NBER WORKING PAPER SERIES

THE GENDER GAP IN SECONDARY SCHOOL MATHEMATICS AT HIGH ACHIEVEMENT LEVELS: EVIDENCE FROM THE AMERICAN MATHEMATICS COMPETITIONS

Glenn Ellison
Ashley Swanson

Working Paper 15238
http://www.nber.org/papers/w15238

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

This project would not have been possible without Professor Steve Dunbar and Marsha Conley at AMC, who provided access to the data as well as their insight. We thank Stu Schmill, Sofia Caloggero, and Karen Johnson for their help with the MIT data and David Card and Jesse Rothstein for help with data matching. Syndie Kim provided research assistance. Financial support was provided by the NSF (SES-0550897), the Toulouse Network for Information Technology, and Microsoft Research. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions
Glenn Ellison and Ashley Swanson
NBER Working Paper No. 15238
August 2009
JEL No. I2,J16

ABSTRACT

This paper uses a new data source, American Mathematics Competitions, to examine the gender gap among high school students at very high achievement levels. The data bring out several new facts. There is a large gender gap that widens dramatically at percentiles above those that can be examined using standard data sources. An analysis of unobserved heterogeneity indicates that there is only moderate variation in the gender gap across schools. The highest achieving girls in the U.S. are concentrated in a very small set of elite schools, suggesting that almost all girls with the ability to reach high math achievement levels are not doing so.

Glenn Ellison
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive, E52-380A
Cambridge, MA 02142-1347
and NBER
gellison@mit.edu

Ashley Swanson
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02142-1347
aswan@mit.edu

1 Introduction

The gender gap in average math test scores receives a great deal of attention in both academia and the popular press.[1] Without denying the importance of studying the performance of millions of students, we feel that the fact that so much less attention is paid to the gender gap in the upper tail is unfortunate for a couple of reasons. First, upper tail outcomes are potentially relevant to various important topics including the underrepresentation of women in math and science careers. Second, and more importantly, the gender gap in mean scores is sufficiently small so as to be of little practical importance, whereas the gender gap in the upper tail can be quite large. In this paper, we explore the gender gap among high-achieving high school students using a new data source: the American Mathematics Competitions. Among our findings are that the gender gap widens dramatically at very high percentiles, that there is some but not a lot of variation in the size of the gender gap across schools, and that the highest-achieving girls are concentrated in a very small set of elite schools.

The American Mathematics Competitions are a series of contests sponsored by the Mathematical Association of America. The annual series begins with two contests, the AMC 10 and AMC 12, which are held at over 3,000 U.S. high schools. The AMC 12 is open to any interested high school student and contains 25 multiple-choice questions on precalculus topics: algebra, probability, geometry, and trigonometry.[2] Approximately 125,000 U.S. students participate. We feel that the AMC contests are a potentially valuable new data source because they are much better than standard tests at distinguishing among high-achieving students. (We presents two types of evidence to support this contention: statistical evidence showing that the AMC tests remain calibrated even at very high percentiles; and sample questions from various tests which we think make the differences evident.) The AMC contests also have a big disadvantage as a research tool: the participants are a highly nonrepresentative, self-selected sample. This influences the analyses we do, and leaves us with multiple potential explanations for some findings.

We begin our analysis of the gender gap in AMC scores in Section 3 by simply graphing the male-female ratio at different score levels. One basic fact is that the high-achievement

[1] One statistic on the size of the academic literature is that Hyde and Linn (2006) identified 46 meta-analyses of the topic which together summarize more than 5000 studies. Stories on Hyde et al. (2008) and/or Guiso et al. (2008) appeared in the New York Times, Wall Street Journal, National Public Radio, ABC News and many other outlets.

[2] The AMC 10 is restricted to students in grade 10 and below. It is similar, but is somewhat less difficult and avoids trigonometry and pre-calculus topics.

gender gap in the AMC data is somewhat larger than has been reported in most previous studies: whereas male-female ratios of about 2:1 have been reported in studies of 99^{th} percentile performers on several standardized tests, we find a ratio of 4:1 among students scoring at least 100 on the AMC 12.[3] The most striking finding of Section 3 is a new finding that exploits the AMC's calibration at higher percentiles: we find that the gender gap widens dramatically as one moves to higher and higher percentiles and exceeds 10 to 1 at the upper end of our data.

Section 4 examines variation in the gender gap across schools. Differences in the gender gap across schools (or other student groups) are of interest both because they provide insight into the causes of the gender gap and because they may help identify policies that might narrow the gap.[4] We focus on students scoring above 100 and 120 on the AMC 12 and begin with the most basic question: how much variation is there in the gender gap across schools? Our first analysis, which examines the degree to which female high scorers are clustered, provides another striking new finding: there is statistically significant variation in the gender gap across schools, but the magnitude of the variation is moderate and almost all high achieving high schools appear to have a substantial gender gap.[5] We also use regressions to look for systematic relationships between the magnitude of the gender gap and school and region demographics like parental education, income, and percent Asian. These tests yield little evidence of systematic variation.

Section 5 brings in data from other math contests: the U.S. Mathematical Olympiad (USAMO) and International Mathematical Olympiad (IMO). These contests are of less intrinsic interest – they require proof-writing skills not taught in most high schools and are only taken by extreme high achievers – but provide an opportunity to get additional insights both in the U.S. and worldwide. One potentially important observation from the U.S. data is that the highest-scoring boys and the highest-scoring girls appear to be drawn from very

[3]This score is at the 94th percentile in the AMC data, which is probably around the 99th percentile for the SAT-taking population. See Feingold (1992) and Hedges and Nowell (1995) for discussions of a number of datasets including the National Assessment of Educational Progress (NAEP) and Hyde et al. (2008) and Guiso et al. (2008) for recent papers using examining state proficiency tests and Programme for International Student Assessment (PISA), respectively.

[4]One closely related paper here is Andreescu et al.'s (2008) study of participants at the International Mathematics Olympiad which states that "Girls were found to be 12% to 24% of the children identified as having profound mathematical ability when raised under some conditions; under others, they were 30-fold or more underrepresented."

[5]More formally and precisely, our main statistical model supposes that the number f_i of female high scorers in a school is distributed Binomial(N_i, p_i), with p_i itself being distributed Beta(α, β), and estimate the mean and standard deviation of p_i. We find a mean of 0.18 and a standard deviation of 0.05 when using a threshold score of 100 on the AMC 12. An example for the "almost all" fact is that 127 of the 131 schools with eight or more students scoring above 100 on the AMC 12 have more boys than girls reaching this threshold.

different pools. Whereas the boys come from a variety of backgrounds, the top-scoring girls are almost exclusively drawn from a remarkably small set of super-elite schools: as many girls come from the top 20 AMC schools as from all other high schools in the U.S. combined. This suggests that almost all girls with extreme mathematical ability are not developing their talent to the degree necessary to do very well on the Olympiad contests. Our IMO analysis reexamines a dataset on the gender of IMO participants developed by Andreescu et al (2008). Here, our main observation is similar to our finding on the gender gap across schools and quite the opposite of what Andreescu et al. emphasize: we note that there is strikingly little variation across countries in the (very large) magnitude of the gender gap.

As noted above, there is a large literature on the gender gap in average test scores. In earlier decades, boys took substantially more math courses in high school and studies tended to find a nontrivial gender gap in average test scores.[6] The gap in coursetaking has largely gone away and more recent studies on universally administered tests typically find that there is now a fairly small gap in average test scores.[7]

The literature on the gender gap among high achievers is smaller. Feingold (1992) includes a nice survey going back to the 19th century. Benbow and Stanley's (1980) report of a 4.5 to 1 male-female ratio among 7th grade "talent search" students scoring at least 600 on the math SAT was highly publicized, but also seems to have hampered further research.[8] Several recent studies report male-female ratios of around 2:1 at the 99th percentile: there is a 2.1:1 male-female ratio among students scoring 800 on the math SAT; Hyde et al. (2008) report a similar figure using data on state proficiency tests; Xie and Shauman (2003) present estimates in this range from a number of (older) tests; and studies of more recent

[6]Hyde et al.'s (1990) meta-analysis of 100 studies from 1963-1988, for example, reported that boys scored 0.29 standard deviations higher on tests of complex problem solving.

[7]Hyde et al. (2008) report an 0.06 standard deviation gap on 11th grade proficiency tests. The gap on the NAEP is about 0.1 standard deviations. PISA scores show a gender gap in almost all covered countries that is usually a little bigger than this. See Freeman (2005), Perie (2005), OECD (2006), and Guiso et al. (2008) for more on the NAEP and PISA gender gaps. Fryer and Levitt (2009)'s results suggest a somewhat larger gap: they examine data from another representative test, the Early Childhood Longitudinal Study, and find that a gender gap of 0.2 standard deviations has already emerged by 5th grade. The SAT gender gap has not narrowed to the same extent – over the past 40 years the drop is just from 40 points to 34 – but selection into test-taking makes interpretation of mean scores difficult.

[8]*Science*'s editors allowed Benbow and Stanley to state (with essentially no other evidence) that

> We favor the hypothesis that sex differences in achievement in and attitude toward mathematics result from superior male mathematical ability.

Much of the subsequent literature seems more focused on refuting this hypothesis than on understanding the gender gap. Papers that could discuss male-female ratios among high achievers sometimes seem to shy away from doing so or to minimize findings. For example, Hyde et al. (2008) say "even at the 99th percentile the gender ratio favoring males is small for whites ..." when the ratio is in fact above 2:1 in their data and Guiso et al. mention in their text that the male-female ratio is less than 1:1 for 99th percentile students in Iceland, but do not mention that this ratio is above 1.6:1 in 36 of the 40 countries in their study.

"talent-search" cohorts also find gender gaps closer to this level.[9]

Our paper contributes to this literature in a few ways. One of these is to provide a more detailed look at how the gender gap widens as one moves beyond the 99th percentile. A number of papers report gender ratios at multiple percentiles such as the 90th and 95th or 95th and 99th, but giving more detail than this is rare and in any case most papers cannot give meaningful statistics on percentiles above the 99th because the tests they are based on do not draw meaningful distinctions at such levels.[10] Another of our main contributions is to examine variation in the gender gap across schools and across countries. We are not aware of any papers presenting analyses of the idiosyncratic variation of the high achievement gender gap across schools.[11] Machin and Pekkarinen (2008) is a recent paper with a similar message to ours on cross-country differences: they report that there is a greater variance in the male population on PISA in 34 of 40 countries. A few other cross-country analysis have emphasized heterogeneity rather than similarity of the gender gap. Feingold (1994) gathered together results from various studies and noted that findings did not appear to be consistent across countries. Guiso et al. (2008) say that their finding that the gender gap is smaller in more gender-equal countries carries over to the tail.[12] Andreescu et al. (2008) is closely related to part of our paper as noted above.

