Is the Science and Engineering Workforce Drawn from the Far Upper Tail of the Math Ability Distribution?

Catherine J. Weinberger

Institute for Social, Behavioral
and Economic Research and
Department of Economics
University of California
Santa Barbara 93106
weinberg@econ.ucsb.edu

Draft—comments welcome
September 2005 version

Abstract: Low representation in the extreme upper tail of the mathematics test score distribution is often assumed to explain the small numbers of women in engineering, mathematics, computer science and physical science (EMS) college majors and careers. However, this study finds that fewer than one-third of the college-educated white men in the EMS workforce had SAT-Math scores above the threshold previously presumed in the literature. The lower-scoring male EMS college graduates have more than an empty credential; they enjoy the same earnings advantage, relative to other college graduates with the same scores, as high scoring EMS majors. This study also finds that white women enter EMS fields at no more than half the rate of men with the same mathematics test scores. Both the large gender gap and the low ability threshold for EMS entry are robust to modeling mathematics test scores as a noisy measure of ability.

The early stages of this research were supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9255347. This paper was completed while the author was a National Academy of Education Spencer Postdoctoral Fellow. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies.

I would like to thank seminars participants at the NBER Conference on Diversity in the Science and Engineering Workforce, U Mass Amherst Department of Economics, and UCSB Department of Sociology, Lee Badgett, Bill Bielby and, especially, Peter Kuhn, for helpful comments on earlier drafts of this paper.

"Individuals with the most potential for high academic achievement in mathematics and science . . . are generally considered to be those students who represent the top few centiles in ability, especially mathematics ability."

Benbow and Arjmand (1992)
Journal of Educational Psychology

" . . . substantially fewer females than males . . . score in the upper tails of the mathematics and science ability distributions and hence are poised to succeed in the sciences."

Hedges and Nowell (1995)
Science

" . . . if we wish to increase the proportion of women in the sciences, we must do something to change the underlying mathematical-attribute distribution."

Paglin and Rufolo (1990)
Journal of Labor Economics

In recent years, fears of an imminent shortage of scientists and engineers prompted discussion about policies to encourage more young people to pursue careers in science and engineering (Koshland, 1988; Widnall, 1988; Atkinson, 1990). One component of this discussion is the idea that the low representation of women and minorities in science and engineering careers might reflect underutilized potential among members of these groups (Berryman, 1983, Oakes, 1990).

An alternative view is that the demographic composition of the science and engineering workforce reflects the underlying distribution of talent, or preparation, as measured by mathematics test scores. Although gender differences in mean scores on standardized math tests have declined, a high proportion of those in the upper tail are male (Benbow and Stanley, 1980, 1982, Feingold, 1988, 1996, Hyde, Fennema and Lamon, 1990, Hedges and Nowell, 1995, Stumpf and Stanley, 1998). Regardless of whether this gender difference has roots in biology or socialization, scholars in fields ranging from Psychology to Economics have argued that the strong representation of men among those with math scores in the "top few centiles" fully explains the low participation of women in science and engineering college majors and careers (Paglin and Rufolo, 1990, Benbow and Arjmand, 1992, Hedges and Nowell,1995). [1]

Both the idea that the underutilization of women is inefficient and the apparently opposing view that increasing the representation of women would have negative impacts on productivity share a common assumption. In both views, the pool of individuals with the ability to succeed in science and engineering careers is assumed to be small relative to the number of jobs to be filled. This assumption has not been questioned. Estimating the size of the pool of capable individuals (conditional on current levels of high school preparation) is vital to understanding the efficiency implications of the composition of the science and engineering workforce.

[1] The biology vs. socialization question is far from settled (Benbow and Lubinski, 1997, Hyde, 1997). Whatever the reason, twice as many boys as girls had math scores in the top 5 percent during the high school years, consistently over the 20 year period 1972-1992 (Xie and Shauman 2003).

The focus of this inquiry is the segment of the U.S. workforce employed in science and engineering occupations such as electrical engineer, chemist, or computer programmer. College-educated professionals in the fields of engineering, mathematics, computer science or physical sciences (EMS) represent 2-3% of the U.S. workforce, and more than one-quarter of all college-educated workers. If participation in this segment of the workforce is truly limited to those in the top few centiles of the ability distribution, then a high rate of participation by individuals with the requisite mathematical ability is vital. This paper examines the relationships between mathematics test scores, gender and entry into EMS careers, and also explores the extent to which scarce ability constrains the set of individuals capable of entering EMS careers.

Certainly, there is a positive correlation between high school mathematics test scores and later entry into EMS higher education and careers (Fiorito and Dauffenbach, 1982; Blakemore and Low, 1984). However, finding a positive correlation is not the same thing as showing that very high mathematics test scores are a prerequisite for success in EMS fields. Regression analysis identifies the correlation, but can not distinguish the proportion of EMS participants drawn from the extreme upper tail of the ability distribution. In this study I specifically estimate the proportion of male EMS participants who had very high math scores as high school seniors, and find that it is surprisingly low.

Regardless of effects on the productivity of the EMS workforce, gender differences in EMS career entry are responsible for a substantial portion of the gender pay differential among college graduates (Eide, 1994; Brown and Corcoran, 1997; Weinberger, 1998, 1999, 2001). Hence the demographic composition of the EMS workforce has both equity and efficiency implications.

Previous studies based on non-representative samples show gender differences in EMS participation conditional on mathematics test scores (Benbow and Arjmand, 1992, Turner and Bowen, 1999), and find that female engineers have higher average mathematics test scores than male engineers (Jagacinski, 1987; McIlwee and Robinson, 1992). Studies based on the representative high school seniors samples also find gender differences in EMS participation conditional on a wide array of observable background characteristics, including math scores and high school math and science coursework (Frehill, 1997; Xie and Shauman, 2003).

The analysis of this paper goes further than previous studies, modeling ability as an unobserved quantity. Maximum likelihood techniques are used to simultaneously estimate the relationship between scores and true ability, and between true ability and EMS participation.

This study uses data from nationally representative samples of 1972 and 1980 high school seniors, followed longitudinally through the college years to answer the following questions: Did most of the men who pursued EMS careers have very high mathematics test scores, or unobserved ability, as high school seniors? and: How robust is the gender differential in EMS participation to modeling math scores as a noisy measure of EMS ability?

Although Paglin and Rufolo (1990, p. 138) speculated that the high ratio of men to women among bachelor's degree graduates in engineering and physical sciences is explained by the small number of women with scores above 650 (out of 800) on the mathematics portion of the SAT exam (SAT-M), I find that fewer than one-third of the college-educated white men in the EMS workforce had high school SAT-M scores above 650. Another one-third scored below the average humanities major. Alternative specifications of the relationship between math scores, unobserved ability, and EMS participation yield estimates of the ability threshold for EMS participation, none of which are above the 78th percentile of the white males' ability distribution. The threshold for EMS entry is far lower than previously believed.

I also find that the lower scoring men who completed EMS college degrees enjoy the same earnings advantage, relative to college graduates with the same scores, as high scoring EMS participants. This suggests that the lower scoring EMS graduates are not simply recipients of an empty credential.

Finally, I find a very robust 50-70% gender gap, conditional on ability, in the probability of completing an EMS college degree with an EMS major. This gender gap is robust to a wide range of specifications of the relationship between the observed mathematics test score and EMS participation, including specifications where scores are a noisy measure of EMS ability. The gender gap in EMS participation can be explained away only by imposing the assumption of an absurdly large gender differential in unobserved ability.

The results of this analysis strongly reject the previously presumed restriction of EMS participants to the "top few centiles" of ability, and indicate that the demographic composition of the science and engineering workforce cannot be explained by the distribution of attributes measured by these standardized tests of mathematics.

Data and Definitions

The data come from two national longitudinal studies of U.S. students. The studies drew representative samples of U.S. high school seniors in 1972 and 1980. The individuals in both studies took a cognitive test of mathematics during the senior year of high school. In later years, the seniors were resurveyed and asked about their educational and occupational attainments. These are the same data used by Hedges and Nowell (1995) to argue that gender differences in the variability of ability explain the gender gap in science and engineering participation, and by Murnane, Willet and Levy (1995) and Grogger and Eide (1995) to show the strong correlation between math scores and later earnings. I use data on race, sex, high school math scores, and later EMS participation, as measured either by completion of an EMS college degree or by employment in an EMS occupation. Following Paglin and Rufolo (1990), the EMS category used in this analysis includes only those fields of science associated with the highest math scores and earnings. EMS educational attainment is defined as having completed a bachelor's degree in an EMS field (with a major in engineering, mathematics, computer science, or physical sciences) by the resurvey date. Employment in an EMS occupation is conservatively defined as reporting employment in a professional engineering, mathematics, computer, chemistry or physics occupation. While these EMS categories are highly aggregated, combining individuals from very different undergraduate institutions and fields of study, this level of abstraction permits analysis of questions raised in the existing literature. [2]

For practical reasons, the analysis focuses on white men and women, and on EMS participation at the bachelor's degree level. The numbers of black, Asian and Hispanic high

[2] The next draft will explore the robustness of the results to using more inclusive measures of science participation.

school seniors in these surveys who later graduated college with EMS majors are very small.[3] Similarly, a far larger study would be required to map a representative sample of high school seniors to EMS Ph.D.'s, or to the even smaller set of exceptional Ph.D. scientists.[4]

Research by sociologists, psychologists and other social scientists describes many factors, other than mathematics test scores, that influence decisions to pursue careers in science and engineering. These include cultural norms surrounding appropriate roles for women, the "chilly climate" in the college classroom, and perceptions, or misperceptions, about ability (Tobias, 1978, 1990; Betz and Hackett, 1981, 1983; Hall and Sandler, 1982; Lunneborg, 1982; Eccles, 1987; Ware and Lee, 1988; Seymour and Hewitt, 1994; Arnold, 1994; Hanson, 1996; Lapan, Shaughnessy and Boggs, 1996; Hyde, 1997; Betz, 1997; Steele, 1997a; Leslie, McClure and Oaxaca, 1998). Economists expect gender differences in career choices based on differences in preferences and anticipated labor force withdrawals (Polachek, 1978, 1981; Blakemore and Low, 1984; Daymont and Andrisani, 1984), and due to economic costs of deviating from socially prescribed roles (Badgett and Folbre, 2003). Yet, for the purpose of this study, all factors other than math scores are lumped into a single category.

