RESEARCH REPORT 2013-4

CollegeBoard

Development of a New ReadiStep™ Scale Linked to the PSAT/NMSQT® Scale

By YoungKoung Kim, Amy Hendrickson, Priyank Patel, Gerald Melican, and Kevin Sweeney

YoungKoung Kim is a psychometrician
at the College Board.

Amy Hendrickson is a senior psychometrician
at the College Board.

Priyank Patel is a research statistician II
at the College Board.

Gerald Melican is a chief psychometrician
at the College Board.

Kevin Sweeney is vice president, psychometrics
at the College Board.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools. For further information, visit www.collegeboard.org.

MEASUREMENT

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Tables

Figures

Executive Summary

The purpose of this report is to describe the procedure for revising the ReadiStep™ score scale using the field trial data and to provide technical information about the development of the new ReadiStep scale score. In doing so, this report briefly introduces the three assessments — ReadiStep, PSAT/NMSQT®, and SAT® — in the College Board Pathway system, describes the sample obtained in the field trial, discusses the procedure for linking ReadiStep to PSAT/NMSQT, and presents the results of the new ReadiStep scale score.

Introduction

The ReadiStep assessment is a norm-referenced, standards-based test for eighth-grade students that measures academic skills in reading, writing, and mathematics. This assessment is the first step in the College Board's College Readiness Pathway — an integrated series of assessments that includes the PSAT/NMSQT and the SAT. The content of ReadiStep was designed to be aligned with both the PSAT/NMSQT and SAT (College Board, 2009). Additionally, ReadiStep was originally scaled to have a corresponding numerical scale score range of 2.0 to 8.0 to accompany the SAT scale of 200 to 800 and the PSAT/ NMSQT scale of 20 to 80. Although placed on the range of 2.0 to 8.0, this initial ReadiStep scale was not linked to the PSAT/NMSQT scale; thus, for example, a score of 5.0 on the ReadiStep assessment had no inherent relationship to a score of 50 on the PSAT/NMSQT.

The original design for ReadiStep was to develop a prediction relationship with the PSAT/NMSQT when sufficient numbers of test-takers had been given both examinations. Thus, the ReadiStep scale was initially developed independently of the PSAT/NMSQT and SAT scales (Antal, 2009). The original ReadiStep scale was derived to meet the seven principles of a well-aligned scale (Dorans, 2002). The intent was to create a scale that would be normally distributed with a mean of 5.0 and a standard deviation of 1.0 in order to allow for easy interpretation; this decision would allow the use of standard normal tables to interpret percentiles, for example. Again, the original concept was to have ReadiStep predict PSAT/NMSQT scores. It became apparent after the test was launched that the three scales (ReadiStep, PSAT/NMSQT, and SAT) could be vertically aligned.

A result of these scaling decisions was that the initial ReadiStep scores could not be considered interchangeable with scores on either the PSAT/NMSQT or the SAT. This gap in the linkage between the ReadiStep scale and those of the PSAT/NMSQT and SAT posed challenges in interpreting score changes over the course of the College Readiness Pathway system from ReadiStep to PSAT/NMSQT and SAT. Most obviously, using the original ReadiStep scale, the college readiness benchmarks for the Pathway system did not follow an intuitive progression. Consequently, without rescaling the ReadiStep assessment to provide a coherent alignment with the PSAT/NMSQT and the SAT, it was difficult and confusing to make direct comparisons of scaled scores between and among the Pathway assessments.

Since the launch of ReadiStep, with the accumulation of data and reanalysis of the importance of growth interpretations, there has been a realization that the linkage of ReadiStep to the PSAT/NMSQT should be stronger than a prediction, and to the extent it is possible, ReadiStep should be placed on the same scale as the PSAT/NMSQT. Because of differences in the assessments (e.g., ReadiStep is rights-only scored and the PSAT/NMSQT is formula scored) and the populations they serve, it would be difficult to achieve this goal from a strictly technical perspective. However, the coherence between the scales could be greatly improved by revising the ReadiStep scale. In an effort to fully integrate ReadiStep into the College Readiness Pathway system, the College Board conducted a field trial in the fall of 2011 to establish a linked scale for the ReadiStep, PSAT/NMSQT, and SAT assessments.

College Board Pathway Assessments

ReadiStep™

The ReadiStep assessment is intended to be a low-stakes assessment tool designed to provide teachers with early feedback to help students — primarily in the eighth grade — identify the skills they need to improve to be college ready. Students also are given feedback to help them identify the skills they need to improve to prepare for the SAT and success in college. ReadiStep is group administered in the fall with a short administration window.

The assessment includes three sections: critical reading, mathematics, and indirect writing. The total testing time is two hours. Each section can be administered separately in a 40-minute period. The critical reading section includes 45 items, the writing section includes 50 items, and the mathematics section includes 36 items. All of the items employ a multiple-choice, four-option format, and rights-only instruction and scoring. Pretest items are also built into each form.

The critical reading section includes both sentence completion and passage-based questions. The writing section has the same three types of indirect writing questions found on the PSAT/NMSQT and SAT. The mathematics section is divided into two parts: calculator allowed and calculator not allowed.

PSAT/NMSQT®

The PSAT/NMSQT assessment is a norm-referenced test designed primarily for 10th- and 11th-grade students that measures critical reading, mathematics, and indirect writing skills. Its primary intended uses are as a low-stakes assessment in preparation for taking the SAT and as a high-stakes assessment for 11th-grade students in order to determine eligibility for participation in the National Merit Scholarship Competition. The PSAT/NMSQT assessment is jointly sponsored by the College Board and the National Merit Scholarship Corporation. In line with ReadiStep, students are provided feedback to help them identify the skills they need to improve in order to prepare for the SAT and success in college. The PSAT/NMSQT is group administered on two days each fall — a Wednesday and the following Saturday — with separate forms.

The PSAT/NMSQT includes three test areas: critical reading, mathematics, and writing. The assessment is administered in one 2-hour and 10-minute session, plus breaks. The critical reading section contains a total of 48 passage-based reading and sentence completion questions. The writing section includes identifying sentence errors, improving sentences, and improving paragraphs, for a total of 39 items in one 30-minute section. The mathematics section consists of a total of 38 items covering numbers and operations, algebra, geometry and measurement, and data analysis, statistics, and probability. Students are allowed to use a calculator on all mathematics items, though one is not required. All but the mathematics grid-ins are five-option, formula-scored multiple-choice questions.

The PSAT/NMSQT inherits much of its content and psychometric properties from the SAT. Except for the mathematics section, the content specifications between the SAT and PSAT/NMSQT are the same; the SAT contains some third-year-level math that is not included on the PSAT/NMSQT. The reported PSAT/NMSQT score scale ranges from 20 to 80 in increments of 1, for a total of 61 points. The PSAT/NMSQT score scale is primarily maintained through the maintenance of the SAT score scale. The parent SAT forms are equated to previous SAT forms. Once the form is given as a PSAT/NMSQT form, each PSAT/NMSQT form is then equated back to its parent SAT form.

SAT®

The SAT assessment is a norm-referenced test designed primarily students in grades 10–12 that measures critical reading, mathematics, and writing skills. The primary intended use is to help college admission officers make fair and informed admission decisions. Thus, the SAT is a high-stakes assessment and is group administered seven times a year.

The SAT includes three sections: critical reading, mathematics, and writing. The assessment is administered in one 3-hour, 45-minute session, plus breaks. There are 10 subsections: three critical reading, two writing multiple choice, one essay, three mathematics, and one variable section used for equating and pretesting. This last subsection can be from any of the three main subject areas.

The three critical reading subsections contain 67 items: 48 passage-based reading items and 19 sentence completions. The writing section includes identifying sentence errors, improving sentences, and improving paragraphs, for a total of 49 items in one 25-minute section and one 10-minute section. Mathematics is divided into two 25-minute sections and one 20-minute section, covering numbers and operations, algebra, geometry, and measurement, and data analysis, statistics, and probability, with 34 multiple-choice items and 10 student-provided-response items (SPRs), for a total of 44 items. Students are allowed to use a calculator on all mathematics items, although one is not required. The multiple-choice questions are five-option, formula scored. The SAT scores are reported on a 200- to 800-point scale in 10-point increments.

Sample

The overall goal of the field trial sampling was to obtain a nationally representative group of students in the United States at the eighth-, ninth-, and 10th-grade levels to provide information on the growth trajectory of these three grade levels and to link the performance of students in these three grades through the available College Board assessments — ReadiStep, the PSAT/NMSQT, and the SAT. The focus of this report is limited to the eighth-grade field trial sample because students in the eighth grade make up the primary test-taking population for ReadiStep, and the data from this sample were chosen to be used for the rescaling. To determine the characteristics of those in the target population (eighth-grade students) and the schools they attend, the following four types of school-level demographic information were taken from data from the National Center for Education Statistics (NCES): school type (i.e., public or private), geographic region, proportion of underrepresented minority students, and location of school (i.e., urban, suburban, rural).

In order to fairly characterize the school sample, public schools that had eighth-grade enrollments of at least 25 students were considered for part of the data sample. For private schools, schools that had at least one eighth-grader enrolled were used to recruit the field trial sample. Table 1 shows the school characteristics for the eighth-graders identified from the NCES data.[1]

1. The information for public and private schools was separately examined and then combined based on the ratio of public and private school enrollments. For public school information, the NCES Common Core of Data 2009-10 Public Elementary/Secondary School Universe Survey was used. For private school information, the 2007-08 Private School Universe Survey (PSS) data were used. According to the NCES, 88% of the total enrollments from prekindergarten through eighth grade in the U.S. are in public schools, and 12% are in private schools. The data source is available at the NCES website: http://nces.ed.gov/fastfacts/display.asp?id=65.

Data Collection Design

Random-groups design. In gathering the data, two data collection designs — random-groups design and single-group design — were considered. For the random-groups design, the test forms for ReadiStep and the PSAT/NMSQT were randomly assigned to field trial schools intending to administer the tests to eighth-grade students. Because ReadiStep and the PSAT/NMSQT have different scoring instructions — rights-only scoring instructions for ReadiStep versus formula scoring instructions for the PSAT/NMSQT — the random assignment of the test forms had to be carried out at the school level. Based on the random-groups design, data were collected from 3,911 and 4,006 eighth-grade students for ReadiStep and the PSAT/NMSQT, respectively.

Single-group design. The single-group design was included as a complement to the potential limitation created by random assignment at the school level instead of at the individual student level in the random-groups design. For the single-group design, the following two types of samples were considered: (1) students recruited through the field trial ("Field Trial Single Group"); and (2) students who happened to take both ReadiStep and the PSAT/NMSQT in October 2011 ("All Single Group"). For the Field Trial Single Group, four schools agreed to administer both tests (ReadiStep and the PSAT/NMSQT) to their eighth-grade students, resulting in a sample of 237 students. In addition, 1,355 eighth-grade students were identified who happened to take both ReadiStep and the PSAT/NMSQT during the field trial period. This serendipitous group of students combined with the Field Trial Single Group formed the All Single Group.

Data Cleaning Procedure

In order to create a sample that best reflects the student population free of unwanted or unintentional biases, the study sought to remove students who exhibited certain patterns of responses that indicated they were uncharacteristic of the population of students who would be taking ReadiStep under operational conditions. Depending on the data collection designs, ReadiStep Form B, the Wednesday October 2011 operational PSAT/NMSQT form, or a combination of both exams were administered to students within each school. After receiving the students' item-response data for each exam, the data were cleaned in order to minimize the impact on the linking results of students who did not take the exam seriously in the field trial.

Random-groups design data. The following data cleaning rules were applied to the ReadiStep data in order to screen and remove the following types of students from the sample:

- Any student who did not have three valid section scores (critical reading, mathematics, and writing);
- Any student who omitted more than 75% of the items within each section; and
- Any student who chose the same response more than 90% of the time within each section.

In addition to the three data cleaning rules above, the PSAT/NMSQT data sample had one additional rule: remove any student who skipped all 10 student-provided response (SPR) items in the second mathematics section of the PSAT/NMSQT. Applying the three data cleaning rules to the ReadiStep data with a sample size of 3,911 yielded a sample of 3,866 for critical reading, 3,867 for math, and 3,857 for writing. Regarding the final sample of

the PSAT/NMSQT, the original field trial sample of 4,006 was reduced to 3,981 for critical reading, 3,974 for math, and 3,959 for writing, respectively.[2] The demographic characteristics and descriptive statistics for the final samples obtained after data cleaning are shown in Tables 1, 2, and 3.