One motivation for studying the gender gap among high-achieving high school students is that the phenomenon may be related to the underrepresentation of women in scientific fields.[13] There is a vast literature on this topic motivated both by concern for women – the fact that the lack of women in technical fields appears to be a significant contributor to the gender gap in wages – and by concern for scientific progress.[14] Xie and Shauman (2003) provide a nice overview and discuss research into dozens of factors that may be important

[9]See Brody and Mills (2005).

[10]One exception is Andreescu et al. (2008): its finding of a 13.5:1 male-female ratio among International Math Olympiad participants can be thought of as a data point on an extremely high percentile of the distribution.

[11]The Hyde et al. (2008) finding that there is no gender gap among 99th percentile Asian Americans in Minnesota can be seen as a related statistic being used to portray male-female ratios at the high end as not consistent. Fryer and Levitt (2009) note that the gender gap in their data (which concern mean scores of fifth graders) is very consistent across demographic groups.

[12]Their online appendix, however, includes a chart with a message more similar to ours: Figure S2A appears to indicate that the male-female ratio at the 99th percentile is above 1.6:1 in 36 of the 40 countries they study.

[13]See Weinberger (2005) for a contrarian view arguing that technical professionals are not necessarily high achievers on standardized math tests. Another important caveat is that a gender gap in high school test scores may not carry over into a gender gap in performance in college courses. Leonard and Jiang (1999), for example, present evidence that SAT scores underpredict women's grades.

[14]See Paglin and Rufolo (1980), Brown and Corcoran (1997), and Blau and Kahn (2000) among others for discussions of the gender gap in pay.

here (including differential math preparation in secondary school).[15]

Our paper is obviously very different from the recent papers that have attracted attention in the popular press with their "girls are doing great" message. But we do not see our results as entirely discouraging. Recognizing that there is a larger gender gap among high achieving students may be an important first step to finding solutions relevant to this part of the distribution. We won't have much to say on policies to narrow the gender gap, but hope that our findings will improve understanding and spur further research.

2 Measuring High Math Achievement: The AMC and Other Tests

In this section we provide an introduction to the AMC contests and note advantages and disadvantages relative to other more commonly used tests of math achievement. The primary advantage of the AMC tests is that they are designed to identify and distinguish between students at high achievement levels. The primary drawback is that they are administered to a self-selected and highly nonrepresentative subset of the population.

2.1 The American Mathematics Competitions

The Mathematical Association of America has sponsored the American Mathematics Competitions (AMC) since 1950. They are given in over 3,000 high schools and at a number of other locations.[16] Some schools have hundreds of students participate, but it is far more common for the AMC exams to be taken by a couple dozen self-selected students. Our primary focus will be on the AMC 12, which is taken by about 125,000 students in a typical year. Test takers are roughly evenly distributed between grades 11 and 12.[17] The test is offered on just two dates each year: the first, in the second week of February, is referred

[15] A number of well-designed studies examine various potential links. For example, Bettinger and Long (2005) and Carrel, Page, and West (2009) study the impact of the gender gap in current faculty on student achievement and future coursetaking.

[16] Although this is a small fraction of the number of public and private high schools in the U.S. (about 15%), the AMC is much more likely to be offered in high-achieving high schools, which probably makes it available to a much higher fraction of the top students in the U.S. than a simple count of high schools would suggest. To provide some evidence on this we gathered data on the distribution of National Merit Semifinalists and AMC availability in Massachusetts, a state with relatively high AMC participation, and Oklahoma, a state with relatively low AMC participation. In Massachusetts, we found that nearly 80% of the 2009 National Merit Semifinalists attend high schools that offered the AMC in 2007. In Oklahoma, even though a very low percentage of high schools participate in the AMC, we still found that 50% of the 2009 National Merit Semifinalists attend high schools that offered the 2007 AMC. Data on national merit semifinalists were obtained from articles published in *Education Station* and *Boston Globe* on September 10, 2008.

[17] About 10% of test takers are in grades 10 and below.

to as the 12A and the second, which occurs 15 days later, is referred to as the 12B. High scorers (typically students who score above 100) are invited to participate in subsequent AMC contests. The AMC series of contests are the most prestigious high school contests in the U.S. and some elite colleges (including MIT, Yale, Brown, and Cal Tech) have places for students to report AMC scores on their application forms.

The AMC contests are designed to distinguish among students at high performance levels. The AMC 12 asks students to solve 25 problems in 75 minutes. The questions increase substantially in difficulty from the beginning of the test to the end and require in varying degrees knowledge of pre-calculus mathematics and/or problem solving skills. Figure 1 presents some sample questions: those numbered 13 through 17 on the 2007 AMC 12A. Some require specific knowledge like the equation for a line or trigonometric identities. Others do not. Most require some creative problem solving. The AMC scoring rule incorporates a mild anti-guessing penalty: students get 6 points for a correct answer, 1.5 points for a blank answer, and zero points for an incorrect answer. The range of possible scores is 0 to 150.

We examine the gender gap at various AMC score levels, but will sometimes focus on students scoring 100 or higher on the AMC 12. There were two main reasons for this choice: we believe that it roughly corresponds to a 99th percentile score on the math SAT (i.e. 780-800) which facilitates comparisons with other studies; and it is a round number which is even more focal in the AMC world because it is the cutoff for automatic advancement to the AIME.[18]

Note that when we say that a 100 on the AMC 12 is roughly equivalent to a 99th percentile SAT score we do not mean that students scoring 100 on the AMC 12 will usually get an 800 on the SAT. Getting an 800 on the SAT is a random event – one needs to make zero mistakes in the course of answering 54 questions at a rate of 78 seconds per question. Indeed, the College Board reports that only 15% of students who retook the math SAT after scoring 800 scored 800 on the retake. The average retake score was 752 (which is only 11 points higher than the average retake score of students who scored 760 on their previous attempt). Instead, we mean that we would guess that a typical student who scored 100 on the AMC 12 and a typical student who scored 800 on the math SAT would be expected to do about as well if given another similar math test.

[18]The sample questions in Figure 1 also provide some sense of the level of achievement corresponding to a score of 100. Fourteen correct answers and eleven blanks gives a score of 100.5. For this reason most students scoring 100 or higher on the AMC 12A got each of the first three sample questions correct. The last two were answered correctly by 44% and 64% of students scoring at least 100.

13. A piece of cheese is located at $(12, 10)$ in a coordinate plane. A mouse is at $(4, -2)$ and is running up the line $y = -5x + 18$. At the point (a, b) the mouse starts getting farther from the cheese rather than closer to it. What is $a + b$?

(A) 6 **(B)** 10 **(C)** 14 **(D)** 18 **(E)** 22

14. Let a, b, c, d, and e be distinct integers such that

$$(6 - a)(6 - b)(6 - c)(6 - d)(6 - e) = 45.$$

What is $a + b + c + d + e$?

(A) 5 **(B)** 17 **(C)** 25 **(D)** 27 **(E)** 30

15. The set $\{3, 6, 9, 10\}$ is augmented by a fifth element n, not equal to any of the other four. The median of the resulting set is equal to its mean. What is the sum of all possible values of n?

(A) 7 **(B)** 9 **(C)** 19 **(D)** 24 **(E)** 26

16. How many three-digit numbers are composed of three distinct digits such that one digit is the average of the other two?

(A) 96 **(B)** 104 **(C)** 112 **(D)** 120 **(E)** 256

17. Suppose that $\sin a + \sin b = \sqrt{5/3}$ and $\cos a + \cos b = 1$. What is $\cos(a - b)$?

(A) $\sqrt{\frac{5}{3}} - 1$ **(B)** $\frac{1}{3}$ **(C)** $\frac{1}{2}$ **(D)** $\frac{2}{3}$ **(E)** 1

Figure 1: Questions 13 thorugh 17 from the 2007 AMC 12A

To provide some evidence on the AMC-SAT correspondence, we gathered data on SAT and AMC scores for a sample of MIT applicants.[19] The first row of Table 1 gives the mean math SAT score that students with junior-year AMC 12 scores in the given range earned when they first took the SAT. The second row gives the percentage of students scoring 800 (again on their first attempt). Students scoring 120 and above on the AMC 12 did much better on the SAT than would a typical student with a previous 800 SAT – they are four times as likely to score 800.[20] Students with AMC 12 scores in the 100-109.5 range are doing somewhat better than the reported retake performances of students with an 800 SAT. Students with AMC 12 scores in the 90-99 range look more like students with a 760 on a previous SAT. We would like to emphasize that these are not unbiased estimates of SAT performance: they are derived from applicants to MIT and, especially in the low AMC ranges one would expect a substantial upward bias because students with SAT scores typical for that AMC score would probably not apply to MIT. We present the data to support our rough assessment of what a 100 on the AMC 12 means and to show that students who score higher on the AMC 12 look stronger on other metrics as well, and do not intend them as an SAT-AMC translation.

	Statistics for students with AMC 12 scores in each range								
	60's	70's	80's	90's	100's	110's	120's	130's	140's
Mean on 1st Math SAT	706	710	711	745	773	774	791	793	800
% with 800 on 1st SAT	0	0	0	19	35	38	65	60	100
Sample size	5	5	12	32	83	21	32	10	5

Table 1: SAT math scores for a sample of students with SAT scores various ranges

Figure 2 contains a histogram of AMC 12 scores for 2007. The average score is typically around 65 out of 150.[21] About 6.5% of students taking the AMC 12 scored 100 or better.

[19]Our AMC data do not include students' names, so we could not match MIT applicants directly to our data. The AMC does, however, publish scores for all students scoring at least 100. Accordingly, we collected data in two ways: we drew a random sample of applicants and looked in the published data to see if their scores were above 100; and we drew another semi-random sample and looked to see if applicants self-reported scores below 100. We dropped three students with SAT scores that were extreme outliers. Each of these students scored at least 100 points lower than any other student with an AMC score in the same 10 point range, which we took to suggest either that there was some problem with the matching or perhaps that the AMC score may not have been valid.

[20]This refers to the College Board's 15% statistic noted earlier. We also collected SAT data on a small random sample of MIT applicants without self-reported AMC scores. This group contained eight students who scored 800 on their first SAT attempt and then retook the SAT. Two of the eight scored 800 on the second attempt. The mean second score was 775.

[21]Interested readers should refer to the Appendix for a detailed description of the data. When students participated in the AMC 12 exam multiple times, the latter of the two scores is included here and in subsequent analyses.

About 0.8% of students scored above 120 on the AMC 12. About 0.06% scored above 140.

