RESEARCH REPORT 2011-7

CollegeBoard

# Examining the Linearity of the PSAT/NMSQT®–FYGPA Relationship

By Jessica P. Marini, Krista D. Mattern, and Emily J. Shaw

**Jessica P. Marini** is a graduate student intern at the College Board.

**Krista D. Mattern** is an associate research scientist at the College Board.

**Emily J. Shaw** is an associate research scientist 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 more than 5,900 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.



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

VALIDIT

**Visit the College Board on the Web: www.collegeboard.org.**

# Contents

## Executive Summary

There is a common misperception that test scores do not predict above a minimum threshold (Sackett, Borneman, & Connelly, 2008). That is, test scores may be useful for identifying students with very low levels of ability; however, higher scores are considered unrelated to higher performance for those above a certain threshold. This study aims to examine whether this is true for the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®), which is used for that very purpose — to differentiate among very high performing students. The linearity of the relationship between PSAT/NMSQT scores and first-year college GPA (FYGPA) was explored in this paper, using a regression approach. This relationship was explored over the entire range of the PSAT/NMSQT score scale, known as the Selection Index, ranging from 60 to 240 as well as the upper end of the score scale (≥ 200), where initial screening decisions are made for scholarship programs conducted by National Merit Scholarship Corporation (NMSC). For the full PSAT/NMSQT scale, the addition of a quadratic term improved model fit; however, the effect size was small as indexed by the change in the squared multiple correlation coefficient ($R^2$) of 0.001. That is, including PSAT/NMSQT Selection Index$^2$ in the model accounted for an additional 0.1% of variance in FYGPA. As for the subset of students who had a PSAT/NMSQT score of 200 or higher, the results indicated a strong linear relationship, which suggests that even among very high-scoring students, the PSAT/NMSQT score scale differentiates between students in terms of academic success measured by grades earned in the first year of college. In sum, the results of this study support the use of the PSAT/NMSQT as a screening tool for selecting Merit Scholarship winners.

## Introduction

The Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®) is an assessment, cosponsored by the College Board and National Merit Scholarship Corporation (NMSC), which offers multiple opportunities to high school juniors, sophomores, and freshmen. Specifically, the PSAT/NMSQT provides students with the opportunity to prepare for the SAT®, enter scholarship competitions, gain information from colleges, and start planning for college and their careers while assessing academic skills. The test has three sections, critical reading, mathematics, and writing, which measure reading, math reasoning, writing, and critical thinking skills that have developed over years of study and experiences inside and outside of the classroom. Scores on each of the three PSAT/NMSQT sections, ranging from 20 to 80, can be compared to scores on the corresponding SAT sections, which range from 200 to 800. When taken during the junior year of high school, the PSAT/NMSQT serves as the qualifying test for National Merit Scholarship Corporation's various scholarship programs. The PSAT/NMSQT Selection Index is the composite score of each of the section scores ranging from 60 to 240. It serves as the initial criterion for the NMSC screening process for its two major scholarship programs, the National Merit Scholarship Program and the National Achievement Scholarship Program. Both of these programs are rigorous academic competitions. For example, the specific goals of the National Merit Scholarship Program set forth in the *Guide to the National Merit Scholarship Program*, include the following:

- identify and honor academically talented U.S. high school students and encourage them to pursue rigorous college studies;
- provide professional services for corporations, company foundations, colleges and universities, and other organizations that wish to sponsor scholarships for outstanding participants in the competition;
- promote a broader and deeper respect for learning in general and for exceptionally talented individuals in particular;
- stimulate increased support for the education of scholastically able students; and
- encourage the pursuit of academic excellence at all levels of education.

NMSC selects approximately 50,000 students from the more than 1.5 million entrants to qualify for recognition in NMSC programs and compete for scholarships. Of these roughly 50,000 students, about 34,000 are selected to receive Letters of Commendation, and about 16,000 are selected to qualify as Semifinalists in the National Merit Scholarship Program.

Only high school juniors who attain a very high score on the Selection Index will advance to become Semifinalists in the Merit and/or Achievement programs. Semifinalists then must meet multiple, additional requirements (submit a completed Semifinalist application with an essay, a recommendation, evidence of leadership and extracurricular activities, and an endorsement from the high school; be in good academic standing; earn SAT scores that confirm PSAT/NMSQT performance, etc.) to progress to Finalist standing. The scholarship recipients in the competitions are selected from the candidates in the Finalist pool.

