structures, whereas Readback items require professional responses in ICAO phraseology in routine contexts. Additionally, in Corrections and Confirmations, the test taker hears a radiotelephony exchange between an air traffic controller and a pilot. The test taker is asked to provide an appropriate response to a message in the exchange by confirming or correcting information, or by providing information that was requested in the message. Some items are fairly routine and others are less routine. The test taker is asked to respond in ICAO phraseology whenever possible. These three tasks ensure that the response latency is measured in various contexts.

Appropriateness and informativeness of responses are measured in Short Answer Questions, Readback, Corrections and Confirmations, and Story Retelling. Every item in each of these tasks has key information content, with which responses can be determined to be appropriate and/or informative. Readback items ask the test taker to respond appropriately and informatively in ICAO phraseology in routine, radiotelephony contexts. The Corrections and Confirmations task requires the test taker to confirm, clarify, or correct the radiotelephony exchange in routine and less predictable situations. In these two tasks, the amount and order of the predefined key information is analyzed to determine if the response is appropriate and informative. In Short Answer Questions and Story Retelling, however, the test taker is required to answer in common English. Responses to Short Answer Question items demonstrate whether or not the test taker can provide the most relevant and appropriate word(s) in the given context with a short spoken response. Unlike Short Answer Questions, Story Retelling can provide evidence for whether or not the test taker can appropriately and informatively relate important and relevant information in the story in an extended spontaneous discourse format using common English.

Responses from these five tasks provide different pieces of information about the quality (immediacy, appropriateness, and informativeness) of the test taker's responses given in both phraseology and common English, which are directly relevant to the ICAO Interactions subskill. Retelling a story effectively requires not only a grasp of the words used in the story but also an understanding of the broader gist of the story's events, and relies upon an awareness on the part of the reteller that certain pieces of information are more relevant for communication than others. Especially in unusual or non-routine situations, a precise response using contextually relevant and specific words to convey key information is more informative and appropriate; a wordy response devoid of relevant content will be less informative and may negatively impact communication. Thus, the ability to effectively describe a situation appropriately using informative language (i.e., specific, relevant, and appropriate content) underlies effective interaction.

Taken together, the tasks in VAET provide direct measures of the test taker's listening and speaking ability, according to the six subskills in the ICAO Rating Scale.

4. Test Development

Item Development

The VAET test development team consisted of numerous spoken language testing experts and relevant experts in aviation language testing. The test developers had extensive experience creating and validating language tests for specific purposes. Table 2 presents the members of the test development team and their qualifications.

The development of the VAET followed a rigorous process that entailed clearly defining the test construct, creating detailed test specifications, writing item specifications to be used by item writers, developing the items, coordinating review of the items by trained aviation professionals, revising the items, creating a data collection system, collecting feedback on the test administration from stakeholders and revising the test instructions, trialing the items with both native and non-native English speaking aviation professionals from a variety of countries, analyzing the data, making a final selection of items, enabling automatic scoring of the items, and validating the test. This process aligns with the test development process recommended by the ICAO Manual (6.5.4) and Luoma (2004).

Table 2. Members of the VAET test development team and their qualifications

Education	Experience
Ph.D., Linguistics	University professor specializing in applied linguistics and language testing, especially in the aviation domain
Ph.D., Psycholinguistics	10+ years of experience developing and conducting research on second/foreign language tests
M.A., TESOL	4+ years of experience developing and conducting research on second/foreign language tests
Ph.D., Psychology	3 years of experience developing and conducting research on second/foreign language tests
Ph.D., Psychology	1.5 years of experience developing and conducting research on second/foreign language tests
Ph.D., Psychology	1.5 years of experience developing and conducting research on second/foreign language tests
M.A., TESOL	1 year of experience developing and conducting research on second/foreign language tests

The items cover a broad range of skill levels and skill profiles. Each VAET item is independent of the other items and presents unpredictable spoken material in English. In the VAET, both ICAO phraseology and common English appear throughout the test. For example, the Repeat section presents common English sentences with both common and aviation content, while the Readback section uses ICAO phraseology. In the Corrections and Confirmations section, ICAO phraseology is used except in those cases in which the test taker is presented with a non-routine situation and must use common English to clarify, correct, or confirm information. The test follows the recommended model in the ICAO Manual, in which a test is “comprised of a mix of both aviation-specific content alongside less aviation-specific content” (Section 6.8.7). Even with common English items, the content of the VAET was designed to address work-related topics in the aviation domain. For the common English tasks such as Repeat and Short Answer

Questions, the ICAO list of priority lexical domains guided the choice of lexical items and topics. In addition, transcripts of conversations from actual pilots and air traffic controllers were used as reference material for creating items in the Readback, Corrections and Confirmations, and Story Retelling sections of the test.

Draft items were reviewed by aviation professionals to ensure that the items are accurate and, when appropriate, conform to ICAO phraseology standards. The item reviewers were all native English speakers that came from dialectically distinct world regions; they were therefore also able to screen test language and test items for geographic and cultural bias. The qualifications and experiences of the item reviewers are summarized in Table 3.

Table 3. Members of the item reviewing team and their qualifications

Occupation	Experience
Program manager for aviation language training	25 years of experience in human resource development and language training in the US
Consultant for ICAO English language competency project	30+ years of experience as Air Traffic Control Officer in the UK and Australia, ICAO PRICESG (Proficiency Requirements in Common English) participant
Commercial Pilot	10+ years of experience as commercial pilot in the UK
International air traffic training program manager	10+ years of experience in international air traffic training in the US
International air traffic training program staff	3+ years of experience in aviation English (program development, materials development, aviation English instructor) Additional 10 years of experience in various aviation projects (international air traffic training, symposiums)

Once items had been reviewed and revised, they were professionally recorded. The voices that present items are both male and female, representing a range of accents and speaking styles. For the recording of the items with ICAO phraseology, seven aviation professionals were recruited: two retired air traffic controllers, two active air traffic controllers, one retired commercial pilot and two active commercial pilots. A professional male voice recorded the examiner prompts for the test.

Field Testing

Each item was field tested with both native and non-native English-speaking aviation professionals. To ensure construct validity, the item specifications require that native English-speaking aviation professionals find the items very easy to understand and to respond to appropriately; otherwise, it is unclear whether the item is measuring language ability or another construct. Native English speakers were roughly defined as individuals who were born and educated in an English-speaking country. A total of 478 native aviation professionals participated in the data collection. Native aviation professionals were recruited from the US, UK, Canada, Australia, and New Zealand.

