Dynamics of the Gender Gap in High Math Achievement
Glenn Ellison and Ashley Swanson
Online Appendix

A Score Adjustments

The adjusted scores for students who took a test other than the AMC 12A in year t are given by formulas of the form

$$AdjustedScore_{ijt} = b_{0jt} + b_{1jt}Score_{ijt},$$

where i indexes the student and $j \in \{10A, 10B, 12A\}$ indexes the test that the student took in year t. For each t from 2000 to 2006, the construction of year t adjusted scores is based on a regression of year $t+1$ AMC 12 scores on dummies for the test taken in year t, interactions between these dummies and the score on the year t test, and a dummy for whether the year $t+1$ score was on the 12B.[1] For each year t test, this regression gives the predicted year $t+1$ AMC 12A score as an affine function of the year t score. We define the adjusted score for each year t test as the score on the year t AMC 12A that has the same predicted year $t+1$ score given the regression estimates.

We cannot adjust 2007 scores in the same way because our dataset does not contain 2008 scores. For the AMC 10 tests, we instead set the slope coefficient $b_{1j2007} \equiv \frac{1}{5}\sum_{t=2002}^{2006} b_{1jt}$ equal to the average of the b_{1jt} for the previous five years, and set $b_{0j2007} = \frac{1}{5}\sum_{j=2002}^{2006} b_{0jt} + \Delta_j$, which is an average of the constants from the previous five years plus an adjustment factor that reflects whether each 2007 test appears to be easier or harder relative to the 2007 AMC 12A than were previous-year AMC 10's relative to their contemporaneous AMC 12A's, in light of data on students who took both tests in each year.[2] To determine the adjustments, we run regressions examining the difference between B-date scores and A-date scores for students who took both tests in each year on dummies for which tests they took,

$$ScoreB_{it} - ScoreA_{it} = c_{10A,t}Dummy10A_{it} - c_{10B,t}Dummy10B_{it} - c_{12B,t}Dummy12B_{it} + \epsilon_{it},$$

and set $\Delta_j = c_{j,2007} - \frac{1}{5}\sum_{t=2002}^{2006} c_{j,t}$. We adjust AMC 12B scores in a somewhat similar manner, but imposing that $b_{1,12B,2007} = 1$ rather than estimating the coefficient.[3] We then set $b_{0,12B,2007} = \frac{1}{5}\sum_{j=2002}^{2006} b_{0,12B,t} + c_{12B,2007} - \frac{1}{5}\sum_{t=2002}^{2006} c_{12B,t} + \left(\frac{1}{5}\sum_{j=2002}^{2006} b_{1,12B,t} - 1\right)\overline{X}$ where $\overline{X} \approx 99.4$ is the mean AMC 12B score among students who scored at least 90 on the 2007 AMC 12 and attend a school that did not offer the 2007 AMC 10A.[4]

To give a feel for the linear adjustments, Table A1 reports the contemporaneous AMC 12A scores corresponding to scores of 100 and 150 on each of the other tests. Recall that 100 is roughly the 95th percentile score on the AMC 12 and 150 is a perfect score. The left and center parts of the Table give the AMC 10A-to-AMC 12 and AMC10B-to-AMC 12 conversions. The median AMC 10 test will have its scores adjusted downward by 13 points at the 100 level and by 18 points at the 150 level. There is some variation around this – the

Table A1: Adjusted Scores for AMC 10A, 10B, and 12B Scores of 100 and 150

	AMC 12A equivalents					
	AMC 10A		AMC 10B		AMC 12B	
Year	100	150	100	150	100	150
2000	83.9	116.7				
2001	64.7	116.1				
2002	90.4	135.2	83.4	131.0	89.0	143.8
2003	92.1	133.2	86.6	126.1	101.9	139.3
2004	94.1	141.0	82.9	131.6	94.4	149.5
2005	91.9	134.7	87.2	133.6	97.6	158.3
2006	94.4	135.5	96.7	131.6	99.5	146.2
2007	79.2	122.6	86.0	129.4	95.9	146.6

AMC 10 seems to have been much easier in its first two years and the 2006 AMC 10 tests appear to have been nearly as hard as the 2006 AMC 12 for students scoring around 100 – but most tests are within one question (6 points) of the average relative difficulty level. Most of the AMC 12B adjustments are also less than the point value of one question.