Using the PSAT/NMSQT as the initial screening criteria for the National Merit Scholarship Program has created some controversy. The National Association for College Admission Counseling (NACAC) indicated that they found the use of a single test score not appropriate for selecting students for a merit-based scholarship (2008). Hayashi (2005) also argued that using the PSAT/NMSQT Selection Index to select the first round of students to compete in the NMSC's scholarship competitions was unfair, adding that having a cutoff discriminates against students who have excellent high school grade point averages (HSGPAs) but have scored poorly on the PSAT/NMSQT. In response to criticisms from NACAC, the NMSC stressed that the entire selection process is quite rigorous and multiple factors including leadership, essays, and student grades are used to make final decisions (McGuire, 2009). They noted that using the PSAT/NMSQT Selection Index as an initial way of examining students is "the most effective, inclusive, and equitable available [way] to consider over 1.5 million students annually on a consistent basis" (McGuire, 2009, p.1).

In addition to the criticisms mentioned above, the PSAT/NMSQT also must respond to and refute commonly held misconceptions about tests and their uses, including whether scores at the top end of the scale meaningfully discriminate among test-takers. In an article reviewing commonly held test misperceptions, Sackett, Borneman, and Connelly (2008) pointed out that there is a misconception that tests are unable to distinguish among high scorers and are only useful for weeding out low scorers. However, research has shown that the ability–performance relationship is linear throughout the entire range of ability, across various domains. This linear relationship provides evidence that tests can distinguish among different ability levels equally and are not only useful for weeding out low-scoring individuals. In other words, with each increase in test scores, there is an associated increase in performance even among very top performers.

There is a misconception that tests are unable to distinguish among high scorers and are only useful for weeding out low scorers.

Studies examining the ability–performance relationship in the occupational domain have concluded that the relationship is linear across the ability scale. For example, Greener and Osburn (1979) examined the ability–performance relationship of 5,900 white-collar managerial and professional employees using the Miller Analogies Test, Non-Verbal Reasoning Ability Test, Guilford-Zimmerman Temperament Survey, a company-developed survey of biographical information, and an overall employee effectiveness rating to predict managerial success. They found a linear relationship between the ability measures and managerial success. Reviewing the literature on the General Aptitude Test Battery (GATB)–job performance relationship, Hawk (1970) found that while some of the 367 studies showed a nonlinear relationship existed between the GATB and job performance, this number did not exceed the number expected by chance, thus indicating a linear relationship exists. Similarly, based on 174 studies ($N$ = 36,614), Coward and Sackett (1990) conducted a meta-analysis to examine whether the relationship between the nine scales of the GATB and job performance is linear. To detect deviations from linearity, they used the $r$ versus eta method that Hawk originally used, as well as power polynomials. Overall, they found that the power polynomial approach had more statistical power in detecting nonlinearity. That being said, the power polynomial approach did not find deviations from linearity at a high rate. They concluded that this low occurrence of nonlinearity, coupled with a highly powerful tool, provided evidence for the argument that the ability–performance relationship is indeed linear.

Research on the linearity of ability–performance relationships has also been conducted in the educational domain with similar findings (Arneson, 2007; Cullen, Hardison, & Sackett, 2004).

That is, research has shown that the relationship between academic tests and performance measures such as student grades, other academic outcomes, or graded performance measures demonstrate linearity. Specifically, based on nearly 50,000 students across 13 universities, a study by Cullen, Hardison, and Sackett (2004) found that the relationship between SAT scores and college GPA by gender and race was linear. Similarly, Arneson (2007) investigated the linearity of the ability–performance relationship among high-scoring students. Three different datasets were analyzed, including data from the College Board with SAT verbal (now critical reading) and math scores and first-year GPA (FYGPA) for about 167,000 students; data from the National Educational Longitudinal study of 1988 (NELS:88) with self-reported grades for approximately 25,000 eighth-graders; and data from Project TALENT with self-reported HSGPA for about 300,000 high school students (grade 9–12). Based on the entire score scale, the results indicated a linear relationship between SAT scores and FYGPA since the addition of a quadratic term (i.e., $SAT^2$) did not result in a statistically significant change in model fit. However, for the NELS:88 and Project TALENT datasets, there was slight evidence of a deviation from linearity for the ability–grades relationship across the entire score scale.