The National Longitudinal Study of the High School Class of 1972 (NLS-72) began with a survey of high school seniors in 1972, and had five follow-up surveys. Information from the base year (1972) survey and the 1972, 1979, and 1986 follow-ups are used in this study. The sample used in the analysis includes 5191 white men and 5317 white women who were1972 seniors, who (in 1972 or 1974) identified themselves as white, who took the mathematics test in 1972, for whom sex was known, and who were resurveyed late in 1979, with 1979 educational attainment known.

For the subsample of 2846 white male seniors both resurveyed and employed in 1986, additional information about occupational and educational attainment, and 1986 hourly earnings is taken from the 1986 fifth follow-up survey. The 1986 earnings regressions are weighted to adjust for selective attrition from the study. The 1970 census three-digit codes used to identify EMS occupations for this cohort are: 006-023 (engineering), 034-036 (mathematics), 003-005 (computer), 045 (chemistry) or 053 (physics or astronomy).

The measure of mathematics test scores available for a representative sample of high school seniors is the score on the Cognitive Test of Mathematics (CTM). The version of the CTM given in 1972 had 25 questions, with a choice between four possible answers on each question. One point was awarded for each correct answer, with a -0.33 point penalty for incorrect guessing.

The 1972 senior survey also includes SAT scores, taken from high school transcripts, where available. Only about one-third of all seniors had SAT-M scores reported, and this is a select

[3] Some analysis of nonwhite EMS participation is described in footnote 18. In a separate paper, math test score information from the high school seniors surveys is merged with data from the much larger Survey of College Graduates (conducted by the U.S. Census Bureau and National Science Foundation) to produce estimates for black, Hispanic and Asian women and men.

[4] Much of the current research pertaining to test scores and the development of scientific talent is concerned with identifying and training individuals with the potential to be exceptional, even among Ph.D. research scientists (Benbow and Stanley, 1980, 1982; Lubinski, Benbow and Sanders, 1993; Stanley, 1996; Benbow and Lubinski, 1997). Ph.D. research scientists comprise less than 0.2% of the U.S. workforce, so very few are included in nationally representative samples of high school graduates.

sample of students with higher-than-average levels of academic ambition and CTM scores. In addition to describing a non-representative sample, the SAT-M is often used for career counseling, and therefore may directly affect EMS participation. Despite these drawbacks, the SAT-M score has the advantage of greater accuracy than the CTM because the SAT exam is longer, more varied, and of more consequence to the test-takers. This analysis relies primarily on CTM scores, but also uses information about the joint distributions to impute SAT-M score probability distributions for a representative sample (see the appendix for further explanation).

The High School and Beyond (HS&B) longitudinal study began in 1980, with surveys of high school sophomores and seniors. Information on the 1980 seniors in the base year (1980) survey and 1986 third follow-up are used in this study. The sample used for this analysis includes the 2117 white men and 2443 white women who were1980 seniors, who identified themselves as white, who took both parts of the mathematics test in 1980, and who were resurveyed early in 1986, with 1986 educational attainment known.

The Cognitive Test of Mathematics given in 1980 is slightly different from the 1972 test, including more questions and a larger range of difficulty. The 1980 test had two parts, with a total of 32 questions. Scores from the two parts of the mathematics test were added together to create a single test score, with a maximum score of 32. Again, one point was awarded for each correct answer, with a -0.33 point penalty for incorrect guessing.

Evidence from another source (NCES, 1994) suggests that there was virtually no change in the distribution of mathematics scores of white male seventeen-year-olds over this period.[5] Where comparisons between the two cohorts are made, the percentile equivalent of the score in the white male distribution is used as a standardized unit of measure.

Transformations of the CTM score were used in some of the analysis. In Figures 1 and 2, fractional scores are rounded to the next largest integer. For the maximum likelihood estimations, the inverse of the mapping from 1980 scores to percentile units is used to generate a comparable 32 point scale for the 1972 seniors, so that two individuals from different cohorts with the same percentile score also have the same rescaled CTM score.[6] Results using this rescaled CTM score were much more robust to changes in specification than were results when the percentile scores were used.

Sample means of the four different math test score measures are reported in Table 1. The first is the raw CTM score, the second is the CTM percentile score (normalized to the percentile distribution of white men in each cohort), and the third is the rescaled CTM score. The means of the percentile and rescaled CTM scores are equal for men in the two cohorts, showing that the normalization was successful. The mean percentile scores are 49 for men in both cohorts (by construction) and 41-42 for women. The lower panel of Table 1 displays mean scores for the 1972 subsample with both CTM and SAT-M scores available. As might be expected, SAT test takers are highly selected, and mean CTM percentile scores for this subsample are much higher than mean scores for the full sample.

---

[5] Based on NAEP scores for 17 year olds, 1973 and 1982.

[6] Because the maximum possible score in 1972 is only at the 94.5th percentile, the rescaled CTM score was adjusted so that those with the maximum 1972 score are assigned the mean score of men in the 1980 cohort with scores at or above the 94.5th percentile.

Table 1 also shows the EMS participation rates for men and for women in each cohort. Among 1972 high school seniors, 4.5% of the men and 0.9% of the women graduated college with an EMS major by the 1979 resurvey. Among 1980 high school seniors, 5.8% of the men and 1.8% of the women completed EMS degrees by the 1986 resurvey. The increase in EMS participation between cohorts, despite the shorter elapsed time before the resurvey, has been noted previously (Hilton and Lee 1988, Grogger and Eide 1995). Estimates of EMS degree attainment drawn from Census data are similar.[7] Recent government statistics show no further increase in the proportion of EMS degrees earned by women between 1984 and 1994, nor between 1994 and 2002 (NCES 1994, 1997, 2003). Women earned 26.6 percent of EMS bachelor's degrees in 2002, compared to 26.3 percent in the High School and Beyond sample studied here.

Note that the 1972 cohort is resurveyed 7 and a half years after high school, while the 1986 cohort is resurveyed only 6 years after high school. Therefore, between cohort comparisons of bachelor's degree completion rates will tend to underestimate the growth in EMS participation. Nonetheless, these data are ideal for the purpose of studying the relationships between ability, gender and EMS entry in two different cohorts.

The Relationship between Math Scores and EMS Participation

Figure 1 shows, for the earlier cohort, EMS participation rates for men and for women at each possible score on the cognitive test of mathematics. Figure 2 depicts these relationships for the later cohort. These graphs show clearly that EMS participation is increasing in mathematics test scores. But, for both cohorts, the EMS participation of white women is much lower than that of white men at all levels of math scores.

Naïve estimates of the gender differential in EMS participation can be generated by assigning each woman with a given score the average participation rate of men with the same score. The mean of this counterfactual variable yields the estimate that the EMS participation of women would be 3.1%, rather than 0.9% in the earlier cohort, and 3.8%, rather than 1.8% in the later cohort if women made the same choices as men with the same math scores. (Table 2, "stepwise model").

Careful examination of Figures 1 and 2 reveals a technical challenge: there is a tradeoff between the accuracy and the precision with which male EMS participation is estimated at any given level of scores. For example, we don't really believe that men who score 12 or 13 in the earlier cohort tend to have higher EMS participation than those who score 14 or 15. There is simply too small a number of observations at any given score to get a very accurate measure of EMS participation. But aggregating data from a broader range of scores yields a less precise estimate of EMS participation at a given score. I therefore make several alternative predictions of the EMS participation of women if they completed EMS college degrees at the same rate as men

[7] Estimates based on census data are: 4.7% of white men and 1.1% of white women in the earlier cohort, and 6.4% of white men and 2.0% of white women in the later cohort completed EMS college degrees by age 27. (Computed as the product of two estimates: The proportion of college graduates in each cohort is estimated from the 1990 Census, and the proportion of college graduates who held an EMS degree by age 27 is estimated for each cohort from the 1993 Survey of College Graduates).

with the same math scores. Each of these predictions estimates the EMS participation rate of men with a given score in a slightly different way, trading off accuracy for precision.

In the second specification, the EMS participation of men is fitted to a quadratic function of the math test score.[8] In the third specification--based on "eyeballing" Figures 1 and 2--the EMS participation of men is fitted to a piecewise linear function, with a kink at the 40th percentile and a jump at the maximum score. In the fourth through seventh specifications, the EMS participation of men is estimated using nonparametric regression, with bandwidths 0.8, 1.2, 1.6, and 2.0, assuming a uniform weighting function. These estimates are all reported in Table 2.

The predicted EMS participation of women under each of these specifications is similar to what is predicted under the stepwise specification. The estimated counterfactual participation of women in the first cohort ranges from 2.7% to 3.2%, compared to an actual participation rate of 0.9%. In the second cohort, the estimated counterfactual participation rates range from 3.5%-3.8%, compared to an actual participation rate of 1.8%.

I now turn to the question of whether EMS participants are drawn from the extreme upper tail of the math score distribution. Figures 3-5 depict the distributions of mathematics test scores for the entire sample (Figure 3), for all bachelor's degree graduates (Figure 4), and for all EMS graduates (Figure 5). Figure 3 shows what we already know: there are more men than women at the upper tail of the math score distribution. Figure 4 shows that this is also true within the (more highly selected) group of college graduates. But, within the even more highly selected group of EMS college graduates, the distributions of men's and women's math test scores are much more similar (Figure 5). In the later cohort, the women's distribution looks nearly identical to the men's, while in the earlier cohort of EMS participants low scoring women are scarce. In fact, among EMS graduates in both cohorts, the women have a higher mean math score and smaller variance than the men. (And, while the variance is slightly higher among the men, virtually none of this gender difference is in the upper tail). Note that if the relationship between mathematics test scores and EMS participation were the same for men and women, then female EMS participants, drawn from a lower test score distribution, would tend to have lower average scores than male EMS participants.[9] The finding that the mathematics test scores of white women with EMS degrees are not lower than those of the men suggests that these women are more cautious about entering unless they have very high scores.

Another striking feature of the data described in Figure 4 is that the distributions for men in the two cohorts are nearly identical. Although the EMS participation of men increased substantially between cohorts, there was absolutely no adverse effect on the pool of participants, as measured by percentile math scores.