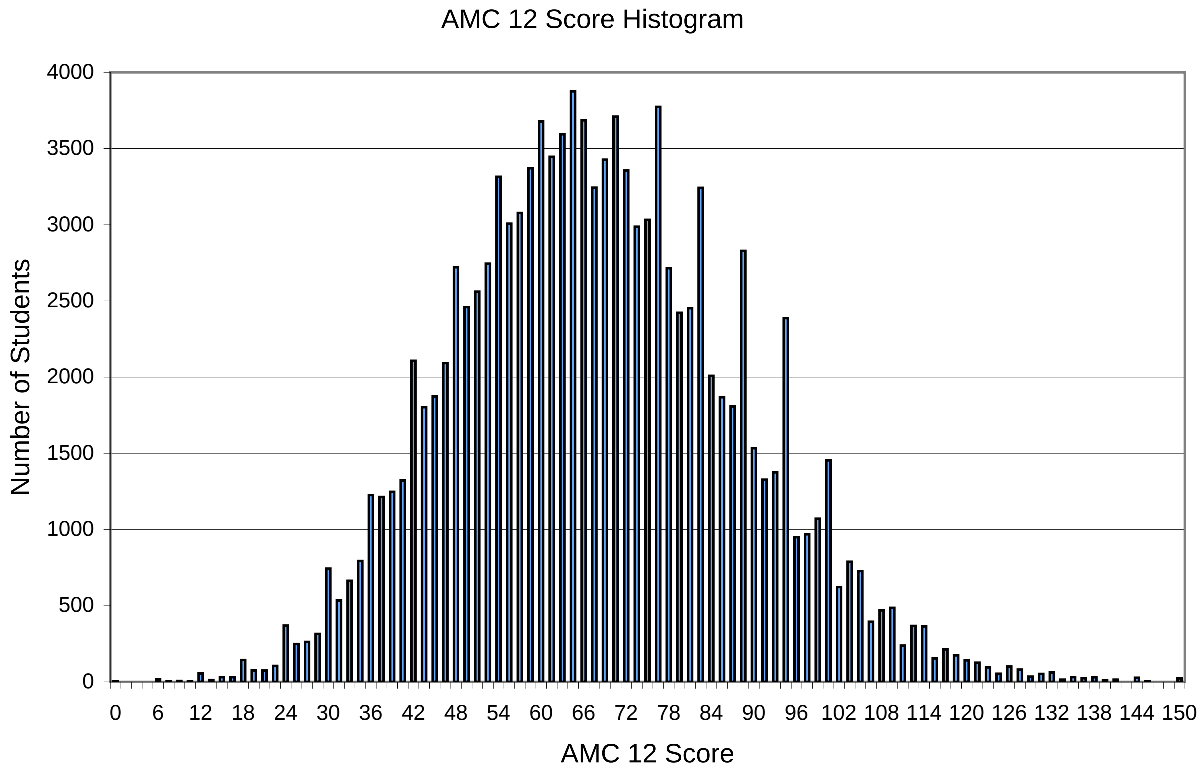


Figure 2: Histogram of 2007 AMC 12 Scores

The AMC is specifically designed to offer challenges even to students at extremely high percentiles. For example, the 2007 AMC 12A included two questions answered correctly by only 20-25% of students who scored at least 100, and three that were answered correctly by fewer than 11% of such students. Data on students who take both the AMC 12A and the AMC 12B provides ample evidence that the test is measuring something that distinguishes students at high percentiles. Students who scored 95 to 105 on the AMC 12A averaged 103 (standard error 11) on the AMC 12B. Students who scored 115 to 125 on the AMC 12A averaged 120 (s.e. 11) on the AMC 12B. And students who scored 138 or higher on the AMC 12A averaged 131 (s.e. 11) on the AMC 12B. Hence, their average scores on the retake are in the 99.6th percentile of the AMC score distribution, which is probably well above the 99.9th percentile in the SAT-taking population.

An obvious limitation of using AMC scores to assess math achievement is that the test is given to a small subset of students. Approximately 4 million U.S. students per year start high school. About 1.5 million of the 1.8 million who are graduating and going on to college take the SAT. Only about 50,000 high school seniors take the AMC 12. But if our assessment of the SAT-AMC correspondence is correct, selection effects are not nearly so severe when looking at high-achieving students: we would guess that one-fourth or one-fifth of the students who score 800 on the SAT take the AMC 12.[22]

There are other AMC contests and we will also present a little data from two of these. The AMC 10 is a contest similar to the AMC 12 open to students in grades 10 and below. It is given in most of the same schools (at the same times) and is also taken by approximately 100,000 students per year. The test is designed so that the mean score is also about 65. Students who took both the AMC 10 and the AMC 12 in 2007 scored about 13 points higher on average on the AMC 10. The gap is somewhat larger at the higher score ranges. For example, students who scored 115 to 125 on the AMC 10A averaged 100 on the AMC 12B.

The American Invitational Math Exam (AIME) is a more demanding contest. Students get 3 hours to work on 15 problems. It is open only to students who have achieved a qualifying score on the AMC 10 or 12.[23] But the problems are sufficiently difficult so that the average score in 2007 was only about 3 out of 15.

2.2 Standard standardized tests

In this section we note that the AMC is substantially different from the standardized tests most commonly used in educational research: the questions on the standard tests tend to be much easier and/or not require much mastery of precalculus mathematics. For this reason, we regard other tests as poorly suited for studying high-achieving students.

The National Assessment of Educational Progress (NAEP) is the primary resource supported by the U.S. Department of Education. It is administered to a representative sample of about 20,000 high school seniors every four years. The NAEP is designed to include "easy", "medium", and "hard" questions. The simplest way to illustrate the limitations

[22]If one does the simplest calculation assuming that all of the juniors who took the AMC 12 in 2007 were as likely to score 800 on the SAT in their final SAT attempt as were the students with the same AMC score in our MIT applicant database, then one would estimate that about 3500 students who took the AMC 12 as juniors in 2007 scored 800 on their final SAT attempt. This would be about 35 percent of the total number of students scoring 800 as reported by the College Board. We would not place much too faith in this estimate, however, because the majority of the students with 800's are estimated to come from the (many) students who had scores in the 80's and 90's on the AMC 12, and our dataset has very few observations (and selection problems) in this range. A similar calculation would say there are another 1500 students who score at least 80 on the AMC 12 and then get 780 or 790 on their final SAT attempt. This alone would be about 20% of the students with 780's or 790's, and again there would be many others from the much larger pool of students scoring in the 60's and 70's on the AMC. Recall also that we estimated that the majority of National Merit Semifinalists attend schools that offer the AMC 12.

[23]Qualifying scores vary slighly from year to year but are roughly 100 (the 95th percentile) for the AMC 12 and 120 (the 99th percentile) for the AMC 10.

of this test is with sample questions. Figure 5 in the Appendix reproduces the five most difficult of the 21 publicly released "hard" questions from 2005. The set includes some questions on middle school topics that require a little bit of reasoning and a couple questions on standard college-preparatory topics that are completely straightforward. It is hard to imagine that such a test could distinguish between students at different high achievement levels.

No Child Left Behind regulations have led states to develop proficiency tests administered to all students. The universal administration is an attractive feature. The lack of difficult questions, however, is again a severe limitation. One nice piece of evidence in this regard is provided by Hyde et al. (2008), who used established criteria to categorize questions on the state proficiency tests they studied into 4 groups on the basis of the level of reasoning required. They noted that most states' tests contained no level 3 or level 4 questions, making the tests less challenging than the NAEP, which they report to have a number of level 3 questions.

The two most common tests for international comparisons are PISA and Trends in International Mathematics and Science Study (TIMSS). PISA is given to 15-year-olds and TIMSS to students in fourth and eighth grades and at the end of high school.[24] Attractive features of these tests include that they are administered in dozens of countries and that samples are representative. PISA is limited in that it again does not appear to be designed to test advanced math skills. Figure 6 in the Appendix contains a sample question from PISA. The TIMMS advanced math test is only administered to students pursuing advanced math courses. The universally administered "mathematics literacy" test is similar to PISA. Figure 7 presents a sample question which appeared on both PISA and TIMSS.

In comparison to the tests discussed above, the SAT reasoning test is more focused on standard college-preparatory mathematics. Although the interpretation of the gender gap in average SAT scores is made difficult by the fact that female participation rates are substantially higher, it is a reasonable source of information on the gender gap at fairly high performance levels because the selection into the SAT is not as big an issue at higher percentiles.[25] We say "fairly high performance levels" rather than "high performance levels" because of a fact noted earlier: students who get a perfect 800 and then retake the SAT only average 752 on the retake. This is in the 97th percentile, so the SAT should be thought

[24]Guiso et al. (2008) is an analysis of the gender gap using PISA data and Mullis et al. (2000) is a report on TIMSS.

[25]There is also an extensive literature questioning whether the SAT is "biased". For example, as discussed in Freedle (2003), several studies have found evidence of persistent biases in the SAT verbal section which adversely affect African Americans.

of as having limited power for distinguishing students in percentiles above the 97th.

3 The Gender Gap on the AMC

In this section we present some data on the gender gap in AMC scores. Our most basic finding is that the gender gap on the AMC is large and widens dramatically as one moves to higher percentiles.

3.1 The gender gap on the SAT and AMC 12

The descriptive statistics in this section focus on the relative number of girls and boys who reach various levels of performance. Specifically, writing $n_f(\tau)$ for the total number of females with scores of at least τ and $n_m(\tau)$ for the number of males with such scores, the graphs in Figure 3 show the fraction female, $n_f(\tau)/(n_f(\tau) + n_m(\tau))$, as a function of τ.