For the sample of non-native English-speaking aviation professionals, participants represented a range of native languages and English proficiency levels. A total of 628 non-native aviation professionals with a wide range of aviation experience were recruited. Their first language

backgrounds were also diversely distributed representing more than 40 first languages. Table 4 shows the demographics of the native and non-native samples.

For both native and non-native samples, most respondents reported that they were active commercial pilots or air traffic controllers at the time of the field-testing with varying degrees of aviation experience. However, respondents do not necessarily currently fly international routes or necessarily communicate with international pilots as air traffic controllers. A number of participants also reported that they were retired aviation professionals, retired or active military pilots, corporate pilots, and flight/air traffic instructors.

Table 4. Demographics of the native and non-native samples

	Native	Non-Native
Number of Participants	478	628
Male : Female	443 (93%) : 35 (7%)	526 (84%) : 102 (16%)
Age Range (mean age)	20 to 80 (mean = 47)	17 to 74 (mean = 40)
Years of Aviation Experience Range	1 to 56	1 to 52
Languages	English (Australia, Canada, New Zealand, UK, US, Other).	Afrikaan, Arabic, Armernia, Bahasa Indonesia, Chinese, Croatian, Czech, Damara, Danish, Dutch, Estonian, Farsi, Finnish, Flemish, French, German, Greek, Hebrew, Hindi, Italian, Japanese, Lithuanian, Luxembougish, Macedonian, Maltese, Norwegian, Polish, Portuguese, Punjabi, Russian, Serbian, Spanish, Swahili, Swedish, Thai, Tigrinya, Tunisian, Urdu, Toruba
Occupation	111 Air Traffic Controller (23%) 259 Pilot, Commercial (54%) 50 Pilot, Corporate (10%) 11 Pilot, Private (2%) 18 Pilot, Retired (4%) 26 Other: Retired Coporate Pilot, Military Pilot, Retired Military Pilot (2%), Flight Instructor, Production Test Pilot, Retired ATC (2%), Student Pilot (1%)	271 Air Traffic Controller (43%) 289 Pilot, Commercial (46%) 34 Pilot, Corporate (5%) 9 Pilot, Private (1%) 25 Other: Accident Investigator, Air Traffic Instructor, Aircraft Dispatcher, Aircraft Maintenance Technician, Flight Instructor, Flight Test Engineer (3%), Retired Military, Student Pilot (2%)

After the field testing, item analyses were performed. For an item to be included in the test, a preponderance of the sample of native speakers presented with the item had to provide correct responses to the item. If an internal threshold for correct responses was not met, the item was excluded from the final item pool.

Item difficulty estimates were determined using a multifacet, one-parameter Rasch model as implemented in the computer program FACETS (Linacre, 2003). The FACETS program estimates rater severity, ability level of the test takers, and item difficulty (Linacre, Wright, and Lunz, 1990). The point biserial correlation between item response and total score was used to determine whether or not an item differentiated abilities well. Only items with a point biserial above a predetermined threshold were classified as gradable and became a part of the final test.

Scoring Development

The data collected during the field test were used to create acoustic models, language models, and pronunciation and fluency models for automatic scoring. In order to build the automated scoring models that meet the ICAO language proficiency requirements, 11 professional raters were recruited to provide human ratings. Note that since VAET is a test specifically designed to meet the ICAO language proficiency requirements, it was deemed critical to have raters who are experienced as aviation English professionals and who are highly knowledgeable about the ICAO language proficiency requirements. All raters have been very active in aviation English teaching, testing, and/or material/curriculum development for many years in different parts of the world (US, UK, Spain, Italy, Central America, Thailand, etc.) and were closely engaged with the ICAO requirements. Two of the raters were participants in the ICAO PRICESG (Proficiency Requirements in Common English Study Group). Three raters have experience as either an air traffic controller or pilot (one of them is still a captain for a commercial airline). The descriptions of the human raters are summarized in Appendix A.

Following the ICAO language proficiency scale, the ICAO expert raters rated a number of test takers' responses collected during the field test for different ICAO subskills. These responses were presented in different random orders to the ICAO expert raters via the Versant telephone grading system or via the Versant web-based grading system. Each response was rated by at least two raters independently. A total of 25,890 independent ratings were produced for the development of the automated scoring system and separately, a total of 20,673 independent ratings were produced for the validation of the automated scoring system, which is described in more detail in Section 7, Reliability and Validity.

A significant number of responses from the field test were also transcribed by trained transcribers to create language models and to optimize the speech processing technologies. To ensure accuracy of transcriptions, people who had some language/linguistic backgrounds were recruited, screened, and trained. Most of the transcribers were in a graduate-level Linguistics degree program or already held a Linguistics degree at the time of the transcription project. A number of responses were randomly presented to the transcribers via the Versant web-based transcription system and each response was transcribed by two transcribers independently. The transcribers produced a total of 84,723 transcriptions for the development of the automated scoring system. They also produced 24,706 transcriptions separately for the validation of the automated scoring system. The qualifications of the transcribers are provided in Appendix B.

5. Scoring Technology

The ICAO Language Proficiency Rating Scale consists of six language skills (Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions). Each skill is measured

separately during the VAET and reported separately using the ICAO Scale Levels 1 through 6. As specified by the standards, the scoring is based on a conjunctive model in which the Final score is determined based on the lowest subscore level. The ICAO Manual provides a detailed explanation of the rating scale (see Appendix C).

The Ordinate patented automated scoring system produces multiple, independent measures from the same set of responses. For this reason, tasks often contribute to more than one subscore. The use of multiple item types for subscores ensures score reliability. An added advantage of evaluating language skills independently is that a subscore is not confounded by features of other language skills. For example, a heavy accent will not affect the evaluation of the content of the test taker's response.

In each section of the test, an incoming response is recognized automatically by a Hidden Markov Model (HMM)-based speech recognizer developed from the HMM Tool Kit (Young, Kershaw, Odell, Ollason, Waltchev, and Woodland, 2000). The acoustic models for the speech recognizer (models of each sound in the language) have been trained on data from a diverse sample of non-native speakers of English. In this way, the speech recognizer is optimized for various types of non-native speakers' accented speech patterns and the machine generally recognizes response words as well as or better than a naïve listener, but does not generally do as well as a trained listener who knows the item content. The speech recognizer also uses language models that represent not only the correct answers, but also the errors and disfluencies that are common for non-native English speakers for each item. For the Structure, Vocabulary, and Comprehension subscores as well as for part of the Interactions subscore, the system checks for the presence or absence of the correct lexical content in the correct sequence.