Although EMS participants are drawn from the upper portion of the math score distribution—90% have scores above the 60th percentile—they are by no means limited to

---

[8] Negative scores are recoded to zero before calculating the square of scores.

[9] Suppose, for example, that a proportion p of the high scoring and a proportion q of the low scoring are EMS participants, and that the high scoring are disproportionately men (say the ratio of men to women is 1 among the low scoring, and k>1 among the high scoring). Then, among EMS participants, there will be equal numbers of low scoring men and low scoring women, but k times as many high scoring men as high scoring women. Therefore, the ratio of hi to low scorers is k times as large among men than among women if, conditional on math scores, men and women are equally likely to be EMS participants.

individuals with math scores in the top few centiles. In both cohorts, more than 25% of both men and women have scores below the 75th percentile. More than half of the men, and nearly half of the women, have scores below the 85th percentile.

What Proportion of the EMS Workforce has SAT-M scores above 650?

Table 1 presents estimates of the proportion of the proportion of the age 32 EMS workforce who had (or would have had) SAT-M scores above 650, or no greater than 550. Selection of the 650 threshold is based on Paglin and Rufolo's (1990) assertion that EMS participants are drawn from the 651-800 range. The 550 threshold is chosen because it is approximately the average score of individuals who later completed a bachelor's degree with a humanities major (Angoff and Johnson, 1990). Estimates are based on white males only because the data presented in Figure 5 indicate that women in the earlier cohort faced more stringent selection criteria than men. Figure 5 shows that the distribution of EMS participant math scores among both men and women in the later cohort was nearly identical to that for men in the cohort studied here.

Estimates of the proportion of the EMS workforce with high SAT-M scores are made both from the subsample for whom SAT-M scores are available, and from the full, representative sample. For individuals with no SAT-M score reported, the status of SAT-M scores relative to the 650 and 550 thresholds is imputed, based on all available information about the joint distributions of SAT-M and CTM scores. (See Appendix).

The columns of Table 3 describe estimates for subsets of the workforce with varying degrees of EMS participation. Column 1 describes the full, representative sample of 1972 white male high school graduates employed in the 1979 workforce. In this group, only 6% scored above the 650 SAT-M threshold, while 76% had scores no greater than 550. Column 2 describes all workers in EMS occupations, Column 3 refers only to college-educated workers in EMS occupations, and Column 4 is restricted to EMS workers with an EMS college degree. Although men with high scores are overrepresented in the EMS workforce, they are not the majority. As might be expected, the proportion of high scorers increases as the definition of the EMS workforce is restricted to those with higher levels of education. But even in the Column 4 group, which includes those who attended graduate school before entering the EMS workforce, no more than 30% had scores above 650, while 32% had scores no greater than the average humanities major.[10] These estimates lead to the surprising conclusion that less than one-third of the EMS work force had SAT-math scores above the threshold previously presumed in the economics and cognitive psychology literatures.

But Are the Lower Scoring EMS Participants Successful in the Labor Market?

We now turn to the question of whether the lower scoring EMS participants were successful in the labor market. The answer to this question might tell us something about unobserved ability, as well as whether it is valuable for lower ability individuals to invest in an EMS education. It is possible that the lowest scoring EMS graduates lack the ability to complete a

[10] The corresponding proportion with scores above 650 was higher for women (0.50, n=22) and for male Ph.D.'s (0.67, n=9).

genuine college course in EMS, but somehow managed to obtain a credential with this title from a less-challenging academic institution. In this case, we expect the low-scoring group to be less successful in the labor market, with earnings similar to college graduates with non-EMS majors.

To answer this question, I fitted an ordinary least squares wage regression to determine whether the return to choosing an EMS college major depends on whether one had low (<575), medium (575-650), or high (>650) SAT-M scores. The dependent variable is the log of 1986 hourly earnings. This specification includes controls for SAT-M scores, educational attainment, and the interaction between them.[11] In column 1 of Table 4, the sample is restricted to white men with both SAT-M and CTM scores available. In column 2, the CTM score is used rather than the SAT-M, holding the sample constant. Again, I test the hypothesis that the return to choosing an EMS college major depends on whether one had low (<75th percentile), medium (75th-90th), or high (>90th percentile) scores. The score ranges are chosen so that the same proportion of EMS graduates fall into the low (or high) range of either score. In column 3, the same CTM regression is performed using the full representative sample of high school graduates.

There is, as expected, a positive return to math scores, to completion of a college degree, and to the interaction between them in all three specifications. This analysis also finds that the wage premium associated with EMS college majors does not vary systematically with the math score. In each specification, the hypothesis that low, medium and high scoring EMS participants all enjoy the same returns to choosing an EMS major cannot be rejected.[12] The result that the lowest scoring EMS graduates enjoy the same economic returns as graduates with high scores is extremely robust.[13] These regressions suggest that the lower scoring EMS graduates become productive members of the EMS workforce, earning about twice the wage premium associated with completing a bachelor's degree in other fields.

Modeling Measurement Error in Ability

It has been clearly established that individuals with math scores below the 90th percentile are not automatically excluded from EMS career paths. This section pushes the analysis still further to investigate whether there might be some unobserved component of "EMS ability," complementary to observed math scores, that can explain the entry of low scoring individuals into EMS while preserving the "top few centiles of ability" explanation of the gender differential in EMS participation.

Under this model, all EMS participants must have ability above a certain threshold, but many individuals have math scores below their true EMS ability. Among those with true ability above the threshold, the probability of participation is modeled to be increasing in ability, since it is likely that either the ease of completing EMS coursework or the range of attainable EMS occupations is increasing in ability. EMS participation is assumed to be zero for individuals with

[11] Hause (1972) argues that the return to a college education must be increasing in ability, since the opportunity cost of college attendance is higher for the more able.
[12] An additional specification compared individuals with high scores on either test to those with high scores on neither the SAT-M nor the CTM and again found virtually equal returns to both groups (.157 and .151). This rules out the possibility that equality of outcomes is a statistical artifact of combining high ability students who had a bad test day with students who truly have lower test-taking ability as well as low scores.
[13] The F-statistics are all three less than 0.3, compared with rejection thresholds ten times as large.

ability below the threshold, H, and to increase linearly in ability above H.[14] I also introduce a parameter, γ_{72}, to capture the fact that participation is lower in the 1972 cohort. We have already determined that the math scores of white men, both in the full population and among EMS participants, have the same distributions in both cohorts. This justifies the assumption that participation, conditional on ability, is proportionally higher (across all ability levels) in the second cohort. It also means that cohort membership, correlated with participation but uncorrelated with scores, serves as an ideal exclusion restriction to identify whether we are in a world with a low threshold or in a world with a high ability threshold but low correlation between math scores and ability.

The model just described is clarified below:

$A_i = S_i + u_i$ Equation 1

A_i = true EMS ability of individual i
S_i = math score of individual i
u_i = unobserved EMS ability of individual i

$$P_i(A_i) = \begin{cases} m(A_i - H) & \text{if } A_i > H \text{ and i in the 1980 cohort} \\ (1 + \gamma_{72})\, m\,(A_i - H) & \text{if } A_i > H \text{ and i in the 1972 cohort} \\ 0 & \text{if } A_i \leq H \end{cases}$$ Equation 2

P_i = Probability of EMS participation, conditional on A_i
H = ability threshold for EMS participation
m = slope of the relationship between ability and participation above H
γ_{72} = a "taste" parameter (empirically observed to be negative)

$$Q_i = (1 + \gamma_{72} cohort_{72}) \left\{ \int_H^{H+\frac{1}{m}} f_S(z) m(z - H) dz + \int_{H+\frac{1}{m}}^{\infty} f_S(z) dz \right\}$$ Equation 3

Q_i = Probability of EMS participation, conditional on i's math score (S_i)
$f_S(.)$ = probability distribution function of A_i (ability distribution, given the math score S_i)

Using maximum likelihood techniques, it is possible to simultaneously estimate the likely ability threshold for EMS participation and the likely dispersion of ability around math scores. This model will help to distinguish between two worlds: a world in which the relationship between math scores and EMS participation can be taken at face value and a world in which EMS participants are drawn from a small pool of individuals with very high ability, many of whom have scores well below their true ability. The model cannot determine whether the unobserved ability in question is a measure of performance below capability on the test, or of personal qualities that cannot be measured by a standardized test of mathematics. Nor can it tell us whether the supply of EMS ability can be increased through educational policies. But this model

[14] A specification in which participation is allowed to jump by a constant amount at the threshold resulted in an (impossible) negative jump, supporting the assumed zero constant.

can tell us something new about the current size of the eligible pool of potential EMS participants, and the gender composition of that pool.

The model was estimated under two different assumptions about the distribution of unobserved EMS ability. In one set of specifications, unobserved ability had the normal distribution. In a second set of specifications, the distribution of unobserved ability follows an exponential distribution. Because the log-likelihood was maximized at a higher value under the exponential specification, these are the results presented.[15]

In the exponential specification, no one has ability lower than scores. A fraction λ has ability equal to scores, and the density of unobserved ability falls off exponentially at rate λ with distance from the score. Formally, the density of ability, conditional on the math score (S), is:

$$f_S(z) = \begin{cases} \lambda e^{(-\lambda(z-S))} & \text{if } z \geq S \\ 0 & \text{if } z < S \end{cases} \qquad \text{Equation 4}$$

If λ is near 1, then scores are a very good proxy for ability. If λ is near zero, scores convey very little information about EMS ability.

Under this model, the likelihood of observing EMS participation, given a man's cohort and math score S_i, is:

$$Q_i = (1+\gamma_{72}cohort_{72})\left\{\int_{\max(score,H)}^{H+\frac{1}{m}} \lambda e^{-\lambda(z-score)} m(z-H)dz + \int_{H+\frac{1}{m}}^{\infty} \lambda e^{-\lambda(z-score)} dz\right\}$$

$$= \begin{cases} (1+\gamma_{72}cohort72)\left(\theta_2 e^{-\theta_1}\right)\left(1-e^{-\theta_3}\right) \| \theta_1 > 0 \\ (1+\gamma_{72}cohort72)\left(\theta_2(1-\theta_1)-e^{-\theta_3}\right) \| \theta_1 \leq 0 \leq \theta_3 \\ (1+\gamma_{72}cohort72) \| \theta_3 < 0 \end{cases} \qquad \text{Equation 5}$$

where $\theta_1 = \lambda(H - score)$, $\theta_2 = m/\lambda$, and $\theta_3 = (\theta_1 + 1/\theta_2)$

This model allows us to estimate λ and H, and to calculate an estimate of where the threshold lies within the true ability distribution.[16]

---

[15] The cohort and gender coefficients are similar in the normal and exponential specifications. Also, a variant of the normal specification with the dispersion allowed to be different above and below the score estimated that most of the dispersion of unobserved ability is above the score, as in the exponential specification described here.

