

by Dr. Bryan Nankervis

Gender Inequity in the National Merit Scholarship Program



Abstract

This article investigates gender inequity in the National Merit Scholarship Competition. Results suggest the competition favors males due to their higher mean score on the mathematics section and greater variability on all sections of the PSAT, which shares differential validity concerns with the SAT in terms of gender. These instruments are designed to predict first-year college success, but have historically under-predicted the abilities of girls compared to boys. Reasons for this disparity are explored and through a statistical analysis utilizing data provided by the College Board, it is predicted that most National Merit semifinalists and finalists are male despite the fact that the majority of participants in the competition are female. This study underscores the problematic practice of using PSAT cutoff scores as the sole measure of merit in the first two phases and SAT scores in the final stage of the National Merit Scholarship Competition, a selection process that runs contrary to best practices in the use of standardized tests.

Each year, approximately 15,000 high school seniors are chosen as National Merit finalists as a result of their scores on the PSAT and SAT exams. They are eligible for generous scholarships at many of the nation's premier universities and are heavily recruited for their talents and the prestige they bring to campuses. However, the PSAT shares validity concerns with the SAT, which over time, has under-predicted the college success of females compared to males in terms of mathematics (Bridgeman and Wendler, 1991; Wainer and Steinberg, 1992; College Board, 2001; College Board, 2008), and is not specifically designed for determining "meritorious" students. Despite these shortcomings, PSAT scores are used as the sole criterion in the first two phases of the National Merit selection process.

This reliance on the results from a single instrument runs contrary to the College Board's own guidelines for use of its standardized exams (College Board 2010) and has been criticized by the National Association for College Admission Counseling (NACAC, 2008). Some institutions, including the University of California System and the University of Texas at Austin, have divested from the National Merit Scholarship Program as a result of this questionable use of standardized test scores, preferring instead to rely upon

a comprehensive review of students' educational backgrounds to award scholarships (BOARS, 2005; *Chronicle*, 2009).

This article will provide an overview of the National Merit selection process and discuss possible reasons for the differential validity of the PSAT and SAT. This article then quantifies gender inequities within the National Merit selection process by using PSAT data supplied by the College Board to demonstrate how male and female statistical dissimilarities (in particular on the mathematics section and generally in terms of variability) can lead to differences in score distributions with the PSAT. These differences, in combination with the use of simple cutoff scores, most probably lead to a majority of male National Merit semifinalists, despite the fact that many more females take part in the competition. Further, since SAT scores, which also generally favor males, play a role in the last phase of the selection process, it is likely that the ratio of males to females is even greater among National Merit finalists (the College Board does not release actual numbers of semifinalists and finalists by gender).

How the National Merit Scholarship Program Works

The National Merit Scholarship Corporation (NMSC) was founded in 1955 to oversee the National Merit program and partnered with

the PSAT in 1971 (NMSC, 2006). This partnership continues today. Because SAT scores play a role in the selection of National Merit scholars, finalists and semifinalists must take the SAT to compete—which gives the College Board’s SAT a competitive edge over other standardized tests.

Each year, in October, about three million high school students take the PSAT. More than 1.5 million are juniors and thus eligible to vie for prestigious National Merit Scholarships (NMSC, 2012). The following April, NMSC identifies students for advancement (commended level) in the competition by use of a nationwide cutoff score applied to students’ combined total scores from the reading, mathematics and writing sections of the PSAT (NMSC, 2012). Then in September, about 16,000 semifinalists are announced as a result of employing a second cutoff score particular to each state. The state cutoff scores are set to allow a number of semifinalists proportional to the state’s percentage of nationwide graduating seniors. For example, about eight percent of the nation’s graduating seniors live in Texas, so a cutoff score is chosen that allows for the same percentage of National Merit semifinalists to come from that state. The next February, approximately 1,000 students are eliminated based on their high school performance or relatively low SAT scores. The remaining students are designated as National Merit finalists and are eligible for more than 8,300 National Merit Scholar monetary awards (NMSC, 2012). These students are highly sought after with some institutions offering full four-year scholarships. The number of National Merit Scholars can directly affect an institution’s overall national ranking and therefore its prestige.

Differential Validity/Prediction of the SAT

All PSAT test questions come from the SAT (College Board, 2006) and the College Board maintains that because the PSAT is an accurate predictor of SAT scores, it therefore shares the same statistical validity of the SAT (BOARS, 2005). However, it has been demonstrated that the SAT consistently under-predicts the college success of females compared to males. In particular, a review of 37 studies (beginning in 1974) on the differential validity/prediction of the SAT conducted by the College Board concluded that “...the general finding of these studies is one of underprediction of women’s college grades” (College Board, 2001, 25).