The Pronunciation and Fluency subscores, and the immediacy aspect of the Interactions subscore are based on speech timing information and spectral information. In the Verant automated scoring system, words, pauses, syllables, phones, and even some subphonemic events are identified and extracted from the recorded signal for measurement. These base measures are automatically generated and then are recalibrated according to models of expert ratings. By rescaling the machine scores, the system ensures that the Pronunciation and Fluency subscores optimally predict expert judgments.

The information that forms the basis of each subscore is disjoint from the information underlying the other component subscores. That is, each subscore is based on different, independent aspects of the spoken response material.

6. Compliance with ICAO Language Proficiency Requirements

By March 2008, all states must ensure that personnel demonstrate spoken English proficiency as described in Level 4 of the ICAO Rating Scale. Pearson endorses the use of VAET scores for making valid decisions about an aviation professional's ICAO Level in English, especially Levels 4 and 5, 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 VAET themselves or by having trusted third parties administer the test.

The ICAO manual also provides five holistic descriptors that characterize the kinds of linguistic behaviors that proficient speakers should exhibit. Test takers who score well on the VAET (Level 4 or above) should meet these criteria:

First, proficient speakers shall “communicate effectively in voice-only (telephone/radiotelephone) and in face-to-face situations.” In the VAET administration, test takers interact with the Versant testing system over the telephone. This voice-only administration paradigm enhances authenticity by offering a close simulation to radiotelephony communication. In both telephone and radiotelephony interactions, no facial cues or body language can aid communication. “Communications without such cues are considered to be more difficult and challenging, requiring a higher degree of language proficiency, than face-to-face interactions” (ICAO, 2004, Section 2.7.1). Given that speakers need a higher degree of language proficiency for voice-only communication, it follows that test takers whose VAET score is high should also be able to communicate as effectively, if not more effectively, in face-to-face situations.

Second, proficient speakers shall “communicate on common, concrete and work-related topics with accuracy and clarity” (ICAO, 2004, Section 2.7.2). The content of the VAET was designed to address common, concrete, and work-related topics in the aviation domain. The scoring of the VAET is such that both accuracy and clarity are assessed across multiple tasks. For accuracy, the scoring algorithms of the VAET evaluate whether or not the test taker understood the prompt and responded with appropriate content. For clarity, the VAET assesses whether or not the test taker speaks intelligibly and fluently. 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 content, and misapplied syntactic structures can also hinder communication. These different dimensions of spoken communication are measured independently in the VAET.

Third, proficient speakers shall “use appropriate communicate strategies to exchange messages and to recognize and resolve misunderstandings (e.g. to check, confirm, or clarify information) in a general or work-related context” (ICAO, 2004, Section 2.7.3). The tasks in the VAET are designed specifically to engage test takers in checking, confirming and clarifying information. For example, the Readback task requires test takers to listen to a radiotelephony message and readback the information in ICAO phraseology. The task provides a means for test takers to show their understanding of how to confirm information using standard communication protocols. In the Corrections and Confirmations task, the test taker listens to a short dialogue between an air traffic controller and pilot and continues the dialogue. For some of the items, the test taker must resolve a misunderstanding by correcting information that was misinterpreted by one of the speakers. Other items elicit affirmative confirmations to questions, while still others request clarification. The test taker must decide how to respond to the dialogue appropriately and in a timely fashion.

Fourth, proficient speakers shall “handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine work situation or communicative task with which they are otherwise familiar” (ICAO, 2004, Section 2.7.4). To handle the linguistic challenges associated with unexpected situations, test takers must not only have access to a vocabulary that extends beyond memorized phraseology, but they must be able to process language in an automatic fashion. In normal everyday conversation, language processing components that function automatically are extremely efficient. Speakers go from building a clause structure to phonetic encoding, for

example, in 40 milliseconds (Van Turenout, Hagoort, and Brown, 1998). Automaticity is required in order for the speaker to be able to pay attention to what needs to be said rather than to how the encoded message is to be structured. The VAET presents both routine and unexpected items to measure the test taker’s control of core language processing. By presenting items at a conversational pace and measuring response time in addition to accuracy of the content in the response, the VAET measures basic encoding and decoding of oral language as performed in real time. This performance predicts a more general spoken language facility, which is essential in successful communication during a complication or unexpected turn of events.

Finally, proficient speakers shall “use a dialect or accent which is intelligible to the aeronautical community” (ICAO, 2004, Section 2.7.5). The VAET estimates the intelligibility of the test taker’s speech through the Pronunciation subscore. The algorithms for Pronunciation scoring focus on the test taker’s manner of speaking as opposed to the content of the test taker’s responses. The algorithm is designed to predict how members of the aeronautical community would judge the pronunciation of test takers. Thus, if a test taker is unintelligible to many listeners because of a heavy accent or dialect, the test results will reflect this in the Pronunciation subscore.

7. Reliability and Validity

Two indicators of a test’s quality are the test’s reliability and the validity of the interpretations derived from the scores (Alderson, Clapham and Wall, 1995). The statistical data presented in this section provide evidence of the reliability and validity of the VAET.

Responses and test scores from a randomly selected sample of participants were analyzed for the reliability and validity studies. Table 5 presents the demographics of the participants. The mixed group was the non-native sample (92%) plus additional native English speakers (8%) to represent performances at the high end of the scale. None of the responses from the participants in Table 5 were used to create the models for the automatic scoring of the test or to scale the test.

Table 5. Demographics of test takers whose responses and test scores were analyzed in the reliability and validity studies

	Non-Native	Mixed
Number of Participants	129	140
Male : Female	101 (78%) : 28 (22%)	110 (79%) : 30 (21%)
Age Range (mean age)	24 to 70 (mean = 43)	24 to 70 (mean = 44)
Years of Aviation Experience Range	1 to 50	1 to 50
Languages	Afrikaans, Chinese, Dutch, French, German, Italian, Norwegian, Portuguese, Serbian, Spanish, Danish, Hebrew, Japanese, Luxembourgish, Punjabi,	Afrikaans, Chinese, Danish, Dutch, English, French, German, Hebrew, Italian, Japanese, Luxembourgish, Norwegian, Portuguese, Punjabi, Serbian, Spanish,

	Swedish	Swedish
Occupation	60 Air Traffic Controller (46%)	64 Air Traffic Controller (45%)
	58 Pilot (45%)	63 Pilot (46%)
	4 Pilot, Corporate (3%)	5 Pilot, Corporate (4%)
	2 Pilot, Other: Private, Retired (2%)	2 Pilot, Other: Private, Retired (1%)
	5 Other: Air Traffic Instructor, Pilot Instructor (4%)	6 Other: Air Traffic Instructor, Pilot Instructor (4%)

Reliability

The reliability of a test refers to the precision, consistency and stability of its test scores. One estimate of score consistency is the standard error of measurement (SEM). The SEM provides an estimate of the amount of error, due to unreliability, in an individual's observed test score. Luoma (2004) states, "The SEM shows how far it is worth taking the reported score at face value" (p. 183). The SEM of the VAET on the Final score is 3.4.