Arneson's 2007 study, which also examined the linearity of the ability–performance relationship among high scoring students, is particularly relevant to the current study. In the upper tail of the SAT-FYGPA relationship, defined as one standard deviation above the mean SAT score and higher (mean = 1140, SD = 169), slight deviations from linearity were detected. It should be pointed out that even though the quadratic term was statistically significant, the effect size was small, with a $\Delta R^2$ of 0.000. In the NELS:88 and Project TALENT datasets, the top of the ability distributions did not deviate from linearity.

Research on gifted children can also shed light on the question of whether test scores meaningfully discriminate among very high-scoring students. Research started by Keating and Stanley (1972) investigated gifted students' performance on the SAT as part of the Study of Mathematically Precocious Youth (SMPY). They collected data from seventh- and eighth-graders who qualified as being gifted and were given the opportunity to take the SAT during their seventh- or eighth-grade year. Benbow (1992) analyzed the top 1 percent of the scorers in the first SMPY cohort (1972–1974) and found that the top quartile of the top 1 percent earned significantly higher college GPAs compared to those of the bottom quartile of the top 1 percent. Statistically significant results were also found for variables such as intellectual level of college attended and math/science GPA among other results. Wai, Lubinski, and Benbow (2005) extended Benbow's (1992) study by using the SMPY data to examine the relationship between SAT scores and occupational outcome measures. The first two cohorts of SMPY data, which included data collected 20 years after the original collection, were analyzed. Similar to the analyses conducted by Benbow (1992), the top and bottom quartile of the top 1 percent of scorers in seventh and eighth grades were compared in relation to earned doctorates, primary income, number of patents, and tenure 20 years later. A significant difference was found for each of the variables, indicating that not only can comparisons be made in such a selective sample, but also that scores from adolescence are predictive of occupational outcomes 20 years later.

The current study will build on previous research by examining the linearity of the ability–performance relationship for another highly visible and utilized educational exam, the PSAT/NMSQT®. Specifically, this study will examine whether the relationship between the PSAT/NMSQT Selection Index (CR + M + W) and FYGPA is linear and whether the PSAT/NMSQT differentiates between high-scoring students. Although the Selection Index is one of many tools to identify students for merit-based scholarships, this research is extremely important. Results will inform the validity of using PSAT/NMSQT scores to discriminate among students at the very top of the score scale in terms of educational outcomes. If the relationship statistically and practically deviates from linearity, then alternative methods for selecting students may need to be explored. However, previous research predicts a linear relationship throughout the entire ability range, including at the very high end of the score scale range.

The current study will build on previous research by examining the linearity of the ability–performance relationship for another highly visible and utilized educational exam, the PSAT/NMSQT®.

# Method

## Participants

The data used for this study are from three cohorts of first-time, first-year students entering college in the fall of 2006, 2007, or 2008 and were collected for the purpose of conducting national validity research on the SAT and other College Board tests (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Patterson, Mattern, & Kobrin, 2009); for detailed information see the original Kobrin et al. (2008) study. The total sample across the three cohorts included 654,696 students from 177 colleges and universities across the United States. To be included in analyses, students had to have a PSAT/NMSQT score and a FYGPA (N = 444,193). As shown in Table 1, the sample was 54.5 percent female. As for race/ethnicity, 15.9 percent were underrepresented minorities, 9.7 percent were Asian, and 64.4 percent of the students were white, while 10.0 percent were missing this information. These findings suggest that the current sample is fairly representative of the national population of students enrolled in four-year institutions in terms of both gender and race/ethnicity (Snyder & Dillow, 2010). As for institutional characteristics, the majority of students attended a public, very large, moderately selective institution.