B Variance Calculation

To provide some estimates of the relative importance of measurement error and true performance increases we compare performance changes over multiple years. Suppose that year-t performance $y_{it} = a_{it} + \epsilon_{it}$ reflects both student i's true ability a_{it} and an additive mean-zero measurement error ϵ_{it}. Suppose that ability evolves according to $a_{it+1} = \alpha_0 + \alpha_1 a_{it} + u_{it+1}$. And suppose that the measurement errors ϵ_{it} are independent of all other terms. We then have

$$\begin{aligned} \text{Var}(y_{it+1} - \alpha_1 y_{it}) &= \text{Var}(\epsilon_{it+1}) + \alpha_1^2 \text{Var}(\epsilon_{it}) + \text{Var}(u_{it}) \\ \text{Cov}(y_{it+1} - \alpha_1 y_{it}, y_{it} - \alpha_1 y_{it-1})) &= -\alpha_1 \text{Var}(\epsilon_{it}) + \text{Cov}(u_{it+1}, u_{it}) \end{aligned}$$

It is natural to assume that the true improvements u_{it} are positively correlated, as some students are presumably working harder on improving than others. In this case, the covariance term as provides a lower bound on the measurement error variance:

$$\text{Var}(\epsilon_{it}) \geq -\frac{1}{\alpha_1} \text{Cov}(y_{it+1} - \alpha_1 y_{it}, y_{it} - \alpha_1 y_{it-1}),$$

The lower bound will be close to the true value if $Cov(u_{it+1}, u_{it})$ is small. If we use $\log(Rank_{it})$ as the performance measure y_{it}, estimate α_1 via an IV regression of y_{it+1} on y_{it} using y_{it-1} as an instrument run on the sample of students who took the AMC for three consecutive years, and compute the above variances and covariances on the same sample, we get a lower bound estimate of $\text{Var}(\epsilon_{it})$ that corresponds to ϵ_{it} having a standard deviation of 0.62. This indicates that a substantial portion of the apparent year-to-year variation in scores is due to the measurement error of the test as a measure of underlying achievement.

If we assume that this lower bound also applies to $\text{Var}(\epsilon_{it+1})$, we can also plug into the formulas above to get an upper bound on the standard deviation of u_{it}, which describes the heterogeneity in students' true improvement from year to year. Again, this should be close to the true value if the covariance of year-to-year improvement is low. This estimate corresponds to u_{it} having a standard deviation of 0.37, which suggests that there is substantial heterogeneity in students' true improvement from year to year, albeit not nearly as much as naively looking at year-to-year changes in scores might suggest.

C Decomposition Appendix

As discussed in the main text, our analysis focuses on changes in the fraction μ^f_{Xt} of students in achievement group X at time t who are female. For example, the group X might be the top 5000 scorers. We relate this to various aspects of differences in the boys' and girls' transition matrices. To define these, we write μ^f_{rt} for the fraction female at rank r at time t. We define a_{rX} as the fraction of students at rank r at time t who achieve a score in X at time $t+1$. We will define this both for numerical rank groups $r \in X$ and for the set of students who do not participate at time $t+1$, which we denote by $r = NP$. Write a^f_{rX} and a^f_{rNP} for the analogous objects for female students. Write N_{rt} and N_X for the number of students at rank r at time t and the number with ranks in the set X. Write μ^{new}_{Xt+1} for the fraction of students in group X at time $t+1$ who had not participated at time t and $\mu^{f,\text{new}}_{Xt+1}$ for the fraction of students in group X at time $t+1$ who are female and who had not participated at t. Finally, define each of the component Δ factors in Proposition 1 as follows:

$$
\begin{aligned}
\Delta^{\text{drop}}_X &= \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(a_{rNP} - a^f_{rNP})\frac{a^f_{rX}}{1-a^f_{rNP}} N_{rt} \\
\Delta^{\text{cont}}_X &= \frac{1}{N_X}\sum_{r \in X} \mu^f_{rt}(1-a_{rNP})\left(\frac{a^f_{rX}}{1-a^f_{rNP}} - \frac{a_{rX}}{1-a_{rNP}}\right) N_{rt} \\
\Delta^{\text{grow}}_X &= \frac{1}{N_X}\sum_{r \notin X, r \neq NP} \mu^f_{rt}(1-a_{rNP})\left(\frac{a^f_{rX}}{1-a^f_{rNP}} - \frac{a_{rX}}{1-a_{rNP}}\right) N_{rt} \\
\Delta^{\text{entry}}_X &= \mu^{f,\text{new}}_{Xt+1} - \mu^f_{Xt}\mu^{\text{new}}_{Xt+1} \\
\Delta^{\text{mech}}_X &= \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt} a_{rX} N_{rt} - \mu^f_{Xt}(1-\mu^{\text{new}}_{Xt+1})
\end{aligned}
$$