Single-group design data. In order to identify students who participated in the single-group design, ReadiStep and PSAT/NMSQT data were matched using identifying information of the students, including name, school code, date of birth, and address. Applying similar rules to the random-groups design data, data were removed from the single-group design sample for the following students:

- Any student who did not have three valid section scores (critical reading, mathematics, and writing) for each exam;
- Any student who omitted more than 75% of the items within each section for each exam;
- Any student who chose the same response more than 90% of the time within each section for each exam;
- Any student who skipped all 10 mathematics and 4 SPR items on the PSAT/NMSQT; and
- Any student who did not have scores on each exam.

For the All Single Group, the data cleaning procedure removed about 20% from the original sample of 1,936. A majority of students were removed because the All Single Group sample was required to have scores on both the PSAT/NMSQT and ReadiStep; 388, 377, and 407 students were removed for critical reading, mathematics, and writing, respectively, and among those, more than 200 students were excluded because they were missing scores from either one or both of the exams. For the Field Trial Single Group, only two students out of 237 students were excluded from the analysis.

Descriptive Statistics

Table 1 presents the demographic characteristics of the field trial samples from the random-groups and single-group designs, both of which were used for the linking analysis. The random-groups design sample for ReadiStep had a slightly higher proportion of diverse schools (i.e., schools where the percentage of underrepresented minority students is 50% or higher) than was expected. The random-groups design sample for the PSAT/NMSQT had a

2. The following are the numbers of students who were removed for each criterion.

For ReadiStep:

- Any student who did not have three valid section scores; 42 students were removed.
- Any student who omitted more than 75% of the items within each section; three students for critical reading, two students for mathematics, and 12 students for writing were removed.
- Any student who chose the same response more than 90% of the time within each section; none were removed.

For the PSAT/NMSQT:

- Any student who did not have three valid section scores; 21 students were removed.
- Any student who omitted more than 75% of the items within each section; four for critical reading, five for mathematics, and 26 students for writing were removed.
- Any student who chose the same response more than 90% of the time within each section. One student for mathematics was removed, and none were removed for critical reading and writing.
- Any student who skipped all 10 mathematics 4 SPR items on the PSAT/NMSQT; five students in mathematics were removed.

lower proportion of suburban schools and a higher proportion of urban schools compared to the NCES data. In addition, both samples have more students from Southern states and fewer students from Midwestern states than those in the NCES sample.

However, overall, the random-groups design samples for both ReadiStep and the PSAT/NMSQT appeared fairly similar to those of the NCES targets, which implies that the field trial recruitment of schools was successful in achieving the representative sample for our study. On the other hand, the demographic characteristics for the two single-group design samples — the Field Trial Single Group sample and the All Single Group sample — seemed different from the NCES targets. This observed discrepancy was expected, however, because only a small number of schools were recruited for the Field Trial Single Group sample, and random selection was not used for the All Single Group sample.

Tables 2 and 3 show the scale score distributions of ReadiStep and the PSAT/NMSQT for the random-groups, All Single Group, and Field Trial Single Group samples. The initial ReadiStep scale, which ranged from 2.0 to 8.0, was used to make these calculations. As reference groups, the score distributions of both the ReadiStep and PSAT/NMSQT exams for the eighth-grade 2011 fall operational data sample are also presented. For ReadiStep, the random-groups and single-group means were similar to the operational means. The standardized mean differences between these two design groups and the operational data sample were less than 0.20 for all three sections, indicating small differences. The All Single Group design exhibited higher standardized differences than 0.20 but less than 0.50, which is considered a medium effect size. These results are not unexpected, as the eighth-grade students in the All Single Group design who chose to take the PSAT/NMSQT tended to be higher skilled. On the other hand, the PSAT/NMSQT scores for all the field trial design groups seemed to be consistently lower than those of the operational data. In particular, the standardized mean differences between the random-groups design sample and the operational data were above 0.4. These results were not unexpected, as eighth-grade students who choose to take the PSAT/NMSQT typically tend to be an extremely able sample affecting the PSAT/NMSQT operational means. The All Single Group design showed lower standardized differences than 0.4. The PSAT/NMSQT scores for the single-group design samples may be lower than those of the operational data because the samples consisted of only a few states and did not include states that usually exhibit high PSAT/NMSQT scores, such as states in New England.

Reliability and Standard Error of Measurement

Random-groups design. The reliability and standard error of measurement (SEM) for ReadiStep and PSAT/NMSQT were computed for the field trial samples. The Kuder-Richardson KR20 (1937) was calculated for the critical reading, mathematics, and writing sections of ReadiStep. Since PSAT/NMSQT items are formula scored, Dressel-KR20 (1940) coefficients were computed for critical reading, mathematics, and writing sections. In addition, variance-components reliability estimates were computed for raw scores on the total test scores. The reliability indices are presented in Table 4; scores for critical reading, mathematics, and writing for both ReadiStep and PSAT/NMSQT from the random-groups design samples indicate reasonable reliabilities, ranging from 0.85 to 0.88 for ReadiStep and from 0.80 to 0.82 for the PSAT/NMSQT. The standard errors of measurement ranged from 2.45 to 2.90 for ReadiStep and from 2.79 to 3.56 for the PSAT/NMSQT.

Single-group design. The reliabilities of ReadiStep and the PSAT/NMSQT from the All Single Group sample were also reasonable and slightly higher than those of the random-groups design sample (ranging from 0.87 to 0.90 for ReadiStep and from 0.80 to 0.83 for the PSAT/NMSQT). The reliabilities of ReadiStep and the PSAT/NMSQT from the Field Trial

Single Group sample (which ranged from 0.79 to 0.88 for ReadiStep and from 0.74 to 0.76 for the PSAT/NMSQT) were lower than the ones from the other field trial design samples. The standard errors of measurement for the All Single Group sample ranged from 2.41 to 2.85 for ReadiStep and 2.77 to 3.50 for the PSAT/NMSQT. The standard errors of measurement for the Field Trial Single Group sample ranged from 2.48 to 2.86 for ReadiStep and 2.84 to 3.63 for PSAT/NMSQT. The standard deviations for the All Single Group sample tended to be smaller than the other samples, and although the reliabilities were a little lower, the standard errors of measurement were similar.

The correlations of raw scores between ReadiStep and the PSAT/NMSQT were also examined for the single-group design samples. The correlations for the critical reading, mathematics, and writing sections in the All Single Group design sample were 0.75, 0.76, and 0.74, respectively. When corrected for unreliability, the correlations were 0.88, 0.89, and 0.88, respectively.

The correlations between ReadiStep and the PSAT/NMSQT in the All Single Group sample were lower than those in the Field Trial Single Group sample. The correlations were 0.69, 0.68, and 0.71, for the critical reading, mathematics, and writing sections, respectively. When corrected for unreliability, the correlations were 0.86, 0.88, and 0.87, respectively.

Since the Field Trial Single Group design sample does not appear to be a nationally representative sample because of the small number of schools (four) and test-takers (235) and because the standard errors of measurement and correlations corrected for unreliability were similar for ReadiStep and the PSAT/NMSQT, this report focused on only two linking samples — the random-groups design sample and the All Single Group sample — for the purposes of creating a linking procedure, and thus the Field Trial Single Group sample analyses are not presented through the rest of the paper.

Method of Linking ReadiStep to PSAT/NMSQT

Raw Score Distributions and Presmoothing

Before linking ReadiStep to the PSAT/NMSQT, several necessary steps were performed in preparation for the analysis. First, the raw score distributions of both ReadiStep and the PSAT/NMSQT were examined for completeness. Next, the relative frequency distribution of the raw scores of both the ReadiStep and the PSAT/NMSQT groups were obtained. Finally, the relative frequency distributions were smoothed using the polynomial loglinear method. This section discusses loglinear smoothing, describes the details of this presmoothing procedure, and presents the results of presmoothing.

Loglinear smoothing. Loglinear smoothing is a commonly used presmoothing technique that uses polynomial loglinear models. These models are discussed in detail in Darroch and Ratcliff (1972), Haberman (1974), Holland and Thayer (1987), and Rosenbaum and Thayer (1987). The most attractive feature of the method is the moment preservation property, which means that a specified number of moments of the smoothed distribution are the same as those for the unsmoothed score distribution. The polynomial loglinear model takes the following form:

$$log[N_X f(x)] = \omega_0 + \omega_1 x + \omega_2 x^2 \ldots + \omega_c x^c,$$

where C represents the highest polynomial degree. The ω parameter is estimated by the maximum likelihood estimation method. The choice of C is important and, therefore, the

model fit statistics for various choices of *C* are usually evaluated using various statistical techniques such as likelihood-ratio chi-square goodness-of-fit statistics or the likelihood ratio difference chi-square test.

Random-groups design. Table 5 provides summary statistics of the raw score distributions of ReadiStep and the PSAT/NMSQT for the random-groups design sample. The critical reading and mathematics scores of ReadiStep were positively skewed, whereas the writing scores were the closest to the normal distribution. In terms of kurtosis, the ReadiStep raw score distributions of all three tests were flatter than the normal distribution. The raw score distributions for all three sections of the PSAT/NMSQT had positive skewness and positive kurtosis, indicating that the tests were difficult for the students and that the distributions had higher peaks and heavier tails than those of the normal distribution. Using the polynomial loglinear model, each raw score frequency distribution for ReadiStep and the PSAT/NMSQT was smoothed. The polynomial degrees used for the loglinear smoothing were decided based on a chi-square difference test, which compares likelihood ratio chi-square fit statistics. Table 6 presents the chi-square statistics with the associated degrees of freedom. For ReadiStep, the polynomial degrees of 6, 5, and 6 were decided for critical reading, mathematics, and writing, respectively. For the PSAT/NMSQT, the polynomial degrees of 5, 6, and 5 were decided for critical reading, mathematics, and writing, respectively. Figures 1–6 compare the smoothed and empirical raw score distributions for each section of ReadiStep and each section of the PSAT/NMSQT. As shown in Figures 2, 4, and 6, in particular, the loglinear smoothing helped to reduce the "teeth" that were exhibited in the raw score distributions of the PSAT/NMSQT, where the frequencies were much lower than those of the neighboring raw scores because of the use of rounded formula scores (Holland & Thayer, 2000).

Single-group design. Table 7 reports summary statistics for the raw score distributions of ReadiStep and the PSAT/NMSQT from the single-group design sample. The critical reading section score for ReadiStep was positively skewed and the writing score was negatively skewed. The mathematics score, however, was the closest to the normal distribution. In terms of kurtosis, the ReadiStep raw score distributions of all three tests had higher peaks and heavier tails than those of the normal distribution. Since the raw score distributions for all three sections were positively skewed, the PSAT/NMSQT seemed to be difficult for the students. In terms of kurtosis, the distributions also had higher peaks and heavier tails than those of the normal distribution.

In order to understand the relationship between the ReadiStep and the PSAT/NMSQT scores for the All Single Group sample as well as to detect the outliers, the scatterplots of the PSAT/NMSQT raw scores against ReadiStep raw scores were examined (Figures 7a–7c). The plots suggest the existence of a ceiling effect for ReadiStep, in particular for critical reading and mathematics. In other words, the very highest-performing eighth-grade students were capped by the raw score scale of the ReadiStep test. The plots imply that it might be difficult to differentiate among these students. Given that the All Single Group sample includes the students who voluntarily took ReadiStep and the PSAT/NMSQT during almost the same time window, as well as those who tend to be highly motivated and high performing, the observed ceiling effect is not surprising. However, the ceiling effect can be a drawback when placing ReadiStep on the PSAT/NMSQT score scale using this single-group design sample.

After obtaining the raw score frequency distribution for each test, loglinear presmoothing was performed. For the single-group design, the cross-product moments in the joint distribution of ReadiStep and the PSAT/NMSQT were considered in the smoothing procedure. Table 8 provides the chi-square statistics with the associated degrees of freedom. Using the likelihood ratio chi-square difference test, the following models with these polynomial

degrees were selected: (1) for critical reading, a model that maintains four moments each for ReadiStep and the PSAT/NMSQT, respectively, and one cross-product between the two exams; (2) for mathematics, a model that maintains five moments each for ReadiStep and the PSAT/NMSQT and one cross-product between the two exams; and (3) for writing, a model that maintains four moments each for ReadiStep and the PSAT/NMSQT and one cross-product between the two exams.