More recently, a study by the College Board revealed that all three sections of the SAT (reading, writing and mathematics, separately or combined) were lesser predictors of early college

success than the high school grade-point average and that the mathematics section had the greatest levels of over-prediction for males and under-prediction for females in terms of freshman-year grade point average (College Board, 2008). Previous studies (conducted by College Board/ETS personnel) of students with similar mathematical backgrounds and grades in like college mathematics courses have revealed that males on average score higher (a third of a standard deviation) than females on the SAT quantitative section.¹

Why Gender Inequities Exist in the National Merit Selection Process

For many years the PSAT consisted of only a reading and mathematics section. While males and females historically performed similarly on the reading portion, males consistently averaged higher scores on the mathematics portion leading to higher total scores. Over the years, it has been suggested that as a result, when the first cutoff score (nationwide) and second cutoff score (state level) were employed, the majority of semifinalists were males. NMSC does not publish a count of semifinalists by gender, but in 1994, FairTest estimated that 60 percent were male even though more females took the PSAT (1999).

At that time, FairTest filed a complaint with the US Department of Education’s Office for Civil Rights claiming that the College Board illegally discriminated against females in the way it administered and designed the PSAT. To settle the case, a writing section (supposedly more favorable to females) was added to bring total scores for males and females closer together. After the addition of the writing section, FairTest used the names of semifinalists (published by NMSC) to identify gender and the male/female split was estimated to be 52 percent to 42 percent with six percent unknown (FairTest, 1999). Thus, the addition of the writing section had seemingly narrowed, but not resolved the alleged gender gap.

The prevailing reasons for any overrepresentation of males in the National Merit selection process today are their higher mean scores on the mathematics section of the PSAT and their generally greater variability on all sections of the exam. Nationwide in 2010 (see Table 1), males and females had similar mean reading scores and females had a slight advantage in writing, but in mathematics, males outscored females by 2.6 points on average (approximately a quarter of a standard deviation). More importantly though, males had a much larger standard deviation in mathematics and were more variable than females on both the reading and writing sections (College Board, 2011).

¹ Bridgeman and Wendler, 1991; and Wainer and Steinberg, 1992 not only point to gender differences, but also conclude (in peer-reviewed journals) that there is gender bias on the SAT.

For many years the PSAT consisted of only a reading and mathematics section. While males and females historically performed similarly on the reading portion, males consistently averaged higher scores on the mathematics portion leading to higher total scores. Over the years, it has been suggested that as a result, when the first cutoff score (nationwide) and second cutoff score (state level) were employed, the majority of semifinalists were males.

Table 1. PSAT Mean Scores/Standard Deviations by Gender, 2010–2011

Section	Male	Female
Reading	47.2/11.9	47.4/11.1
Mathematics	50.3/12.2	47.7/11.0
Writing	44.7/11.8	46.1/11.3

In the past, studies have shown that males have higher mean scores, resulting in a larger concentration of males in the population of students who surpassed the cutscore (Hedges and Friedman 1993). More recent data continue to show evidence of such disparities. This undoubtedly leads to a larger number of high-scoring males in the distribution of total scores (mathematics, reading and writing sections combined) since they have a higher average score and larger standard deviation than females. Further, as larger cutoff scores are employed, as in the NMSP selection process with a nationwide cutoff score followed by a higher state cutoff score, the ratio of males to females that remain eligible in the competition likely grows even larger.

Since national cutoff scores are generally around 200 (total for reading, mathematics and writing sections, each of which has a maximum of 80 points available) and state cutoff scores are higher (the average for all states is 210), students must average about 70 points per section to qualify as a National Merit semifinalist. Table 2 shows the exact number of students, by gender, who scored 70 or better on the individual sections of the PSAT in 2010–2011, as reported by the College Board. Note that while males have a slight edge in the reading section and females are better represented in the writing section, there are almost twice as many males scoring at 70 or above on the mathematics section. It is this disparity that leads to the underrepresentation of females as semifinalists and finalists in the National Merit Competition. The data in Table 2 demonstrate that while the addition of the writing section has improved women’s representation in the National Merit Competition, it has not leveled the playing field.