Proof of Proposition 1 Proposition 1 states that the change in the fraction female in group X can be written as

$$\mu^f_{Xt+1} - \mu^f_{Xt} = \Delta^{\text{drop}}_X + \Delta^{\text{cont}}_X + \Delta^{\text{grow}}_X + \Delta^{\text{entry}}_X + \Delta^{\text{mech}}_X.$$

With all transition probabilities like a_{rX} representing the realized fraction of students at rank r at time t who score in the top X at $t+1$, and taking the sum over all ranks that have at least one female student at time t, we have:

$$\begin{aligned}
\mu^f_{Xt+1} - \mu^f_{Xt} &= \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt} a^f_{rX} N_{rt} + \mu^{f,\text{new}}_{Xt+1} - \mu^f_{Xt} \\
&= \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt} a_{rX} N_{rt} + \mu^{f,\text{new}}_{Xt+1} - \mu^f_{Xt} + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(a^f_{rX} - a_{rX}) N_{rt} \\
&= \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt} a_{rX} N_{rt} - \mu^f_{Xt}(1 - \mu^{\text{new}}_{Xt+1}) + \mu^{f,\text{new}}_{Xt+1} - \mu^f_{Xt}\mu^{\text{new}}_{Xt+1} \\
&\quad + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(a^f_{rX} - a_{rX}) N_{rt} \\
&= \Delta^{\text{mech}}_X + \Delta^{\text{entry}}_X + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(1 - a_{rNP}) \left(\frac{a^f_{rX}}{1 - a_{rNP}} - \frac{a_{rX}}{1 - a_{rNP}} \right) N_{rt} \\
&= \Delta^{\text{mech}}_X + \Delta^{\text{entry}}_X + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(1 - a_{rNP}) \left(\frac{a^f_{rX}}{1 - a^f_{rNP}} - \frac{a_{rX}}{1 - a_{rNP}} \right) N_{rt} \\
&\quad + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(1 - a_{rNP}) \left(\frac{a^f_{rX}}{1 - a_{rNP}} - \frac{a^f_{rX}}{1 - a^f_{rNP}} \right) N_{rt} \\
&= \Delta^{\text{mech}}_X + \Delta^{\text{entry}}_X + \Delta^{\text{cont}}_X + \Delta^{\text{grow}}_X + \\
&\quad + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt} \left((1 - a^f_{rNP})\frac{a^f_{rX}}{1 - a^f_{rNP}} - (1 - a_{rNP})\frac{a^f_{rX}}{1 - a^f_{rNP}} \right) N_{rt} \\
&= \Delta^{\text{mech}}_X + \Delta^{\text{entry}}_X + \Delta^{\text{cont}}_X + \Delta^{\text{grow}}_X + \frac{1}{N_X}\sum_{r \neq NP} \mu^f_{rt}(a_{rNP} - a^f_{rNP}) \left(\frac{a^f_{rX}}{1 - a^f_{rNP}} \right) N_{rt} \\
&= \Delta^{\text{mech}}_X + \Delta^{\text{entry}}_X + \Delta^{\text{cont}}_X + \Delta^{\text{grow}}_X + \Delta^{\text{drop}}_X \qquad \square
\end{aligned}$$

The decomposition in Proposition 1 is an accounting identity that will hold exactly in the data for any one year if one defines the top X so that it has exactly X students in each year, sets all transition probabilities like a_{rX} to be the actual fraction of students at rank r at t who scored in the top X at $t+1$, and is consistent in what one plugs in for the multiple occurrences of conditional transition probabilities like $a^f_{rX}/(1-a^f_{rNP})$ that are undefined because the denominator is zero.[5] It will also hold exactly in data from multiple years if one uses appropriate weighted averages.

We instead implement the decomposition by estimating the transition probabilities both for the full population and for girls as smooth functions of the initial year rank via local linear regressions with $\log(Rank)$ as the right-hand-side variable.[6] This makes all of the transition probabilities continuous in rank and provides a natural definition for the conditional probabilities, avoiding any indeterminacies. We do this separately for students in 9th, 10th, and 11th grades, pooling the data for all six cohorts within each regression.

