ONLINE SUPPLEMENTARY MATERIALS

Technical Appendix

Density Checks

We see no visual evidence of manipulation of the forcing variable in Figure A-1. The distribution of scores around the pass-fail cutoff on the MCAS science examination is smooth.

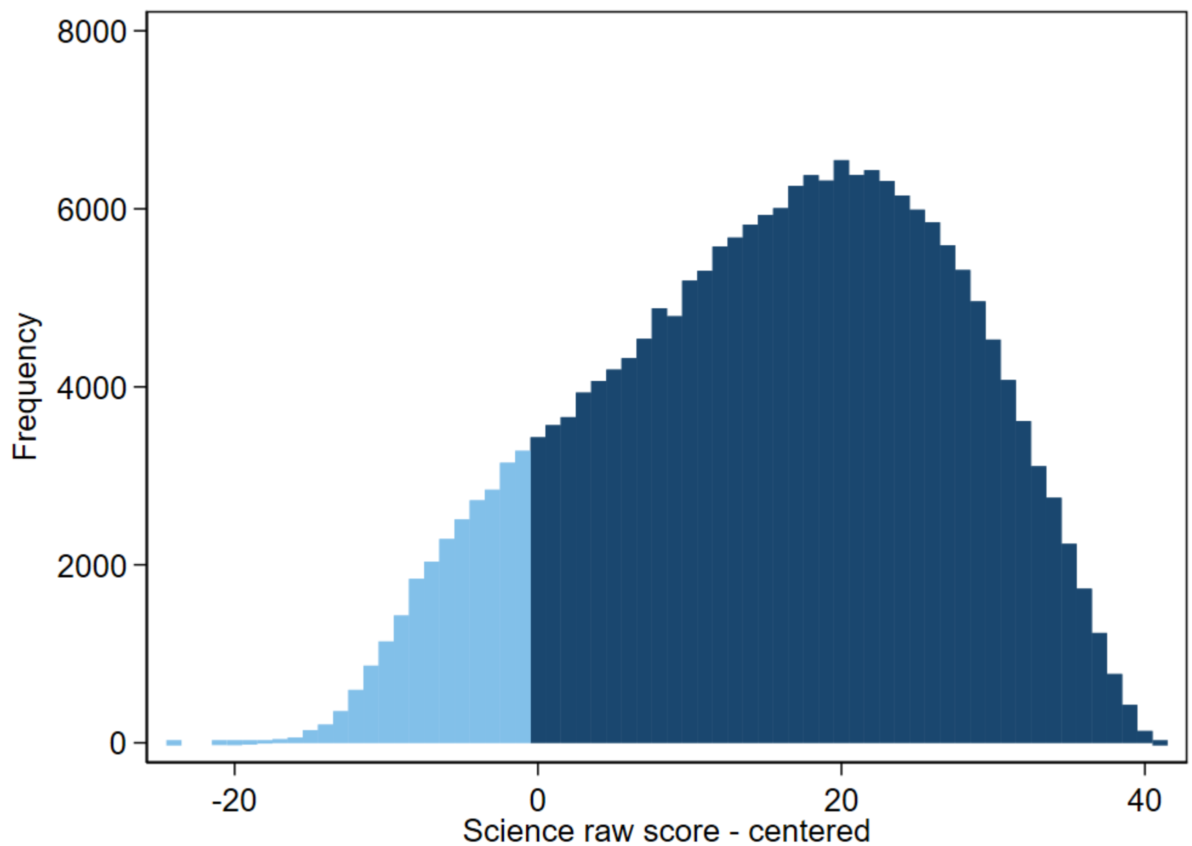

Figure A-1. Test score density around the pass-fail cutoff, 10th grade Massachusetts Comprehensive Assessment System (MCAS) science test, 2010-12 cohorts.

We also perform density tests described in McCrary (2008) and Frandsen (2017) and find no evidence of manipulation on the forcing variable. Using a bin size of 1 for the McCrary test, we obtain a discontinuity estimate of -0.001 with a standard error of .0116, meaning we fail to reject the null hypothesis of no manipulation on the forcing variable.

Frandsen (2017) argues that the McCrary density test can be misleading when the forcing variable is discrete, as in our case, and suggests an alternate test. The Frandsen test depends on the choice of the bound coefficient, k. With 24 support points within one standard deviation of the passing cutoff, we use values of k ranging from .002 to .01 and consistently fail to reject the null with this test as well.