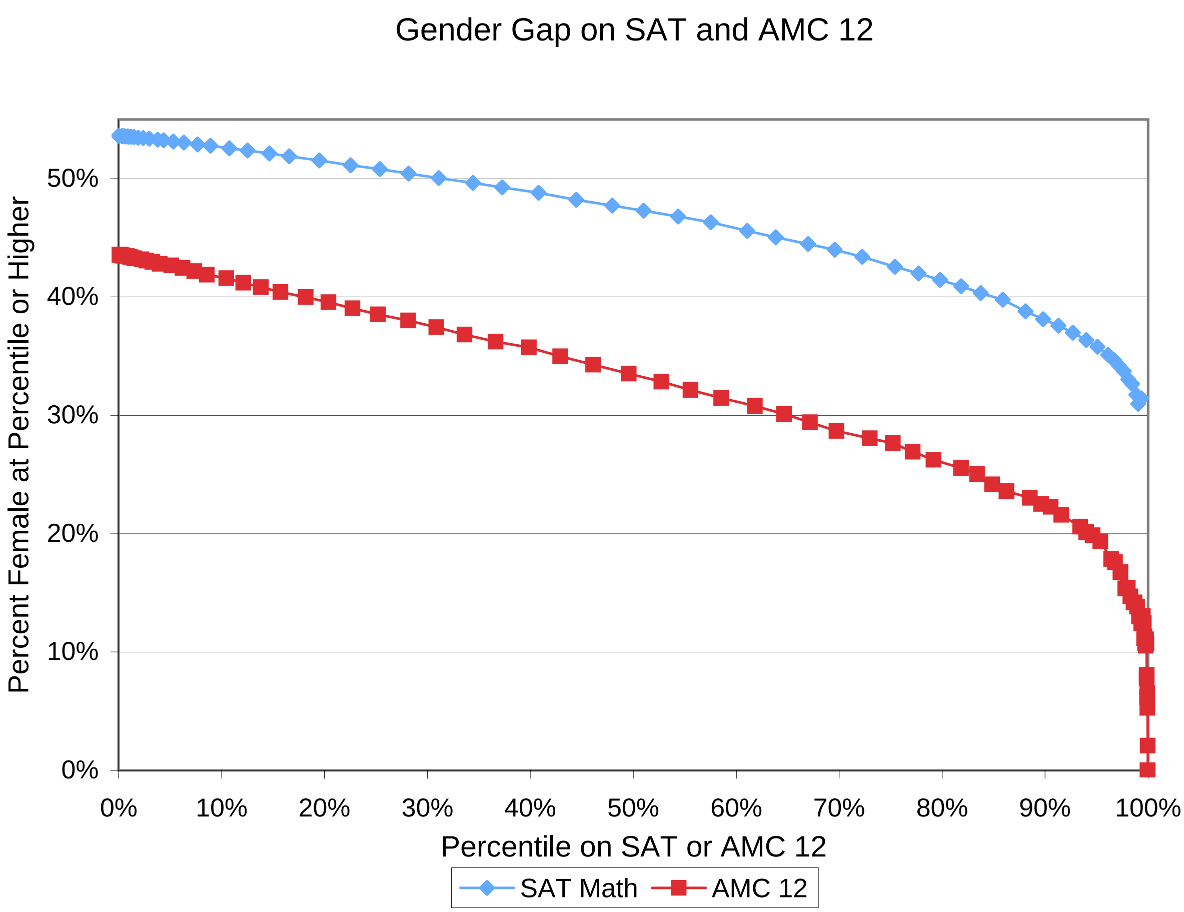


Figure 3: Gender Gaps on the Math SAT and on the American Mathematics Contests

The top curve is a benchmark derived from data on the math SAT scores of 2007 college-bound seniors.[26] A rescaling of math SAT scores is on the x-axis and the fraction female among students scoring at each level or higher is on the y-axis. The x-axis is scaled to be linear in percentile ranks and nonlinear in the actual scores. The fact that the curve starts out above 0.5 on the left side reflects that more girls take the SAT: the raw numbers are about 800,000 vs. 700,000.[27] The ratio dips below unity around the 30th percentile, reflecting that the number of boys and girls achieving scores in excess of 460 are approximately equal. The ratio drops substantially at higher SAT scores. Approximately 200,000 boys and 150,000 girls receive scores of at least 600. The ratio declines most steeply at the highest percentiles and reaches 0.31 at 800; i.e., there are more than 2.1 times as many boys as girls with scores of 800.[28]

The bottom curve of Figure 3 is one of our main results. It is constucted in a similar way using data on American students taking the 2007 AMC 12.[29] The scaling convention is mechanically identical to that of the SAT curve – the x-axis is linear in percentile ranks within the population of AMC takers. The populations taking the two tests are quite different, however, so readers should keep in mind that the percentiles have very different meanings.

Several aspects of the graph are noteworthy:

- The female-to-male ratio is below unity even at the left-most point of the graph: 43% of AMC 12 test-takers are female.

- The AMC curve is at least as steeply downward sloping as the SAT curve. The fraction female drops below 31% at an achievement level that is only moderately high. (The 60th percentile score on the AMC 12 is 70.5, which would roughly correspond to a score in the high 600's on the SAT.) There is a 4.2 to 1 male-female ratio in the pool of students scoring 100 or higher.

- The AMC curve turns sharply downward at the percentiles above those that the SAT can measure. The male-female ratio reaches 6.2 to 1 for students in the 99th percentile

[26] Data is taken directly from the table "SAT Percentile Ranks for Males, Females, and Total Group, 2007 College Bound Seniors–Mathematics," available at http://www.collegeboard.com/prod_downloads/highered/ra/sat/composite_CR_M_W_percentile_ranks.pdf.

[27] Both this difference and the fact that the selection differs by ethnicity (and presumably other factors) makes interpreting differences in mean scores difficult.

[28] The College Board reports an 800 as being in the 99+ percentile because approximately 10,000 of the 1.5 million students who take the math SAT score 800 on their last attempt. The number of students who score 800 at least once, however, is presumably substantially higher.

[29] This figure excludes data from schools outside the U.S.

of the AMC population (1,213 students with scores of 114 or higher) and 12 to 1 in the 99.9th AMC percentile (116 students with scores of 135 or higher). The top 36 scorers were all male.

The first observation on its own is not surprising and does not indicate that there is a strong gender-related selection into taking the AMC 12. Most AMC takers must come from the high end of the SAT population and the population of students with SAT scores of 600 or above is also 43% female.

The second observation is unexpected. If AMC takers were drawn in a gender-independent way from the population of students with SAT scores above 600 and scores on the two tests were similar, then the AMC curve would resemble a stretched out version of the right portion of the SAT graph: it would start lower but decline at one-half or one-third of the observed rate so that it met the SAT curve at the right endpoint. Instead, the AMC curve is steeper. Such steepness could come from two sources. First, there may be differential selection into AMC-taking, with girls of moderate achievement being *more* likely to take the exam than comparably accomplished boys, and girls of high achievement being substantially less likely to take the AMC than comparably accomplished boys. Second, it could be attributed to gender-related differences in knowledge and problem-solving skills: among students who know the SAT material equally well it may be that girls are less likely to know the additional material covered on the AMC 12 and less likely to have developed the problem-solving skills the AMC requires. One or both of these effects must be substantial: the male-female ratio among students scoring 100 or more on the AMC 12 is twice as large as the ratio among students scoring 800 on the SAT.

The fact that the male-female ratio reaches 2.1:1 at an AMC 12 score threshold that is well below the 800 SAT threshold suggests that there is a substantial gender-related selection effect. We regard this as an important finding, and not just a confounding factor that makes estimating population characteristics difficult. Math contests are one of the primary means by which high-ability students are motivated to develop the knowledge and problem-solving skills that contribute to success in many technical fields. If math contests are less appealing to girls than to boys, then this will be a reason why fewer girls are reaching very high achievement levels.

In the case of students scoring 100 or higher it is plausible that the difference between the SAT results (a 2.1 to 1 ratio at the 800 level) and the AMC results (a 4.2 to 1 ratio at the 100+ level) is mostly due to gender-related selection. What would be required is that girls who would score 100 or higher are only half as likely to take the AMC 12 as comparably accomplished boys. A portion of the effect may also be attributable to the AMC test being more accurate at identifying the 99th percentile of the SAT-taking population (as opposed to the SAT's mixing in many 97th percentile students). A portion could also be attributable to differences between the tests.[30]

At higher score ranges, however, it becomes increasingly implausible that gender-related selection into taking the AMC could account for much of the effect: why would girls capable of scoring 130 or higher be only one-quarter as likely as boys who would do this well to take the test? Indeed, the knowledge and problem-solving skills needed to get a 130 are sufficiently high so that we feel that almost all students (male or female) who have acquired such skills are probably taking the AMC 12, making gender-related selection nonexistent (for measurements of what the gender composition of high scorers would be under universal administration). As a result, we interpret the body and right tail of the curve in combination as telling us that many very talented girls are simply not taking the AMC 12, and that there is an additional effect in which a smaller fraction of girls who do become involved in math contests develop the mathematical knowledge and problem-solving skills necessary to achieve extremely high scores.

We would like to emphasize that nothing in the data requires an assumption of different distributions of *ability* in the full male and female populations. What we see the data as suggesting is that there is a gender-related selection into AMC participation which disproportionately reduces the number of high-ability girls, and some additional effect that results in fewer girls developing their knowledge and problem solving skills to the most elite level.

3.2 The gender gap on various AMC contests: AMC 10, AMC 12, and AIME

In this section we compare data from the AMC 12 and data from two other AMC contests: the AMC 10 and AIME. Recall that the AMC 10 is a slightly easier contest taken by students in grades 10 and below, and that the AIME is a more challenging contest open only to students with high scores on the AMC 10 and AMC 12. Figure 4 presents data from the three contests. We have changed a couple of things to facilitate comparisons: the x-variable in these graphs is simply the AMC score rather than the percentile; and the y-variable is the fraction female among students at each score level (rather than at the

[30] Hyde, Fennema, and Lannon (1990), for example, discuss evidence suggesting that the gender gap was larger on tests requiring complex problem solving.

score level or higher as in our previous graph).[31] The AMC 12 scores are the red squares. Two things that are easier to see in this graph are that the population of students receiving just about every score below 58.5 is more than 50% female, and that the fraction female declines fairly smoothly as we move through the range in which most of the data lies.

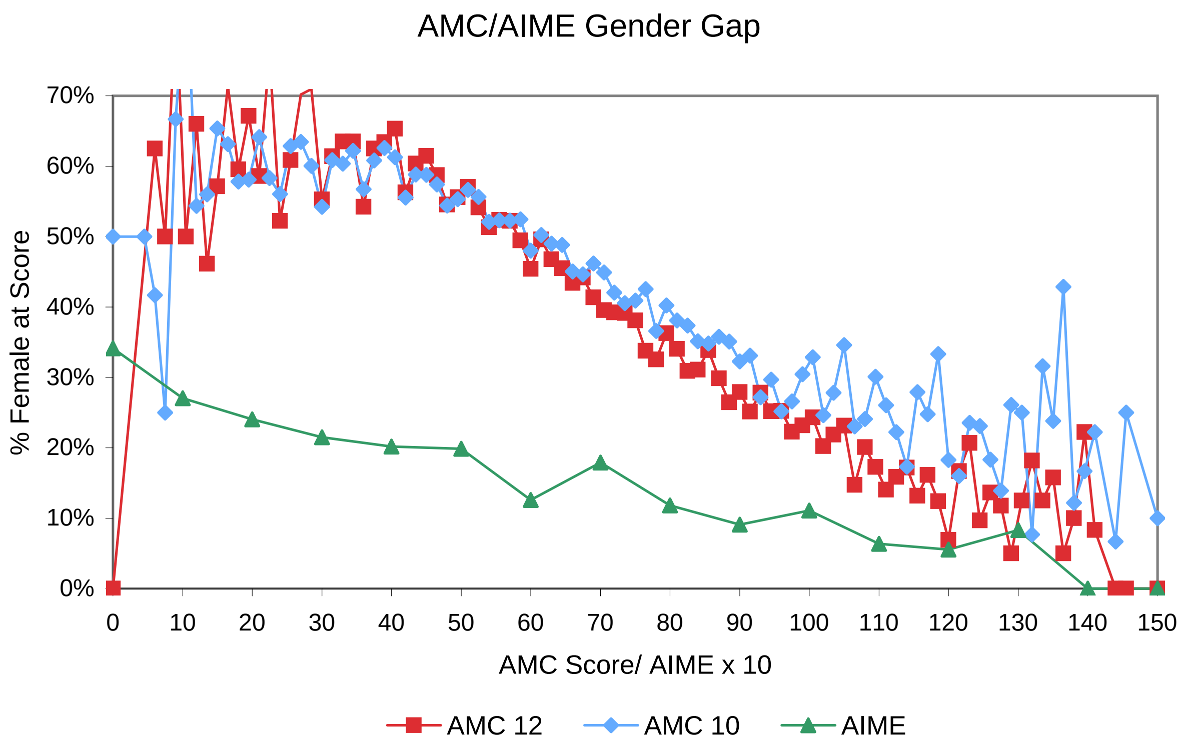


Figure 4: Gender Gaps on the AMC 10, AMC 12, and AIME

The AMC 10 data is fairly similar to the AMC 12 data. One difference is that the fraction female drops off somewhat more slowly at scores above 70 and that there are girls with scores very close to the maximum possible. This could be due to differences in the tests. For example, the last few questions on the AMC 12 are substantially harder than any AMC 10 questions; and the AMC 12 includes trigonometry and other precalculus topics. But, the most obvious difference between the tests is that the AMC 10 is being taken earlier in high school, which suggests that the effects that lead to the gender gap in high school build throughout the high school years.