Equipercentile Linking

Using the smoothed percentile rank from the loglinear smoothing procedure, the raw ReadiStep and PSAT/NMSQT scores that had the same percentile rank were identified through equipercentile linking. All linking analyses were carried out using Equating Recipes (Brennan, Wang, Kim, & Seol, 2009). This section discusses the equipercentile linking procedure and presents the results of equipercentile linking with the random groups and single-group design samples.

Equipercentile linking method. The equipercentile linking method is preferable to the linear linking method when a sample size is large and if the two tests to be linked have score distributions with different shapes. As presented previously, the sample sizes for both random-groups and single-group design samples were reasonable, and the shape of the ReadiStep score distributions was different from those of the PSAT/NMSQT. Thus, the equipercentile linking method was chosen to place ReadiStep on the PSAT/NMSQT scale.

General procedure. When X and Y are different tests that measure similar constructs, define F as the cumulative distribution of the scores on test X, G as the cumulative distribution of the scores on test Y, F^{-1} as the inverse function of F, and G^{-1} as the inverse function of G. The equipercentile linking function, $e_Y(x)$, that provides the scores on X on the scale of Y associated with the percentile rank of $G(x)$ can be written as:

$$e_Y(x) = G^{-1}[F(x)],$$

where $F(x)$ is the percentile rank for score x, and $G^{-1}(*)$ is the inverse of the percentile point function for Y and provides the raw score for Y for a given percentile. Similarly, the equipercentile linking function, $e_X(y)$, that provides the score on the scale of X associated with the percentile rank of $F(y)$ can be written as:

$$e_X(y) = F^{-1}[G(y)],$$

where $G(y)$ is the percentile rank for score y, and $F^{-1}(*)$ is the percentile point function for X. The full description of the equipercentile method can be found in Kolen and Brennan (2004).

Results of equipercentile linking. The equipercentile linking of ReadiStep to the PSAT/NMSQT resulted in ReadiStep raw scores on the PSAT/NMSQT raw score scale. In order to achieve the "raw-to-scale" score conversion, the PSAT/NMSQT conversion table for the 2011 Wednesday test form was applied. The unrounded PSAT/NMSQT equivalents for the ReadiStep raw scores for the three sections are reported in Tables 9–11. The tables include the results from both random-groups and single-group design samples. According to the tables, the conversion for the single-group design sample produced higher mean scale scores for ReadiStep than those of the random-groups design for mathematics and writing, but lower for critical reading.

The difference in the conversion line between the single-group and random-groups design samples can be described more clearly using graphs. Figures 8–13 display the unrounded linking

conversion line for each section as well as the differences in the conversion lines between two samples. Overall, the linking results from the random-groups design sample were close to the ones from the single-group design sample. However, the unrounded PSAT/NMSQT equivalents for the highest ReadiStep raw scores in the random-groups sample seemed to be consistently higher, and for lower- to middle-range, raw scores seemed to be consistently lower than the ones in the single-group design sample across all three sections.

New Scale

After extensive analyses, the College Board Research department presented the methodology and results to the Pathway Linking Advisory Committee[3] and the senior management of both the ReadiStep and the PSAT/NMSQT programs. The Research department and the advisory committee recommended that the conversion obtained from the linking analysis based on the eighth-grade random-groups design sample be used. The decision to use the eighth-grade random-groups design sample over those of other grades and designs was based on the following reasons: (1) eighth-grade students make up the majority of ReadiStep test-takers and (2) the sample was large and representative. Furthermore, the random-groups design sample was collected based on the master plan of the College Board field trial for the linking study, while the single-group design sample was obtained from a nonrandom procedure.

Decision on New Scale Score

Using the conversion line from the eighth-grade random-groups design sample as the basis for the new ReadiStep score scale, a few possible scale options were explored. When ReadiStep was launched, the initial plan was to use a scale from 2.0 to 8.0, which is similar to the 20- to 80-point and 200- to 800-point scales for the PSAT/NMSQT and the SAT, respectively. With the ReadiStep scale aligned to the PSAT/NMSQT, however, the 2.0- to 8.0-point scale was no longer a viable option because in the field trial sample, a significant number of students scored below 2.0, and no students reached a score of 8.0. The results reflect the predictably lower skill level of eighth-grade students, who have yet to develop the academic preparation and skills found among the 10th-grade PSAT/NMSQT takers.

If the 2.0- to 8.0-point scale were to be used, the major drawback would be the inability to exhibit growth for those students with true ReadiStep scores below 2.0 who eventually take the PSAT/NMSQT and the SAT after they start high school. For example, a student who received a ReadiStep score of 2.0 in the eighth grade (whose true score was below 2.0) and a PSAT/NMSQT score of 20 in the 10th grade exhibits no gain. However, if the ReadiStep scale is extended below 2.0 to more accurately gauge the student's score, then his or her PSAT/NMSQT score of 20 would generally reflect a score improvement. Therefore, to better capture the growth of test scores from eighth to 10th grade, it was decided to lower the final ReadiStep minimum possible score. With respect to the upper limit of the scale, because some topics on the SAT and the PSAT/NMSQT are usually covered in advanced course work later in high school, most eighth-graders are not capable of SAT scores of 800 (or PSAT/NMSQT of 80). Therefore, it was also decided to lower the final ReadiStep maximum reported score.

The following are a few of the more promising possibilities: (1) a scale of 1 to 7 with increments of 0.10, resulting in 61 score points; (2) a scale of 1 to 7 with increments of 0.20, resulting in 31 score points; (3) a scale of 1.5 to 7 with increments of 0.10, resulting

3. The Pathway Linking Advisory Committee is made up of external psychometric consultants who are experts in the practices of linking and growth modeling.

in 56 score points; and (4) a scale of 1.4 to 7 with increments of 0.2, resulting in 29 score points. Among the possible scale options, a scale of 1.5 to 7 was chosen for raw scores of 1 and above. A raw score of zero was set to a scale score of 1.0. Increments of 0.10 were selected to provide greater score precision (compared to using increments of 0.20) and to be consistent with the PSAT/NMSQT and the SAT scales.

Adjustment for Final Conversion

Finalizing the scales required several adjustments to the conversion in order to (1) prevent the scale scores from exceeding the range of possible scores; (2) construct new scale scores for the raw scores at the low end of the range where the same converted scores were repeated because of a lack of data; and (3) provide flexibility in equating other forms in the future. The adjustments were as follows:

- The linking conversion lines below the 5th percentile and above the 95th percentile were replaced by a straight line through a "doglegging" procedure.
- To allow for equating flexibility, raw scores of both 1.0 and 2.0 were converted to 1.50. In addition, a raw score of 0.0 was converted to 1.0 to differentiate the students who did not answer any items correctly and extended the final lower bound of the ReadiStep base scale to 1.0.
- At the top end, a score of $k - 1$ was set to 7.0 to allow for expansion over forms in which k was the number of items for each section of ReadiStep. With this modification, raw scores of both $k - 1$ and k were converted to 7.0 on the base scale.

Applying New Scale — Conversion Tables, Descriptive Statistics, and Norms

The final new ReadiStep scales after the aforementioned adjustments are reported in Tables 12–14. The tables display the new scale score conversions for all three forms in comparison with the old scale score conversions. The new ReadiStep scale of 1.0 to 7.0 with increment of 0.10 was set for the base form, Form B, which was administered in the 2011 field trial. In the 2008 field trial, which was used to set the original ReadiStep scale of 2.0 to 8.0, the relationship among the three existing ReadiStep forms was established by equating Forms A and C to the base Form B through random-groups equating design. In the 2011 field trial, the known relationship among the three previous forms from the conversion tables obtained from the 2008 field trial was used to transfer the scores for Forms A and C to the new scale of 1.0 to 7.0, which was established for Form B.

The ReadiStep norms, which had been based on the fall 2008 field trial data consisting of all three forms, were recalculated using the new scale scores (Table 15). The new scale was also applied to the ReadiStep operational data, which were collected in the fall of 2011. Table 16 provides the descriptive statistics of the 2011 administration after applying the new scale. Figures 14–16 show the scale score distributions for critical reading, mathematics, and writing for the 163,936 fall 2011 ReadiStep test-takers. The means for critical reading, mathematics, and writing were 3.5, 3.6, and 3.4, respectively. In addition, it appears that all three sections were positively skewed.

The overall change in scores of eighth-graders who took ReadiStep and 10th-graders who took the PSAT/NMSQT was examined by applying the new ReadiStep scales to data from the eighth-graders who both completed ReadiStep as part of the 2008 field studies that were used to establish the initial ReadiStep scale and who also took the PSAT/NMSQT in 2010 as 10th-graders. Table 17 shows the descriptive statistics for both matched and unmatched

samples from the 2008 ReadiStep and the 2010 PSAT/NMSQT data. The matched group has a higher mean for all three sections. On the PSAT/NMSQT scale (i.e., multiplying the ReadiStep scale score by 10), the matched sample showed about 3-, 6-, and 4-point gains in critical reading, mathematics, and writing, respectively.

Conclusion

The current study describes the procedure used to place ReadiStep scores on the same scale as the PSAT/NMSQT and the SAT. Extensive analyses were conducted on possible scale scores using various data collection designs (random-groups design and single-group design) and analyzing various test-taker populations (eighth, ninth, and 10th grades). Based on these analyses and on recommendations and suggestions from a variety of groups, including the Pathway Linking Advisory Committee, the ReadiStep and the PSAT/NMSQT programs, the ReadiStep scale was set at 1.0 to 7.0 with increments of 0.10, based on the eighth-grade random-groups design sample. To capture the vertical relationships among the major test-taker populations — eighth-graders for ReadiStep, and 10th-, 11th, and 12th-graders for the PSAT/NMSQT and the SAT — the new ReadiStep scale was modified to 1.0 to 7.0 from 2.0 to 8.0. Lowering the minimum allows for estimating growth for students in the extremely low range. Thus, the new scale can assign an actual scale score to low-performing ReadiStep students instead of assigning a 2.0 to all of them. Lowering the maximum was consistent with the results of the field trial.

The new ReadiStep scale is now directly linked to the PSAT/NMSQT so that one can readily identify ReadiStep and PSAT/NMSQT scores that have the same percentile rank. For example, a score of 4.2 on ReadiStep has the same percentile rank as a 42 on the PSAT/NMSQT, and these scores indicate approximately the same level of overall achievement. The new ReadiStep scale enhances the interpretability of the scores from the College Board Pathway system. By virtue of the new scale, the College Board benchmarks for college readiness follow a logical progression from eighth grade to high school graduation. In addition, since ReadiStep scores are interpretable in PSAT/NMSQT and SAT units, explaining the changes in performance from test to test is now more appropriate and ultimately more beneficial to key stakeholders in the education sector. For example, ReadiStep scores, whether viewed individually or in aggregate at the school, district, or state level, can be more easily understood at the early start of the college planning process.

References

Antal, J. (2009). *Test analysis report: ReadiStep scaling and equating.* Unpublished Statistical Report. New York: The College Board.

Brennan, R. L., Wang, T., Kim, S., & Seol, J. (2009). Equating Recipes. [Computer Software]. Center for Advanced Studies in Measurement and Assessment, University of Iowa, Iowa City, IA.

College Board. (2009). *Technical information on the ReadiStep assessment.* New York: The College Board.

Darroch, J. N., & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics. 43*, 1470–1480.

Dorans, N.J. (2002). Recentering and realigning the SAT score distributions: How and why. *Journal of Educational Measurement, 39,* 59-84.

Dressel, P. L. (1940). Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika*, *5*, 305–310.

Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics*, *30*, 589–600.

Holland, P. W., & Thayer, D. T. (1987). *Notes on the use of log-linear models for fitting discrete probability distributions.* Educational Testing Service Research Report 87-31. Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, *25*, 133–183.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*, 151–160.

Rosenbaum, P. R., & Thayer, D. (1987). Smoothing the joint and marginal distributions of scored two-way contingency tables in test equating. *British Journal of Mathematical and Statistical Psychology, 40*, 43–49.

Table 1.