## Measures

**PSAT/NMSQT scores.** Official junior-year PSAT/NMSQT scores were obtained from College Board records. The PSAT/NMSQT consists of three sections: critical reading ($M = 53.63$, $SD = 9.45$), mathematics ($M = 55.90$, $SD = 9.58$), and writing ($M = 54.74$, $SD = 9.99$), each with a score scale range of 20 to 80. The PSAT/NMSQT Selection Index ($M = 164.27$, $SD = 25.33$) ranges from 60 to 240 and is the sum of the individual scores on the three PSAT/NMSQT sections.

**First-Year GPA (FYGPA).** Participating institutions provided FYGPA for all first-year, full-time students. The range, across cohorts and institutions, is 0.00 to 4.27 ($M = 3.02$, $SD = .69$).

## Analysis

**Graphical Analysis.** A scatterplot of mean FYGPA by PSAT/NMSQT Selection Index was produced to visually inspect the relationship. The plot was examined for evidence of a linear relationship.

**Table 1**

Sample Distribution by Relevant Study Variables

| Variable | | N | Percent |
|---|---|---|---|
| Cohort | 2006 | 136,795 | 30.8 |
| | 2007 | 141,664 | 31.9 |
| | 2008 | 165,734 | 37.3 |
| Gender | Female | 242,297 | 54.5 |
| | Male | 201,896 | 45.5 |
| Ethnicity | American Indian or Alaska Native | 1,952 | 0.4 |
| | Asian, Asian American, or Pacific Islander | 42,947 | 9.7 |
| | Black or African American | 25,346 | 5.7 |
| | Hispanic | 31,912 | 7.2 |
| | White | 285,989 | 64.4 |
| | Other | 11,721 | 2.6 |
| | Missing | 44,326 | 10.0 |
| Control | Private | 134,130 | 30.2 |
| | Public | 310,063 | 69.8 |
| Selectivity | under 50% | 82,215 | 18.5 |
| | 50 to 75% | 278,553 | 62.7 |
| | over 75% | 83,425 | 18.8 |
| Size | Small: 750 to 1,999 undergraduates | 19,146 | 4.3 |
| | Medium to Large: 2,000 to 7,499 undergraduates | 85,772 | 19.3 |
| | Large: 7,500 to 14,999 undergraduates | 94,465 | 21.3 |
| | Very large: 15,000 or more undergraduates | 244,810 | 55.1 |
| Total | | 444,193 | 100.0 |

**Power Polynomial Approach.** The main goal of this study was to examine the relationship between FYGPA and PSAT/NMSQT Selection Index scores. There was a natural hierarchy inherent in the data with students within schools within cohorts. Means were calculated by cohort across institutions for FYGPA and PSAT/NMSQT Selection Index. There was little variability in means by cohort year; therefore, no correction for cohort was applied. However, means were also calculated by institution across cohorts, and there were notable differences in FYGPA. To adjust for grading differences across institutions, a dummy variable was created for each of the 177 colleges/universities included in the sample to control for institutional level differences. Specifically, for each of these institutions, an indicator variable was created. Then a linear regression between FYGPA and PSAT/NMSQT Selection Index with the school indicators ($n = 176$) was run. The coefficients of each of the indicator variables were added to the FYGPA as a correction for school difference. Then, based on the corrected FYGPA, linear and quadratic models were run to examine the PSAT/NMSQT-FYGPA relationship. Improvement in model fit of the quadratic model compared to the linear model was examined based on $\Delta R^2$, as was done by Arneson (2007) and Coward and Sackett (1990).

The linearity of this relationship was also examined at the upper end of the PSAT/NMSQT Selection Index score scale. It is at this range of the score scale that scholarship selection decisions are made. Therefore, it is even more imperative to examine linearity at this range of the score scale to determine whether there is evidence to support its current use. A score of 200 was chosen because the threshold to qualify for program recognition, the first hurdle of the multihurdle PSAT/NMSQT selection process identifying roughly 50,000 "high scoring" students nationally each year, is usually around this value. The sample included 38,719 students, with PSAT/NMSQT Selection Index scores of 200 and higher ($M=210.12$, $SD=8.39$).