[16] One seminar participant asked whether changes in the level of EMS interest would affect the estimated values of l and H. The answer is no. Changes in EMS interest would be captured by the g parameter. (This was verified by simulating changes in interest, randomly coding a fraction of EMS participants as non-participants).

The results of this analysis are presented in Table 5, Column 1. The dispersion of unobserved ability estimated by this model is quite large (λ=.20), and the estimated threshold is H=30.9 . The men of the 1972 cohort are estimated to be 21 percent less likely to be EMS participants than those in the 1980 cohort. This is close to the true difference in participation (5.8-4.5)/5.8=.22, a reassuring check on the accuracy of the model.

A better understanding of the meaning of the threshold estimate can be gained by using the parameter estimates to calculate the proportion of men with ability above the threshold for EMS entry. Given the specified model, all men with scores above the threshold H have ability above H with certainty. In addition, an individual with a lower score has ability above H with probability $e^{-\lambda(H-score)}$. The sum of these probabilities, across all individuals in the sample, yields an estimate of the proportion with true ability above the threshold H. This simple calculation yields the estimate that the threshold H lies at the 78^{th} percentile of the true ability distribution. An even simpler alternative calculation notes that the mean of unobserved ability is $1/\lambda$, so the average man at the ability threshold has math score= H - $1/\lambda$= 26, which is at the 78^{th} percentile. Looking back at Figure 2, the few stragglers with scores below 26 are estimated by this model to be high ability students who had a bad test day. But, although the 78^{th} percentile is toward the upper end of the ability distribution, it is far from the "top few centiles."

Another calculation sheds light on the gender composition of EMS participants. Plugging the column 1 parameter estimates into Equation 5 yields a prediction, for each individual, of the probability of EMS participation. The means of these predicted probabilities are identical to actual participation for men in each cohort. However, the means of predicted probabilities for women are again much higher than women's actual participation (.031 predicted vs. .009 actual for the 1972 cohort, .039 predicted vs. .018 actual for the 1980 cohort). These estimates are close to the highest counterfactual EMS participation rates estimated earlier using the naive models of Table 2. Modeling math scores as a noisy measure of ability does not change the conclusion that women are less than half as likely as equally able men to be EMS participants.

Further Analysis of the Gender Differential

The histogram of Figure 5 suggested that women may be even more cautious about entering EMS when their scores are not very high. It seems likely that the mean predicted probabilities of EMS entry described in the previous paragraph will be even farther above actual participation among lower scoring women. The calculations presented in Table 5b confirm that suspicion. Women with higher math scores are more strongly represented among EMS participants, but both high and low scoring women are underrepresented compared to men with the same math scores.

The gender difference in EMS participation can also be estimated by including a gender parameter directly into the MLE model. This is not as clean an experiment as including the cohort parameter because gender is correlated with math scores. This approach is also less than ideal because we know that the gender effect on EMS participation is not proportional across all ability levels. Nonetheless, specifications without, then with, a gender*cohort interaction

(Columns 2 and 3 of Table 5) both display a large gender differential in EMS participation conditional on EMS ability.[17]

Every specification described so far assumes that the dispersion of unobserved ability is the same for men and women. While it is reasonable to assume that the distribution of unobserved ability around scores is the same for men across cohorts, this is unlikely to be true when comparing men to women. Assuming there are more high-ability men than high-ability women, if it were the case that men and women with equal ability had the same distribution of scores, then a man with a given score would tend to have higher unobserved ability than a women with the same scores. This is because the man is more likely to be a high-ability individual who did not test well. However, research on standardized testing suggests that men and women with equal ability do not have the same distribution of scores. On average, women do not test as well as men with the same ability (Linn, et al. 1987, Rosser 1989, Byrnes and Takahira 1993). In fact, this effect is so strong that women appear to have more unobserved ability—as measured by later performance in the classroom—than men at a given level of math scores (Kimball 1989, Bridgeman and Wendler 1991, Kessel and Linn 1996). Therefore, the specifications presented so far will tend to underestimate the gender gap in EMS participation conditional on ability.

Of course, any possible gender differential could be explained by gender differences in an unobserved ability of some sort. In fact, this is the model underlying the notion that the demographic composition of the EMS workforce reflects the underlying distribution of attributes. In an illustrative example, equal participation conditional on ability is imposed. This exercise allows us to estimate the gender differential in the dispersion of unobserved ability required to sustain this model. The results are that men would have exactly the same dispersion of unobserved ability estimated in the model of Column 1 (λ= .20), while women must have far less unobserved ability (λ_{Female}=.845*.196+.196=.36) (Table 5, Column 4). In order to sustain the premise of this model, we have to believe that women with math scores at the men's 90th percentile have the same average EMS ability as men scoring at the 78th percentile of male high school seniors. The gender differential in participation can be explained away if we assume that men have quite a bit of "the right stuff," but this is no longer an argument based on gender differentials in math scores.

Addressing Additional Measurement Issues

In Table 6, various measurement issues are addressed. The first is how much the parameter estimates change when the model is estimated one year at a time (Columns 1 and 2). Despite losing the ideal exclusion restriction (cohort membership), the estimates are similar to those in Table 5, Column 3. The second issue is whether some of the dispersion in unobserved ability is related to the ability to pay for college. In Columns 3 and 4, the samples are restricted to college graduates only, and λ actually falls (estimated dispersion increases). According to this estimate, the dispersion of unobserved ability is at least as high within the group of students able to afford

[17] When the model of Column 2 is estimated for nonwhite women, rather than white women, the estimated coefficient is much smaller and is not statistically significant (-.13, with s.e. .13, compared to -.54 with s.e. .04 for white women). For nonwhite men, the estimated coefficient is zero (.02, with s.e. .12). With one exception, no statistically significant differences from white men are found for black, Hispanic or Asian women or men. The exception is Hispanic women, with participation patterns closer to those of white women.

college. The third issue is whether the estimates of the position of H in the true ability distribution and of gender differentials in participation conditional on ability would change if we had a better measure of mathematics ability. In columns 5 and 6, the samples are restricted to those for whom both the CTM and SAT-M scores are available. Within this sample, the estimated gender differential does not change when SAT-M scores are substituted for CTM scores. The estimated threshold for EMS participation is surprisingly low, 578 SAT-M points. And the estimated dispersion of unobserved ability is so high that 57 percent of men in this sample are estimated to have true EMS ability above the threshold. This estimate of the proportion of men above the EMS ability threshold is more than 50 percent higher than that obtained using CTM scores with the same sample and model. This finding means that, if SAT-M scores were available for a representative sample, the estimated true ability threshold would be far below the 78th percentile estimated using CTM scores.

There remains little room for doubt that many individuals with scores, and ability, well below the "top few centiles" complete EMS college degrees and enter well-paid occupations.

Summary and Conclusions

It was previously widely believed that entry into bachelor's degree level careers in engineering, mathematics, computer science, or physical sciences (EMS) was limited to individuals who had ability (as measured by high school mathematics test scores) in the top few centiles (Paglin and Rufolo, 1990, Benbow and Arjmand, 1992, Hedges and Nowell, 1995). This belief was so well accepted that no empirical analysis testing this assumption has yet been published.

This study shows that bachelor's degree level EMS participants are drawn from throughout the upper 40% of the mathematics test score distribution. While it is true that there is a high correlation between high school mathematics test scores and EMS participation, many EMS participants had high school mathematics test scores that were not exceptional. For example, fewer than one-third of college-educated white men working in EMS occupations had high school SAT-M scores above 650 (out of 800), while more than one-third had SAT-M scores below 550—the score of the average humanities major—at the 76th percentile of all white men.

The lowest scoring white men with EMS college degrees have the same earnings advantage, relative to college graduates with the same scores, as high scoring EMS participants. This suggests that EMS education is equally beneficial for this lower ability group.

When an unobserved component of ability is explicitly modeled, white male EMS participants are estimated to come from the upper 22% of the "true" ability distribution. It can not be determined from this analysis whether the unobserved component of ability is due to under-performance on the math test, or whether the unobserved ability important to EMS success is something that simply cannot be measured by these tests.

The result that relatively few EMS participants were drawn from the top few centiles of the mathematics test score distribution is robust to basing the definition of EMS participation on either educational or occupational attainment, and to using scores from two different standardized mathematics tests. Because EMS participants are not drawn from the very top of the distribution, there are many non-participants with comparable math scores or ability.

Between the early 1970's and the early 1980's, the EMS participation of white women increased substantially, but remained less than half that of white men with the same mathematics test scores. This result is robust to modeling math scores as a noisy measure of ability. White women at all levels of achievement are underrepresented in the EMS workforce relative to white men with the same mathematics test scores or ability.

The contrast between the results of this study and previously held convictions illustrates the stability of erroneous beliefs. For at least twenty years, thoughtful scholars in many academic fields have maintained a mistaken idea of the implications of observed distributions of mathematics test scores for the proportion of women capable of success in science and engineering careers. It is likely that ordinary employers and potential mentors are prone to similar mistakes. This "meta-finding" suggests a continued role for policies to ensure equitable educational and employment opportunities to members of traditionally underrepresented groups.