Robustness to bandwidth

We perform the usual checks for robustness to bandwidth selection, presented in Table A-1. Panel 1 shows the estimated effect of barely passing the examination on the probability of graduating from high school within two years of the test, while Panel 2 shows the estimated impact on college graduation. The optimal bandwidth in our analysis is typically $h^*=2$. We present estimates at bandwidths 1.5 and 2 times as large ($h=3$ and $h=4$). We show estimated impacts for all students as well as for the key demographic groups on which we find heterogeneous effects in the paper (female students for high-school graduation, and higher-income students for college graduation).

The impact estimates, overall and by student group, are consistent in sign across the range of bandwidths, with the exception of college graduation for all students at $h=4$. While coefficients generally become smaller in magnitude as bandwidth increases, our main results are statistically significant at both $h^*=2$ and $h=3$. In particular, we replicate the key finding of a causal effect for female students on high-school graduation and higher-income students on college graduation at multiple bandwidths.

Table A-1. Estimated causal effects of passing the 10th grade exit examination in science for students on the probability of high-school and college graduation for students at the margin of passing, for different bandwidths by subgroup.

	Bandwidth		
Group	2	3	4
Panel 1: High-school graduation (on-time)			
All students	**0.0376****	0.0293**	0.0114
	(0.00464)	(0.00579)	(0.0135)
Females	**0.0697****	0.0527**	0.0265
	(0.0010)	(0.0074)	(0.0197)
Males	**0.00147**	0.00306	-0.00543
	(0.00924)	(0.00685)	(0.00923)
Panel 2: Any college graduation			
All students	**0.0126***	0.0144**	-0.000453
	(0.00344)	(0.00255)	(0.00818)
Low-income students	**0.00646**	0.00910*	-0.0112
	(0.00402)	(0.00294)	(0.0107)
Higher-income students	**0.0308****	0.0307**	0.0347**
	(0.00132)	(0.00134)	(0.00203)
N	17089	23867	30658

** $p<0.01$, * $p<0.05$, + $p<0.1$

Notes. Standard errors clustered on raw score point are in parentheses. Results using the optimal bandwidths appear in bold.

Visualizations

We include visualizations of discontinuities at the passing threshold, in addition to those presented in Figure 3 of the paper. In the top panel of Figure A-2, we display the mean high-school graduation outcomes for all students near the threshold, with on-time graduation at left and five-year graduation at right. A discontinuity at the passing cutoff is clearly visible in both. In the bottom panel, we show the same plots for graduation from a four-year college on the left and graduation from any college on the right. There is much less evidence of discontinuity at the threshold for these outcomes, though the plot for any college graduation has a small “bump” at the passing cutoff.