## Results

Figure 1 shows a scatterplot of mean FYGPA by PSAT/NMSQT Selection Index. The mean FYGPA was calculated for each PSAT/NMSQT scale score. From this plot, an apparent linear trend in the data is revealed. In the figure, there are deviations from linearity at the lower end of the PSAT/NMSQT scale, occurring at about 100 and below. This is due to the fact there are fewer cases at the bottom of scale and thus more sampling error. Specifically, 1,946 students earned a PSAT/NMSQT of 100 or lower, which translates to less than 0.5 percent of the sample distributed across the bottom quartile of the score scale range. Since the plot represents mean FYGPA, the less data existing per PSAT/NMSQT score point, the more possible fluctuation. No scholarship decisions are made at this range of the score, and therefore departures for linearity are less of an issue.

**Figure 1**
Mean FYGPA by PSAT/NMSQT Selection Index Score

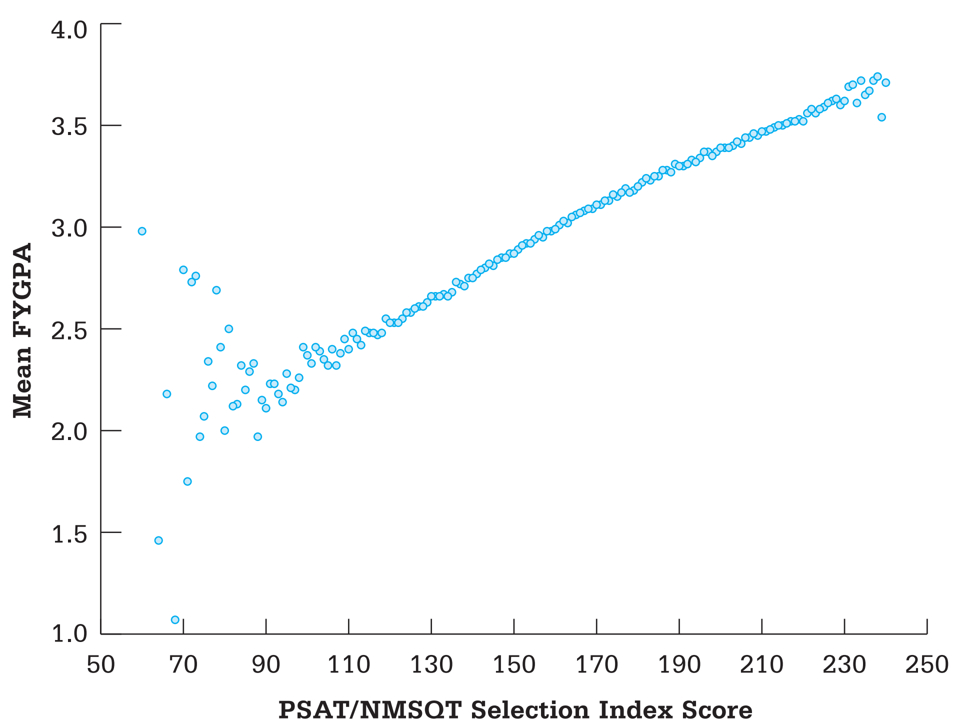



Table 2 provides the coefficients and $\Delta R^2$ values from the linear regressions run between the adjusted FYGPA and PSAT/NMSQT. In the full dataset, the coefficient of the quadratic term is small but significant ($B$ = -0.000, $p < .001$) and the $\Delta R^2$ associated with the addition of the quadratic term was .001. The addition of the quadratic terms does improve model fit; however, the effect size is small, accounting for only an additional 0.1 percent of the variance in FYGPA. While statistically significant, the sample size for the study is very large, and thus has the power to detect very small differences based on power alone. In the analysis based on the top end of the PSAT/NMSQT Selection Index score scale (Model 2), the $\Delta R^2$ was not significant (.000), indicating a linear relationship.

