

Running Head: A REVIEW OF PHONE PASS

Has PhonePass® Passed the Review?

Eve Ribeiro

Tony Becker

Georgia State University

Second Language Assessment and Evaluation

Dr Sara Weigle

## Introduction

Phone Pass is an automated language proficiency test devised by Ordinate Corporation that takes place over the phone. It offers three kinds of test: SET 10, SET 5 and Junior SET. It assesses the ability of non-native English speakers to understand spoken English and to respond to the utterances.

The aim of this paper is to describe the test and provide an assessment of Phone Pass. Since limited online information was provided on SET 5 and Junior SET, we will concentrate on SET 10.

### *Primary Uses*

Phone Pass can be used for academic purposes (i.e. entrance exams, placement, exit, or benchmarking purposes, qualify international teaching assistants, certification) or employment purposes (i.e. evaluate employees for promotions, international duties, recruitment).

### *Administration*

The test administrator provides the test-taker with a paper or electronic (PDF) sample. The first part provides written, general introduction and test procedures and the second part requires an identification number that must be entered on the telephone keypad at the beginning of the test. Once the test is completed, the test-taker simply hangs up. SET-10 lasts approximately 10 minutes, the SET-5 and Junior SET for approximately 5 minutes.

### *Registration and Pricing*

The registration fee varies if the test is taken within the U.S.A. (\$40) or abroad (price varies). Ordinate can be contacted via email, fax or telephone by an individual, a company or an institution. Discount is available for large groups and the minimum quantity for orders is 20.

### *Score Reports*

Test-takers and administrators can view their final score online an hour after test completion and print it for their own records. The service contractor also receives a report, which is calculated as a weighted average of the *five* skills:

- Listening Vocabulary – understanding spoken words
- Repeat Accuracy – repeating utterances verbatim
- Pronunciation – consonants, vowels, and word-stress
- Reading Fluency – rhythm, phrasing/timing in reading aloud
- Repeat Fluency – rhythmic phrasing in repeating sentences

The Phone Pass test reports five general sub-scores on a scale from 2 through 8 (lower numerical value = lower test results): Listening Vocabulary, Repeat Accuracy, Pronunciation, Reading Fluency and Repeat Fluency.

All spoken responses are collected using a special-purpose speech processing system. This system performs the recognition, alignment, and scoring of the spoken responses. The machine interprets Phone Pass test scores by combining component measures that operate by two techniques: analysis of correct/incorrect responses and numerical output from the speech processing system.

### General Description

Phone Pass is a test of English language proficiency for general language ability. The test is comprised of 61 items, divided into five parts (located in Appendix). The automated recordings are presented in three distinct North American dialects with a natural pace and intonation. Both male and female voices are very audible and clear.

In *Part A*, of the demonstration test, examinees are supposed to read aloud the numbered statements directly from the test form. There are three different contexts that are not thematically linked. During the test, examinees are instructed by the computer to read the sentences at random. The sentences are relatively simple in structure and vocabulary. This part collects a sample of pronunciation and reading fluency.

In *Part B*, examinees repeat, verbatim, increasingly long sentences without the support of the written form. The examinee needs to recognize the words in the stream of continuous speech, retain them in their short-term memory and reproduce them.

In *Part C*, examinees listen to a spoken question and respond with a single word or a short phrase. The questions are not content specific. They ask for basic information or inferences requiring logic, time, sequence or lexical content knowledge. In order to respond to the question, the examinee needs to recognize the words in phonological and syntactic context and make inferences on the proposition.

In *Part D*, examinees are presented with a sequence broken into three jumbled phrasal word groups. They are supposed to reorder the increasingly long chunks into a syntactically correct sentence. The task requires local syntactic skill and receptive lexical processing.

Finally, *Part E* challenges examinees to speak on a given topic for 20 seconds, uninterrupted, for collecting spontaneous speech. Some of the topics, for instance, in the SET-10 demo test were: “Given the choice, would you prefer to live in the city or in the countryside? Why?”, and “If your family was given a fairly large amount of money, what do you think should be done with it?”

Table goes about here

### Strengths and Weaknesses

#### *Validity*

The PhonePass test is one of the first automated, speech assessment tests delivered over the telephone, and Bernstein (1998) provides an extensive report of the SET-10 tests’ validity and reliability, located on the website [<http://www.ordinate.com>]. Validation studies for the SET-10 were conducted using a norm group of 4,000 native English speakers, from the U.S. and U.K., as well as 3,500 non-native speakers of English from Asia, Europe, and South America. Such sample-group sizes may initially appear adequate, but one should take into account the considerable high-stakes of the tests’ uses, as well as the large number of people who apply for employment, teaching assistantships, etc. each year in the U.S. and U.K. After you consider the millions of people included in this tabulation, the sample size seems relatively small, and possibly inadequate, for such important determinations.

Although many critics agree about the great lengths to which Ordinate has gone through to research the validity of their test, one particular reviewer (Fulcher 2000) argues, “The problem with tests like PhonePass is that it relies for its validity on the correlation with direct measures such as the Oral Proficiency Interview (OPI)” (p.97). Ordinate has strong concurrent validity for

the SET-10 test with other well-established language examinations, but this might not be enough. The major concern appears to be that PhonePass draws correlations between its test and other listening and writing tests, and experts warn that these correlations are possibly irrelevant since oral proficiency appears more closely related to reading skills (Frase, 2000). In defense of this argument, Ordinate provides a strong correlation for PhonePass with TSE (Test of Spoken English), and perhaps Ordinate should place more emphasis on correlating its results with measures from other types of speaking, and oral proficiency tests. However, this can only become more likely when the use of computer adapted tests become more commonplace, and such suggestions will be more relevant once other CAT's, like the TOEFL Speaking section, are finally introduced.