Demographic Characteristics of U.S. Eighth Grade Population and Field Trial Sample

			Random-Groups Design		Single-Group Design	
		NCES Population	RS	PN	FT SG Sample	All SG Sample
Location	Rural	0.24	0.25	0.23	0.18	0.07
	Suburban	0.48	0.52	0.34	0.19	0.51
	Urban	0.28	0.23	0.42	0.63	0.43
Diversity (% minority)	<=50	0.60	0.51	0.58	0.27	0.14
	>50	0.40	0.49	0.42	0.73	0.86
CB Region	Middle States	0.22	0.22	0.21	0.00	0.38
	Midwestern	0.14	0.04	0.07	0.00	0.00
	New England	0.04	0.07	0.02	0.00	0.00
	Southern States	0.23	0.40	0.41	0.37	0.14
	Southwestern	0.12	0.15	0.16	0.00	0.36
	Western	0.24	0.11	0.13	0.63	0.12
School Type	Private	0.12	0.18	0.19	0.27	0.06
	Public	0.88	0.82	0.81	0.73	0.94

Table 2.

Eighth Grade ReadiStep Old Scale Score (2 to 8 Scale) Distributions, Based on Various Field Trial and Operational Samples

		N	Mean	*SD*	Skewness	Kurtosis	Min	Max	SMD*
Fall 2011 Operational Data	Critical Reading	124,155	4.56	1.18	-0.11	-0.26	2.00	8.00	
	Math	123,701	4.98	1.04	0.02	0.06	2.00	8.00	
	Writing	124,203	4.93	1.10	-0.08	0.13	2.00	8.00	
Random-Groups Design	Critical Reading	3,866	4.78	1.11	-0.21	-0.10	2.00	8.00	0.19
	Math	3,867	4.91	0.99	-0.02	-0.05	2.00	8.00	-0.07
	Writing	3,857	5.00	1.07	-0.06	0.12	2.00	8.00	0.06
All Single-Group Design	Critical Reading	1,548	4.83	1.17	-0.31	-0.18	2.00	7.80	0.23
	Math	1,559	5.37	1.02	-0.07	0.23	2.00	8.00	0.38
	Writing	1,529	5.24	1.10	-0.10	-0.02	2.00	8.00	0.28
FT Single-Group Design	Critical Reading	235	4.88	5.00	1.01	-0.54	0.52	2.00	0.09
	Math	235	5.17	5.20	0.80	-0.14	1.29	2.00	0.05
	Writing	235	5.25	5.20	1.05	-0.26	0.28	2.00	0.08

*Note: Standardized mean difference (SMD) was computed by $(M_{FT} - M_{OD})/\sigma$, where M_{FT} is the mean of the field trial sample, M_{OD} is the mean of the operational data sample, and σ is the pooled standard deviation (i.e., $\sqrt{(\sigma^2_{OD} + \sigma^2_{FT}) / 2}$).

Table 3.

Eighth Grade PSAT/NMSQT Scale Score Distribution, Based on Various Field Trial and Operational Samples

		N	Mean	*SD*	Skewness	Kurtosis	Min	Max	SMD*
Operational Data	Critical Reading	51,937	40.33	9.03	0.17	0.10	20.00	80.00	
	Math	51,949	39.51	8.98	0.43	0.43	20.00	80.00	
	Writing	51,764	38.27	8.34	0.27	0.35	20.00	80.00	
Random-Groups Design	Critical Reading	3,981	36.49	9.08	0.37	0.15	20.00	80.00	-0.42
	Math	3,974	35.14	8.49	0.53	0.42	20.00	74.00	-0.50
	Writing	3,959	34.74	8.08	0.39	0.31	20.00	67.00	-0.43
All Single-Group Design	Critical Reading	1,548	37.93	8.97	0.38	0.33	20.00	80.00	-0.27
	Math	1,559	37.80	8.40	0.63	1.69	20.00	80.00	-0.20
	Writing	1,529	36.70	7.83	0.29	0.29	20.00	71.00	-0.19
FT Single-Group Design	Critical Reading	235	38.27	7.83	0.07	-0.19	20.00	58.00	-0.24
	Math	235	36.44	7.11	0.50	1.04	20.00	65.00	-0.38
	Writing	235	36.92	7.14	0.31	0.06	20.00	60.00	-0.17

*Note: Standardized mean difference (SMD) was computed by $(M_{FT} - M_{OD})/\sigma$, where M_{FT} is the mean of the field trial sample, M_{OD} is the mean of the operational data sample, and σ is the pooled standard deviation (i.e., $\sqrt{(\sigma^2_{OD} + \sigma^2_{FT}) / 2}$).

Table 4.

Reliability Coefficients and Standard Errors of Measurement for Eighth Grade Field Trial Samples

			Random Groups			All Single Group			Field Trial Single Group		
Exam	**Test Section**	**Reliability Type**	***N***	**Reliability**	**SEM**	***N***	**Reliability**	**SEM**	***N***	**Reliability**	**SEM**
ReadiStep	Critical Reading	*KR20*	3,866	0.88	2.78	1,548	0.90	2.76	235	0.86	2.82
	Math	*KR20*	3,867	0.85	2.45	1,559	0.87	2.41	235	0.79	2.48
	Writing	*KR20*	3,857	0.88	2.90	1,529	0.89	2.85	235	0.88	2.86
PSAT/ NMSQT	Critical Reading 1	Dressel-*KR20*		0.60	2.49		0.64	2.43		0.54	2.51
	Critical Reading 2	Dressel-*KR20*		0.73	2.55		0.73	2.52		0.63	2.62
	Total Critical Reading	Variance Components	3,981	0.80	3.56	1,548	0.81	3.50	235	0.74	3.63
	Math 1	Dressel-*KR20*		0.70	2.21		0.72	2.16		0.57	2.26
	Math 2	Dressel-*KR20*		0.70	1.70		0.70	1.74		0.65	1.72
	Total Math	Variance Components	3,974	0.82	2.79	1,559	0.83	2.77	235	0.75	2.84
	Writing	Dressel-*KR20*	3,959	0.80	3.24	1,529	0.80	3.23	235	0.76	3.34

Table 5.

Eighth Grade ReadiStep and PSAT/NMSQT Raw Score Distributions: Random Groups Design

	N	Mean	*SD*	Skewness	Kurtosis	Min	Max	Median
ReadiStep								
Critical Reading	3,866	18.98	8.17	0.29	-0.70	0.00	40.00	18.00
Math	3,867	14.96	6.32	0.33	-0.64	0.00	32.00	14.00
Writing	3,857	24.42	8.38	-0.01	-0.68	3.00	44.00	24.00
PSAT/NMSQT*								
Critical Reading	3,981	9.65	8.03	0.82	0.72	-9.00	47.00	8.00
Math	3,974	6.28	6.63	0.89	0.60	-7.00	36.00	5.00
Writing	3,959	7.31	7.17	0.82	0.40	-7.00	34.00	6.00

*Note: The PSAT/NMSQT is a formula-scored examination; "formula scored" equals the number of right answers minus one-quarter the number of wrong answers.

Table 6.

Eighth Grade Random Groups Design Smoothing Polynomial Degree

Exam	Polynomial Degree (*m*)	Critical Reading		Math		Writing	
		Chi-Square	*df*	Chi-Square	*df*	Chi-Square	*df*
ReadiStep	2	280.3917	38	274.9503	30	146.7515	43
	3	198.8772	37	179.0613	29	142.5233	42
	4	58.97518	36	27.8799	28	54.5309	41
	5	51.86512	35	**22.6052**	27	50.7580	40
	6	**44.17622**	34	22.3977	26	**40.9632**	39
	7	43.96832	33			40.8928	38
PSAT/NMSQT	2	788.97952	58	700.7680	43	1110.0292	47
	3	436.94274	57	416.7771	42	773.0893	46
	4	304.3451	56	180.6667	41	536.4688	45
	5	**291.3069**	55	119.1045	40	**519.5783**	44
	6	291.2862	54	**103.2834**	39	518.8966	43
	7			103.0840	38		

Note: Bolded values indicate the polynomial degree chosen.

Table 7.

Eighth Grade ReadiStep and PSAT/NMSQT Raw Score Distributions: Single Group Design

	N	Mean	*SD*	Skewness	Kurtosis	Minimum	Maximum
ReadiStep							
Critical Reading	1,548	19.49	8.54	0.17	2.18	1.00	39.00
Math	1,559	17.93	6.59	0.00	2.25	1.00	32.00
Writing	1,529	26.33	8.55	-0.18	2.28	5.00	45.00
PSAT/NMSQT							
Critical Reading	1,548	10.88	8.12	0.80	3.67	-7.00	47.00
Math	1,559	8.31	6.74	0.80	3.94	-6.00	38.00
Writing	1,529	9.01	7.23	0.60	2.96	-6.00	37.00

Table 8.

Eighth Grade Single Group Design Smoothing Polynomial Degree

Polynomial Degree	Critical Reading		Math		Writing	
	Chi-Square	*df*	Chi-Square	*df*	Chi-Square	*df*
3 and 1 Cross Product	1319.6924	2493	890.6492	1,510	1289.0163	2292
4 and 1 Cross Product	**1229.8040**	**2491**	854.1625	1,508	**1180.0982**	**2290**
5 and 1 Cross Product	1225.5028	2489	**823.5691**	**1,506**	1177.2152	2288
6 and 1 Cross Product	1213.7844	2487	822.8250	1,504	1173.1831	2286

Note: Bolded values indicate the polynomial degree chosen.

Table 9.

Eighth Grade ReadiStep Unrounded Raw to Scale Score Conversion: Critical Reading

RS Raw	Random Groups	Single Group	RS Raw	Random Groups	Single Group
0	16.3159	16.3159	23	40.6879	40.4680
1	16.3159	16.3159	24	41.4979	41.3428
2	16.3159	16.3159	25	42.2668	42.1514
3	16.3159	16.3159	26	43.0826	42.9253
4	16.3159	17.5591	27	44.1101	43.8968
5	16.5774	20.1391	28	45.2221	44.9754
6	19.0502	22.4215	29	46.4454	46.0701
7	21.4953	24.1820	30	47.5553	47.2043
8	23.3774	25.6904	31	48.7526	48.2663
9	24.8854	27.3937	32	50.4533	49.7678
10	26.4911	29.0263	33	51.9617	51.4918
11	28.1380	30.6031	34	53.6785	53.1293
12	29.7447	31.6632	35	55.7703	55.3153
13	31.2579	32.4034	36	57.6857	57.5575
14	32.0738	33.1257	37	60.1990	60.5258
15	32.8622	34.0736	38	62.7454	63.7664
16	33.8449	35.0529	39	66.3401	68.5136
17	34.9499	36.0025	40	71.6904	76.0312
18	36.0499	36.8557			
19	37.0157	37.5379	Mean	37.8550	36.3812
20	37.7992	38.2015	*SD*	9.0131	9.2103
21	38.5561	38.9224	Skew	0.1913	0.2296
22	39.5228	39.8567	Kurt	3.1696	3.2242

Table 10.

Eighth Grade ReadiStep Unrounded Raw to Scale Score Conversion: Math

RS Raw	Random Groups	Single Group	RS Raw	Random Groups	Single Group
0	6.6385	6.4544	19	39.9450	38.6352
1	9.4113	9.4659	20	41.1033	39.6813
2	12.2237	12.4046	21	42.3016	40.7491
3	15.0809	15.3167	22	43.5638	41.8566
4	17.8314	18.1498	23	44.9010	43.0183
5	20.1359	20.5246	24	46.3692	44.2550
6	22.3086	22.5484	25	47.9769	45.5657
7	24.2912	24.4226	26	49.7727	46.9717
8	26.0148	26.0321	27	51.7963	48.5000
9	27.5731	27.4750	28	54.1450	50.1978
10	29.0161	28.7704	29	56.8432	52.1223
11	30.4055	30.0181	30	59.8995	54.4767
12	31.7344	31.2047	31	63.3911	57.9638
13	33.0033	32.3381	32	67.6887	67.9252
14	34.2226	33.4439			
15	35.3993	34.5122	**Mean**	**35.0777**	**37.8177**
16	36.5496	35.5470	***SD***	**8.7081**	**8.3920**
17	37.6868	36.5704	**Skew**	**0.3605**	**0.3682**
18	38.8145	37.6001	**Kurt**	**3.3325**	**3.8931**

Table 11.

