

Ordinate® SET Benchmarking Kit Analysis

New Jersey Court

March 2005



Introduction

This report presents and explains the results of the Benchmarking Kit grading and methodology for determining SET cutoff scores. The cutoff score is based on requirements for English skills as determined by one or more graders familiar with the levels of spoken English skills required for satisfactory performance of required tasks.

Grader Statistics

Four graders graded 116 responses uttered by 58 non-native speakers of English. These included native speakers of 9 different languages at various levels of English proficiency.

Table 1 shows grader statistics for each grader.

Table 1. Grader Statistics

Grader	Name	Benchmarking PIN	Mean μ	Standard deviation std	Number of grades
1	Grader 1	7867 1668	2.69	3.00	116
2	Grader 2	4765 9882	2.74	3.00	116
3	Grader 3	1975 4587	3.78	2.91	116
4	Grader 4	2131 5609	1.71	2.72	116

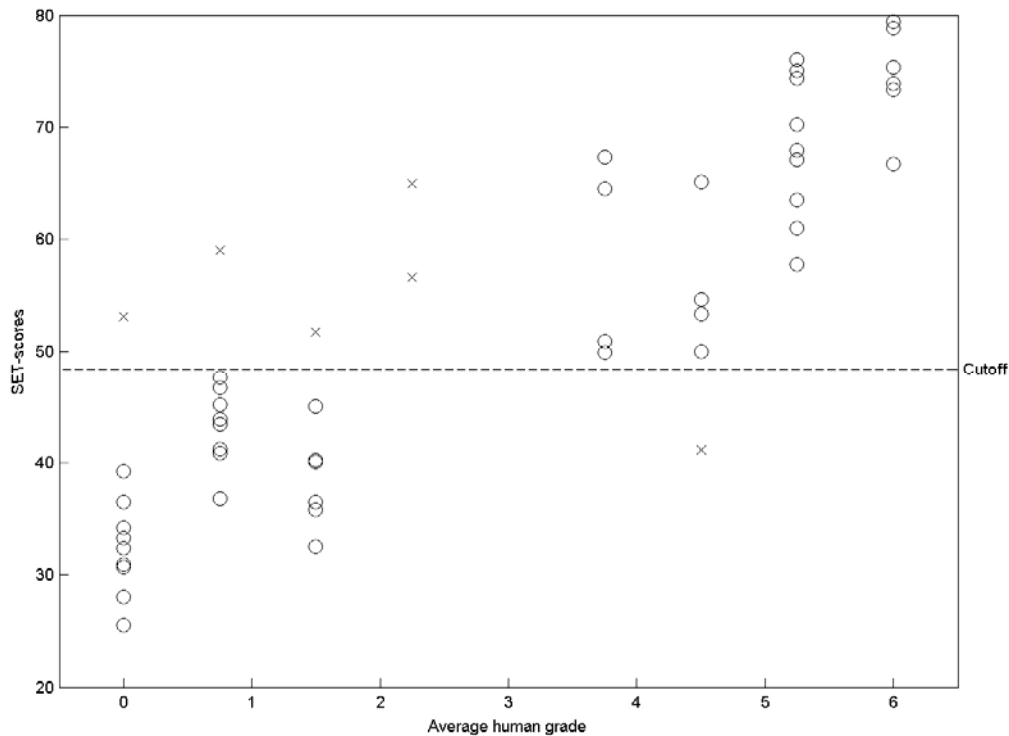
An average human grade (AHG) per examinee was calculated, and the examinee set was divided into two subsets, *unacceptable-ahg* (examinees judged as unacceptable by human graders) and *acceptable-ahg* (examinees judged as acceptable by human graders). All examinees with an AHG below 3 were assigned to 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 SET score for the desired threshold.

Figure 1 shows results of the analysis. The horizontal dashed line at 48.4 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. SET-10 Overall, N=58.



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 an Ordinate SET 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 an Ordinate SET test.

Conclusion

In this report we recommend a methodology for determining an optimal SET cutoff score. If this method is to be used, an optimal threshold based on judgments from the four New Jersey Court graders would be 48.4; thus, a score of 49 is acceptable and a score of 48 is not acceptable. The 48.4 threshold maximizes the total number of correctly classified examinees.

If it is more important to ensure that acceptable candidates are not incorrectly rejected, then the lower threshold should be used (accept scores of 44 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 60 or greater). Table 2 summarizes threshold score values.

Table 2. Summary of threshold scores

Optimal threshold	48.4
Lower threshold (minimize false rejections)	Accept 44 or greater
Upper threshold (minimize false acceptances)	Accept 60 or greater

Disclaimer

The Ordinate 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 Ordinate SET 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 SET-10 tests in the Ordinate white paper: "Validity of the PhonePass Test in Employee Selection."