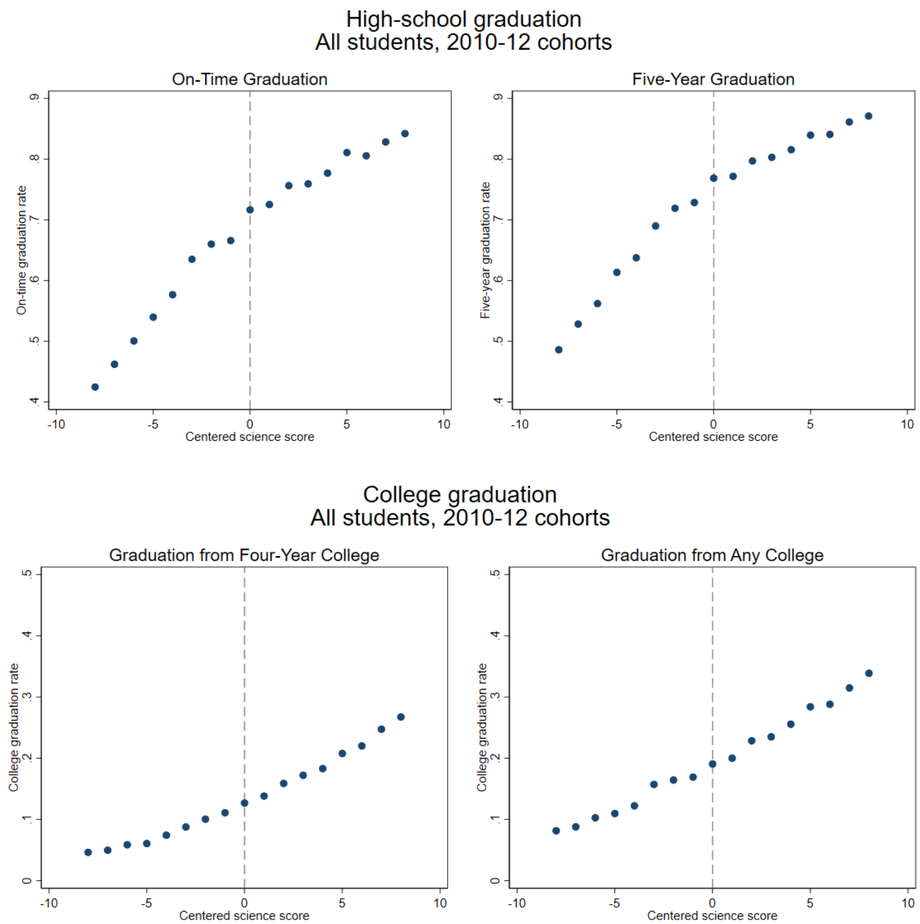

Figure A-2. Sample mean probabilities of on-time high-school graduation (top panel) and college graduation (bottom panel) at score points near the passing threshold on the science high-school exit exam, all students

In Figure A-3, we present a similar set of plots for males and females, while Figure A-4 includes plots by family income. The high-school graduation plots by gender, which also appear in the main text in Figure 3, reveal a more noticeable visual discontinuity for females than for males. In the college graduation plots, however, there is no evidence of any disruption at the passing cut-off for either group.

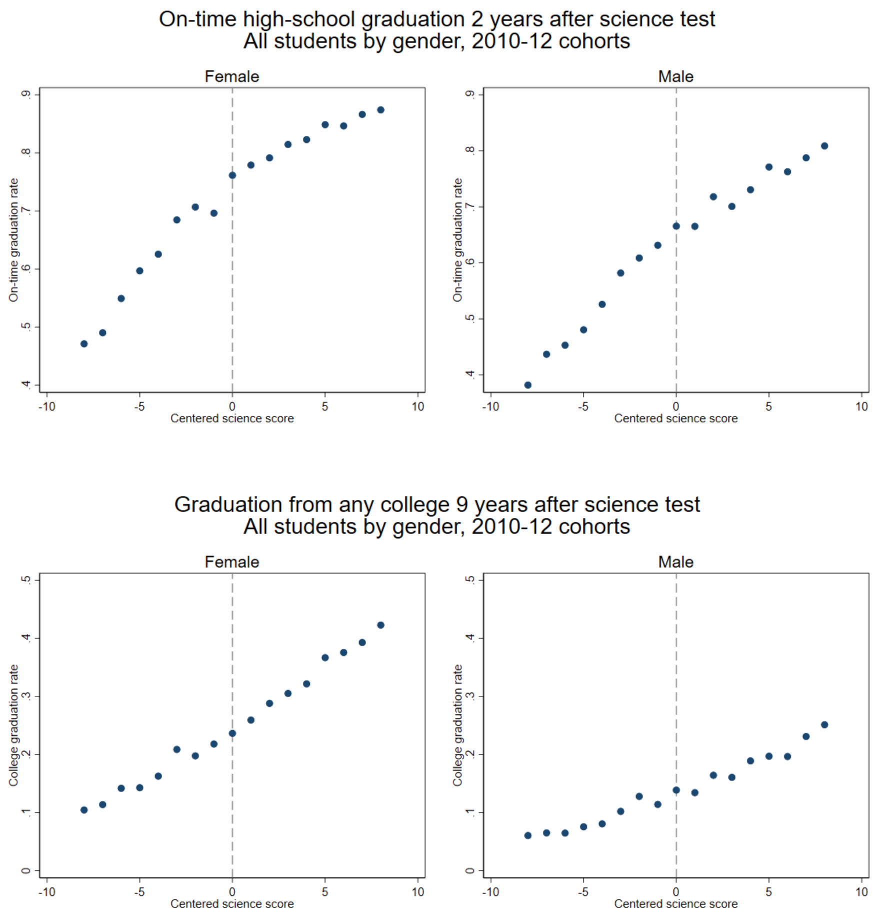

Figure A-3. Sample mean probabilities of on-time high-school graduation (top panel) and college graduation (bottom panel) at score points near the passing threshold on the science high-school exit exam for female and male students

In the top panel of Figure A-4, we note obvious discontinuities in the high-school graduation plots for both low-income and higher-income students. The jump at the passing threshold appears larger in the plot on the left, which aligns with the larger impact estimates for low-income students in Table 3. However, the bottom panel, copied here from Figure 3 in the main paper for reference, reveals a clear discontinuity in college graduation only for higher-income students.