Eighth Grade ReadiStep Unrounded Raw to Scale Score Conversion: Writing

RS Raw	Random Groups	Single Group	RS Raw	Random Groups	Single Group
0	17.8134	17.8134	26	35.2780	35.4811
1	17.8134	17.8134	27	36.3228	36.4043
2	17.8134	17.8134	28	37.4103	37.3714
3	17.8134	17.8134	29	38.2571	38.1473
4	17.8134	17.8134	30	38.9075	38.7755
5	17.8134	17.8134	31	39.5514	39.3623
6	17.8134	17.8134	32	40.4984	40.1450
7	17.8134	17.8134	33	41.8387	41.3113
8	17.8134	17.8134	34	43.0642	42.6178
9	17.8134	18.2665	35	44.0034	43.6395
10	19.4865	20.2162	36	45.5121	44.7651
11	21.0098	22.0257	37	47.1599	46.4700
12	22.4880	23.5562	38	48.5072	47.9625
13	23.9660	25.7468	39	50.5551	49.1665
14	25.8905	27.2517	40	52.7974	51.5207
15	27.2427	27.8752	41	54.9107	53.4606
16	27.8285	28.5128	42	57.4992	55.1538
17	28.4321	29.1292	43	61.0208	57.2991
18	29.0545	29.9946	44	65.0081	60.4966
19	29.9004	30.8687	45	70.7101	63.1850
20	30.8332	31.8913			
21	31.9408	32.7603			
22	32.8712	33.4310	**Mean**	**34.7017**	**36.6505**
23	33.6356	33.9307	***SD***	**8.1402**	**7.8852**
24	34.1299	34.3689	**Skew**	**0.2927**	**0.1596**
25	34.6284	34.8010	**Kurt**	**3.3362**	**3.2910**

Table 12.

ReadiStep New Versus Old Scale Score: Critical Reading

	New Scale Scores			Old Scale Scores		
Raw Scores	Form A	Form B	Form C	Form A	Form B	Form C
0	1.0	1.0	1.0	2.0	2.0	2.0
1	1.5	1.5	1.5	2.0	2.0	2.0
2	1.5	1.5	1.5	2.0	2.0	2.0
3	1.6	1.6	1.6	2.0	2.0	2.0
4	1.8	1.8	1.8	2.0	2.0	2.0
5	1.9	1.9	1.9	2.2	2.2	2.2
6	2.0	2.0	2.0	2.6	2.6	2.6
7	2.1	2.1	2.1	2.8	2.8	2.8
8	2.3	2.3	2.3	3.0	3.0	3.0
9	2.5	2.5	2.4	3.4	3.4	3.2
10	2.6	2.6	2.5	3.6	3.6	3.4
11	2.8	2.8	2.7	3.8	3.8	3.6
12	3.0	3.0	2.9	4.0	4.0	3.8
13	3.1	3.1	3.0	4.2	4.2	4.0
14	3.3	3.2	3.2	4.4	4.2	4.2
15	3.3	3.3	3.3	4.4	4.4	4.4
16	3.4	3.4	3.4	4.6	4.6	4.6
17	3.6	3.5	3.5	4.8	4.6	4.6
18	3.6	3.6	3.6	4.8	4.8	4.8
19	3.7	3.7	3.7	5.0	5.0	5.0
20	3.9	3.8	3.8	5.2	5.0	5.0

Table 12. (cont.)

ReadiStep New Versus Old Scale Score: Critical Reading

	New Scale Scores			Old Scale Scores		
Raw Scores	Form A	Form B	Form C	Form A	Form B	Form C
21	3.9	3.9	3.9	5.2	5.2	5.2
22	4.0	4.0	4.0	5.2	5.2	5.2
23	4.0	4.0	4.0	5.4	5.4	5.4
24	4.1	4.1	4.1	5.4	5.4	5.4
25	4.2	4.2	4.2	5.6	5.6	5.6
26	4.3	4.3	4.3	5.6	5.6	5.6
27	4.4	4.4	4.4	5.8	5.8	5.8
28	4.5	4.5	4.5	5.8	5.8	5.8
29	4.6	4.6	4.6	6.0	6.0	6.0
30	4.7	4.7	4.7	6.0	6.0	6.0
31	4.8	4.8	4.8	6.2	6.2	6.2
32	4.9	5.0	5.0	6.2	6.4	6.4
33	5.1	5.1	5.1	6.4	6.4	6.4
34	5.5	5.5	5.5	6.6	6.6	6.6
35	5.7	5.8	5.8	6.6	6.8	6.8
36	6.0	6.1	6.1	6.8	7.0	7.0
37	6.3	6.4	6.4	7.0	7.2	7.2
38	6.6	6.7	6.7	7.2	7.4	7.4
39	6.9	7.0	7.0	7.6	7.8	7.8
40	7.0	7.0	7.0	8.0	8.0	8.0

Table 13.

ReadiStep New Versus Old Scale Score: Math

	New Scale Scores			Old Scale Scores		
Raw Scores	Form A	Form B	Form C	Form A	Form B	Form C
0	1.0	1.0	1.0	2.0	2.0	2.0
1	1.5	1.5	1.5	2.0	2.0	2.0
2	1.5	1.5	1.5	2.0	2.0	2.0
3	1.6	1.7	1.7	2.2	2.4	2.4
4	1.8	1.9	1.8	2.6	2.8	2.6
5	1.9	2.0	2.0	2.8	3.0	3.0
6	2.1	2.2	2.2	3.2	3.4	3.4
7	2.3	2.4	2.4	3.4	3.6	3.6
8	2.6	2.6	2.6	3.8	3.8	3.8
9	2.8	2.8	2.9	4.0	4.0	4.2
10	2.9	2.9	3.0	4.2	4.2	4.4
11	3.0	3.0	3.1	4.4	4.4	4.6
12	3.2	3.2	3.3	4.6	4.6	4.8
13	3.3	3.3	3.3	4.8	4.8	4.8
14	3.4	3.4	3.4	5.0	5.0	5.0
15	3.6	3.5	3.6	5.2	5.0	5.2
16	3.7	3.7	3.7	5.2	5.2	5.2
17	3.9	3.8	3.9	5.4	5.2	5.4
18	3.9	3.9	4.0	5.4	5.4	5.6
19	4.0	4.0	4.0	5.6	5.6	5.6
20	4.2	4.1	4.2	5.8	5.6	5.8
21	4.2	4.2	4.2	5.8	5.8	5.8
22	4.5	4.4	4.5	6.0	5.8	6.0
23	4.5	4.5	4.6	6.0	6.0	6.2
24	4.6	4.6	4.6	6.2	6.2	6.2
25	4.8	4.8	4.8	6.4	6.4	6.4
26	5.0	5.0	5.1	6.4	6.4	6.6
27	5.4	5.4	5.5	6.6	6.6	6.8
28	5.8	5.8	5.9	6.8	6.8	7.0
29	6.1	6.2	6.2	7.0	7.2	7.2
30	6.5	6.6	6.6	7.2	7.4	7.4
31	6.9	7.0	7.0	7.6	7.8	7.8
32	7.0	7.0	7.0	8.0	8.0	8.0

Table 14.

ReadiStep New Versus Old Scale Score: Writing

	New Scale Scores			Old Scale Scores		
Raw Scores	Form A	Form B	Form C	Form A	Form B	Form C
0	1.0	1.0	1.0	2.0	2.0	2.0
1	1.5	1.5	1.5	2.0	2.0	2.0
2	1.5	1.5	1.5	2.0	2.0	2.0
3	1.6	1.6	1.6	2.0	2.0	2.0
4	1.6	1.6	1.6	2.0	2.0	2.0
5	1.7	1.7	1.7	2.0	2.0	2.0
6	1.8	1.8	1.7	2.2	2.2	2.0
7	1.8	1.8	1.7	2.4	2.4	2.0
8	1.9	1.9	1.8	2.6	2.6	2.4
9	2.0	2.0	1.9	2.8	2.8	2.6
10	2.0	2.0	2.0	3.0	3.0	3.0
11	2.1	2.1	2.1	3.2	3.2	3.2
12	2.2	2.2	2.2	3.4	3.4	3.4
13	2.4	2.4	2.3	3.6	3.6	3.4
14	2.6	2.6	2.5	3.8	3.8	3.6
15	2.7	2.7	2.6	4.0	4.0	3.8
16	2.8	2.8	2.7	4.0	4.0	4.0
17	2.8	2.8	2.7	4.2	4.2	4.0
18	2.9	2.9	2.8	4.4	4.4	4.2
19	3.0	3.0	2.9	4.4	4.4	4.2
20	3.1	3.1	3.0	4.6	4.6	4.4
21	3.2	3.2	3.1	4.6	4.6	4.4
22	3.3	3.3	3.2	4.8	4.8	4.6
23	3.4	3.4	3.3	4.8	4.8	4.6
24	3.4	3.4	3.3	5.0	5.0	4.8
25	3.5	3.5	3.4	5.0	5.0	4.8
26	3.5	3.5	3.4	5.2	5.2	5.0
27	3.6	3.6	3.5	5.2	5.2	5.0
28	3.7	3.7	3.6	5.4	5.4	5.2
29	3.8	3.8	3.7	5.4	5.4	5.2
30	3.9	3.9	3.8	5.6	5.6	5.4
31	4.0	4.0	3.8	5.8	5.8	5.4
32	4.0	4.0	3.9	5.8	5.8	5.6
33	4.2	4.2	4.1	6.0	6.0	5.8
34	4.2	4.3	4.1	6.0	6.2	5.8
35	4.4	4.4	4.3	6.2	6.2	6.0
36	4.6	4.6	4.5	6.4	6.4	6.2
37	4.7	4.7	4.6	6.6	6.6	6.4

Table 14. (cont.)

ReadiStep New Versus Old Scale Score: Writing

	New Scale Scores			Old Scale Scores		
Raw Scores	Form A	Form B	Form C	Form A	Form B	Form C
38	4.9	4.9	4.8	6.8	6.8	6.6
39	5.2	5.2	5.1	7.0	7.0	6.8
40	5.6	5.6	5.5	7.2	7.2	7.0
41	5.9	5.9	5.8	7.4	7.4	7.2
42	6.3	6.3	6.2	7.6	7.6	7.4
43	6.6	6.6	6.5	7.8	7.8	7.6
44	7.0	7.0	7.0	8.0	8.0	8.0
45	7.0	7.0	7.0	8.0	8.0	8.0

Table 15.

ReadiStep Percentile Ranks

Raw Score	Critical Reading	Writing	Mathematics
1.0	-	-	-
1.5	0.0	0.0	0.0
1.6	0.0	0.0	0.1
1.7	0.1	0.0	0.2
1.8	0.1	0.2	0.4
1.9	0.2	0.7	0.9
2.0	0.4	1.3	1.7
2.1	1.0	3.0	2.9
2.2	1.9	4.5	3.7
2.3	1.9	6.4	5.5
2.4	3.4	7.1	6.6
2.5	4.1	8.8	9.5
2.6	6.5	9.6	9.5
2.7	8.5	12.2	14.6
2.8	9.9	16.3	14.6
2.9	12.5	21.8	18.7
3.0	13.9	25.5	24.9
3.1	18.2	29.1	31.1
3.2	21.5	33.0	33.5
3.3	24.8	36.7	37.7
3.4	31.3	42.2	46.1
3.5	36.2	50.2	51.4
3.6	39.3	56.9	53.0
3.7	45.8	60.8	56.7
3.8	50.1	65.0	61.4
3.9	52.6	70.8	62.9
4.0	57.8	74.8	68.7
4.1	65.3	79.9	74.3
4.2	69.1	82.5	75.8
4.3	72.6	85.5	81.4
4.4	76.1	87.5	81.4
4.5	78.9	89.2	82.6
4.6	81.3	90.2	86.3
4.7	83.9	92.8	90.1
4.8	86.6	94.2	90.1
4.9	89.0	94.9	92.4
5.0	89.8	95.9	92.4
5.1	91.5	95.9	93.7
5.2	93.5	96.7	94.4

Table 15. (cont.)