Although not as convincing, yet another argument against the validity of PhonePass is found in the definition of "fluency," in which Fulcher (1996) finds:

A discrepancy between Bernstein and that within the applied linguistics field. Bernstein (1998) simply refers to fluency as rhythm, phrasing and pausing, which is an understatement, when compared to the complexity given language fluency within applied linguistics. (p.98)

A universal agreement needs to be forged here, in order for the SET-10 to gain greater credence among linguists and applied linguists, but this cannot be done unless fluency is properly defined by all parties involved.

Perhaps there is no greater threat to the validity of the test than its predictive validity. This point is critical for Ordinate, but at the same time difficult to assess, since PhonePass is still relatively new. Still, Ordinate might be best suited to provide some longitudinal evidence of how performance on this test might predict performance in schools, workplace, etc. If not, it seems

somewhat difficult to prove that someone who performs well on PhonePass will perform equally as well in a proposed setting. Also, Ordinate might consider offering some type of predictive value for scores on the SET-10 Demo Test, and how they relate to scores on the official SET-10 test. Perhaps, as this test gains more exposure and further feedback it will give greater knowledge of the test's predictive power, as well as continue to set the example for computer adaptive testing.

Overall, PhonePass provides strong evidence for ascertaining validation of its scoring method, and is only hampered slightly by the issues discussed above. Ordinate has obviously done extensive testing and has recently refined its diagnostic subscoring, proving that PhonePass will continue to improve the test and retain its excellent validity measures.

### *Reliability*

Ordinate provides extensive information for their reliability measures for scoring, including how the machine SET-10 scores compare to the human-generated scoring, .97 and .98 overall. However, the table for this figure indicates that only one human grader is compared to one machine score for the reliability analysis, and this seems insufficient for any credible conclusions regarding this topic. It would be more effective to report the reliability of several human and machine score-types, rather than just one. In order to add further strength to their reliability measures, Ordinate also provides the correlation value,  $r = .97$  overall, for the relationship between the generated machine and human scores for 288 non-native speakers of English. This figure indicates that non-native English speakers' performance correlates very closely from human scoring to machine scoring. This information provides convincing evidence that the machine scoring used for the SET-10 test is consistent and ultimately indicates reliable information.

*Fairness*

Care was taken to ensure that test takers who are not familiar with phone-automated systems are not unfairly disadvantaged when taking PhonePass. First, PhonePass directions are clearly explained by the automated system, and there are plenty of chances to practice with the system. Ordinate offers a free service in which test-takers can call and practice with PhonePass, and they can even receive feedback within one hour of calling. Second, PhonePass also provides written instructions for test-takers, and these are downloadable from the website. In addition, open-ended responses are not scored, but rather, they can be accessible for human review by authorized listeners. This ensures that test-takers' responses are not included in overall scoring, and also eliminates the danger of biasness against a test-taker's preferences or opinions.

*Administration*

The SET-10 test is administered over the phone and requires little time to complete. The test takes approximately 15 minutes to complete, and is convenient for companies who want to screen people in larger numbers. An even greater advantage for PhonePass test-takers and administrators is that the test can be taken from any phone, at any time, and scores can be received within one hour of test completion. Therefore, test-takers can complete the test during times which best suit their schedule and needs, and they can obtain results within a very reasonable time. Such advantages help to decrease anxiety for test-takers, and also help alleviate costs, paperwork, or other administrative drawbacks for employers, universities, etc.

*Summary*

To summarize, the Ordinate PhonePass is a well-supported test: test-takers and test-users can easily find out a great deal about it, and the test is available in most parts of the world.

Furthermore, the test is relatively quick to administer and scoring is done via computer and accessible within hours of test completion. Ordinate's own evidence shows that the test is reliable and there do not appear to be very many weaknesses in this area. However, there are problems relating to validity and construct definitions. Most of these concerns are within reason, and there is realistic room for improvement about the facts that:

- the validity measures rely too heavily upon correlations to other language tests;
- there isn't agreement between Ordinate and applied linguists on the definition of "fluency";
- the test fails to provide any predictive values for how test takers perform on the demo test and the actual test;
- the test also fails to provide any longitudinal predictions for the performance of those who have taken the SET-10;

In conclusion, with time and further efforts like Ordinate's, to make computer adaptive testing more accessible and efficient, tests like these will lend a great deal of support to language assessment. As the evolution of computerized language assessment continues to take place, so too does the realization of more effective and cost-effective methods of evaluating language performance worldwide.

## References

Bernstein, J. (1998). Preliminary report: Comparison of language proficiency interview (LPI), PhonePass Spoken English Test (SET-10), and the Test of English for International Communication (TOEIC). October 27, 1998, 11 pages. Ordinate Corporation. Menlo Park, California.

Frase, L.T. (2000). Adapting to change in the new millennium. T.H.E. (Technological Horizons in Education), 27, 6, 47.

Fulcher, G. (2000). [Computers in Language Testing](#). In Brett, P. and G. Motteram (Eds) 2000, A Special Interest in Computers. Manchester: IATEFL Publications, 93 – 107. Reprinted in electronic format with the kind permission of IATEFL.

Ordinate (2000). SET-10 Test description and validation summary. Merlo Park, CA: Author.

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

## Appendix

The five components for the SET-10 test.

Part A	Read sentences aloud	8 items
Part B	Repeat sentences	16 items
Part C	Short-answer questions	24 items
Part D	Sentence-build	10 items
Part E	Respond to open questions	3 items