Another estimate of internal score consistency is reliability. Split-half reliability is calculated from data on a single administration of the test. Scores from equivalent halves of each subtest representing approximately parallel forms are correlated. Bachman (2004) states, "The reliability of a test is partly a function of its length, in terms of items it has, so that long tests can generally be expected to be more reliable than short tests" (p.162). For this very reason, the Spearman-Brown Prophecy Formula, a standard procedure to correct the split-half underestimation, was applied to derive the corrected reliability coefficient.

For the Final score and all subscores on the VAET, the split-half reliabilities were calculated using the data from the mixed sample described in Table 5. The reliabilities were calculated both for the automated scoring system and for the human ratings. The human scores were calculated from human transcriptions and human ratings provided by the ICAO expert raters. Table 6 below displays the reliabilities for the Final score and all subscores on the VAET.

Table 6. Reliability coefficients of the Final score and subscores

Score	Machine Reliability	Human Reliability
Pronunciation	0.97	0.98
Structure	0.82	0.84
Vocabulary	0.89	0.89
Fluency	0.95	0.97
Comprehension	0.88	0.87
Interactions	0.88	0.89
Final	0.93	0.92

Reliability values range from 0 (no consistency) to 1 (perfect consistency). The higher the reliability coefficient, the greater confidence one can place in the consistency and precision of the scores. The high Final reliability scores on the VAET are good indicators that the computerized

assessment will be consistent for the same test taker assuming no changes in the test taker's language proficiency level. As can be seen from Table 6, the Versant automated scoring system can produce scores as reliably and consistently as the human judges.

Correlations Among Subscores

Table 7 presents the correlations among the VAET subscores and the Final score for the Mixed group in Table 5. The subscores correlate with each other to some extent by virtue of presumed general covariance within the test taker population between different component elements of spoken language skills. The correlations are, however, significantly below unity, which indicates that the different subscores do in fact measure different aspects of the test construct.

For example, the Pronunciation subscore correlates highly with the Fluency subscore; however, the Pronunciation subscore does not correlate as highly with the Vocabulary, Comprehension, and Interactions subscores. This is because Pronunciation and Fluency are both related to acoustic characteristics of the test taker's responses, whereas Vocabulary, Comprehension, and Interactions are more related to the content aspect of the test taker's responses. For example, a high correlation between Vocabulary and Interactions seems logical since the Interactions subskill requires speech production that is appropriate and informative in content. Namely, knowledge of vocabulary is required in order to produce appropriate and informative content and it is therefore reasonable to think that test takers who have good vocabulary knowledge may be able to produce appropriate and informative content as measured in Interactions.

Table 7. Correlations among subscores and the Final score (n=140)

	Pronunciation	Structure	Vocabulary	Fluency	Comprehension	Interactions	Final
Pronunciation	--	0.72	0.75	0.90	0.74	0.79	0.87
Structure	0.72	--	0.70	0.69	0.73	0.73	0.86
Vocabulary	0.75	0.70	--	0.67	0.91	0.91	0.90
Fluency	0.90	0.69	0.67	--	0.69	0.73	0.81
Comprehension	0.74	0.73	0.91	0.69	--	0.87	0.88
Interactions	0.79	0.73	0.91	0.73	0.87	--	0.90

Validity

Test validity is the extent to which a test measures what it is intended to measure. The degree of validity depends on the evidence supporting the interpretation of test scores. Three lines of such evidence are presented below: content validity, a demonstration of machine scoring accuracy, and construct validity.

Content Validity

Content validity is the degree to which the test items represent the content that the test is designed to measure. Content-related evidence of validity was provided by both expert judgment and empirical item analysis. As described in the Content Development section above, each item

was reviewed by subject matter experts to ensure content relevancy and conformity to expected usage. In addition, a series of item analyses were performed to statistically analyze data from both native and non-native speakers of English. If an item did not meet certain pre-determined internal criteria, then it was unclear if the item was measuring spoken English or another underlying ability, and the item was thus not included in the test. Other statistical analyses were performed to ensure that each item is effective at discriminating language ability. For more information, see the Content Development section above.

Accuracy of Machine Scores

The VAET is different from most other assessments because of its use of technology to automatically score spoken performances. As evidence of the accuracy of Versant's automated scoring, a study was conducted to compare automatically-generated scores and scores generated from expert human raters and transcribers on the same items in the test. A high correlation between the two scores suggests that the machine scoring is similar to scores produced by experts.

Using the Mixed sample, responses were transcribed by the trained human transcribers and rated independently by the ICAO expert raters for pronunciation and separately for fluency on the ICAO scale. Again, the qualifications and experience of the human transcribers and the ICAO expert raters are provided in the Appendices A and B.

Using human transcriptions and human ratings of the VAET test items, test taker responses were analyzed and scored without automatic speech processing technologies. These human-generated scores were then compared with the machine-generated scores to determine the accuracy of the automatic scoring. Figure 3 shows a scatter plot of the Final machine-generated scores and Final human-generated scores and Table 8 shows the correlation for the Final score and each subscore between the machine-generated scores and human-generated scores.