Table A2: Decomposition of Declines in Fraction Female, with Confidence Intervals

Grade Level	Achievement Level	Change in % Female	Decomposition of decline				
			Drop	Cont	Grow	Entry	Mech
Average	Top 5000	-3.1	-0.4	-1.2	-3.6	-1.1	3.5
		[-3.2,-2.9]	[-0.5,-0.4]	[-1.3,-1.1]	[-3.8,-3.5]	[-1.2,-1.0]	[3.4,3.6]
$9 \rightarrow 10$	Top 5000	-4.6	-0.1	-1.4	-2.9	-2.2	2.0
		[-5.1,-4.2]	[-0.2,0.0]	[-1.5,-1.2]	[-3.1,-2.7]	[-2.5,-1.9]	[1.9,2.2]
$10 \rightarrow 11$	Top 5000	-2.3	-0.4	-1.1	-3.7	-0.8	3.8
		[-2.7,-1.9]	[-0.5,-0.3]	[-1.2,-1.0]	[-3.9,-3.5]	[-1.0,-0.5]	[3.7,4.0]
$11 \rightarrow 12$	Top 5000	-2.3	-0.9	-1.1	-4.3	-0.4	4.8
		[-2.7,-1.9]	[-1.0,-0.7]	[-1.3,-1.0]	[-4.5,-4.1]	[-0.5,-0.2]	[4.5,4.9]
Average	Top 500	-1.9	-0.3	-1.0	-4.5	-0.4	4.1
		[-2.4,-1.5]	[-0.4,-0.1]	[-1.2,-0.7]	[-4.8,-4.1]	[-0.7,-0.1]	[3.9,4.5]
Average	Top 50	-0.5	-0.3	-0.8	-3.0	0.2	3.1
		[-1.6,0.5]	[-0.6,0.3]	[-1.7,0.1]	[-4.1,-2.2]	[-0.4,0.7]	[2.4,4.2]

Notes: Table reports 90 percent confidence intervals from nonparametric bootstrap, resampling at the student level and holding ranks fixed across 2,000 draws.

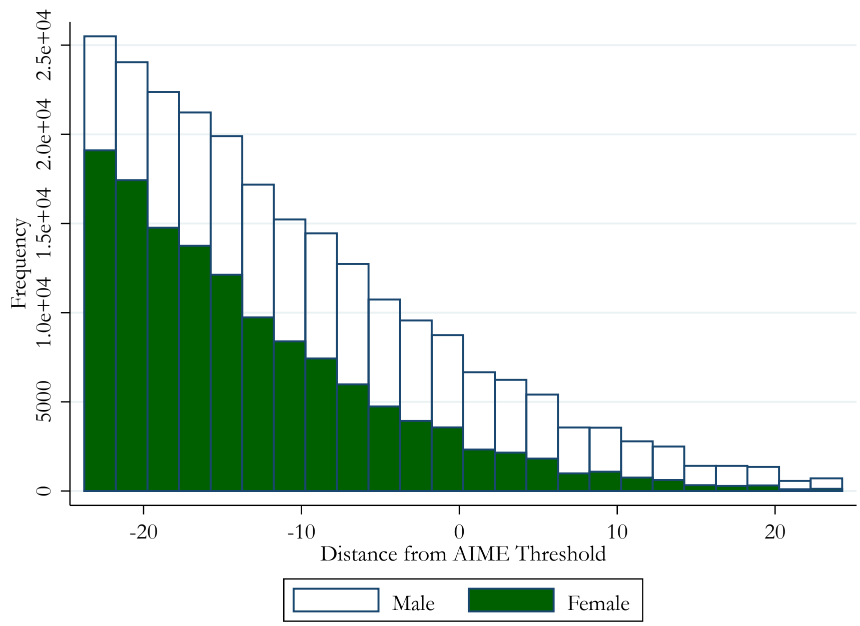
D Other Tables and Figures

Table A3: Average Achievement Gains as a Function of Initial Achievement, OLS and IV Specifications

	Dep. Var.: $\log(Rank_{i,t+1}) - \log(Rank_{it})$			
	$10th \rightarrow 11th$		$11th \rightarrow 12th$	
Specification	OLS	IV	OLS	IV
$-(\log(GradeRank_{it}) - \log(5000))$	0.16***	-0.07***	0.25***	-0.05***
	(0.002)	(0.004)	(0.002)	(0.004)
Constant	-0.27***	-0.50***	-0.02***	-0.33***
	(0.004)	(0.005)	(0.004)	(0.005)
Number of Observations	81430	81430	100270	100270
Root MSE	0.87	0.92	0.94	1.01

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table reports the results of OLS and IV regressions of growth in absolute performance as a function of initial performance relative to cohort. $\log(GradeRank_{it})$ instrumented with $\log(GradeRank_{i,t-1})$.

Figure A1: Regression Discontinuity Support – Histogram by Gender



Notes: Figure reports the count of male and female students in each relative score bin (the regression discontinuity running variable). Running variable is distance between the student's score on the first test he or she took in a given year and the AIME cutoff for that test-year.