**Table 2**

Coefficients and Effect Sizes of the Change in Model Fit Based on the Power Polynomial Analysis

| | First-Year GPA | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 (Full Sample) | | | | Model 2 (High Scorers) | | | |
| | linear | | quadratic | | linear | | quadratic | |
| Variable | *B* | *SE B* | *B* | *SE B* | *B* | *SE B* | *B* | *SE B* |
| Constant | 2.961** | 0.001 | 2.980** | 0.001 | 3.421** | 0.003 | 3.421** | 0.004 |
| PSAT/NMSQT Selection Index | 0.011** | 0.000 | 0.011** | 0.000 | 0.008** | 0.000 | 0.008** | 0.000 |
| PSAT/NMSQT Selection Index$^2$ | | | -0.000** | 0.000 | | | – 0.000 | 0.000 |
| $R^2$ | | 0.149 | | 0.150 | | 0.016 | | 0.016 |
| $\Delta R^2$ | | | | 0.001* | | | | 0.000 |

Note: *B* = raw regression coefficient; *SE B* = standard error of the raw regression coefficient; $R^2$ = squared multiple correlation; $\Delta R^2$ = incremental validity. * $p < 0.01$, ** $p < 0.001$

## Discussion

Congruent with previous research on the ability–performance relationship, the results of the current study found that the PSAT/NMSQT–FYGPA relationship appears to be linear with only slight deviations from linearity for the full score scale. This was evidenced both graphically (see Figure 1) and based on the regression analyses, in particular among the top end of the distribution. The results suggest that even among very high-scoring students, the PSAT/NMSQT score scale differentiates among students in terms of academic success as measured by first-year college grades. That is, higher PSAT/NMSQT scores are related to higher grades earned in college, supporting its use as a screening tool in the overall process of selecting Merit Scholarships.

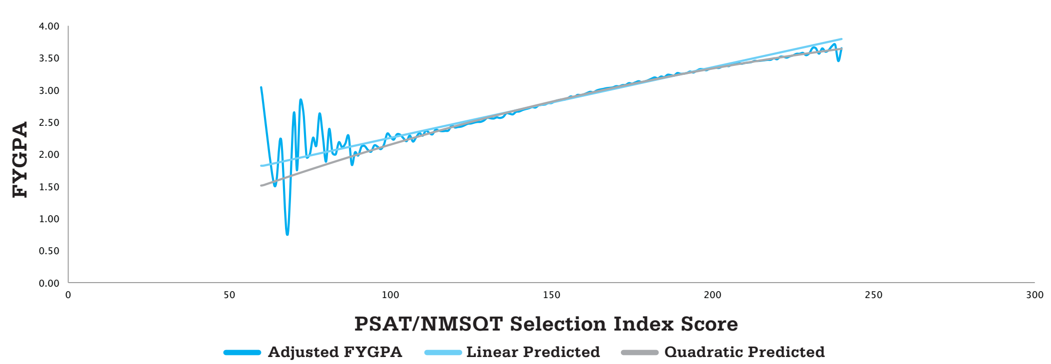
Higher PSAT/NMSQT scores are related to higher grades earned in college, supporting its use as a screening tool in the overall process of selecting Merit Scholarships.

The results did indicate some slight departures from linearity. In the full sample model (Model 1), the quadratic term was statistically significant and according to the $\Delta R^2$, the quadratic model provided a better fit than the linear model, and the coefficient of the quadratic term was also significant. That being said, the coefficient for the quadratic term was very small ($B = 0.000$, $p < .001$). Likewise, the $\Delta R^2$ was also small, accounting for only 0.1 percent of the variance in FYGPA. These significant results are partly a function of a large sample size. With a sample size of over 400,000 students, there is power to detect the smallest differences, even when they are not practically meaningful as shown by a very small effect size of 0.001 (Cohen, 1988). That is, the quadratic terms were statistically significant, but the predicted values of FYGPAs based on a linear versus quadratic model were quite similar.

Figure 2 shows plots of the mean adjusted FYGPA by PSAT/NMSQT score along with the predicted values from both the linear and the quadratic model for the entire score scale range. The difference between the predicted FYGPA at each PSAT/NMSQT Selection Index score scale point for the linear model compared to the quadratic model was computed, and the average absolute difference was 0.08. For example, for a PSAT/NMSQT Selection Index score of 170, the linear model predicts a FYGPA of 3.02 compared to 3.04 for a quadratic model, which is a difference of 0.02. For a score of 180, the respective values are 3.13 and 3.15, which is also a difference of 0.02. It should additionally be pointed out that the 10-point change in PSAT/NMSQT resulted in the same change in predicted FYGPA for the linear and quadratic models (0.11).