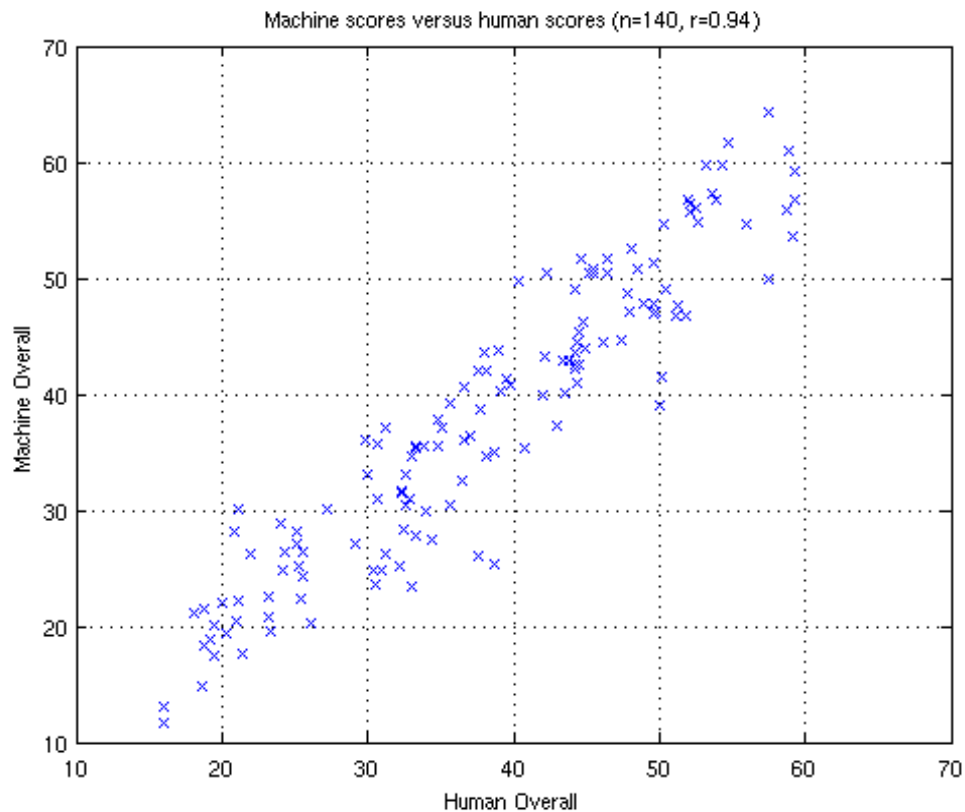


Figure 3. Versant Aviation English Test Overall scores versus human-generated Overall scores (n=140, r=0.94).

The scatter plot in Figure 3 above illustrates the relation between machine-generated and human-generated scores. The data points show a clear linear trend, indicating that for every case evaluated, the scores generated by machine and by trained transcribers and ICAO expert raters are closely aligned. In addition, a very strong correlation was found for each of the subscores. These correlations and scatter plot suggest that the machine-generated scores for the VAET systematically correspond with human-generated scores, providing evidence that the automatically generated scores are virtually the same as those of ICAO expert raters.

Table 8. Correlations between the VAET and human scores

Score	Machine - Human Correlation
Pronunciation	0.89
Structure	0.93
Vocabulary	0.95
Fluency	0.82
Comprehension	0.94
Interactions	0.93

Final	0.94
-------	------

Construct Validity

Construct validity is the extent to which test scores can be interpreted as a measure of the intended construct. The construct of the VAET test is facility with spoken English in the aviation domain. Two lines of evidence for construct validity are (1) strong correlations with other measures of spoken aviation English, and (2) the classification differences between native versus non-native English-speaking aviation professionals

Correlation of Scores from the VAET and Human ICAO Ratings

Evidence for construct validity is a high correlation between the automatically generated scores of the VAET and other measures of spoken aviation English on the ICAO scale. To gather ICAO ratings, six ICAO expert raters provided estimates of test takers' ICAO levels by analyzing the 30-second response samples on Story Retelling and Open Questions from the test takers in the Mixed sample described in Table 5.

Five responses were rated independently for each test taker, and each response was rated by at least two different raters independently, producing a total of at least 10 independent ratings per test taker. ICAO Estimate ratings were analyzed using a multi-faceted, one-parameter Rasch model as implemented in the computer program FACETS (Linacre, 2003). The analysis assigned a single ICAO score (represented on a continuous Logit scale) for each test taker. This ICAO Estimate was then correlated with the automatically generated VAET scores for the same set of test takers.

The correlation between the human ICAO Estimates and VAET automatically-generated Final scores was 0.91. The scatter plot is presented in Figure 4.

The VAET scores correlate very highly with the ICAO Estimates, suggesting that both the VAET test and the ICAO Estimates are measuring spoken aviation English ability.

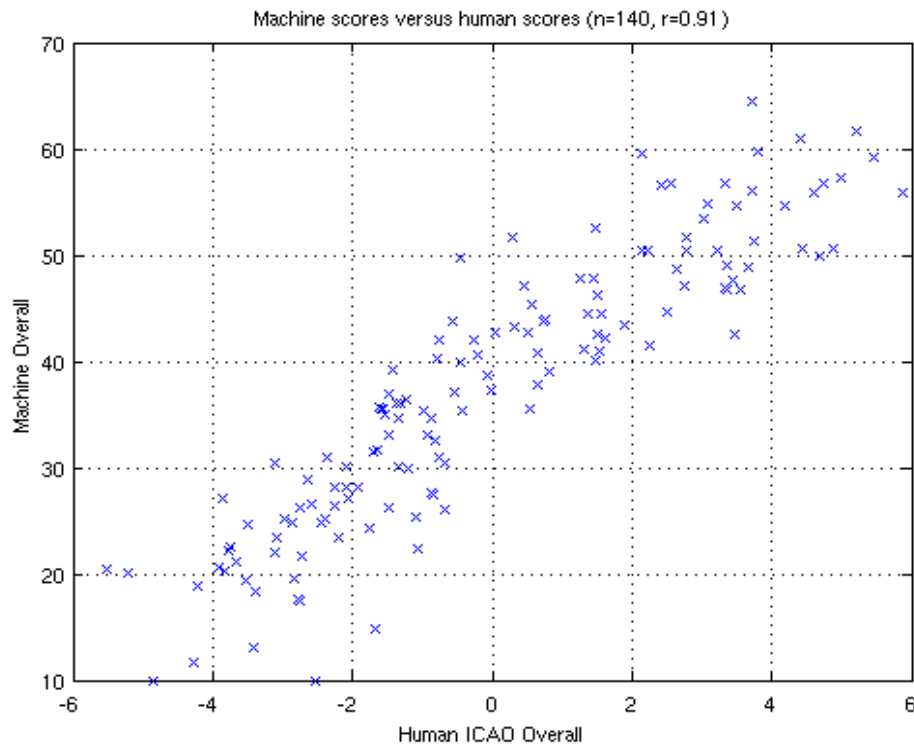


Figure 4. ICAO Estimates on 30-second response samples as a function of VAET Final scores.

Classification Differences between Native and Non-Native Speakers

Because the construct of the VAET test is an amalgam of proficiency in English and facility with ICAO phraseology, it is reasonable to expect that most aviation professionals who are native speakers of English will do relatively well on the VAET test, while professionals who are non-native speakers may vary more widely according to their English proficiency level. In order to determine whether this is the case, an analysis was performed to identify the percentages of native and non-native speakers who were classified at each ICAO level.