Table 2. Students Scoring at 70 or Above on PSAT Sections, 2010–2011

Section	Male	Female
Reading	20,871	19,554
Mathematics	39,229	19,648
Writing	15,284	19,980

Method

Data for this study is taken from the College Board website. Numbers of participants, means and standard deviations are provided in the PSAT/NMSQT state summary reports for high school juniors. The analysis conducted here is based on the assumption that male and female test scores are normally distributed. The PSAT is composed of questions from and modeled after the SAT, which was realigned in 1995 to more closely adhere to a normal distribution (College Board, 2002). Comparisons of percentile ranks corresponding to raw scores on the PSAT (pages 5–11 in each state summary report) with predicted percentiles based on a normal distribution for all students and by gender on each section of the PSAT (reading, mathematics and writing) reveal little evidence of marked departure from normality in the distribution of scores (College Board, 2011) overall or by either subgroup.

Thus, given a mean and standard deviation for any group, a p-value, particular to any exam score, can be found using a normal distribution table (or statistical software). A p-value is the probability a member of the group will achieve a score greater than or equal to a given score (i.e. cutoff score) and it represents the percentage of the total area under a normal curve to the right of the given score. Accordingly, respective means, standard deviations and a common cutoff score are used to predict percentages of males and females remaining in the competition after each of the first two phases of the National Merit selection process. These percentages can in turn be utilized (by multiplying by the original number of males or females) to approximate the actual number of males and females that scored at or above the cutoff scores. These same predictions

The PSAT, like the SAT, is designed solely to predict early-college success, yet the NMSC uses this test for a completely different purpose. Determining “merit” using a score from a single test that is not validated for such a purpose is ill-conceived and runs contrary to best practices in the use of standardized tests.

can be easily reproduced by any statistical software package that works with normal distributions.

Results

Table 3 lists the predicted percentages by gender for high school juniors scoring at or above the initial nationwide cutoff score for the 2010 administration of the PSAT. The predicted breakdown by gender in column four is based on true participation numbers (College Board 2011) and reflects the probable male and female representation after the nationwide cutoff score is used to narrow the field in the National Merit Competition. The nationwide cutoff score in 2010–2011 was 202 (out of a possible 240 points) for the combined PSAT sections of math, reading and writing (NMSC, 2013). Note that the predicted rates by gender at the commended level are a virtual reverse of the initial participation rates. This is due in part because males had a one point higher combined mean score than females nationally (see Table 1), but to a greater degree because of the larger combined standard deviation for the males (a differential of 2.5 points).

Table 3. Predicted Breakdown by Gender at or Above the National Cutoff Score, PSAT 2010–2011

	Participants	Predicted Percentage At or Above Cutoff Score	Predicted Number Commended
Male	729,907 (46.7%)	4.79%	34,963 (54.9%)
Female	834,276 (53.3%)	3.44%	28,699 (45.1%)

In phase two of the competition, a second cutoff score, particular to each state, is employed to select semifinalists. Table 4, focusing on the three most populous states and the 2010–2011 administration of the PSAT, illustrates how the practice of using a second, higher cutoff score at the state level in the National Merit selection process increases levels of gender inequity. Even though most participants were female, the majority of students remaining eligible after the first round (nationwide cutoff score) are predicted to be male. This inequity is exacerbated after the second round due to higher cutoff scores employed at the state level to the extent that about three out of every five semifinalists in each state are predicted to be male. This is despite the fact that females actually have a higher combined mean score in both Texas and New York. Here, the sole reason for males’ greater participation after employing the state cutoff score is their larger overall variability (combined standard deviations from each section of the test).

Discussion

This analysis illuminates and quantifies inequities in the National Merit Scholar Competition. The results suggest that the practice of using cutoff scores in the first two phases of the National Merit Scholarship competition most probably leads to an underrepresentation of females as semifinalists. Because of males’ much greater mean on the mathematics section and variability on all sections of the PSAT, the practice of using a simple cutoff score

Table 4. Predicted Breakdown by Gender at or Above the state Cutoff Score, PSAT 2010–2011

State (cutoff score)		Participants	Predicted At or Above National Cutoff Score	Predicted Number Commended	Predicted At or Above State Cutoff Score	Predicted Number of Semifinalists
Texas (219)	Male	99,204 (48.5%)	2.24%	2,222 (56.2%)	0.61%	605 (60.1%)
	Female	105,447 (51.5%)	1.64%	1,729 (43.8%)	0.38%	401 (39.9%)
New York (219)	Male	73,888 (48.0%)	3.08%	2,276 (54.8%)	0.96%	709 (58.0%)
	Female	80,120 (52.0%)	2.34%	1,875 (45.2%)	0.64%	513 (42.0%)
California (221)	Male	80,364 (45.5%)	5.70%	4,580 (54.2%)	1.82%	1,463 (58.0%)
	Female	96,435 (54.5%)	4.02%	3,877 (45.8%)	1.10%	1,061 (42.0%)

gives them an advantage in the first phase of the competition. Though historically male PSAT test-takers are in the minority, it is predicted they comprise about 55 percent of the group surviving the first cutoff score at the national level in the 2010–2011 competition. This inequity generally grows as a higher cutoff score is employed at the state level to determine semifinalists. For example, in Texas, it is predicted that 60 percent of its semifinalists are males. This overrepresentation of males no doubt increases in the final phase of the competition since SAT total scores, which are historically higher for males (also due to the mathematics section), are used in the selection process of National Merit finalists.