The AIME is taken only by students who have first achieved a very high score on the

[31] AIME scores have been multiplied by ten so that they also range from 0 to 150. This figure excludes data from schools outside the U.S.

AMC 10 or 12.[32] It emphasizes the ability to solve hard problems over speed: the exam is 3 hours long and the median participant in 2007 only solved 3 of the 15 problems. The pool qualifying to take the AIME is 22% female. The graph illustrates that there is a remarkably smooth decline in the percentage female at higher score levels.

4 Cross-Sectional Patterns in the Gender Gap

In this section we use data on the clustering of high-scoring girls to obtain estimates of how the gender gap varies across schools. Such variation is of interest for at least two reasons: it may provide insight into the sources of the gender gap; and an examination of schools where the gender gap is relatively small and large may suggest policies that might narrow the gap. Our most basic conclusion is that there is variation in the gender gap across schools, but that the magnitude of the variation is not very large.

4.1 Does the gender gap vary across schools?

Our first exercise is to ask how much variation there is in the gender gap across schools. We use no data on school characteristics, and instead infer whether there is unobserved heterogeneity by examining the clustering of high-scoring girls. Intuitively, if one only had data on the single highest score in each school in a single year it would be impossible to tell whether all schools had identical environments in which there was a 17% chance that the high scorer would be a girl, or whether 34% of schools had no gender gap whereas the remaining 66% of schools had zero chance of producing a female high scorer. But with data on multiple high-scoring students from each school the two scenarios would produce very different patterns: in the former case we'd find that most schools had one or two girls among their top ten; whereas in the latter we'd see lots of zeros, fours, fives, and sixes.

More formally, we use data on the number of boys and girls in each school with AMC 12 scores above 100 (and 120). We suppose that the environment of school i is such that the number of high scoring girls will be distributed as $f_i \sim \text{Binomial}(N_i, p_i)$, where N_i is the total number of high scoring students at the school and p_i is a parameter that reflects how the environment affects the gender gap. We are interested in variation in p_i across schools. We estimate this by assuming that the p_i are themselves independent realizations from from a $\text{Beta}(\alpha, \beta)$ distribution and estimate the parameters of this distribution by maximum likelihood.[33]

[32] Qualifying scores are approximately 120 on the AMC 10 and 100 on the AMC 12.

[33] The beta distribution is assumed because it makes the estimation easy. The integrals over the unobserved

The first column of Table 2 presents estimates of the mean and variance of the p_i derived from data on the number of girls and boys scoring above 100 on the AMC 12.[34] The point estimates are that the p_i are drawn from a distribution with a mean of 0.18 and standard deviation of just 0.05. The standard deviation is statistically significant, so that we can conclude that there are some schools where the gender gap is relatively large and others where it is relatively small. But the more important thing to take away is the magnitude of the standard deviation: it indicates the variation in the gender gap across schools is not very large.

The second column presents estimates derived from data on the number of boys and girls scoring above 120 on the AMC 12. Here, the p_i are estimated to be drawn from a distribution with mean 0.11 and standard deviation 0.04. The standard error for the latter estimate is somewhat larger due to the fact that the number of girls scoring above 120 is quite small.

	Sample of Students			
	AMC 12 > 100		AMC 12 > 120	
Parameter	Est.	SE	Est.	SE
E(p_i)	0.18	(0.01)	0.11	(0.01)
St.Dev(p_i)	0.05	(0.01)	0.04	(0.05)
Number of schools	1,306		273	
Number of students	4,583		532	

Table 2: Estimates of the variation in the gender gap across schools

If these estimates seem a bit mysterious, we can point out that the universality of the gender gap is plainly apparent in the raw data on high achieving schools. In the case of students scoring over 100 on the AMC 12 there are a number of schools with many high-scorers. For example, there are 131 schools which had eight or more students scoring above 100. At 127 of these 131 schools, boys outnumbered girls among the high scorers.[35] This

parameter p_i have simple closed form solutions and the likelihood becomes

$$\text{Prob}\{f_i|N_i,\alpha,\beta\} = \binom{N_i}{f_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(f_i+\alpha)\Gamma(N_i-f_i+\beta)}{\Gamma(N_i+\alpha+\beta)}.$$

[34] The data include 1,306 schools. In addition to schools with no students scoring above 100, we drop schools outside the U.S., schools we were not able to identify in the NCES data (see Appendix), and single-sex schools.

[35] The exceptions are one private school and three very strong but otherwise unremarkable public schools: at Holmdel High School (Holmdel, NJ) 8 of the 16 high scorers were girls, at Canton High School (Canton, MA) 5 of the 9 high scorers were girls, and at Lawton Chiles High School (Tallahassee, FL) 5 of the 9 high scorers were girls. At the private Hotchkiss School (Lakeville, CT) 6 of the 11 high scorers were girls.

requires a substantial and nearly universal gender gap – at a school with eight high scorers one would expect to have four or more girls among them more than one quarter of the time even if p_i were just one-third.[36] The estimates for the population of students scoring above 120 on the AMC 12 reflect that there are few clusters of high-scoring girls: only eight schools have more than one girl with a score above 120.

We conclude that the factors that are contributing to the gender gap are felt quite broadly. There are some schools where the gender gap is somewhat bigger or smaller, but the differences are not large.

4.2 Where is the gender gap larger and smaller?

Although the results above indicate that there is not much variation in the gender gap across schools, it still may be interesting to investigate where it is relatively large and small. Patterns could potentially lead to insights on the causes of the gender gap, and might suggest policies that could narrow the gap. In this section we examine the determinants of the gender gap using simple school-level regressions. For each public school which could be matched to NCES data, we computed the fraction female among students in the school who scored at least 100 (and 120) on the AMC 12.[37] Table 3 reports estimates from regressions of these variables on a number of characteristics of the school and of the zip code in which it is located.

As might be expected given that our earlier results suggest that most of the variation in the data is Poisson-style noise rather than true variation in the gender gap, we fail to find many strong patterns. The regression looking at the fraction female among students having AMC 12 scores of 100 or more has 904 observations but only one significant coefficient: the fraction female is estimated to be higher in schools with a higher percentage of students with sufficiently low income to qualify for free lunch. Variables reflecting the race and ethnicity of the school population and the education and income of the zip code have small insignificant effects. The regression examining the fraction female among students with AMC 12 scores of 120 or higher has somewhat greater explanatory power: the R^2 is 0.08 rather than 0.01. But sample size is smaller here – we can only include schools with at least one student scoring 120 or higher – and perhaps as a result we now have no significant estimates at all.

[36] This comment only applies to high-achieving schools. Our estimates have no power to detect variation in the gender gap at schools with zero or one high scorer.

[37] The sample excludes schools we were unable to match to the NCES data (see Appendix), non-coed schools, private schools, charter schools, and magnet schools.

	Dep. var.: Fraction Female			
	AMC12 > 100		AMC12 > 120	
Variable	Coef.	t-stat	Coef.	t-stat
Constant	-0.19	(0.90)	0.14	(0.35)
Number in school > 100/120	0.002	(1.39)	0.01	(1.36)
School frac. Asian	-0.05	(0.39)	0.15	(0.56)
School frac. white	0.08	(0.75)	0.02	(0.09)
School frac. black	-0.15	(1.11)	-0.25	(1.00)
School frac. female	0.50	(1.44)	0.09	(0.12)
Title 1 school	-0.02	(0.64)	-0.07	(1.34)
School free lunch pct	0.30	(2.22)	0.37	(1.32)
Adult frac. BA	0.04	(0.21)	-0.27	(0.79)
Adult frac. Grad	0.00	(0.00)	-0.23	(1.05)
ZIP median income	0.00	(0.89)	0.00	(1.46)
ZIP frac. urban	0.07	(1.10)	-0.18	(1.27)
ZIP frac. white	-0.08	(0.63)	0.03	(0.13)
Number of obs.	904		185	
R^2	0.01		0.08	

Table 3: Patterns in the gender gap among high scorers on the AMC 12 across schools

One estimate that is consistent across the two columns although it falls short of being statistically significant in either is that the gender gap appears to be somewhat narrower in schools that have many high achievers on the AMC 12. Table 4 looks at this relationship more closely by simply tabulating the number of boys and girls in schools that have different numbers of students scoring above 100 on the AMC 12. The left columns show that the percent female among students scoring at least 100 on the AMC 12 rises from about 15% to about 21% as we move from the lowest to the highest bins. The right columns examine students scoring at least 120 on the AMC 12. Here, the relation between school quality on the gender gap is even more pronounced. The lowest three bins contain 237 boys and only 21 girls. The highest bin is 16.3% female. If one tested the hypothesis that the gender gap is smaller in schools with 5 or more students scoring 100 or higher on the AMC 12 via a comparison of means one would find that these differences are approximately significant at the 5% level. But obviously this hypothesis was only generated after looking at the data.

5 Evidence from the extremes

In this section we examine evidence on students at extremely high-achievement levels: we examine students who achieve the very best scores and are chosen to represent their

Number of students with AMC 12 > 100	Gender composition of high scoring students within bin					
	AMC 12 > 100			AMC 12 > 120		
	Female	Male	% Female	Female	Male	% Female
1	65	364	15.2	9	120	7.0
2	70	296	19.1	7	61	10.3
3-4	91	473	16.1	5	56	8.2
5-6	71	265	21.1	7	21	25.0
7+	344	1290	21.1	15	77	16.3

Table 4: Patterns in the gender gap among high scorers on the AMC 12 across schools

countries in international competition. In contrast to our earlier results, we would not argue that the gender gap in these data are inherently important: the population studied here is much, much smaller. Instead, the motivation for this section is that understanding what is going on at the extremes may help us understand the gender gap at less extreme percentiles. Indeed, the U.S. data reveal one striking contrast that may turn out to be an important observation.

5.1 U.S. data: The U.S. IMO and CGMO teams

The AMC series includes one additional invitational test beyond the AMC 12 and AIME, the United States of America Mathematical Olympiad (USAMO). The rules for invitations have varied from year to year, but roughly their effect is that 300 to 500 of the students with the highest AIME scores take the USAMO. The USAMO is very different from the AMC and AIME: whereas the AMC and AIME exams mostly require knowledge of material from the standard college-preparatory curriculum, the USAMO is a proof-oriented contest and therefore relies critically on skills that would not be obtained in coursework in just about any U.S. high school.