Atkinson, Richard C. "Supply and Demand for Scientists and Engineers: A National Crisis in the Making," Science, 248, p. 425-432, April 1990.
Badgett, M.V. and Nancy Folbre. "Job Gendering: Occupational Choice and the Marriage Market." Industrial Relations 42 (2) (2003): 270-298.
Benbow, Camilla and Stanley, Julian, "Sex Differences in Mathematical Ability: Fact or Artifact?" Science, 210(4475) 1262-64 Dec 1980.
Benbow, Camilla and Stanley, Julian, "Consequences in High School and College of Sex Differences in Mathematical Reasoning Ability: A Longitudinal Perspective," American Educational Research Journal, 1982, 19(4), 598-622.
Benbow, Camilla and Minor, Lola, "Mathematically Talented Males and Females and Achievement in the High School Sciences," American Educational Research Journal, 1986, 23(3), 425-436.
Benbow, Camilla and Lubinski, David, 1997. "Mathematics: Is Biology the Cause of Gender Differences in Performance?" in *Women, Men and Gender: Ongoing Debates*, Ed. Mary Roth Walsh, Yale Univ. Press, 1997.
Benbow, Camilla and Arjmand, Olya, "Predictors of High Academic Acheivement in Mathematics and Science by Mathematically Talented Students: A Longitudinal Study," Journal of Educational Psychology, 1990, 82(3), 430-441.
Berryman, Sue E., 1983. "Who will Do Science? Trends and their Causes" in Minority and Female Representation among Holders of Advanced Degrees in Science and Mathematics. A Special Report. Rockefeller Foundation, New York, New York.
Betz, Nancy E. and Hackett, Gail, "The Relationship of Career-Related Self-Efficacy Expectations to Perceived Career Options in College Women and Men," Journal of Counseling Psychology, 1981, 28 (5):399-410.
Betz, Nancy E. and Hackett, Gail, "The Relationship of Mathematics Self-Efficacy Expectations to the Selection of Science-Based College Majors," Journal of Vocational Behavior, 1983, 23(3), 329-345.
Betz, Nancy, "What Stops Women and Minorities from Choosing and Completing Majors in Science and Engineering?," in *Minorities and Girls in School*, David Johnson, ed., Sage, California, 1997.
Blakemore, Arthur E. and Stuart A. Low, 1984. "Sex Differences in Occupational Selection: The Case of College Majors." Review of Economics and Statistics, 66(February): 157-163
Bridgeman, Brent and Wendler, Cathy, "Gender Differences in Predictors of College Mathematics Performance and in College Mathematics Course Grades,"Journal of Educational Psychology, 1991, 83(2).
Brown, Charles and Mary Corcoran, "Sex-Based Differences in School Content and the Male-Female Wage Gap," Journal of Labor Economics, 1997, 15, 431-465.
Byrnes, James P. and Takahira, Sayuri, "Explaining Gender differences on SAT-Math Items," Developmental Psychology, 1993, 29(5), p. 805-810.
Daymont, Thomas and Paul Andrisani, 1984. "Job Preferences, College Major and the Gender Gap in Earnings." Journal of Human Resources, 19(Summer):408-428.
Eccles, Jaquelynne, "Gender Roles and Women's Achievement-Related Decisions," Psychology of Women Quarterly, 1987, 11, 135-172.
Eide, Eric, "College Major and Changes in the Gender Wage Gap." Contemporary Economic Policy, 1994, 12(April): 55-64.
Feingold, Alan. 1996. Cognitive Gender differences: Where are they and why are they there?" Learning and Individual differences. 8(1):2-32.
Feingold, Alan. 1988. Cognitive Gender differences are disappearing. American Psychologist. 43(2):95-103.
Fiorito, Jack and Robert Dauffenbach, 1982, "Market and Nonmarket Influences on Curriculum Choice by College Students," Industrial and Labor Relations Review, 36(1), 88-101.
Frehill, Lisa. 1997, "Education and Occupational Sex Segregation: The Decision to Major in Engineering." *The Sociological Quarterly*, 38(2):225-249**.**
Grogger, Jeff and Eric Eide, 1995. "Changes in College Skills and the Rise in the College Wage Premium." Journal of Human Resources, 30(Spring):280-310.
Hall, Roberta and Bernice Sandler, "The Classroom Climate: A Chilly One for Women?" Project on the Status and Education of Women. Washington, D.C.: Association of American Colleges, 1982.
Hanson, Sandra, *Lost Talent, Women in the Sciences*, Temple Univ. Press, 1996.
Hedges, Larry V. and Amy Nowell, "Sex Differences in Mental Test Scores, Variablility, and Numbers of High-Scoring Individuals," Science, 1995, 269, 41-45.
Hilton, Thomas and Lee, Valerie, "Student Interest and Persistence in Science: Changes in the Educational Pipeline in the Last Decade," Journal of Higher Education, 1988, 59(5), 510-526.
Hause, John C. 1972. "Earnings Profile: Ability and Schooling" *Journal of Political Economy*. 80(3, Part II): S108-S138

Hyde, Janet, “Gender Differences in Math Performance: Not Big, Not Biological,” in *Women, Men and Gender: Ongoing Debates*, Ed. Mary Roth Walsh, Yale Univ. Press, 1997.
Hyde, Janet, Elizabeth Fennema, and Susan Lamon. 1990. “Gender Differences in Mathematics Performance: A Meta-Analysis.” Psychological Bulletin, 101(2):139-155.
Jagacinski, Carolyn. 1987. “Engineering Careers: Women in a Male-Dominated Field.” *Psychology of Women Quarterly*, 11(1): 97-110.
Kessel, Cathy and Marcia C. Linn. 1996. “Grades or Scores: Predicting Future College Mathematics Performance,” Educational Measurement: Issues and Practice, 15(4):10-14.
Kimball, M. M., "A New Perspective on Women's Math Achievement," Psychological Bulletin, 105(2), p. 198-214, 1989.
Koshland, Daniel E. This untitled editorial opens: "The threat of serious shortage of scientific personnel looms in the years ahead..." Science , 239(4847), p. 1473, March 25, 1988
Lapan, Richard T., Peter Shaughnessy. and Kathleen Boggs, “Efficacy Expectations and Vocational Interests as Mediators between Sex and Choice of Math/Science College Majors: A Longitudinal Study,” Journal of Vocational Behavior, 49, 277-291, 1996.
Leslie, Larry L., Gregory T. McClure, and Ronald Oaxaca. 1998. “Women and Minorities in Science and Engineering, a Life Sequence Analysis.” *The Journal of Higher Education* 69(3):239-276.
Linn, Marcia C., Tina DeBenedictis, Kevin Delucchi, Abigail Harris and Elizabeth Stage. 1987. “Gender Differences in National Assessment of Educational Progres Science Items: What Does ‘I Don’t Know’ Really Mean?” Journal of Research in Science Teaching 24(3):267-278.
Lubinski, David, Benbow, C.P. and C.E. Sanders, “Reconceptualizing Gender Differences in Achievement among the Gifted,” in *International Handbook of Research and Development of Giftedness and Talent*, ed. K.Heller, F. Monks, H. Passow, Pergamon Press, 1993, 693-708.
Lunneborg, Patricia W., “Role Model Influences of Nontraditional Professional Women,” Journal of Vocational Behavior, 20(3), 276-281, 1982.
McIlwee, Judith and J. Gregg Robinson. 1992. *Women in Engineering, Gender, Power and Workplace Culture*. Suny Press, Albany, New York.
Momsen, Kent. 1974. “Black PhD’s in the Academic Marketplace.” Journal of Higher Education, 45(4).
Murnane, Richard J., John B Willet, and Frank Levy, 1995. “The Growing Importance of Cognitive Skills in Wage Determination,” Review of Economics and Statistics, 77(May): 251-66.
National Center for Education Statistics, 1994. "The Condition of Education," U.S. Department of Education, Washington, D.C.
National Center for Education Statistics, 1997. "The Condition of Education," U.S. Department of Education, Washington, D.C.
National Center for Education Statistics, 2003. “Digest of Education Statistics” U.S. Department of Education, Washington, D.C.
National Research Council, Committee on Women in Science and Engineering. "Science and Engineering Programs: On Target for Women?" National Academy Press, Washington, D.C., 1992.
National Research Council, Committee on Women in Science and Engineering. "Women in Science and Engineering: Increasing Their Numbers in the 1990s" National Academy Press, Washington, D.C., 1991.
National Science Foundation, *Meeting the National Need for Scientists to the Year 2000*, Commission on Professionals in Science and Technology, Washington, D.C., 1989.
National Science Foundation, *Women and Minorities in Science and Engineering*, Washington, D.C., 1990.
National Science Foundation, *Future Scarcities of Scientists and Engineers: Problems and Solutions,* Washington, D.C., 1990.
Oakes, Jeannie, 1990. *Lost Talent: The Under-participation of Women, Minorities, and Disabled Persons in Science*. NSF and Rand.
Paglin, Morton and Anthony M. Rufolo. 1990. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." Journal of Labor Economics 8(January):123-144.
Polachek, Solomon. 1978. "Sex Differences in College Major." Industrial and Labor Relations Review 31(July):498-508.
Polachek, Solomon W., 1981. "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure." Review of Economics and Statistics, 63(February):60-69.
Rosser, Phyllis *The SAT Gender Gap : Identifying the Causes*, (Center for Women Policy Studies, Washington, D.C., 1989).
Seymour, Elaine and Hewitt, Nancy, *Talking about Leaving: Factors Contributiong to High Attrition Rates Among Science, Mathematics and Engineering Undergraduate Majors* Bureau of Sociological Research, University of Colorado, Boulder, CO, April 1994.

Stanley, Julian C., in *Intellectual Talent: Psychometric and Social Issues*, C. P. Benbow and D. J. Lubinski, Eds. (Johns Hopkins Press, Baltimore, 1996), pp. 225-235.
Steele, Claude M., "A Threat in the Air, How Stereotypes Shape Intellectual Identity and Performance," American Psychologist, 1997a, 52(6), 613-629.
Stumpf, Heinrich and Julian C. Stanley, 1998. "Stability and change in gender-related differences on the College Board Advanced Placement and Achievement Tests" *Current Directions in Psychological Science.* Vol 7(6):192-196.
Tobias, Sheila, *Overcoming Math Anxiety*, Houghton Mifflin, Boston, 1978.
Tobias, Sheila, *They're not Dumb, They're Different: Stalking the Second Tier,* Research Corporation, Tucson, AZ, 1990.
Turner, Sarah and William Bowen. 1999. "Choice of Major: The Changing (Unchanging) Gender Gap," Industrial and Labor Relations Review, 52(2):289-313.
U.S. Office of Education. 1943. *National Survey of the Higher Education of Negroes.* Washington D.C., U.S. Government Printing Office.
Ware, Norma and Lee, Valerie, "Sex Differences in Choice of College Science Majors," American Educational Research Journal, 1988, 25(4), 593-614.
Weinberger, Catherine J., 1998 "Race and Gender Wage Gaps in the Market for Recent College Graduates," Industrial Relations 37(1), 67-84.
Weinberger, Catherine J., 1999, "Mathematical College Majors and the Gender Gap in Wages," Industrial Relations 38(3), 407-413.
Weinberger, Catherine J. 2001 "Is Teaching More Girls More Math the Key to Higher Wages?" in Squaring Up: Policy Strategies to Raise Women's Incomes in the U.S., edited by Mary C. King. University of Michigan Press.
Widnall, Sheila E. "AAAS Presidential Lecture: Voices from the Pipeline." Science, 241, p. 1740-1745, September 1988.
Xie, Yu and Kimberlee A. Shauman. 2003. Women in Science: Career Processes and Outcomes. Cambridge, MA: Harvard University Press.