In Figure 2, the two lines had the most deviation at the very bottom of the score scale; however, this finding does not have much bearing on the current study because merit decisions are not made at this low range of the score scale. On the other hand, there are deviations at the very top of the score scale (≥ 200); however, the mean adjusted FYGPA data points were closer to the linear regression line (mean absolute difference = 0.04) than the quadratic

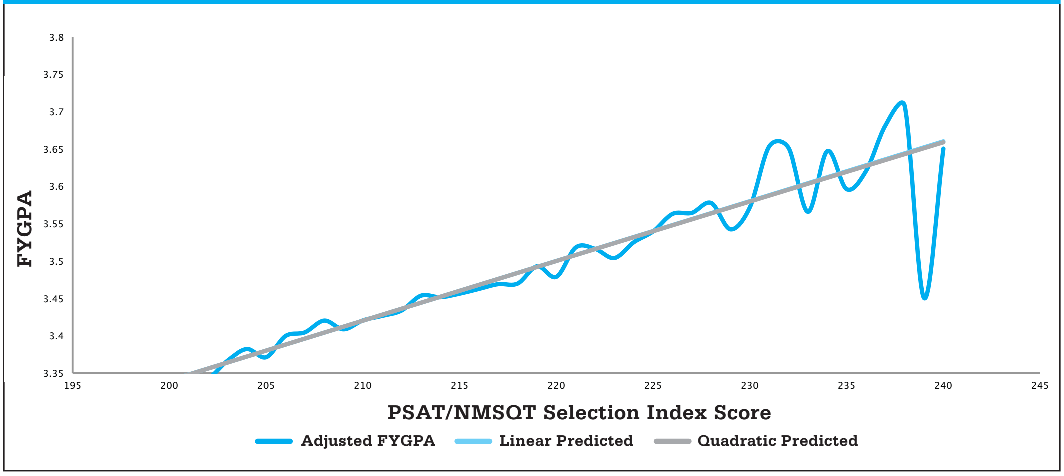
**Figure 2**
Mean adjusted FYGPA by PSAT/NMSQT Selection Index Score with linear and quadratic predicted values for the full sample.




regression line (mean absolute difference = 0.06). These results reiterate the fact that the quadratic term does not add greatly to the prediction of FYGPA. In fact, the linear model appears to more closely approximate the mean adjusted FYGPA for the top or higher-achieving end of the scale, which is where merit-based decisions are made.

Similar to Figure 2, Figure 3 shows plots of the mean adjusted FYGPA by PSAT/NMSQT score along with the predicted values from both the linear and the quadratic model; however, only high-scoring students (≥ 200) were included in these analyses. Inspection of the graph reveals that the two lines are almost identical. For example, for a PSAT/NMSQT Selection Index score of 200, the linear model predicts a FYGPA of 3.34 compared to 3.34 for the quadratic model, which is a difference of 0.00. For a score of 240, the respective values are 3.66 and 3.66, which is a difference of 0.00. In fact, the average absolute difference in prediction between the models was 0.00. Figures 2 and 3 indicate, specifically for the full sample model, that while differences do exist between predicted values, the differences lie in the hundredth-of-a-score range, with most predictions (more than half in each model examined) less than or equal to 0.05.

**Figure 3**
Mean adjusted FYGPA by PSAT/NMSQT Selection Index Score with linear and quadratic predicted values for the high scorers.




It is apparent from Figure 2 that the data in the lower end of the PSAT/NMSQT scale are "pulling down" the predicted values of the quadratic model (gray line) toward them. However, as stated before, there are very few actual data points in this area (PSAT/NMSQT score of 100 or less is only 0.5 percent of the total data). The predicted values of the linear model, illustrated by the light blue line, seem to visually fit the actual data better than the quadratic model, providing evidence that the linear model fits better in practicality. Figure 3 depicts the same type of plots for the high scorers and shows that both the linear and quadratic models predict the actual data equally as well. Both lines are equal for almost the entire PSAT/NMSQT score scale. They also fall in what looks to be the center of the actual data, indicating that the linear model predicts just as well as the quadratic model, and no additional information is gained from adding a quadratic term.