Since 1974 the U.S. has sent teams to compete in the International Mathematics Olympiad. In recent years this team has consisted of six students chosen from the USAMO winners. One can combine several years of this data to obtain a noisy estimate of the female-to-male ratio at an extremely high percentile. Over the full 35 year period the U.S. has sent 224 students to the IMO. Five have been female.[38] This gender gap, however, does appear to be narrowing. In the first 24 years there were no female team members. Since then the ratio has been 12 to 1.

In the last two years the U.S. has also sent an eight person team to the China Girls' Math

[38] This consists of three different students, two of whom went twice.

Olympiad. These teams are publicly announced, which provides us with an opportunity to compare the backgrounds of a group of extreme high scoring girls with that of a group of extreme high scoring boys: the CGMO team members are roughly the top scoring girls from the USAMO, whereas the IMO team is roughly the top scoring students regardless of gender.[39] We do not have data on individual USAMO scores, but the highest scoring CGMO team member, Sherry Gong in 2007, was also on the IMO team and hence must have been in the top 12. No other CGMO student, however, was in the top twenty-four and announced cutoffs suggest that the lowest-scoring 2008 CGMO team member was approximately 170th on the USAMO. The CGMO team members are somewhat closer to the IMO team members on the other AMC tests. For example, in 2007 the median CGMO team member had 9 on the AIME and the median IMO team member had 10.

Table 5 presents data on the schools which produced IMO and CGMO members in 2007 and 2008. Specifically, it reports the number of each student's classmates who scored at least 100 on the 2007 AMC 12, the number who scored at least 5 on the 2007 AIME, and the number who qualified to take the 2008 USAMO along with percentile ranks of the schools on these measures.[40]

The bottom half of the table contains statistics on the CGMO team members' schools. The nonrepresentativeness of the schools these girls come from is startling: the median CGMO team member comes from a school at the 99.3rd percentile among AMC participating schools, i.e. from one of the top 20 or so schools in the country. Only three come from schools that are not in the 99th percentile in most measures. And even those three are from schools that had at least one other student qualify for the 2008 USAMO and are at least in the 93rd percentile in terms of the number of high-scorers on the AMC 12.

The male IMO team members, in contrast, come from a much broader set of schools. Some are from super-elite schools and most come from schools that do very well on the AMC 12, but the median student is just from a 93rd percentile school. The majority of the IMO team members had no schoolmates qualify to take the USAMO, whereas all CGMO team members had at least one schoolmate qualify and most had at least four.

The fact that the top boys and girls are coming from such different sets of schools

[39] Neither description is exactly right. The IMO team is chosen from among the 12 high scorers on the USAMO using USAMO scores and another test. The first CGMO teams were based on scores in the previous year, and it is also true that at least one student offered a place on the CGMO team declined.

[40] The percentile rank is always the rank that the school would have without the student in question. We do this because otherwise all schools would have a very high rank on the USAMO qualifier metric. We use 2007 data rather than the most recent data for the AMC and AIME because we need to compute school-level percentiles using our complete dataset and this only runs through 2007.

		School strength: counts and percentiles					
		AMC 12		AIME		USAMO	
Student	High School	> 100		> 5		≥ 0	
U.S. International Math Olympiad Teams: 2007 and 2008							
Sherry Gong	Phillips Exeter Acad.	45	99.9	24	99.9	16	100
Eric Larson	South Eugene HS	8	96.1	1	81-93	0	0-92
Brian Lawrence	Mongomery Blair HS	42	99.8	19	99.8	9	99.9
Tedrick Leung	North Hollywood HS	9	96.9	2	93-97	0	0-92
Arnav Tripathy	East Chapel Hill HS	5	92.0	1	81-93	1	92-98
Alex Zhai	University Laboratory HS	5	92.0	0	0-81	0	0-92
Paul Christiano	The Harker School	21	99.4	13	99.7	4	99.2
Shaunak Kishore	Unionville-Chaddsford HS	0	0.0	0	0-81	0	0-92
Evan O'Dorney	Venture (Indep. Study)	0	0.0	0	0-81	0	0-92
Colin Sandon	Essex HS	3	84.2	2	93-97	0	0-92
Krishanu Sankar	Horace Mann HS	6	93.8	2	93-97	0	0-92
Alex Zhai	University Laboratory HS	5	92.0	0	0-81	0	0-92
Median for Male IMO Team Members		5.5	92.9	1.5	81-93	0	0-92
U.S. China Girls Math Olympiad Teams: 2007 and 2008							
Sway Chen	Lexington HS	16	99.1	8	99.4	4	99.2
Sherry Gong	Phillips Exeter Acad.	45	99.9	24	99.9	16	100
Wendy Hou	Hillsborough HS	7	95.0	0	0-81	1	92-98
Jennifer Iglesias	IL Math & Sci. Acad.	45	99.9	12	99.6	4	99.2
Colleen Lee	Palo Alto HS	26	99.5	12	99.6	6	99.7
Patricia Li	Lynbrook HS	39	99.8	13	99.7	6	99.7
Marianna Mao	Mission San Jose HS	18	99.3	10	99.5	4	99.2
Wendy Mu	Saratoga HS	10	97.5	6	99.1	7	99.8
In Young Cho	Phillips Exeter Acad.	45	99.9	24	99.9	16	100
Jenny Jin	The Taft School	7	95.0	5	98.7	2	98-99
Carolyn Kim	Lawton Chiles HS	6	93.8	2	93-97	1	92-98
Jennifer Iglesias	IL Math & Sci. Acad.	45	99.9	12	99.6	4	99.2
Colleen Lee	Palo Alto HS	26	99.5	12	99.6	6	99.7
Wendy Mu	Saratoga HS	10	97.5	6	99.1	7	99.8
Lynelle Ye	Palo Alto HS	26	99.5	12	99.6	6	99.7
Joy Zheng	Lakeside School	17	99.2	10	99.5	4	99.2
Median for Male IMO Team Members		5.5	92.9	1.5	81-93	0	0-92
Median for CGMO Team Members		18	99.3	10	99.5	4	99.2

Table 5: The 2007 and 2008 U.S. IMO and CGMO teams: statistics on team members' schools

suggests that the gender gap at the very highest levels is in part due to extreme selection effects. It may be that parents of extremely talented girls are much more likely than parents of extremely talented boys to send them to schools with elite math programs. (This would be an independently interesting finding even if it were only strong enough to account for a fraction of the differences between the teams.) It may be that the school-quality measures are endogenous because high-achieving girls dramatically raise the achievement of their classmates. But we feel that the magnitudes are such that it is implausible that there is not a large pool of highly talented girls in the 99% of schools that are not in the top 1% who could also have reached performance levels similar to those of the CGMO team members with the right encouragement and education.[41] It may also be worth noting that almost all of the CGMO team members are Asian-American, which suggests that even within the super-elite schools the U.S. educational system may be missing the opportunity to bring many talented girls up to the highest level.

5.2 International evidence: the International Math Olympiad

The International Math Olympiad was first held in 1959. Over the past 50 years it has grown from a small contest among Soviet-bloc nations to a true worldwide contest among 100 countries. Each country may send up to six high school students. These students are often winners of the country's national olympiad, but the manner in which teams are selected varies.[42] A recent paper by Andreescu et al. (2008) analyzes the gender composition of IMO teams in order to gain insight into the degree to which the gender gap is due to cultural, educational, and other factors that vary across countries. They show that there are statistically significant differences in the gender gap across countries and emphasize the outliers in their discussion

> Girls were found to be 12%-24% of the children identified as having profound mathematical ability when raised under some conditions; under others, they were 30-fold or more underrepresented. Thus, we conclude that girls with exceptional mathematical talent exist; their identification and nurturing should be substantially improved so this pool of exceptional talent is not wasted.

[41] Andreescu et al. (2008) note that most U.S. IMO team members are also selected from a small fraction of the population in that many are Asian, Jewish, children of immigrants, and/or children of parents with advanced mathematical training.

[42] Countries that wish to send a team to the IMO must submit an application which describes the procedure by which the team will be selected and are encouraged to include past copies of the country's national math olympiad.

We see our CGMO results as in complete agreement with their view that there is a substantial pool of exceptionally talented girls that the U.S. is failing to develop. But we see the IMO data very differently. Where they emphasize the statistical significance of differences across countries, we would emphasize that the magnitudes of the differences across countries are quite small. Whatever combination of factors is leading to the gender gap at the extreme appears to be strikingly universal, and it does not seem likely that emulating any other country's educational system would dramatically narrow the U.S. gender gap on such tests.

Our primary IMO data is the same data as in Andreescu et al. (2008): the gender of each student who participated at some point in 1988-2008 as a member of the team from one of 30 high-scoring countries.[43] One basic fact about the IMO is that there is a very large gender gap: only 5.7% of the participants in this sample (185 of 3,246) are female. There has been some narrowing of the gender gap over time: the fraction female increases from 4.3% in 1988-1997 to 6.8% in 1998-2008.[44]

Andreescu et al. (2008) highlight the outliers in the data – the team from Yugoslavia/Serbia and Montenegro is 24% female in the most recent decade whereas Iran, Japan, and Poland sent entirely male teams – and note that a simple Chi-squared test rejects the hypothesis that the variation is entirely random at a very high p-value. Looking at the magnitudes of the differences, however, we would emphasize that the gender gap is if anything strikingly universal. In the 27 countries they consider for the 1998-2008 period, the number of female participants had a mean of 4.6 (out of 66 total participants) and a standard deviation of 3.7.[45] In a model in which each participant was female with independent probability 0.069, the number of female partipants from a country that sent 66 students would have mean 4.6 and standard deviation 2.1. Further, one would expect the actual variance to be greater than that of the independent model for a mechanical reason: some students qualify for the IMO multiple times.[46] Some heterogeneity across countries will be needed to account for the 17 female participants from Yugoslavia/Serbia and Montenegro and the countries sending no young women, but the magnitude cannot be very large.

[43]The data is posted on the IMO website: http://www.imo-official.org, which we presume obtained it from Andreescu et al. Our sample has 26 countries rather than 30 because we combine Germany/East Germany/West Germany and Czechoslovakia/Czech Republic/Slovakia into single countries.

[44]The 6.8% figure is roughly the same as the gender gap at the 139.5 and above level on the 2007 AMC 12 and at the 11 and above level on the 2007 AIME. These scores were achieved by 60 and 115 U.S. students.

[45]Serbia and Montenegro sent 72 participants rather than 66 because Montenegro sent a separate team in 2007 and 2008. Serbia and Montenegro is the only country that sent more than 10 female participants.

[46]For example, while it is true that 8 of Bulgaria's 66 participants were female, the 8 female student-years is attributable to just 3 young women each of whom went multiple times.

To provide a formal estimate we perform the same calculation as in section 4.1. We assume that the probability that each participant from country i is female is p_i and estimate the mean and standard deviation of the p_i under the assumption that these have a Beta distribution across countries. In the the 1998-2008 time period, we estimate the mean of p_i to be 0.065 and the standard deviation of p_i to be 0.031.[47] We would describe this as indicating that there is not a great deal of variation across countries.