Table 1—Sample Means

	(1)	(2)	(3)	(4)
	Men 1972 cohort	Men 1980 cohort	Women 1972 cohort	Women 1980 cohort
Full Sample:				
Raw CTM formula score	14.9	18.7	13.1	16.6
CTM percentile score	49	49	42	41
CTM formula score, rescaled	18.7	18.7	16.8	16.6
College Graduate	.28	.23	.26	.24
EMS College Graduate	.045	.058	.009	.018
Sample size	5191	2117	5317	2443
Subsample with both CTM & SAT-M available				
CTM percentile score	67	n/a	59	n/a
SAT-Math score	520	n/a	480	n/a
Sample size	1675		1651	

CTM is the cognitive test of mathematics given to participants in the NLS72 and HSB studies. Maximum score is 25 points in the 1972 cohort, and 32 points in the 1980 cohort. Percentile score based on the correspondence between the raw CTM score and the white male base year sample. The rescaled CTM score adjusts the 1972 score to the 32-point scale, using the 1972 percentile score and the inverse of the mapping from 1980 raw score to 1980 percentile score.

Table 2—Actual EMS participation and counterfactual EMS participation rates under different specifications of the relationship between observed scores and participation for white men.

	(1)	(2)	(3)	(4)
	1972 men n=5191	1972 women n=5317	1980 men n=2117	1980 women n=2443
Actual Proportion with EMS College Degree by the Resurvey	.045	.009	.058	.018
Counterfactual Proportion with EMS College Degree by the Resurvey, if Participation Matched Men with the Same Math Scores:				
Stepwise Model	.045	.031	.058	.038
Quadratic Model	.045	.032	.058	.038
Spline Model	.045	.031	.058	.038
Kernel Regression, bandwidth =0.8	.041	.027	.056	.035
Kernel Regression, bandwidth =1.2	.041	.027	.055	.035
Kernel Regression, bandwidth =1.6	.041	.028	.055	.035
Kernel Regression, bandwidth =2.0	.040	.027	.055	.035

TABLE 3: Estimated Percentages of White Male Employed Professionals in Engineering, Mathematics, Computer, or Physical Science (EMS) Occupations who had Very High, or Relatively Low, SAT-M Scores as 1972 High School Seniors.

	(1)	(2)	(3)	(4)
	Sample: 1979 Full Labor Force (N=4733)	Sample: In EMS Occupation 1979 (N=224)	Sample: In EMS Occupation with College Degree 1979 (N=161)	Sample: In EMS Occupation either 1979 or 1986 with EMS Degree by 1979 (N=145)
% with SAT-M > 650				
Main Estimate	6	23	27	28
Upper Bound Estimate	--	25	29	30
Subsample with SAT-M Reported	11	30	33	33
% with SAT-M ≤ 550				
Main Estimate	76	46	36	34
Lower Bound Estimate	--	43	33	32
Subsample with SAT-M Reported	63	36	28	26
% with EMS Degree	4	54	75	100
% with SAT-M Reported	32	62	70	70

For the "Main Estimate" individuals missing SAT-M scores were assigned imputed values based on all men with the same CTM score and SAT-M scores non-missing.
For the "Upper/Lower Bound Estimate" missing SAT-M scores were assigned imputed values based only on the white male EMS participants, rather than all men with both scores reported.
The "Subsample with SAT-M Reported" uses no imputed values, but is not representative.

Table 4--Earnings Premia to Men with College Degrees in Engineering, Mathematics, Computer Science or Physical Science (EMS) as a Function of 12th Grade Math Scores.

	Matched Sample with both CTM and SAT-M scores		Representative Sample
	(1)	(2)	(3)
Math test:	SAT-M / 10	CTM	CTM
EMS degree and low math score	0.163 (0.065)*	0.142 (0.064)*	0.186 (0.052)**
EMS degree and med. math score	0.171 (0.055)**	0.161 (0.055)**	0.208 (0.047)**
EMS degree and high math score	0.121 (0.063)	0.164 (0.057)**	0.182 (0.075)*
Bachelor's Degree or higher	0.166 (0.040)**	0.165 (0.035)**	0.173 (0.022)**
Math Score	0.004 (0.003)	0.003 (0.006)	0.006 (0.002)**
BA*(Math Score-mean Math Score)	0.003 (0.003)	0.013 (0.008)	0.008 (0.004)*
Constant	2.269 (0.138)**	2.384 (0.112)**	2.261 (0.028)**
Sample Size	1082	1082	2836
R-sq	0.074	0.074	0.081
Proportion of EMS grads in low score category	.35	.35	.39
Proportion of EMS grads in high score category	.27	.28	.25

* significant at 5%; ** significant at 1% (Standard Errors in Parentheses).
Low (high) scores are defined as SAT-M scores <575 (>650) or CTM scores <75th percentile (>90th percentile).
EMS degrees include all college graduates with undergraduate major in engineering, mathematics, computer science or physical sciences.
Dependent Variable: Mean Hourly Earnings 14 Years After High School; OLS regressions, 1986 response weights.
Sample: White Men with 1972 High School Math Scores Reported and 1986 Hourly Earnings between $1 and $100.

Table 5 --Maximum Likelihood Estimates of the EMS Participation Threshold and Gender Differential

	(1)	(2)	(3)	(4)
	Men Only	Men and Women	Men and Women	Men and Women
Estimated parameters:				
Cohort72	-.21 (.08)	-.23 (.05)	-.20 (.08)	-.28 (.07)
Female		-.54 (.04)	-.52 (.08)	
Female*cohort72			-.04 (.13)	
λ	.20 (.01)	.23 (.02)	.23 (.02)	.20 (.01)
$H * \lambda$	6.07 (.89)	6.41 (.34)	6.41 (.34)	5.58 (.40)
m / λ	.26 (.25)	.16 (.05)	.16 (.05)	.17 (.06)
$(\lambda_{Female} - \lambda)/\lambda$				.85 (.25)
Sample Size	7308	15068	15068	15068
Calculated from parameter estimates:				
Threshold (H)	30.9	28.0	28.0	28.5
Proportion of men with ability > H	.22	.32	.32	.32

H represents the ability threshold for EMS entry, in test score units.
λ is a measure of the dispersion of unobserved ability. (In column 4, for men only).
m is a measure of the rate at which EMS participation increases in ability above the threshold (for men in the 1980 cohort).
(standard errors in parentheses)

Table 5 b – (Actual EMS Participation)/(EMS Participation Predicted with Parameter Estimates of Table 5, Column 1)

	(1)	(2)	(3)	(4)
	Men 1972 cohort	Men 1980 cohort	Women 1972 cohort	Women 1980 cohort
Full Sample	1.00	1.00	.28	.46
Math Score $\geq 90^{th}$ percentile	1.02	.89	.41	.56
Math Score $< 90^{th}$ percentile	.99	1.07	.26	.42

Table 6 --Maximum Likelihood Estimates to Explore Various Measurement Issues

	(1)	(2)	(3)	(4)	(5)	(6)
cohort	1972	1980	1972	1980	1972	1972
Sample Restrictions	none	none	college graduates	college graduates	SAT-M score available	SAT-M score available
Math Score	CTM	CTM	CTM	CTM	CTM	SAT-M
Estimated parameters:						
Female	-.71 (.05)	-.52 (.08)	-.73 (.04)	-.61 (.06)	-.68 (.06)	-.65 (.06)
λ	.22 (.02)	.25 (.03)	.12 (.01)	.16 (.03)	.19 (.02)	.011 (.001)
$H * \lambda$	6.44 (.53)	6.65 (.61)	3.68 (1.03)	3.98 (.58)	5.81 (2.21)	6.35 (.68)
$m \quad / \lambda$	.17 (.10)	.11 (.04)	.31 (.34)	.26 (.07)	.27 (.62)	.12 (.03)
Sample Size	10508	4560	2858	1072	3326	3326
Calculated from parameter estimates:						
Threshold (H)	29.6	26.4	31.1	25.3	31.1	578
Proportion of men with ability > H	.25	.38	.52	.75	.37	.57