Conclusion

The PSAT, like the SAT, is designed solely to predict early-college success, yet the NMSC uses this test for a completely different purpose. Determining “merit” using a score from a single test that is not validated for such a purpose is ill-conceived and runs contrary to best practices in the use of standardized tests. In order to settle a previous discrimination lawsuit and rather than redesign the quantitative section to more fairly measure participants’ abilities, the College Board added a writing section to the PSAT. While this helped to close the gender gap in combined mean scores, an unintended result was an increase in the overall standard deviation gap, thus males continue to be favored in the competition.

Regardless of why instruments such as the PSAT and SAT do a poor job of predicting future student success, it is clear that they are not designed to measure merit. There is, however, limited public inspection of the National Merit Scholarship Program because it functions on the periphery of the domains of colleges and high schools. Further, many universities condone the practices of the NMSC and contribute financially because it is a relatively inexpensive way to attract National Merit Scholars and thereby increase institutional prestige. Meanwhile, most students, parents and high schools do not have the ability or capacity to challenge the program even if they were aware of inequities.

The College Board and National Merit Scholarship Corporation do not release a breakdown of National Merit Scholars by gender, leaving stakeholders questioning this policy for the use of test scores. That is, that a minimum score from a single test should not be used for determining admission or scholarships. Nevertheless, the levels of inequity predicted

by this study highlight the need for the nation’s postsecondary institutions to find more inclusive and equitable ways of determining merit among America’s students.



DR. BRYAN NANKERVIS received his PhD in mathematics education from the University of Texas at Austin in 2006. He is currently researching inequities associated with the SAT and PSAT which result in the underrepresentation of females and minorities at the postsecondary level and in the awarding of scholarships.

REFERENCES

- Board of Admissions and Relations with Schools. 2005. Letter to University of California System schools from Michael T. Brown (Chair) concerning divestment from the National Merit Scholarship Program. Assembly of the Academic Senate, March 1, 2005
- Bridgeman, B. and Wendler, C. 1991. Gender differences in predictors of college Mathematics performance and in college mathematics course grades. *Journal of Educational Psychology*, 83,(2), 275-284.
- The Chronicle of Higher Education. 2009. *University of Texas at Austin won't sponsor National Merit Scholars anymore*. September 1, 2009.
- The College Board. 1998. *SAT and gender differences*. Research summary, Office of Research and Development, RS-04.
- The College Board. 2001. *Differential Validity, Differential Prediction, and College Admission Testing: A Comprehensive Review and Analysis*. Research Report no. 2001-6.
- The College Board. 2002. *The Recentering of SAT Scales and Its Effects on Score Distributions and Score Interpretations*. Research Report no. 2002-11.
- The College Board. 2006. *PSAT/NMSQT Program Fact Sheet*.
- The College Board. 2008. *Differential validity and prediction of the SAT*. Research Report no. 2008-4.
- The College Board. 2010. *Guidelines on the Uses of College Board Test Scores and Related Data*.
- The College Board. 2011. *PSAT/NMSQT 2010-2011 College-Bound High School Juniors*. <http://professionals.collegeboard.com>. (Date of retrieval: 15 December 2011).
- FairTest. 1999. *Test-Makers to Revise National Merit Exam to Address Gender Bias*. <http://www.fairtest.org/examarts/fall96/natmerit.htm>.
- Hedges, L.V. and Friedman, L. 1993. Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research*. 63: 94-105.
- National Association for College Admission Counseling. 2008. *Report of the Commission on the Use of Standardized Tests in Undergraduate Admission*. September, 2008.
- National Merit Scholarship Corporation. 2012. *Information About The National Merit Scholarship Competition*. www.nationalmerit.org/Merit_About_Leaflet.pdf
- National Merit Scholarship Corporation. 2013. *National Merit Scholarship*. <http://www.collegeplanningsimplified.com/NationalMerit.html>
- Wainer, H. and Steinberg, L.S. 1992. Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review*, 62: 323-336.