ReadiStep Percentile Ranks

Raw Score	Critical Reading	Writing	Mathematics
5.3	93.5	97.5	94.4
5.4	93.5	97.5	94.4
5.5	93.5	97.5	95.6
5.6	95.1	97.9	96.1
5.7	95.1	98.4	96.1
5.8	95.6	98.4	96.1
5.9	96.6	98.8	97.2
6.0	96.6	99.2	97.6
6.1	97.2	99.2	97.6
6.2	97.9	99.2	97.9
6.3	97.9	99.3	98.5
6.4	98.4	99.6	98.5
6.5	99.0	99.6	98.5
6.6	99.0	99.7	98.9
6.7	99.2	99.8	99.3
6.8	99.6	99.8	99.3
6.9	99.6	99.8	99.3
7.0	99.7	99.8	99.6

Table 16.

ReadiStep Descriptive Statistics of the 2011 Administration, After Applying the New Scale

Section	*N*	Mean	*SD*	Skewness	Kurtosis	Minimum	Maximum
Critical Reading	163,936	3.5	0.99	0.53	0.58	1.00	7.00
Math	163,936	3.6	1.01	0.80	1.14	1.00	7.00
Writing	163,936	3.4	0.88	0.66	1.29	1.00	7.00

Table 17.

2008 ReadiStep and 2010 PSAT/NMSQT Descriptive Statistics

	2008 ReadiStep			2010 PSAT/NMSQT		
Section	***N***	**Mean**	***SD***	***N***	**Mean**	***SD***
Matched Sample						
Critical Reading	3,916	4.1	0.9	3,916	43.9	10.7
Math	4,076	3.9	1.0	4,076	44.9	10.6
Writing	4,210	3.7	0.8	4,210	41.5	10.5
Unmatched Sample						
Critical Reading	11,587	3.8	0.9	333,529	39.8	10.6
Math	12,189	3.6	0.9	333,529	41.3	10.3
Writing	12,636	3.5	0.8	333,529	38.0	10.0

Figure 1.

ReadiStep smoothed and empirical raw score distributions (random groups design, critical reading, *m* 6).

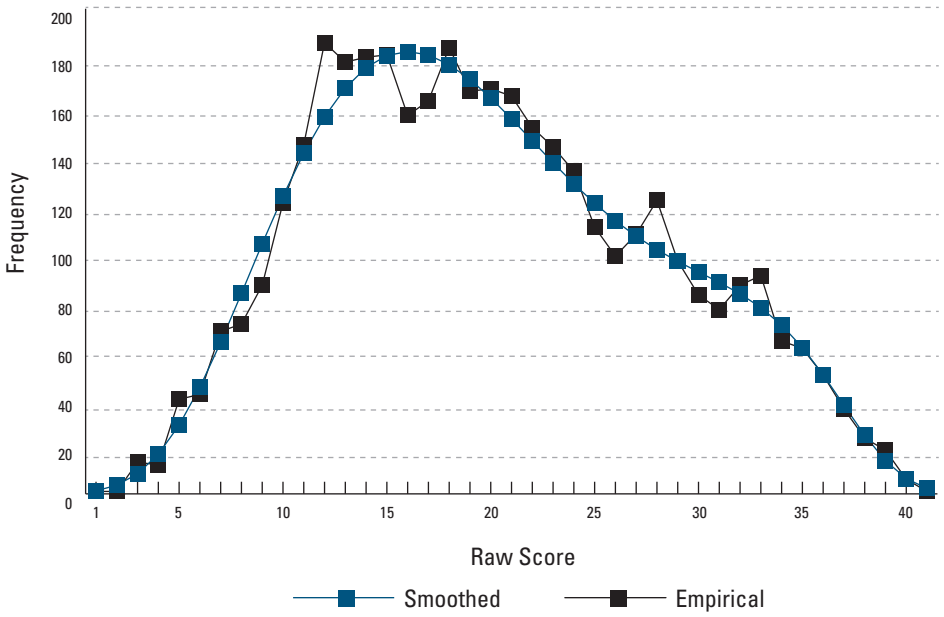


Figure 2.

PSAT/NMSQT smoothed and empirical raw score distributions (random groups design, critical reading, *m* 5).

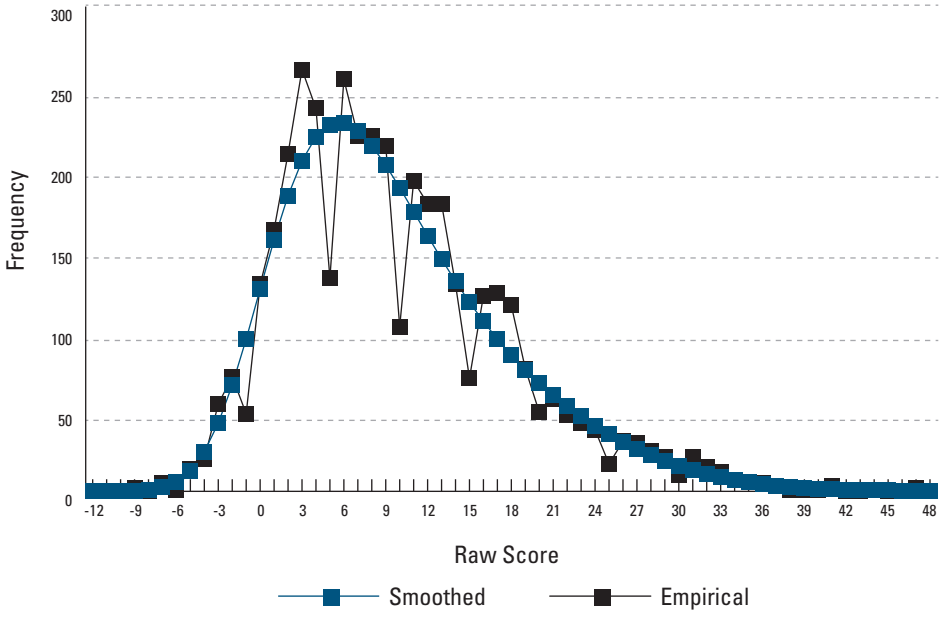


Figure 3.

ReadiStep smoothed and empirical raw score distributions (random groups design, math, *m* 5).

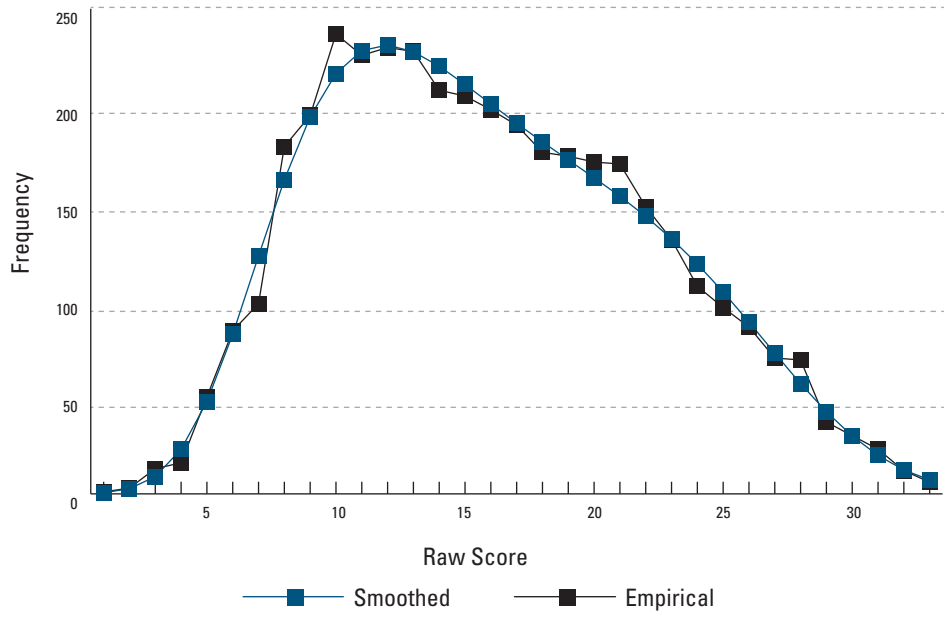


Figure 4.

PSAT/NMSQT smoothed and empirical raw score distributions (random groups design, math, *m* 6).

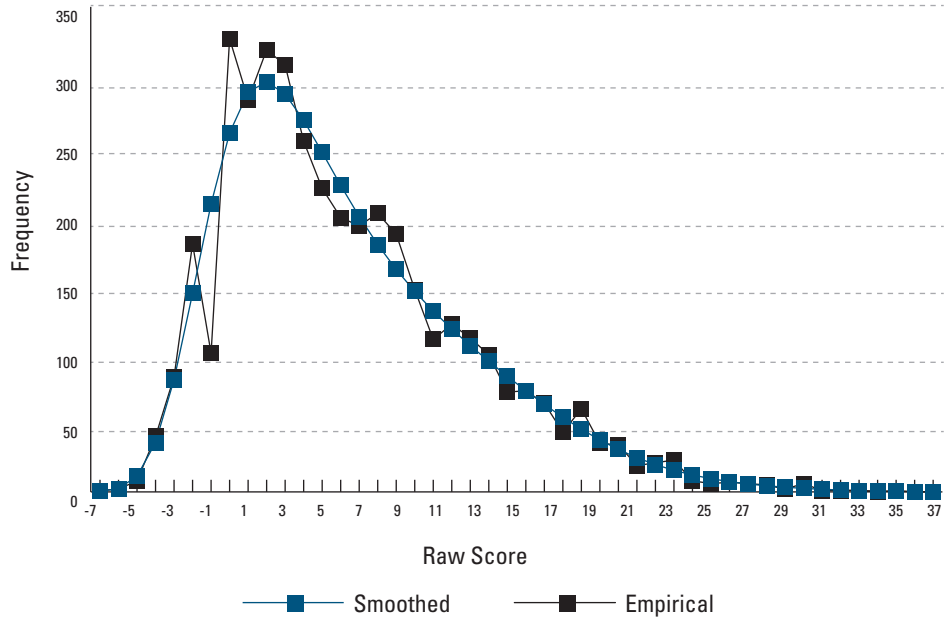


Figure 5.

ReadiStep smoothed and empirical raw score distributions (random groups design, writing, *m* 6).

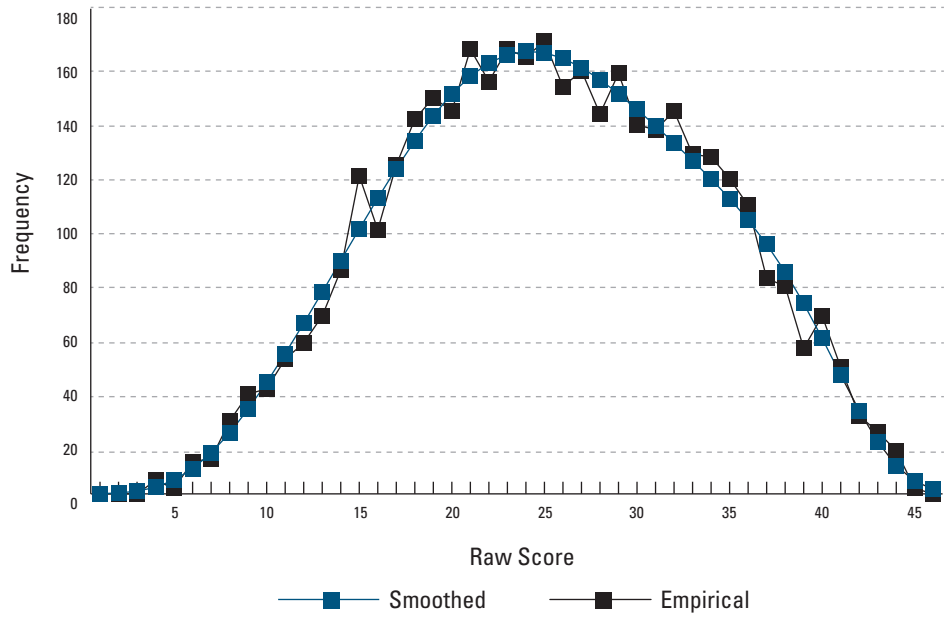


Figure 6.

PSAT/NMSQT smoothed and empirical raw score distributions (random groups design, writing, *m* 5).