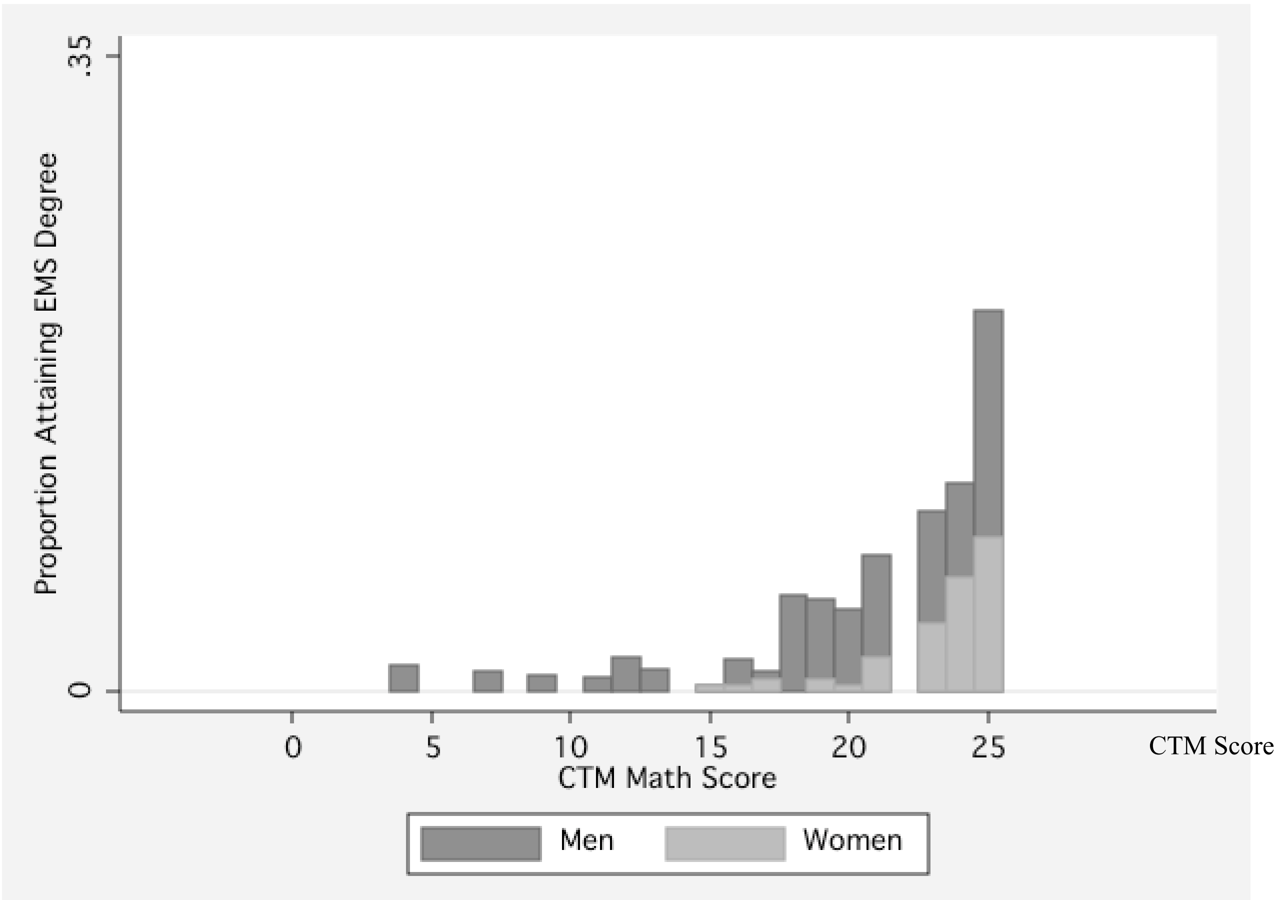


FIGURE 1: Proportion of White 1972 High School Seniors who Later Completed a College Degree in Engineering, Mathematics, Computer Science, or Physical Sciences (EMS), as a Function of 12th Grade Mathematics Test Score (CTM), By Sex. (n=5191 men, 5317 women)

(For comparison: a 1972 CTM score of 18 is at the 60th percentile of white male senior scores, 21 is at the 75th percentile, 24 is at the 90th percentile)

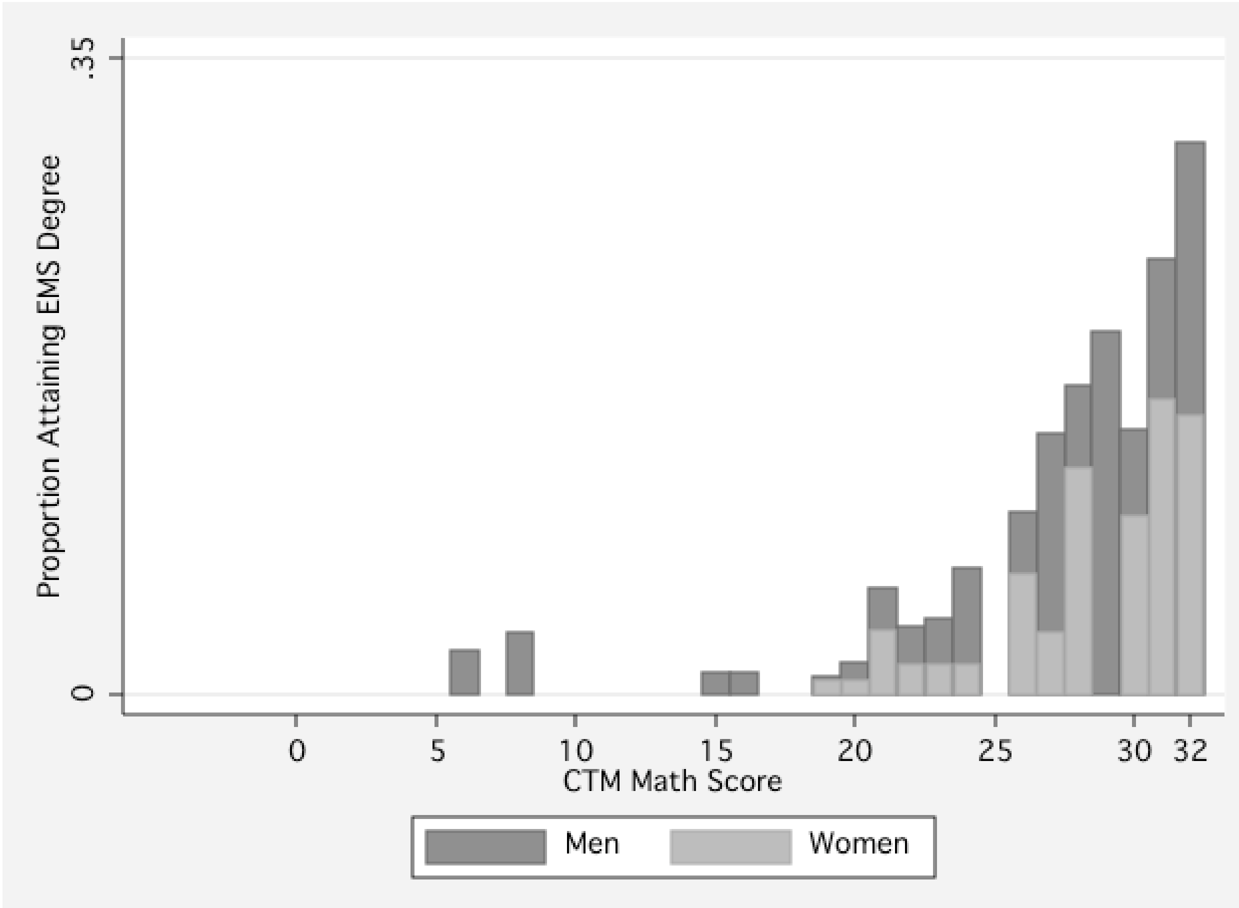
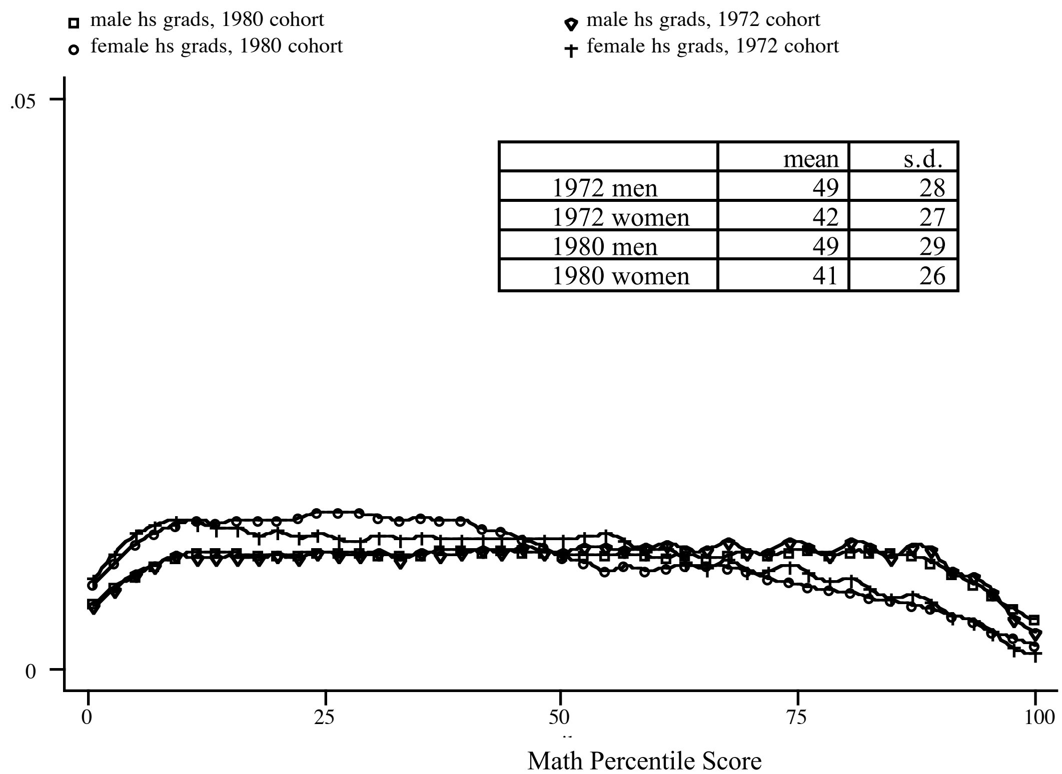


FIGURE 2: Proportion of White 1980 High School Seniors who Later Completed a College Degree in Engineering, Mathematics, Computer Science, or Physical Sciences (EMS), as a Function of 12th Grade Mathematics Test Score (CTM), By Sex. (n=2117 men, 2443 women)

(For comparison: a 1980 CTM score of 22 is at the 60th percentile of white male senior scores, 25 is at the 75th percentile, 29 is at the 90th percentile)



	mean	s.d.
1972 men	49	28
1972 women	42	27
1980 men	49	29
1980 women	41	26

FIGURE 3: High School Math Test Score Density Distributions of White High School Seniors, by Sex and Cohort.