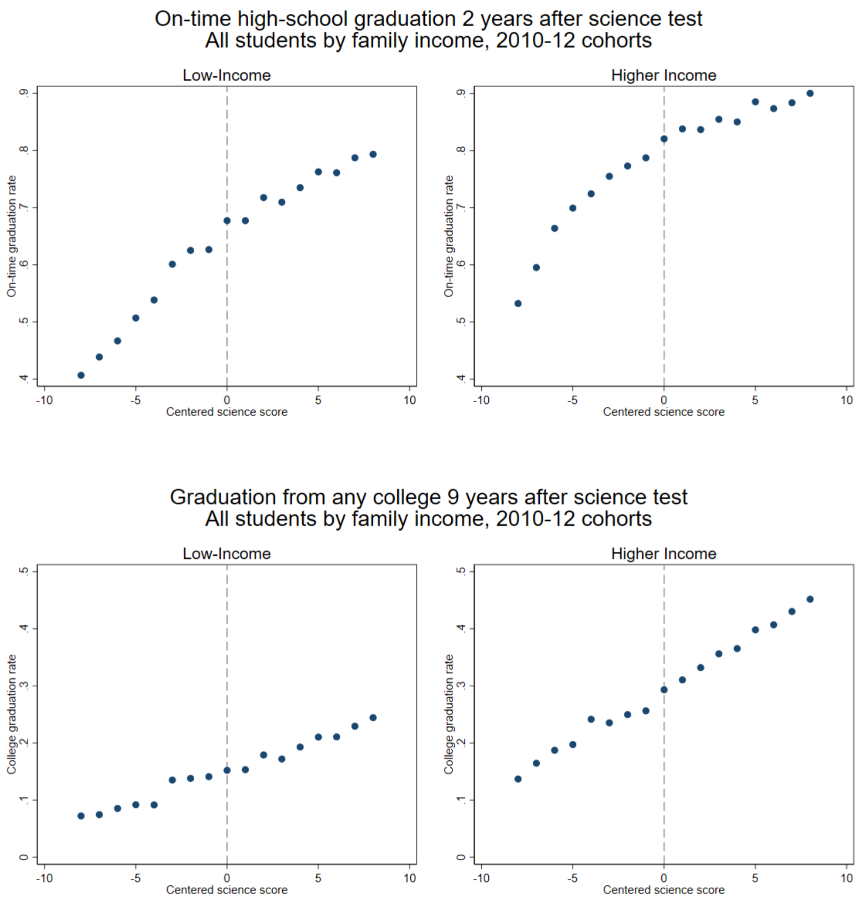

Figure A-4. Sample mean probabilities of on-time high-school graduation (top panel) and college graduation (bottom panel) at score points near the passing threshold on the science high-school exit exam, by family income

Placebo tests

We conduct placebo tests to assess whether students are indeed assigned exogenously relative to the cutoff. Here, we estimate our preferred RD model, treating each point in the test-score distribution as a placebo cutoff. If we have isolated the causal impact of barely passing, as compared to barely failing, the examination, we expect to find larger estimated effects at the actual cutoff than at the placebo cutoffs. This is indeed what we find.

We plot the results from this analysis in two panels of Figure A-5, highlighting our central findings. In the top panel, we illustrate the estimated impact of barely passing on on-time high-school graduation for female students. Given precision, we focus on points in the distribution for which we include more than 10,000 observations and do not include placebo estimates for points within two bandwidths of the actual cutoff. In the bottom panel, we do the same for the impact on college graduation for higher-income students. We compare these placebo estimates (in blue) to the actual impact estimate for students at the cutoff (in orange). For the high-school graduation estimate, we see clearly that the estimated effect at the actual cutoff (0) is far greater in magnitude than the impact at any placebo cutoff. For the impact of college graduation for higher-income students, we see a couple of placebo cutoffs with magnitude estimates that are close to, but still smaller than, the impact at the actual cutoff. Give that we calculate approximately 70 placebo estimates across these two models, it is reassuring that we find zero placebo estimates that are greater in magnitude than our actual impact estimates.