Guiso, Monte, Sapienza, and Zingales (2008) examine the relationship between math test scores and measures of cultural, political, and economic gender equity using PISA data. They find that the gender gap in average scores is smaller in countries with greater gender equity. We looked for a similar effect in the IMO data by regressing the number of female IMO competitors in 2006-2008 on the World Economic Forum's Gender Gap Index for each country.[48] The regression has little explanatory power and the positive point estimate on the gender gap index is not statistically significant.[49]

One other item that may be of interest is the performance of female IMO participants. In particular, it may be interesting to know how women fare once they make it to the IMO. To examine this we regressed IMO scores on country-year fixed effects and a female dummy using the full sample of IMO participants from 2006-2008. The results indicate that young women scored about one point lower than their male teammates on average. This is significant at the 5% level, but fairly small in magnitude.[50]

6 Conclusion

In this paper we've examined data from the American Mathematics Competitions. Girls are somewhat underrepresented among the participants – 47% of AMC 10 takers and 43% of AMC 12 takers are female – which is unfortunate given the important role that math contests play in motivating American high school students to reach high levels of math achievement.

Our main focus, however, is not on the contests themselves but on what they may

[47] The standard errors on these estimates are 0.009 and 0.011, respectively. In the 1988-1997 data we estimate a mean of 0.046 and a standard deviation of 0.028.

[48] The restriction to 2006-2008 allows us to use a much larger sample of 91 countries.

[49] Hyde and Mertz (2009) independently conducted a similar test and obtained a somewhat different result. They examine the relationship between the GGI and the number of female competitiors from 1989-2008 in a 30 country sample and find a positive significant correlation. The difference may be related to Fryer and Levitt's (2009) finding that the Guiso et al. result does not carry over to the broader sample of countries in the TIMSS data.

[50] The mean score is about 14.5. The residual standard deviation of a regression with team-year fixed effects and a female dummy is about 5.3.

tell us about the gender gap at high achievement levels. The AMC contests are able to draw consistent distinctions between the problem solving and precalculus math skills of students that remain valid at very high percentiles, and hence provide an opportunity to learn more about what goes on in the upper part of the distribution. We feel this part of the distribution is important because, in contrast to what goes on around the mean where the gender gap is sufficiently small as to be of little practical importance, the gender gap in the upper tail can be quite large. Policy interventions targeted at the mean may be of little or no help in narrowing the gender gap in the upper tail.

Our first main observation is another verification of the common observation that there is a large gender gap among high math achievers: even restricting attention to girls who choose to participate in these contests, many fewer girls are reaching the highest levels. Indeed, the AMC data portray the gender gap as being somewhat larger than have previous 99th percentile studies: we find a 4.2 to 1 male to female ratio among students scoring at least 100 on the AMC 12. Our most visually striking finding, however, concerns what happens at even higher percentiles. Here the AMC data reveal that the male-to-female ratio rises dramatically as we move beyond the 99th percentile and reaches more than 10 to 1 at the top end.

Another striking observation that comes up in a couple places is the universality of the gender gap. Both in our comparison across schools and in the international comparison of IMO competitors we've seen that there is statistically significant evidence that the gender gap varies from environment to environment, but that the magnitude of this variation is modest. One implication is that that factors contributing to the gender gap are felt quite broadly. It also suggests that there are no simple solutions to the problem of eliminating the gender gap. One should keep in mind, however, that the "modest" variation we've found is suffiently large relative to the fraction top students who are female so as to make it plausible that it may not be too difficult to make large proportionate increases in the number of girls who are doing well. For example, our $18\% \pm 5\%$ estimate suggests that there are many schools with values below 13% and above 23%. Shifting a school from one end to the other would entail a large proportionate increase. For this reason, further studies of the environments where girls are doing relatively poorly and relatively well would seem to be potentially very valuable.

One hopeful observation that comes up in a couple places is that girls appear to do well in top schools. We saw marginally significant evidence in the AMC data of a narrower gender gap at better schools. And the data on CGMO participants showed that the top girls were remarkably concentrated in a few superelite schools. This could in part reflect that parents of high-achieving girls feel a need to move them to such schools. But the results suggest that the number of girls (and boys) at the highest achievement levels could perhaps be increased quite dramatically by increasing the number of schools that provide opportunities for elite achievement.

We have consciously focused on reporting the facts in our data, rather than on attempting to draw out what the data might say about the relative importance of the many different factors that may contribute to the gender gap. Mostly, we do this because our data contain many new facts and do not seem particularly well suited for distinguishing theories. To some extent, however, it also reflects that the issues are sensitive and prior debates on competing theories seem to have not been highly productive.

If asked to speculate, our first remark would be that several elements in our results seem consistent with the view that girls suffer because they are more compliant with authority figures and/or are more sensitive to peer pressure. Most high schools offer math courses to suit students at several different levels. But, even in the highest-level "honors" courses, it is probably unusual to teach material at the level needed to bring students to the 99th percentile. If girls are less likely to complain and get schools to make special accomodations, then we would expect them to be more underrepresented among students with skill levels that are farther beyond those developed in the classroom. Peer pressure would also presumably be more limiting when we look at achievement levels that are only likely to be reached if students join a math team or take online courses. Such explanations could also fit well with the CGMO data: the superelite schools could be places where students can join a community learning advanced material.

Potential explanations for the gender gap in high percentile math scores should also be consistent with the fact that such gaps do not exist in some other areas. For example, the male-female ratio among students scoring 800 on the SAT Critical Reading test is about 1 to 1 and the the ratio on the SAT Writing test is about 0.7 to 1. A compliance/peer-pressure story could be elaborated in any of several ways to do this. One factor could be that the verbal SAT tests have a different relationship with the standard high school curriculum: it may be that they do not test much beyond what is gained from a standard high school English class plus a lot of reading and hence compliance with standard school path is not costly. Peer pressure could also differ: reading could be a much more accepted hobby than joining the math team. Or there could be institutional factors like schools treating boys and girls differently in different fields.

A number of alternative explanations for the gender gap are also possible. One would be a Summers-esqe model in which there is less variance of ability in the female population. This could be consistent both with the increasing gender gap at the highest achievement levels and with the extreme high-achieving girls coming from the extreme high-achieving schools: in a model where achievement is the sum of ability and education, one could posit parameters are such that the variance of education falls between the variance of ability in the male population and the variance of ability in the female population.[51] Our results on the concentration of high-scoring girls, however, suggests to us that there is limited value to trying to putting a lot of effort into estimating "ability" distributions when almost all girls who would be capable of achieving extremely high scores do not do so. Another alternate model would be a model in which a lack of girls in the population of extreme high math achievers is not a bad thing: it might be that the girls who could reach the highest achievement levels tend not to do so because they are more likely to have other skills and interests as well and that they tend to pursue less math-focused paths that lead them to develop portfolios of skills that will be more valuable in the long run. This could be part of what is going on, but we would note that our impression is that the achievement levels needed to score 100 on the AMC do not look very high in comparison to what would be needed to succeed in the economics profession.

To conclude we would like to go back to a more factual posture. The AMC data reveal a very large and widespread gender gap at the high achievement levels. The gender gap at these levels is more striking and significant than gaps in average scores and calls out for further study.

[51] There is some evidence in the psychology literature that gender may serve as a "proxy" variable for underlying cognitive processes related to both ability and education. For example, Byrnes and Takahira (1993) discuss how solving SAT problems can be seen as requiring knowledge and multiple cognitive processes and discuss where in this chain the gender gap appears.

References

Andreescu, Titu, Joseph A. Gallian, Jonathan M. Kane, and Janet E. Mertz (2008): “Cross-Cultural Analysis of Students with Exceptional Talent in Mathematical Problem Solving,” *Notices of the American Mathematical Society*, 55 (10), 1248–1260.

Benbow, Camilla Persson, and Julian C. Stanley (1980): “Sex Differences in Mathematical Ability: Fact or Artifact?,” *Science*, 210, 1262–1264.

Bettinger, Eric P. and Bridget Terry Long (2005): “Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students,” *American Economic Review* 95(2), 152-157.

Blau, Francine D., and Lawrence M. Kahn (2000): “Gender Differences in Pay,” *Journal of Economic Perspectives* 14(4), 75-99.

Brody, Linda, and Carol Mills (2005):, “Talent Search Research: What Have We Learned?,” *High Ability Studies* 16(1), 97-111.

Brown, Charles, and Mary Corcoran (1997): “Sex-Based Differences in School Content and the Male-Female Wage Gap,” *Journal of Labor Economics* 15(3), 431-465.

Byrnes, James P., and Sayuri Takahira, “Explaining Gender Differences on SAT-Math Items,” *Developmental Psychology* 29 (5), 805-810.

Carrell, Scott E., Marianne E. Page, and James E. West, “Sex and Science: How Professor Gender Perpetuates the Gender Gap,” mimeo, University of California, Davis and United States Air Force Academy.

Feingold, Alan (1992): “Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy,” *Review of Educational Research* 62(1), 61-84.

Feingold, Alan (1994): “Gender Differences in Variability in Intellectual Abilities: A Cross-Cultural Perspective,” *Sex Roles* 30(1/2), 81-92.

Frank, Kenneth A., Chandra Miller, Kathryn S. Schiller, Catherine Riegle-Crumb, Anna Strassman Mueller, Robert Crosnoe, and Jenniver Pearson (2008): “The Social Dynamics of Mathematics Coursetaking in High School,” *American Journal of Sociology*, 113 (6), 1645–1696.

Frank, Kenneth A., Chandra Muller, Kathryn S. Schiller, and Catherin Riegle-Crumb, Anna Strassman Mueller, Robern Crosnoe, and Jennifer Pearson (2008): “The Social Dynamics of Mathematics Coursetaking in High School,” *American Journal of Sociology* 113 (6), 1645-1696.

Freedle, Roy O. (2003): “Correcting the SAT’s Ethnic and Social-Class Bias: A Method for Reestimating SAT Scores,” *Harvard Educational Review* 73(1), 1-43.

Freeman, Catherine E. (2005): *Trends in Educational Equity of Girls & Momen: 2004* (NCES 2005-016) U.S. Department of Education, National Center for Education Statistics.

Washington D.C.: U. S. Government Printing Office.

Fryer, Roland G., Jr., and Steven D. Levitt (2009): "An Empirical Analysis of the Gender Gap in Mathematics," mimeo, Harvard University and University of Chicago.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales (2008): "Culture, Gender, and Math" *Science* 320, 1164–1165.

Hedges, Larry V., and Amy Norwell (1995): "Sex Differences in Mental Test Scores, Variability, and Numbers of High Scoring Individuals," *Science* 269 (5220), 41-45.

Hyde, Janet S., Elizabeth Fennema, and Susan J. Lamon (1990): "Gender Difference in Mathematical Performance: A Meta Analysis," *Psychological Bulletin* 107, 139-155.