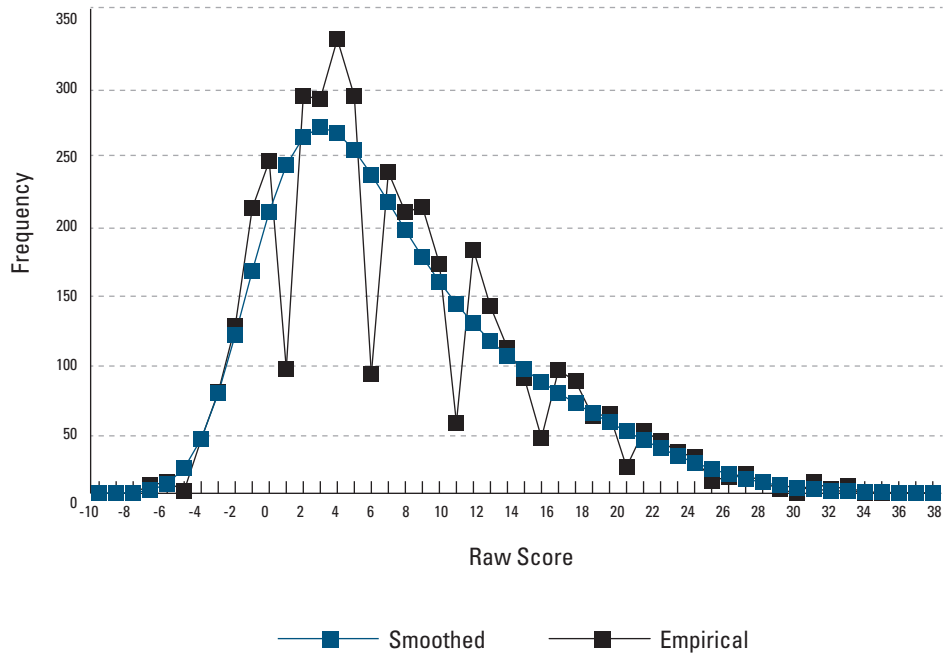


Figure 7a.

Scatterplot of ReadiStep versus PSAT/NMSQT raw scores in single-group design sample — critical reading.

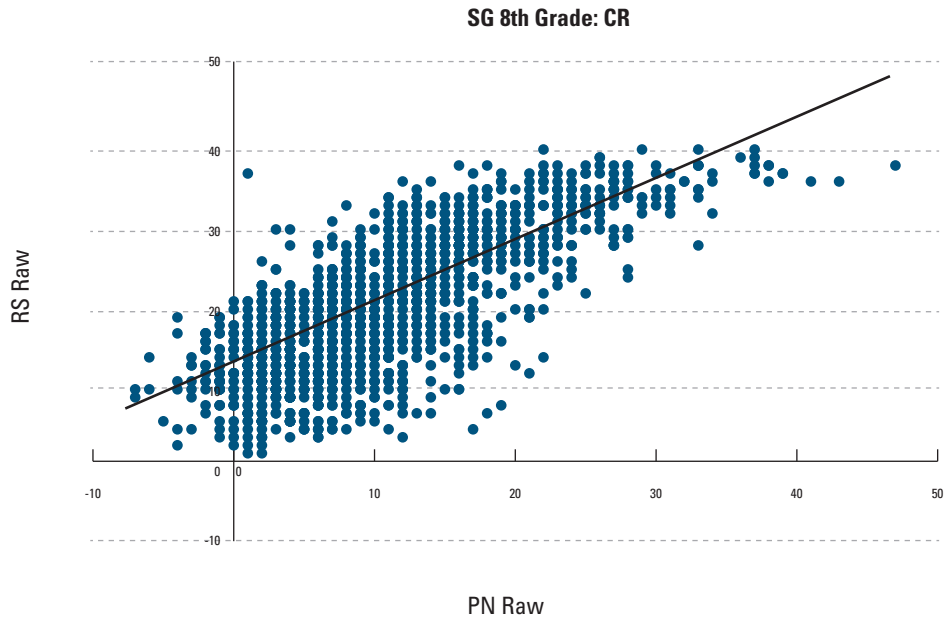


Figure 7b.

Scatterplot of ReadiStep versus PSAT/NMSQT raw scores in single-group design sample — math.

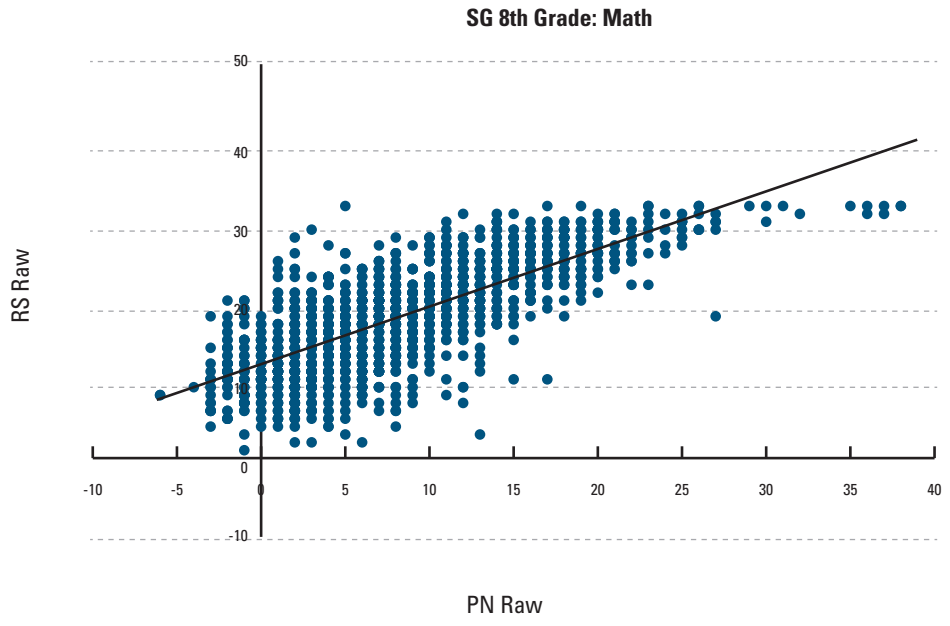


Figure 7c.

Scatterplot of ReadiStep versus PSAT/NMSQT raw scores in single-group design sample — writing.

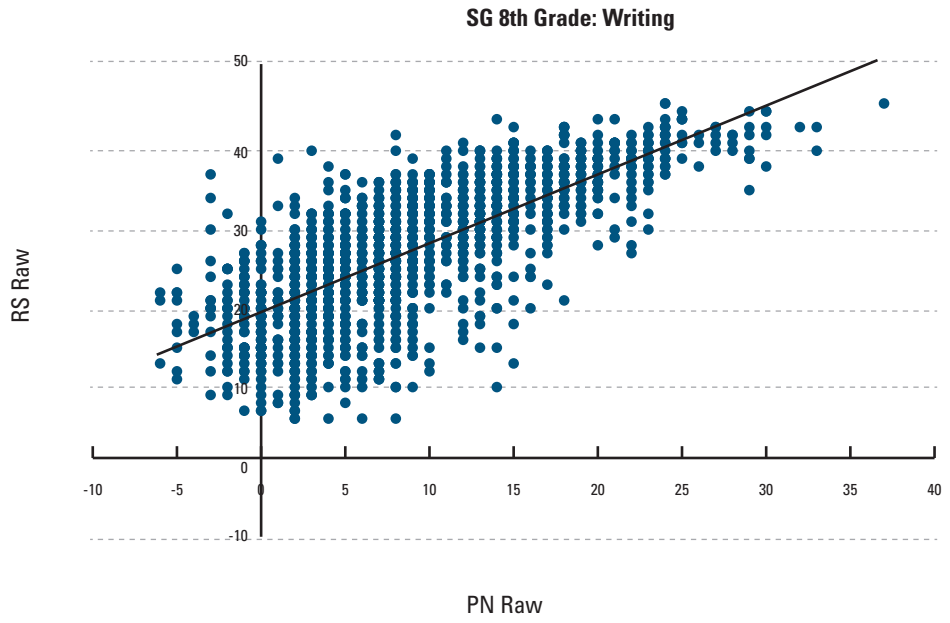


Figure 8.

Unrounded ReadiStep to PSAT/NMSQT conversion lines for the all-single-group and random-group samples — critical reading.

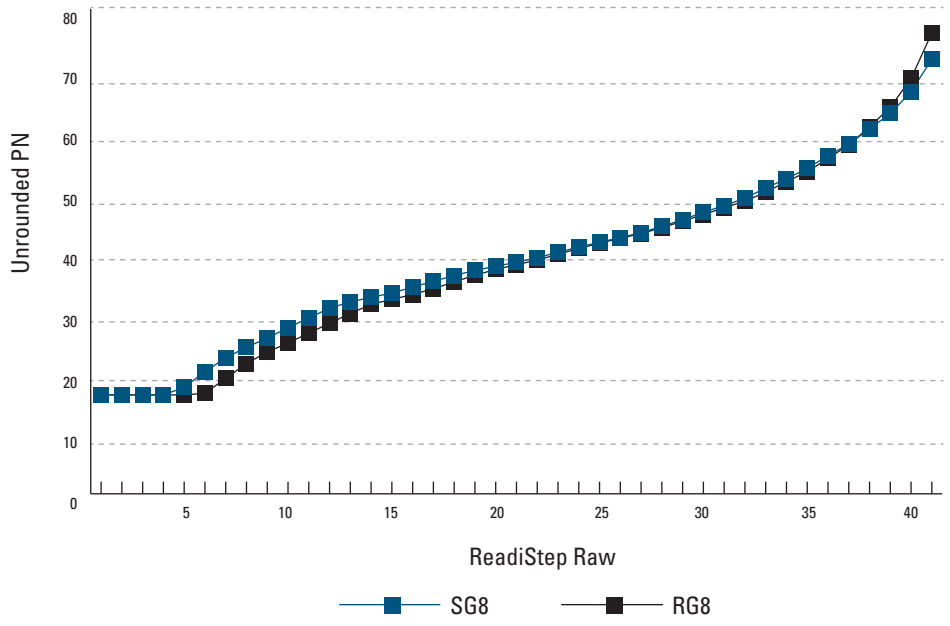


Figure 9.

Difference between single group and random groups ReadiStep to PSAT/NMSQT conversion lines — critical reading.

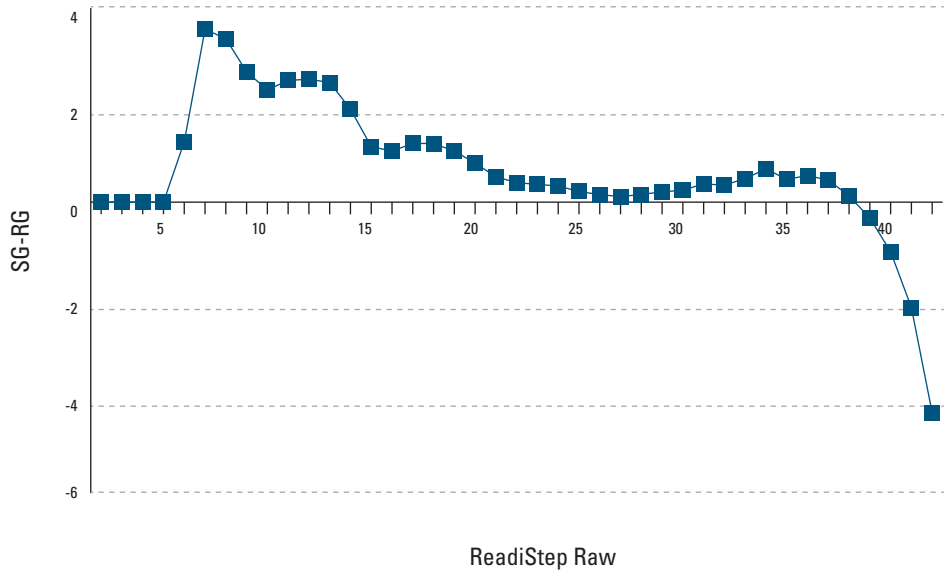


Figure 10.

Unrounded ReadiStep to PSAT/NMSQT conversion lines for the all single group and random groups samples — mathematics.