As with all research, there are limitations to this study that should be noted. The large sample size provides a limitation when examining significance in model fit. The sample provides power to detect the smallest deviations from linearity when they are not meaningful. That is, the quadratic term was statistically significant, but it is not likely to be practically significant. Furthermore, upon deeper inspection, the plot of adjusted and predicted FYGPAs by PSAT/NMQT scores for the entire score scale revealed that the quadratic model may better approximate the bottom end of the score scale range. However, students with those scores are not in consideration for the merit scholarship. On the other hand, the linear model appeared to better approximate the data for the top end of the scale, which includes the students who are under consideration for the Merit Scholarship. An analysis focusing on the higher-scoring subsample confirmed that the linear model better fit the data for these students. These findings underscore the need to include graphical representations of the data along with the quantitative results in order to accurately interpret the findings.

Additionally, more sophisticated models have been developed to handle data that are hierarchically structured. For example, hierarchical linear modeling (HLM) may be a more appropriate method to examine the current research question. In this paper, regression analysis was utilized because results are typically more accessible to a wide range of audience members, and previous studies have used the same method allowing future metaanalyses. Specifically, because the use of PSAT/NMSQT scores as part of the scholarship decision process is of interest to multiple constituents (e.g. students, parents, teachers, educational researchers), it was important that the results be presented in the most straightforward manner possible.

In sum, the current study unequivocally showed that PSAT/NMSQT scores are positively related to FYGPA, even among very high-scoring students, supporting its use as a screening tool in the scholarship decision-making process.

These findings underscore the need to include graphical representations of the data along with the quantitative results in order to accurately interpret the findings.

## References


Arneson, J. J. (2007) *An examination of the linearity of ability–performance relationships among high-scoring applicants*. Unpublished doctoral dissertation, University of Minnesota, Twin Cities Campus.

Benbow, C. P. (1992). Academic achievement in mathematics and science students between ages 13 and 23: Are there differences among students in the top one percent of mathematical ability? *Journal of Educational Psychology*, 84(1), 51–61.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coward, M. W. & Sackett, P. R. (1990). Linearity of ability–performance relationships: A reconfirmation. *Journal of Applied Psychology*, 75(3), 297–300.

Cullen, M.J., Hardison, C.M., Sackett, P.R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, 89(2), 220–330.

Greener, J. M. & Osburn, H.G. (1979). An empirical study of the accuracy of corrections for restriction in range due to explicit selection. *Applied Psychological Measurement*, 3(1), 31–41.

Hawk, J. (1970). Linearity of criterion–GATB aptitude relationships. *Measurement and Evaluation in Guidance*, 2(4), 249–251.

Hayashi, P. (2005). T*he merits of the national merit scholars program: Questions and concerns*. (CSHE Paper CSHE.6.05). Retrieved from University of California, Berkeley, Center for Studies in Higher Education website: http://cshe.berkeley.edu/publications/docs/ROP.Hayashi.6.05.pdf.

Keating, D. P., & Stanley, J.C. (1972). Extreme measures for the exceptionally gifted in mathematics and science. *Educational Researcher*, 1, 3–7.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average.* (College Board Research Report No. 2008-5). New York: The College Board. Retrieved from: http://professionals.collegeboard.com/profdownload/pdf/08-1718_RDRR_081017_Web.pdf.

McGuire, T. E. (2009, March 13). National Merit Scholarship Corporation's response to the National Association for College Admission Counseling's letter dated January 22, 2009. Retrieved from http://www.nacacnet.org/PublicationsResources/Marketplace/research/Documents/NMSCResponse_March2009.pdf.

National Association for College Admission Counseling (2008). *Report of the commission on the use of standardized tests in undergraduate admission.* Retrieved from http://www.nacacnet.org/PublicationsResources/Marketplace/research/Pages/TestingComTestingCommis.aspx.

Patterson, B. F., Mattern, K. D., & Kobrin, J. L. (2009). V*alidity of the SAT for Predicting FYGPA: 2007 SAT Validity Sample.* (College Board Statistical Rep. No. 2009-1). New York: The College Board.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment. *American Psychologist*, 63(4), 215–227.

Snyder, T.D., and Dillow, S.A. (2010). *Digest of Education Statistics 2009* (NCES 2010-013).

National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wai, J., Lubinski, D., & Benbow, C.P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, 97(3), 484–492.

11b-3668