Figure 5 presents the performance of the same sample of 478 native-English and 625 non-native-English aviation professionals that was recruited during the field-testing as summarized in Table 4. The red-squares in Figure 5 indicate the percentage of non-native aviation professionals classified at each ICAO level and the green-circles represent the corresponding percentage of native aviation professionals at each level. It can be seen that more than half of the non-native sample (approximately 55%) fall into either Level 3 or Level 4.

In the native-speaker sample, approximately 70 % of test takers were classified as either Level 5 or Level 6, which suggests that the distributions of ICAO levels are substantially different in the native sample and non-native sample. The data also indicate that the non-native speakers were more widely distributed from Level 1 to Level 6 than the native sample. Only a few native speakers were classified as Level 2 or Level 3.

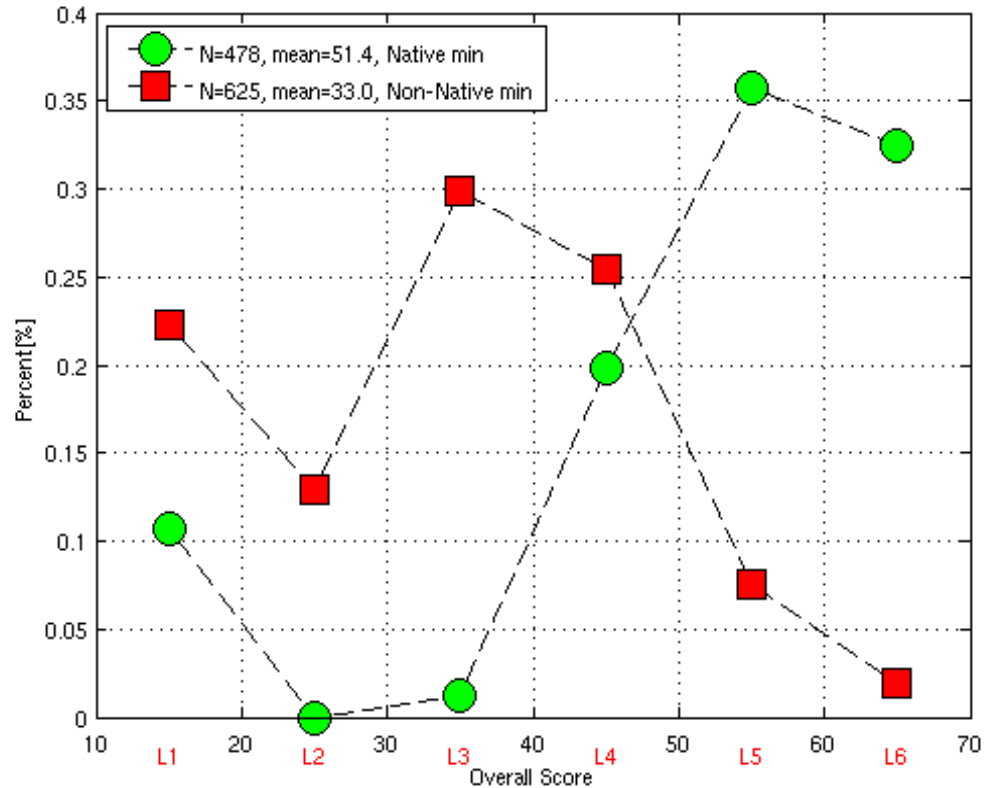


Figure 5. Classification percentages for native and non-native speakers on the Final ICAO level. (Natives n=478, and Non-Natives n=625)

Note that approximately 11% of the test takers in the native sample received Level 1. Review of the sample demographic information suggests that a number of the Level 1 native-English test takers are either retired/inactive pilots or are private pilots with no ICAO experience. Analysis of a sample of responses from the Level 1 native-English speakers found that some of these test takers did not follow the test instructions correctly and simply answered “roger” or remained silent in response to implicit or explicit readback requests.

8. Conclusions

The Versant Aviation English Test is a proficiency test that is designed to directly measure listening and speaking abilities of non-native English-speaking pilots and air traffic controllers. The VAET tasks are intended to measure how well the test taker can understand spoken English on work-related topics and to respond intelligibly in spoken English at a fully-functional pace. The test is delivered automatically over the telephone or via computer and takes about 25 minutes to complete. During the test, the Versant testing system presents a series of spoken prompts in English from a variety of speakers and elicits oral responses in English.

A few minutes after the VAET test has been completed, the Versant system has completed the analyses of the test taker’s responses and it posts scores to www.VersantTest.com. The score

report consists of a Final score and the six subscores from the ICAO Language Proficiency Rating Scale: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions.

For any high-stakes tests, it is crucial that the test be reliable. The reliability of the VAET Final score is 0.93. The high reliability suggests that test scores are internally consistent. Evidence of validity is provided from many sources including expert review of the items, criteria for selecting only items that discriminate abilities among non-native speakers of English, clear separation of performance of native and non-native English-speaking aviation professionals, and strong correlations with scores generated from human experts.

In sum, the available evidence indicates that the VAET complies with the ICAO standards and is both a reliable and valid measure of test takers' facility with spoken aviation English.

9. Pearson: Testing and Technology

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

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

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

Research at Pearson: In close cooperation with international experts, Pearson conducts ongoing research aimed at gathering substantial evidence for the validity, reliability, and practicality of its current products and at investigating new applications for Ordinate technology. Research results are published in international journals.

10. References

- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.
- Caplan, D. & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-126.
- 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.
- International Civil Aviation Organization, (2004). *Manual on the implementation of the ICAO language proficiency requirements, First Edition, Document 9835*.
- 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.
- 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.
- Standards for Educational and Psychological Testing*. (1999). Washington, DC: American Psychological Association.
- Young, S., D. Kershaw, J. Odell, D. Ollason, V. Valtchev and P. Woodland. 2000. *The HTK Book Version 3.0*. Cambridge, England: Cambridge University.
- Van Turenhout, M. Hagoort, P., & Brown, C.M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280, 572-574.

