

Versant English Test Benchmarking Kit Analysis

NCSC

November 2008



Introduction

This report presents and explains the results of the Benchmarking Kit grading and methodology for determining Versant English Test cutoff scores. The cutoff score is based on requirements for English skills as determined by eight (8) NCSC graders familiar with the levels of spoken English skills required for satisfactory performance of required tasks.

Grader Statistics

Eight graders rated 114 responses uttered by 57 non-native speakers of English. These included native speakers of different languages at various levels of English proficiency. Grader 3 (PIN 71426282) and Grader 5 (PIN 99141132) were excluded from the analysis because they did not complete the rating process.

Table 1 shows grader statistics for each grader.

Table 1. Grader Statistics

Grader	Name	Benchmarking PIN	Mean μ	Number of grades
1	NCSC 2008 Grader 1	5923 1772	4.11	114
2	NCSC 2008 Grader 2	5872 5984	2.05	114
3	NCSC 2008 Grader 4	7383 3770	2.32	114
4	NCSC 2008 Grader 6	8142 9507	2.63	114
5	NCSC 2008 Grader 7	9000 5210	3.79	114
6	NCSC 2008 Grader 8	9502 6796	3.05	114
7	NCSC 2008 Grader 9	4623 3833	4.05	114
8	NCSC 2008 Grader 10	2598 7355	4.37	114

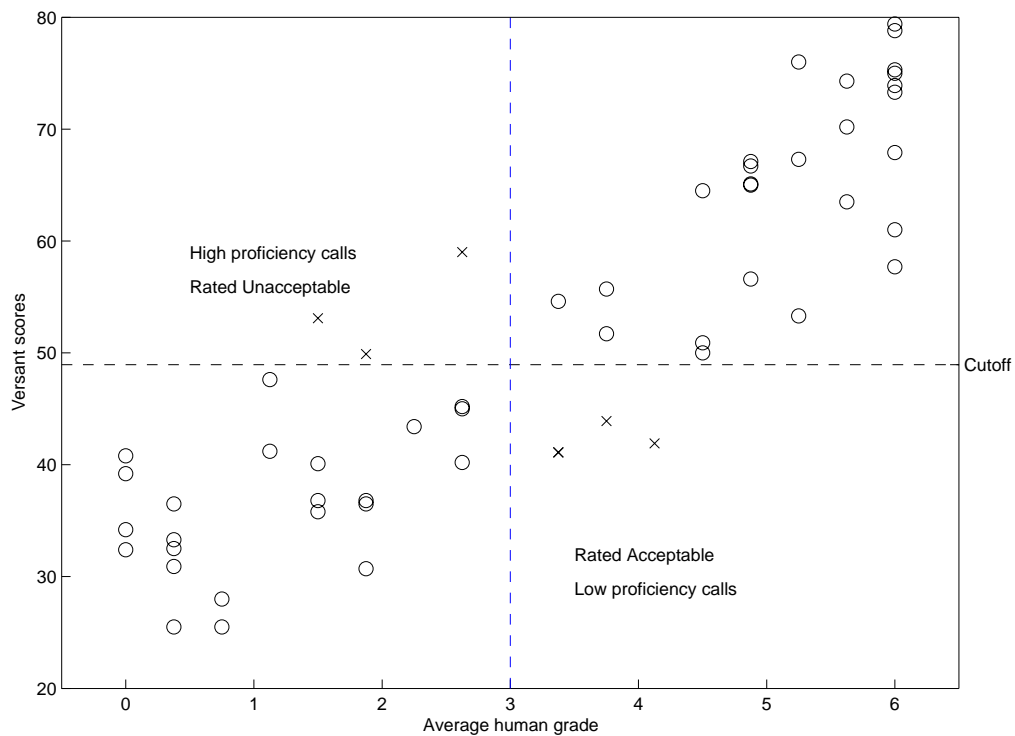
An average human grade (AHG) per examinee was calculated, and the examinee set was divided into two subsets, *unacceptable-ahg* (examinees judged as unacceptable by human graders) and *acceptable-ahg* (examinees judged as acceptable by human graders). All examinees with an AHG below 3 were assigned to the *unacceptable-ahg* set, and ones with an AHG above 3 were assigned to an *acceptable-ahg* set. An AHG of 3 was chosen as a threshold, since it is the mean value of the acceptable (6) and unacceptable (0) grade. Examinees with an AHG equal to 3 were excluded from the analysis because their ability, determined by an AHG, was exactly at the threshold value and it was not clear whether they were acceptable or not.

Methodology for Determining an Optimal Cutoff Threshold

Given the classification of the examinees based on an AHG, a minimum error classifier was used to determine a Versant English Test score for the desired threshold.

Figure 1 shows results of the analysis. The horizontal dashed line at 48.9 is an optimal cutoff score. If one used this threshold, all examinees who scored below the threshold would be judged as unacceptable, whereas examinees with scores above the threshold would be judged as acceptable. Data points marked with an X correspond to examinees whose results appear to not conform to the classification methodology.

Figure 1. AHG vs. Versant English Test Overall, N=57.



For the purposes of this report, the lower bound for the threshold is defined such that 90% of examinees determined to be acceptable by an AHG would also be judged as acceptable by a Versant English Test. Similarly, the upper bound for the threshold is defined so that 90% of all examinees judged as unacceptable by an AHG would also be judged as unacceptable by a Versant English Test.

Conclusion

In this report we recommend a methodology for determining an optimal Versant English cutoff score. If this method is to be used, an optimal threshold based on judgments from the eight NCSC raters would be 48.9; thus, a score of 49 is acceptable and a score of 48 is not acceptable. The 48.9 threshold maximizes the total number of correctly classified examinees. This threshold is optimal to minimize false rejections, i.e. it does not have any false rejection.

If it is more important to ensure that acceptable candidates are not incorrectly rejected, then the lower threshold should be used (accept scores of 43 or greater). Similarly, if it is more important to ensure that unacceptable candidates are not accepted, then the upper threshold should be used (only accept scores of 55 or greater). Table 2 summarizes threshold score values.

Table 2. Summary of threshold scores

Optimal threshold	48.9
Lower threshold (minimize false rejections)	Accept 43 or greater
Upper threshold (minimize false acceptances)	Accept 55 or greater

Disclaimer

The Versant Benchmarking Kit for Workforce Development is designed for use in commercial settings where facility in English as a Second Language is an important qualification.

If you will be using Versant English tests in a commercial setting subject to Equal Employment Opportunity Uniform Guidelines, you should familiarize yourself with EEOC Compliant assessment and selection procedures. The following resources may be useful as a general guide for testing principles, job analysis, validation and legal compliance issues:

- *Standards for Educational and Psychological Testing*. AERA, APA and NCME. Washington, D.C.: American Psychological Association, 1985
- *Principles for the Validation and Use of Personnel Selection Procedures, 3rd Edition*. College Park, MD: Psychology, 1987
- *Code of Fair Testing Practices in Education*. Prepared by the Joint Committee on Testing Practices. Washington, D.C.: American Psychological Association, 1988

You may also want to review norming and validation data for Versant tests in the Pearson white paper: "Validity of the PhonePass Test in Employee Selection."