(Scale is percentile score of all white men in the cohort)

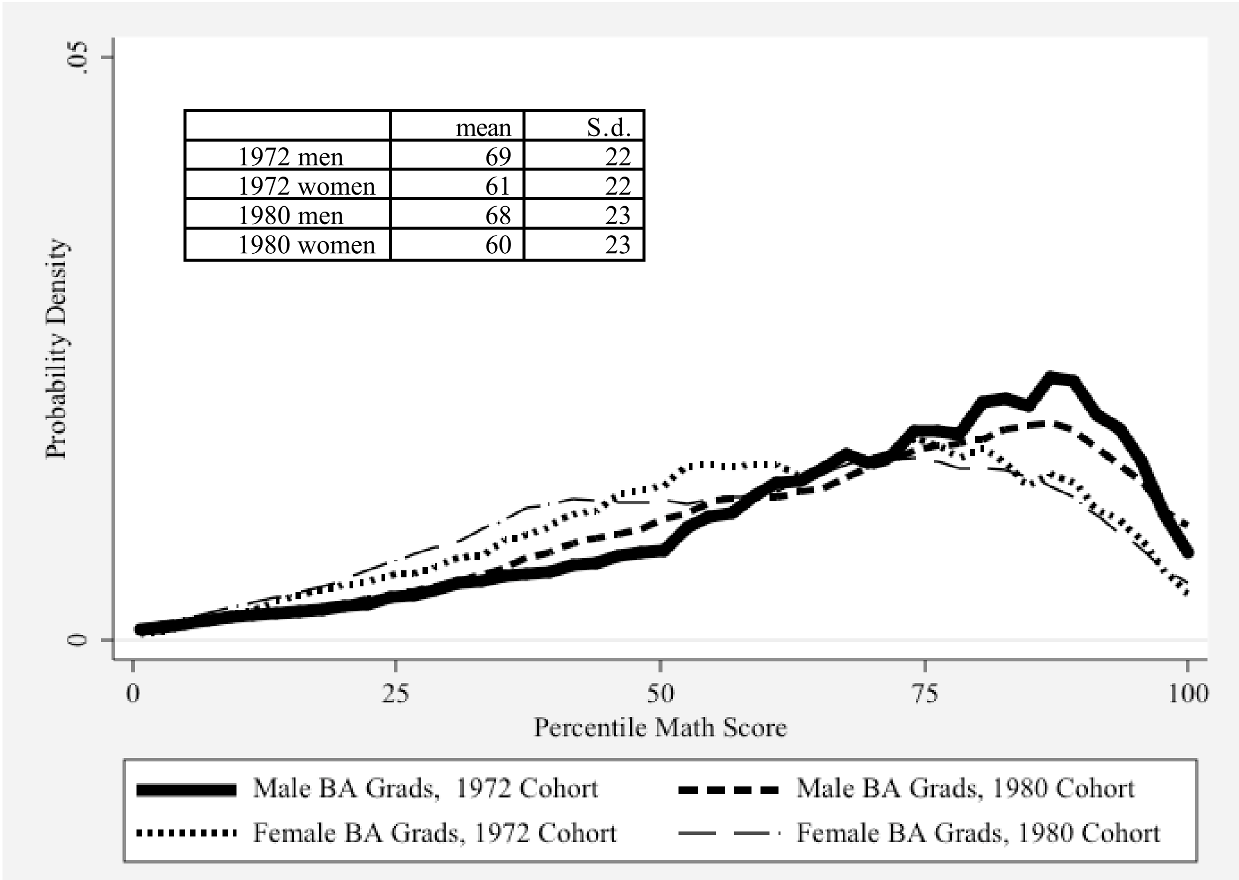
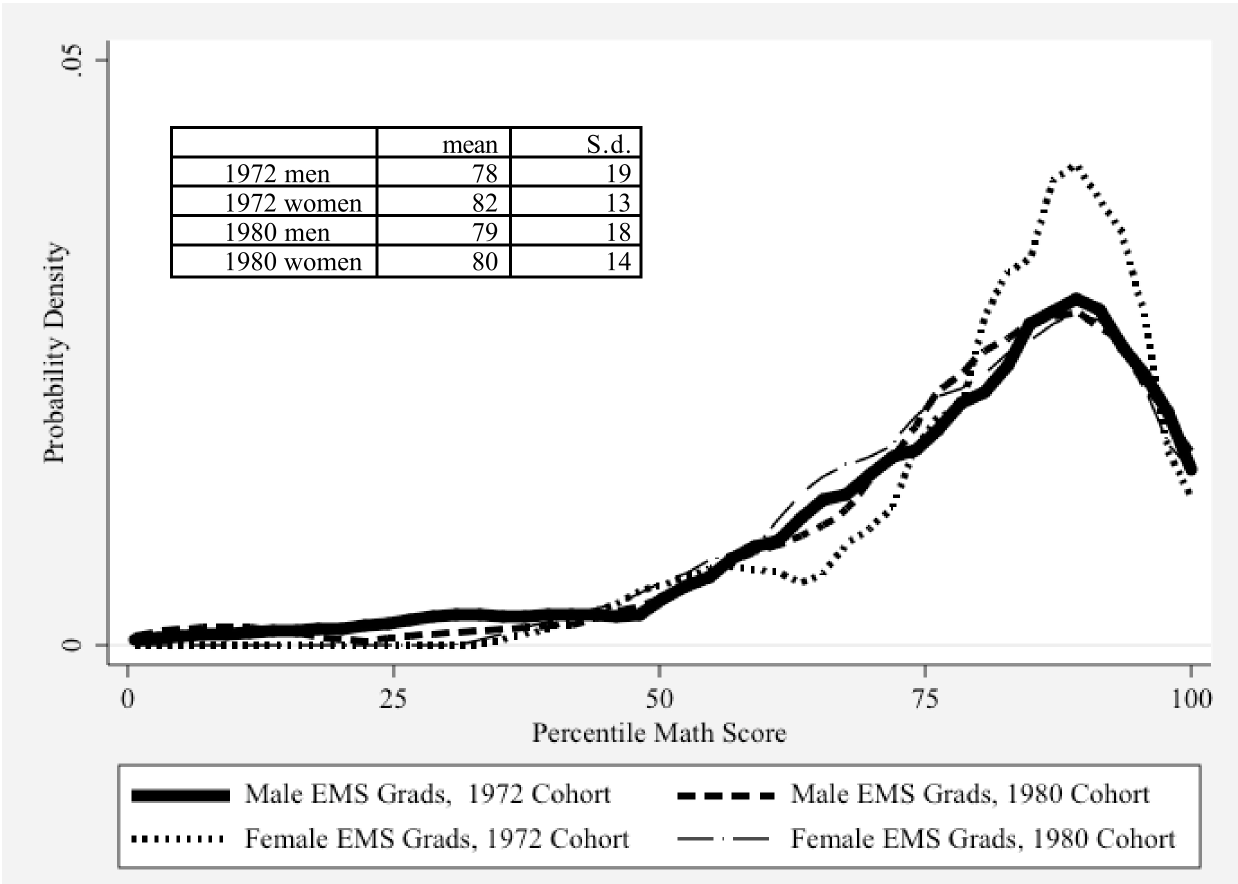


FIGURE 4: 12th Grade Math Test Score Probability Density Distributions of White Men and Women who Later Completed a College Degree, by Sex and Cohort.

(Scale is percentile score of all white men in the cohort)



	mean	S.d.
1972 men	78	19
1972 women	82	13
1980 men	79	18
1980 women	80	14

FIGURE 5: 12th Grade Math Test Score Probability Density Distributions of White Men and Women who Later Completed a College Degree with Major in Engineering, Mathematics, Computer Science or Physical Sciences, by Sex and Cohort.

(Scale is percentile score of all white men in the cohort)

Appendix--Imputed Proportions with High and Low SAT-M Scores

For individuals with no SAT-M score reported, the status of SAT-M scores relative to the 650 and 550 thresholds is imputed. Relying only on data for individuals with SAT-M score reported would lead to biased estimates because those with the greatest ability to earn high scores are the most likely to take the exam. An extension of the indicator variable provides an estimate of the proportion with scores greater than 650 for the entire sample.

The imputed values are created by assigning each man with no SAT-M score reported a conditional probability that he would have scored above 650 (or no more than 550), given his observed cognitive test of mathematics (CTM) score. The conditional probability that an individual with no SAT-M score reported would have earned a score greater than 650 is estimated to be equal to the proportion of all white men with a similar CTM score who earned SAT-M scores greater than 650. In cases where no SAT-M is known, this conditional probability is substituted for either 0 or 1 in the indicator variable. For example, among those with both scores reported, 14.2% of white men with CTM scores between 21.5 and 23 scored over 650 on the SAT-M. Therefore, for individuals with CTM scores between 21.5 and 23, the indicator variable for high SAT-M scores is set to 1 if the SAT-M score is greater than 650, 0.142 if SAT-M is not reported, and 0 if SAT-M is reported but not greater than 650. The mean of this new variable is the "main estimate" of the proportion of the representative sample likely to earn SAT-M scores greater than 650.

The main estimate might still overstate the number of high scorers if those who did not take the SAT would actually have earned lower SAT-M scores than those with similar CTM scores who took the SAT. A second imputed value is calculated for EMS participants, with conditional probabilities based on EMS participants only. There are two reasons to expect that an EMS participant with a given CTM score would report a higher SAT-M score than the typical person with the same CTM score. First, SAT-M scores are used as a basis for counseling students about the career paths that are open to them, so the SAT-M score might have an independent effect on career choices. Second, individuals who pursued EMS careers but reported lower CTM scores are more likely than the typical student to have negative measurement error in their CTM scores. To correct for both of these possibilities, and to get an absolute upper bound on the proportion of the EMS workforce who might have had very high SAT-M scores, imputed estimates for EMS populations were based on conditional probabilities of high or low scores for the subsample of white men with both scores reported who either completed an EMS bachelor's degree before the 1979 resurvey, or who were employed in an EMS occupation either in 1979 or in 1986.

The estimated conditional probabilities are reported in the Appendix Table.

Appendix Table: Conditional Probability Estimates. Proportion of White Men with SAT-M Scores Above (or Below) the Indicated Thresholds, as a Function of Cognitive Test of Mathematics Scores and College Major. 1972 Senior cohort.

	Based on All White Men (N=1944)		Based on White Male EMS Participants Only (N=352)	
	SAT-M ≤ 550	SAT-M > 650	SAT-M ≤ 550	SAT-M > 650
CTM ≤ 10	.978	.000	1.000	.000
10 < CTM ≤ 15	.975	.004	.909	.000
15 < CTM ≤ 18	.925	.004	.938	.000
18 < CTM ≤ 20	.803	.011	.711	.022
20 < CTM ≤ 21.5	.582	.061	.482	.111
21.5 < CTM ≤ 23	.429	.142	.250	.273
23 < CTM ≤ 24	.210	.269	.182	.364