Appendix A

Occupation and experience of ICAO expert raters used in the development process and in the validation experiments

Occupation	Experience
Program manager for aviation language training	25 years of experience in human resource development and language training
Consultant for ICAO English language competency project	30+ years of experience as Air Traffic Control Officer in the UK and Australia ICAO PRICESG (Proficiency Requirements in Common English) participant
Policy, plans, and program officer for government English language center	30+ years of experience in training and developing curriculum for aviation English language training ICAO PRICESG (Proficiency Requirements in Common English) participant
English Consultant for Aeronautical Radio of Thailand Ltd.	10 years of experience teaching English and serving as English consultant with 7 years focused on aviation English teaching
Air Traffic Control Instructor and Aviation English Specialist in Costa Rica	10 + years of experience as Air Traffic Controller (10 years as a supervisor) 5+ years of experience as Air Traffic Control Instructor Started aviation English program for countries in Central America with Central American Safety Agency
Aviation English Consultant in Spain	20 years' experience in Aviation English training, materials development and assessment TESOL qualified and experienced
Aviation English Consultant in US	5+ years of experience in consultation, material development, teaching for aviation English in a few different countries (e.g. China and Mongolia) TESL and TEFL certificates
Aviation English Instructor in US	5+ years of teaching English 2+ years teaching aviation English at a flight school in the US MA in Linguistics; TESOL certificate
University professor in US	PhD in TESL/Psycholinguistics Oral proficiency assessment specialist Many years of teaching, assessment, and curriculum development in aviation English
Captain for commercial airline Aviation English Consultant in Italy	20+ years of experience as commercial airline pilot (10+ years as captain), flight instructor and examiner 7 years of experience as military pilot (fighter), specialized in Human Factors, 3+ years experience as Quality and Safety Manager 2+ years of experience as aviation English consultant in accordance with ICAO standards Certificate of Aviation English Technical Teacher and Rater
Aviation English Consultant in Italy	10+ years of experience in teaching, assessment, material development for aviation English in accordance with ICAO LPR Standards TEFL Certificate and Certificate in Aviation Technical English Teaching and Rating

Appendix B

Qualifications of human transcribers used in the development process and in the validation experiments

Qualification
Ph.D. in English from Oklahoma State University with a concentration in Linguistics
Ph.D. student in Linguistics at Indiana University, Bloomington
Ph.D. student in TESL at Oklahoma State University
Ph.D. student in African Linguistics at Indiana University, Bloomington
Ph.D. student in French Linguistics at Indiana University, Bloomington
MA in TESOL from the University of Pennsylvania
MA in Linguistics from San Jose State University with TESOL and Computational Linguistics certificates
MA in Linguistics from San Jose State University with a Computational Linguistics certificate
MA in Linguistics from the University of South Carolina
MA in Reading from the University of Virginia
MA student in Linguistics at Indiana University, Bloomington
MA student in Linguistics at San Jose State University
MA student in Linguistics from San Jose State University
MA student in Hispanic Linguistics at Indiana University, Bloomington
MA student in TESOL at the School for International Training
MA student in Spanish Translation at the Monterey Institute of International Studies
MA student in Spanish Translation at the Monterey Institute of International Studies
BA in Linguistics from San Jose State University
BA in Literature from the University of California, Santa Cruz
BA in Psychology from San Jose State University

Appendix C

ICAO Language Proficiency Language Rating Scale (Levels 4-6)

ICAO Language Proficiency Rating Scale: Expert, Extended and Operational Levels.						
LEVEL	PRONUNCIATION <i>Assumes a dialect and/or accent intelligible to the aeronautical community.</i>	STRUCTURE <i>Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.</i>	VOCABULARY	FLUENCY	COMPREHENSION	INTERACTIONS
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, rarely interfere with ease of understanding.	Both grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work-related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work-related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and /or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Operational Level 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work-related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete and work-related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstanding by checking, confirming, or clarifying.
<i>Level 1, 2, and 3 are on subsequent pages.</i>						

ICAO Language Proficiency Language Rating Scale (Levels 1-3)

ICAO Language Proficiency Rating Scale: Pre-operational, Elementary and Pre-elementary Levels.

LEVEL	PRONUNCIATION <i>Assumes a dialect and/or accent intelligible on the aeronautical community.</i>	STRUCTURE <i>Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.</i>	VOCABULARY	FLUENCY	COMPREHENSION	INTERACTIONS
<i>Levels 4, 5 and 6 are on preceding page.</i>						
Pre-operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language of regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work-related topics, but range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pausing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work-related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistics or situational complication or an unexpected turn of events.	Responses are sometimes immediate, appropriate, and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language of regional variation and frequently interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolate words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow and often inappropriate. Interaction is limited to simple routine exchange.
Pre-elementary 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.
No evidence 0	Insufficient evidence to decide on score category: Silence, irrelevant, or unintelligible (speech in other language than English) material.					

Note: The Operational Level (Level 4) is the minimum required proficiency level for radiotelephony communication. Levels 1 through 3 describe Pre-elementary, Elementary, and Pre-operational levels of language proficiency, respectively, all of which describe a level of proficiency below the ICAO language proficiency requirement. Levels 5 and 6 describe Extended and Expert levels, at levels of proficiency more advanced than the minimum required Standard. As a whole, the scale will serve as benchmarks for training and testing, and in assisting candidates to attain the ICAO Operational Level (Level 4).

Appendix D

Sample Test Paper: The Versant Aviation English Demo Test



Versant Aviation English Demo Test

Call: 1-800-335-6393 or +1.415.738.3800

Test Identification Number

1234 5678

Introduction:

Thank you for calling the Ordinate testing system.

Please enter your Test Identification Number on the telephone keypad.

Now, please say the city and country you are calling from.

This is the Versant for Aviation English Demo Test. This test has 8 parts.

You will be asked several questions in each part. Now, the test will begin.

Please follow the instructions for Parts A through H.

Part A: Aviation Reading.

Please read the following radiotelephony messages when you are asked.

1. World Air 891, request descent.
2. World Air 891, maintain flight level 280 expect descent shortly.
3. Maintaining flight level 280, World Air 891.
4. World Air 891, descend flight level 120.

Part B: Common English Reading.

Please read the following English sentences when you are asked.

1. Our flight didn't start very well.
2. First there were some minor mechanical issues that needed to be resolved.
3. Then all outgoing flights were delayed because of the weather conditions.
4. When we finally took off hours later, it was a smooth flight all the way.

Part C: Repeat.

Listen to each sentence, then repeat it.

Example: when you hear: "My next flight is on Saturday."
you say: "My next flight is on Saturday."

Part D: Short Answer Questions.

Please give a simple answer to the question.

Example: when you hear: "Where in the airplane do the pilots control the aircraft?"
you say: "cockpit" or "in the cockpit".

**Part E: Readback.**

You will hear several spoken radiotelephony messages. Say an appropriate readback after each one. For each question, a call sign is printed on the test paper for your information. Use this call sign to say an appropriate readback.