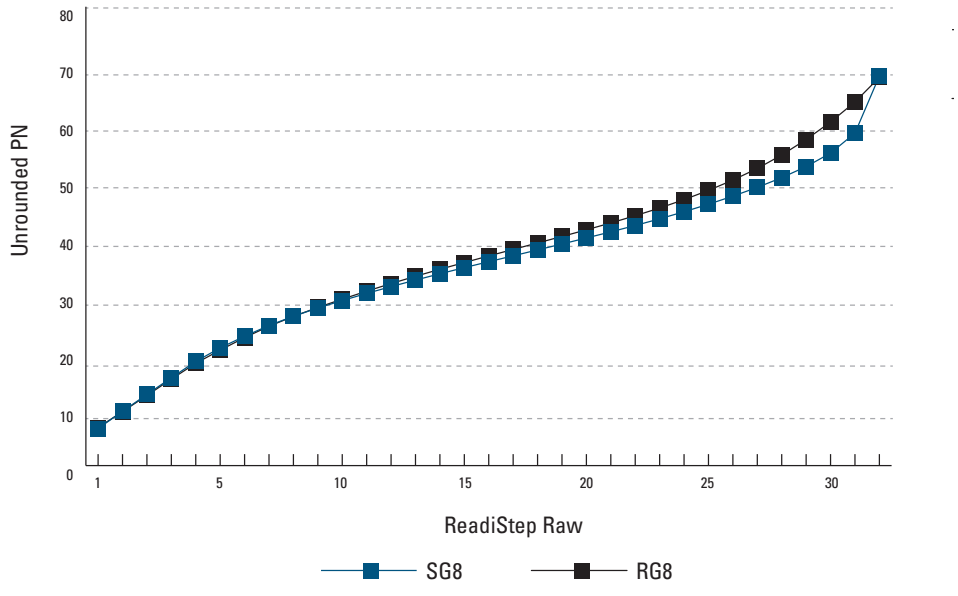


Figure 11.

Difference between single group and random groups ReadiStep to PSAT/NMSQT conversion lines — mathematics.

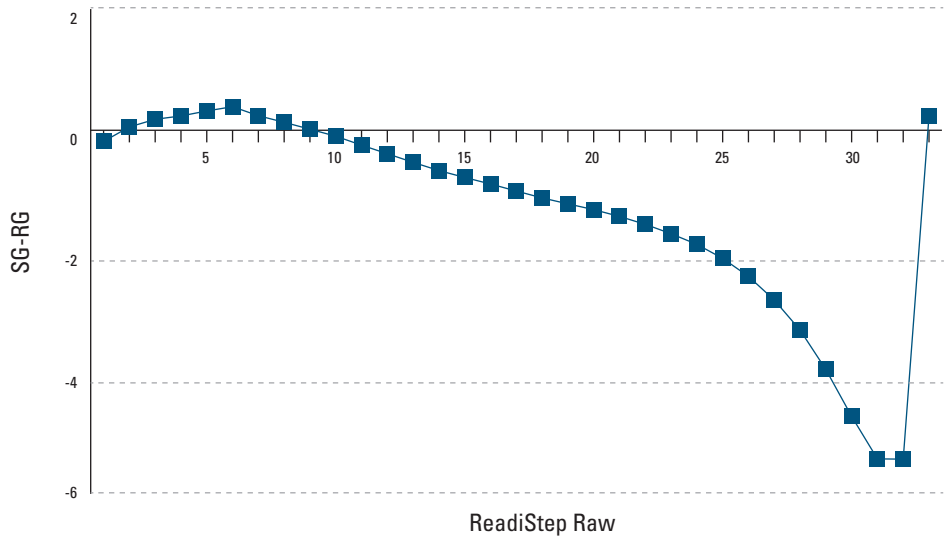


Figure 12.

Unrounded ReadiStep and PSAT/NMSQT conversion lines for the all single group and random groups samples — writing.

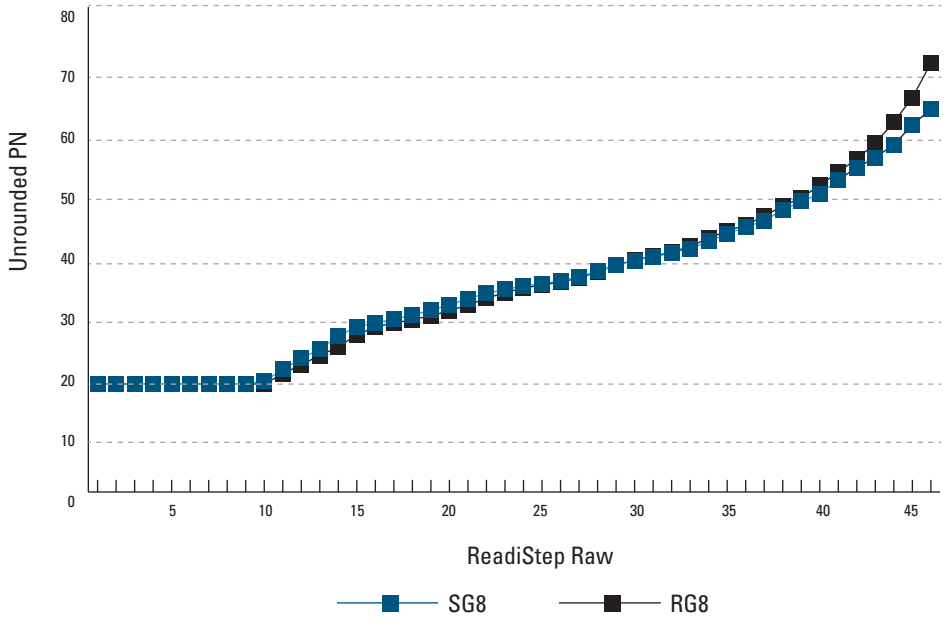


Figure 13.

Difference between single group and random groups ReadiStep to PSAT/NMSQT conversion lines — writing.

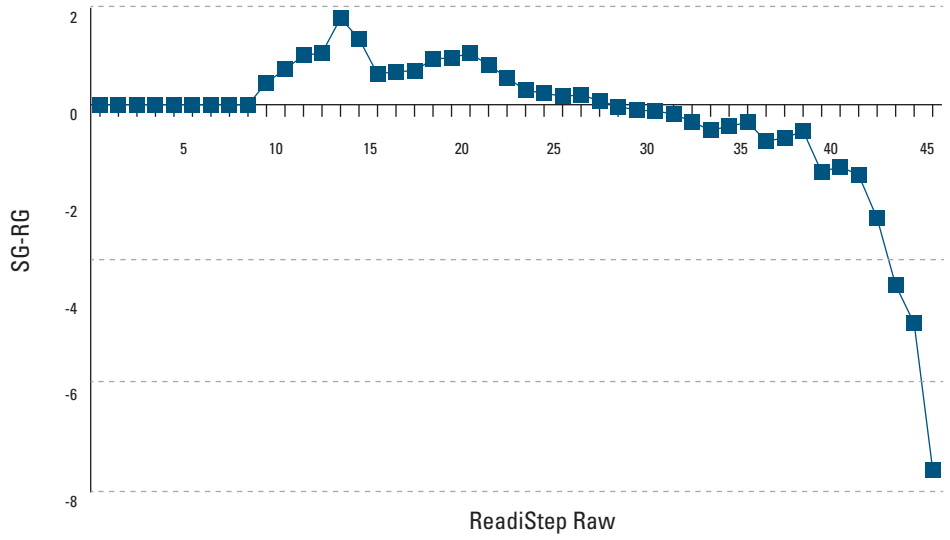


Figure 14.

New ReadiStep scale score distribution: critical reading.

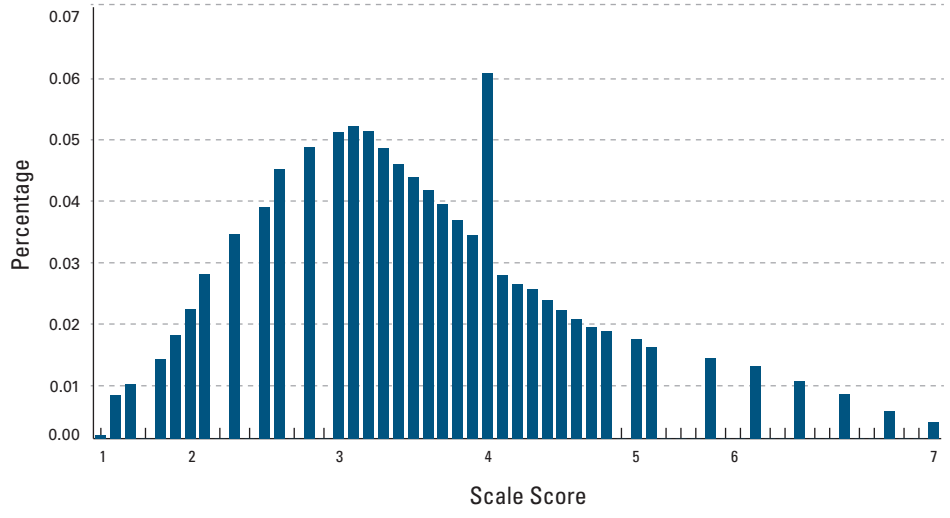


Figure 15.

New ReadiStep scale score distribution: math.

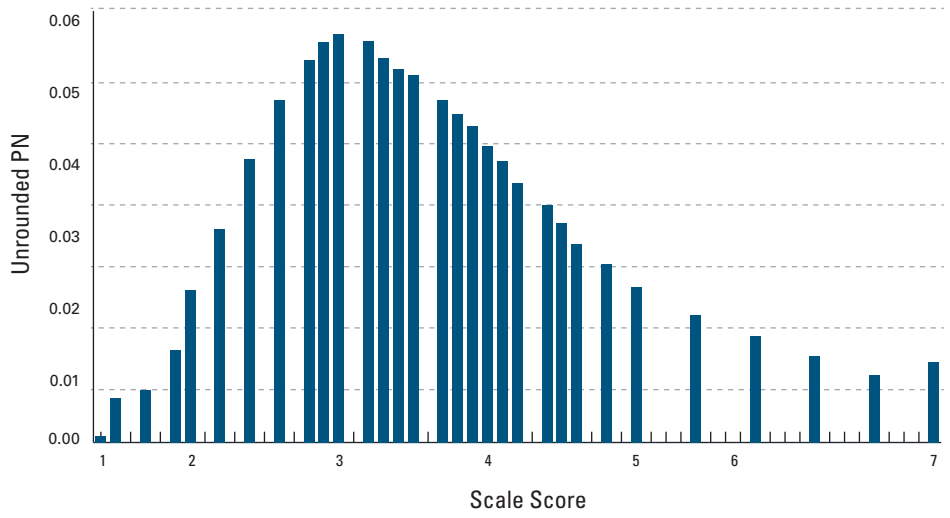
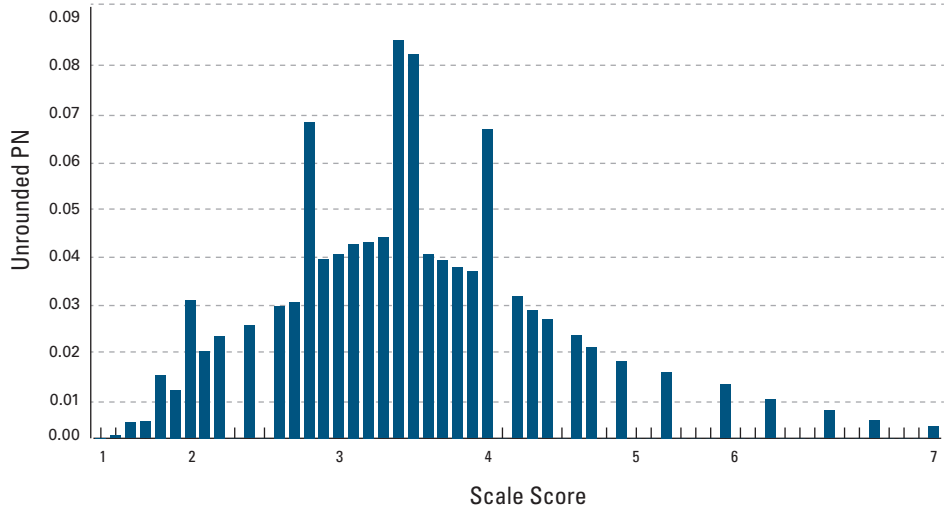


Figure 16.

New ReadiStep scale score distribution: writing.



The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

CollegeBoard

13b-7814