Hyde, Janet S., Sara M. Lindberg, Marcia C. Linn, Amy B. Ellis, and Caroline C. Williams (2008): "Gender Similarities Characterize Math Performance," *Science*, 321, 494–494.

Hyde, Janet Shibley, and Marcia C. Linn (2006): "Gender Similarities in Mathematics and Science," *Science* 314, 599-600.

Hyde, Janet S., and Janet E. Mertz (2009): "Gender, Culture, and Mathematics Performance," *Proceedings of the National Academy of Science* 106(22), 8801-8807.

Leonard, David K., and Jiming Jiang (1999): "Gender Bias and the College Prediction of the SATs: A Cry of Despair," *Research in Higher Education* 40(4), 375-407.

Machin, Stephen, and Tuomas Pekkarinen (2008): "Global Sex Differences in Test Score Variability," *Science* 322, 1331-1332.

Maccoby, Eleanor E., and Carol Nagy Jacklin (1974): *The Psychology of Sex Differences.* Stanford University Press: Stanford, California.

The Mathematical Association of America, American Mathematics Competitions (2007): *58th Annual Summary of High School Results and Awards.*

Mullis, Ina V. S., Michael O. Martin, Edward G. Fierros, Amie L. Goldberg, and Steven E. Stemler (2000): *Gender Differences in Achievement, IEA's Third International Mathematics and Science Study.* International Study Center, Lynch School of Education, Boston College.

Organization for Economic Co-operation and Development (2006): *PISA 2006: Science Competencies for Tomorrow's World.* Paris: OECD.

Paglin, Morton and Anthony Rufolo (1990): "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences," *Journal of Labor Economics* 8(1), 123-144.

Perie, Marianne, Rebecca Moran, and Anthony D. Lutkus (2005): *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics* (NCES 2005-464). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington D.C.: Government Printing Office.

U.S. Dept. of Commerce, Bureau of the Census (2000): *Census of Population and Housing, 2000: Summary File 1*. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census.

U.S. Dept. of Commerce, Bureau of the Census (2000): *Census of Population and Housing, 2000: Summary File 3*. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census.

U.S. Dept. of Education, National Center for Education Statistics: *Common Core of Data: Public School Data, 2006-2007*, available at http://nces.ed.gov/ccd/bat/.

U.S. Dept. of Education, National Center for Education Statistics: *PSS Private School Universe Survey Data, 2006-2007*, available at http://nces.ed.gov/pubsearch/getpubcats.asp?sid=002.

Weinberger, Catherine, "Is The Science and Engineering Workforce Drawn from the Far Upper Tail of the Math Ability Distribution?", mimeo, University of California, Santa Barbara.

Xie, Yu, and Kimberlee A. Shauman (2003): "Women in Science: Career Processes and Outcomes." Harvard University Press: Cambridge, Massachusetts.

Appendix

Our data on AMC 10/12 and AIME participants were provided by Professor Steve Dunbar and Marsha Conley at the American Mathematics Competitions. The AMC data include school state, College Entrance Examination Board (CEEB) code, encrypted name, exam (AMC 10 or 12), contest (A or B), age, gender, grade, score, student state, student zip, and student city for each participant. Occasionally the score for a participant is missing, so that participant was eliminated from the data. Other variables, such as age, gender, and grade, were also occasionally missing from the data, but those participants with missing data were not excluded from the analysis unless the missing variable was relevant to that analysis.

The AIME data are similar to the AMC data. They include school state, CEEB code, encrypted name, age, gender, grade, AMC 10/12 scores, AIME score, student state, student zip, and student city, as well as variables linking entries for students who took the AMC multiple times. According to the Mathematical Association of America (MAA) (2007), there were 8,472 AIME participants total in 2007. After eliminating duplicate participants and observations for which score data were missing, we have data on 8,349 AIME participants.

Unfortunately, students who participated in the AMC 10/12 multiple times in one year were not identified separately in the AMC data. Students were matched to their AIME data by CEEB, name, and AMC 10/12 scores where possible. When determination of duplicate participant entries was not possible using the AIME data, duplicates were eliminated using matches on school state, CEEB, age, gender, grade, and name. Some students' scores in different exams could not be linked by this method due to inconsistency in CEEB, name, age, and grade reporting. However, this fairly conservative method yields an estimate of the total number of AMC 10/12 participants (225,044) which is consistent with the estimate of "over 225,000" reported by the MAA (2007).

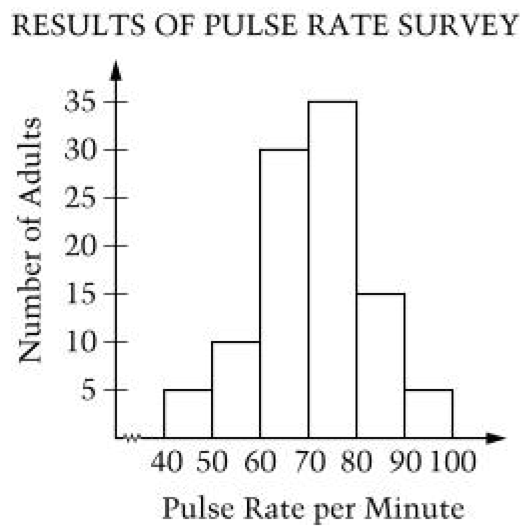
The analyses performed in section 4 employ school- and ZIP code-specific demographic data. Data on public and private U.S. schools for the 2005-2006 school year were downloaded from the National Center for Education Statistics (NCES). The NCES collected private school data from schools that responded to the Private School Universe Survey (PSS). Data on public schools are from the NCES Common Core of Data, which is collected annually from state education agencies. School name, city and state data were linked to the AMC/AIME data using the CEEB code search program provided on the College Board's website. CEEB codes for schools participating in the AMC 10/12 and AIME were matched to NCES data by school name, city, and state. Special thanks are due to David Card and Jesse Rothstein for help in matching the CEEB and NCES data. Of the 3,750 schools with numerical CEEB codes in the AMC data, 3,105 were matched to schools in the NCES data. 255 of the remaining 625 schools do not appear in the NCES data because they did not have official CEEB identifiers and were thus not linked to school data of any kind. It was not possible to match the remaining 370 schools by the identifiers reported by the College Board or NCES.

Data on United States ZIP code demographics are from the U.S. Census.

3. Roxanne plans to enlarge her photograph, which is 4 inches by 6 inches. Which of the following enlargements maintains the same proportions as the original photograph? Justify your answer.

5 inches by 7 inches   5 inches by $7\frac{1}{2}$ inches

14. In a certain restaurant a whole pie has been sliced into 8 equal wedges. Only 2 slices of the pie remain. Three people would each like an equal portion from the remaining slices of pie. What fraction of the original pie should each person receive?

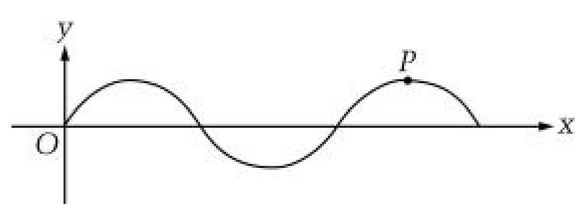


15. The pulse rate per minute of a group of 100 adults is displayed in the histogram above. For example, 5 adults have a pulse rate from 40-49 inclusive. Based on these data, how many individuals from a comparable group of 40 adults would be expected to have a pulse rate of 80 or above?

16. A clock manufacturer has found that the amount of time their clocks gain or lose per week is normally distributed with a mean of 0 minutes and a standard deviation of 0.5 minute, as shown below.

In a random sample of 1,500 of their clocks, which of the following is closest to the expected number of clocks that would gain or lose more than 1 minute per week?

A) 15   B) 30   C) 50   D) 70   E) 90



17. The graph of $f(x) = \sin x$ is shown above. Which of the following is the x coordinate of point P ?

A) $\frac{\pi}{2}$ B) π C) $\frac{3\pi}{2}$ D) 2π E) $\frac{5\pi}{2}$

Figure 5: The most difficult "hard" problems from the 2005 NAEP

THE BEST CAR

A car magazine uses a rating system to evaluate new cars, and gives the award of "The Car of the Year" to the car with the highest total score. Five new cars are being evaluated, and their ratings are shown in the table.

Car	Safety Features (S)	Fuel Efficiency (F)	External Appearance (E)	Internal Fittings (T)
Ca	3	1	2	3
M2	2	2	2	2
Sp	3	1	3	2
N1	1	3	3	3
KK	3	2	3	2

The ratings are interpreted as follows:

3 points = Excellent
2 points = Good
1 point = Fair

Question 1: THE BEST CAR *M704Q01*

To calculate the total score for a car, the car magazine uses the following rule, which is a weighted sum of the individual score points:

Total Score = (3 x S) + F + E + T

Calculate the total score for Car "Ca". Write your answer in the space below.

Total score for "Ca": ....

Question 2: THE BEST CAR *M704Q02*

The manufacturer of car "Ca" thought the rule for the total score was unfair.

Write down a rule for calculating the total score so that Car "Ca" will be the winner.

Your rule should include all four of the variables, and you should write down your rule by filling in positive numbers in the four spaces in the equation below.

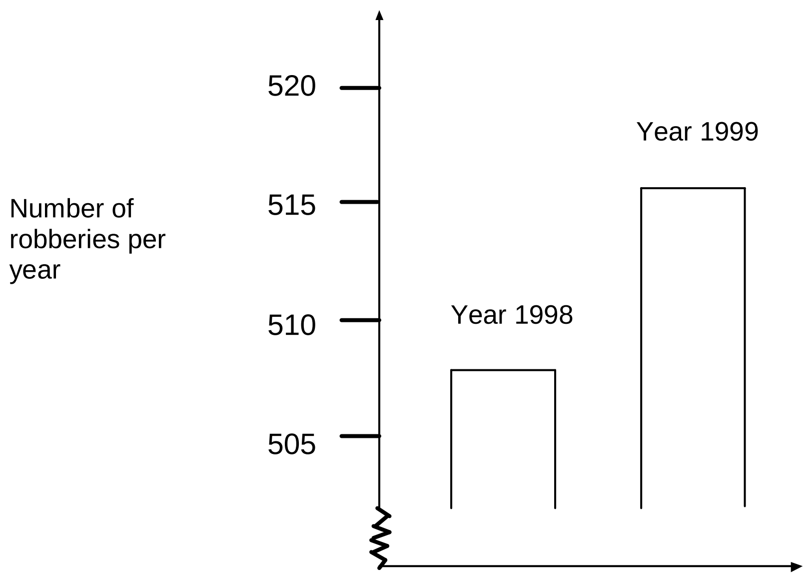
Total score = ………× S + ………× F + ………× E + ………× T.

Figure 6: A sample question from the PISA test

Question 1: ROBBERIES *M179Q01- 01 02 03 04 11 12*

A TV reporter showed this graph and said:

“The graph shows that there is a huge increase in the number of robberies from 1998 to 1999.”



Do you consider the reporter’s statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.

Figure 7: A sample question common to PISA and TIMSS