Example: Cessna 29

When you hear: "Cessna 29, exit taxiway Hotel."

one possible answer is: "Exit taxiway Hotel, Cessna 29." OR

another possible answer is: "Exiting taxiway Hotel, Cessna 29."

1. Coastal Air 315
2. World Air 395
3. Coastal Air 405
4. Global Air 295
5. World Air 135
6. GABCJ

**Part F: Corrections and Confirmations.**

In this task, you will hear a radiotelephony exchange between two speakers, Speaker 1 and Speaker 2. The speakers are a pilot and an air traffic controller, but not necessarily in that order. Speaker 1 speaks first, then Speaker 2 responds next. After you hear the exchange, imagine that you are Speaker 1. Your task is to continue the dialog and reply to Speaker 2's response. If Speaker 2's response includes wrong information, correct that information. If Speaker 2's response includes a question or request, respond appropriately. For each question, a call sign is printed on the test paper.

Example: East Global Air 295

(Speaker 1) "East Global Air 295, contact Atlanta Radar 122.15."

(Speaker 2) "Atlanta Radar 121.5, East Global Air 295."

After you hear this exchange,

you could say: "East Global Air 295, negative, contact Radar 122.15." OR

another possible answer is: "East Global Air 295, I say again, 122.15."

1. Charlie Romeo 4013
2. GABCJ
3. Coastal Wing 927
4. Coast Wing 75
5. World Air 295
6. Coastal Airline 445

Part G: Story Retelling.

Now, you will hear three stories. After each story, you will have 30 seconds to retell it. Each story will be spoken only once. When you hear this beep, you will have 30 seconds to retell it. Try to retell as much of the story as you can. You will hear another beep at the end of the 30 seconds.

Part H: Open Questions.

You will hear two questions. You will have 30 seconds to answer each question. Each question will be spoken twice. When you hear a beep, you will have 30 seconds to answer. You will hear another beep at the end of the 30 seconds.

Appendix E

Test Instructions: The Versant Aviation English Demo Test



Versant Aviation English Demo Test Instructions

The Versant Aviation English Demo Test is delivered over the telephone using the Ordinate® testing system. It is designed to test your ability to understand and use ICAO phraseologies as well as common English. The test is intended to measure the six subskills in the ICAO language proficiency scale (Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions).

Procedure: First, take time to read the whole test paper. If there are words or sentences that you don't understand, you may use a dictionary or ask someone for help. When you are ready to begin the test, use an appropriate telephone to call the telephone number printed on the test paper. When asked, you will enter the Test Identification Number using the buttons on your telephone keypad. You will take the test on your own by following the instructions given over the telephone. Relax, concentrate, and do your best. If you do not know how to respond to a test item, just be silent or say "*I don't know.*"

Test Sections: The test has eight sections (Part A, B, C, D, E, F, G and H) as follows:

Part A: AVIATION READING

You will be asked to read two aviation specific messages in a random order from among those printed in Part A. Read the messages out loud smoothly and naturally.

Part B: COMMON ENGLISH READING

You will be asked to read two sentences in a random order from among those printed in Part B. Read the messages out loud smoothly and naturally.

Part C: REPEAT

You will hear several sentences in this section, one at a time. Repeat each sentence you hear – exactly as you hear it. If you can't repeat the whole sentence, repeat as much of the sentence as you can.

Example: When you hear: "My next flight is on Saturday."
You say: "My next flight is on Saturday."

Part D: SHORT ANSWER QUESTION

In this section, you will hear several short questions. Answer each question with a single word or a short phrase of two or three words.

Example: When you hear: "Where in the airplane do the pilots control the aircraft?"
You say: "cockpit" or "in the cockpit"

Part E: READBACK

You will hear several radio messages in this section, one at a time. The radio messages use ICAO phraseologies. When you hear a message, say an appropriate readback using ICAO phraseologies.

Example: Cessna 29 (The call sign is printed on the test paper.)
When you hear: "Cessna 29, exit taxiway Hotel."
You say: "Exit taxiway Hotel, Cessna 29."
Or: "Exiting taxiway Hotel Cessna 29."





Part F: CORRECTIONS and CONFIRMATIONS

You will hear several radio communication exchanges between two speakers, Speaker 1 and Speaker 2. These speakers are a pilot and an air traffic controller. You need to respond to Speaker 2's response as if you were Speaker 1. Speaker 2's response may contain wrong information, a question, or a request. If the response contains wrong information, correct that information. If the response includes a question or request, respond appropriately. Depending on the situation, you may answer the questions using only ICAO phraseologies or a combination of ICAO phraseologies and common English.

Example: East Global Air 295 (The call sign is printed on the test paper.)

When you hear: (Speaker 1) "East Global Air 295, contact Atlanta Radar 122.15."

(Speaker 2) "Atlanta Radar 122.5, East Global Air 295."

You say: "East Global Air, negative, contact Radar 122.15."

Or: "East Global Air 295, I say again, 122.15."

Part G: STORY-RETELLING

You will hear three brief stories. After each story, retell it as best you can in your own words in English. Each story will be read only once. After you hear a beep, you will have 30 seconds to retell the story.

Part H: OPEN QUESTIONS

You will hear two questions. After you hear a beep, speak your opinion as fully and clearly as you can. You will have 30 seconds to answer. Express your opinion and supporting reasons in clear, coherent English. Speak for the whole 30 second period.

Demographic Questions: After Part H, you will hear "*This is the end of the Versant for Aviation English Test. Thank you for calling the Ordinate Testing System.*" After this message, you will be asked a few questions about yourself. We do not ask for your name or phone number. Please also answer these questions. Your answers to these questions will NOT affect your test results.

When you hear: "*Thank you for calling the Ordinate Testing System.*" again, the test is complete; you may hang up.

How the test is scored: The Versant Aviation English Test scores are based on the exact words that you speak, as well as the pace, fluency, and pronunciation of those words as combined in phrases and sentences. Give quick, smooth, loud responses. Note that some test items have more than one correct answer.

Suggestions: Place your call to the Ordinate testing system on a good telephone in a suitable location. Choose a location that is quiet and where you will not be interrupted. Hold the phone as shown in the figure below and speak in a loud, steady voice. Use a land-line, push-button telephone in good working order that is set to "tone" (not "pulse"). Newer phones are generally better than older phones. Do not use a cordless, cellular phone, or VOIP phone.



NO
too low, too far away



YES
in front of mouth



YES
a good distance