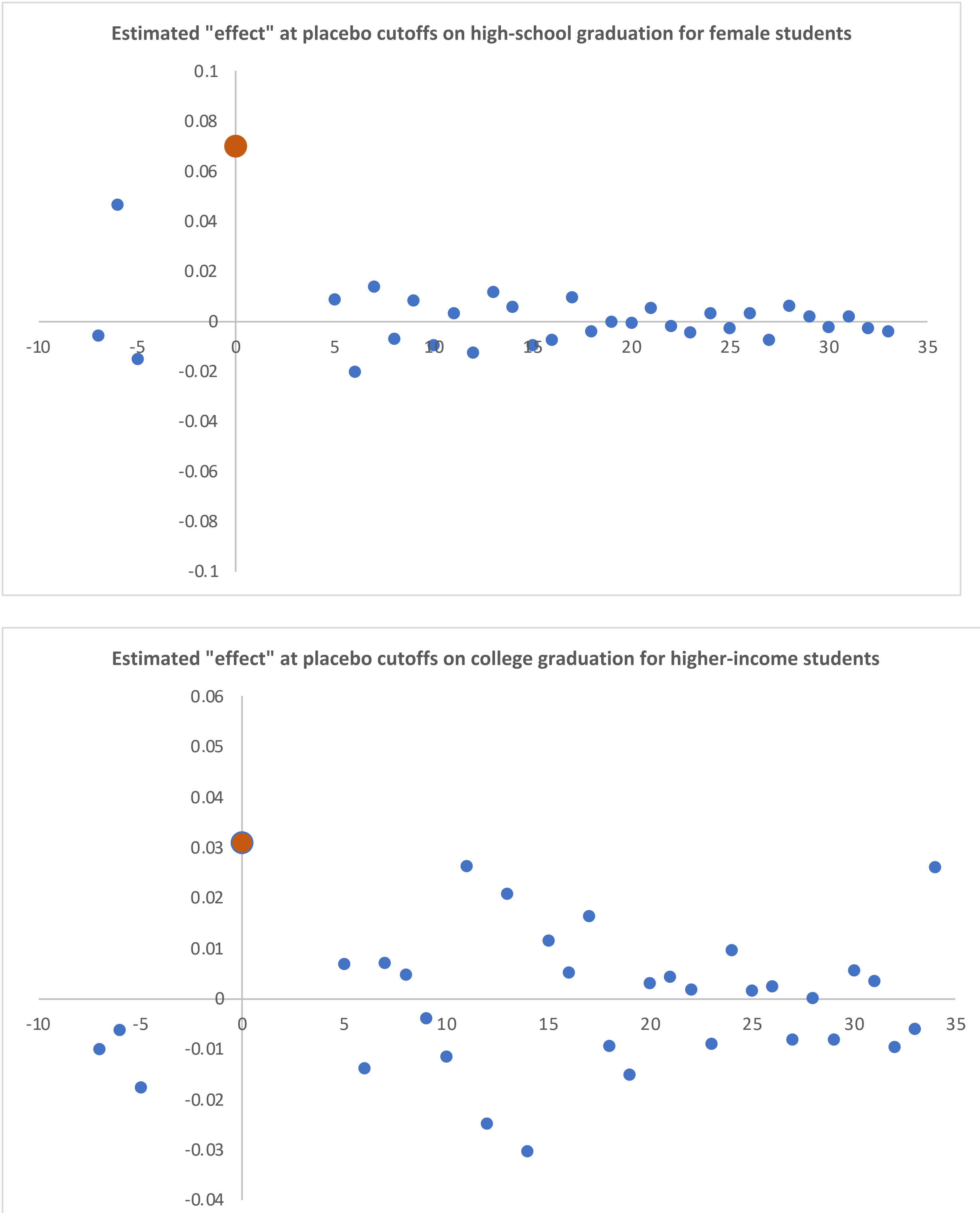

Figure A-5. Estimated impact of barely passing the MCAS at the actual cutoff (0) and different placebo cutoffs throughout the test-score distribution on on-time high school graduation for female students (top panel) and college graduation for higher-income students (bottom panel).

Covariate balance

Our models include many covariates. Given that assessing smoothness of covariates across the cutoff would entail a large number of hypothesis tests, we take a parsimonious approach by creating a single composite of these covariates. We do this by using empirical weights derived from a regression of our high-school graduation outcome on the full set of covariates in our model (other than test scores). In other words, we fit a model that regresses high-school graduation on our pre-treatment covariates. In this model, we use only observations outside of our analytical window around the cutoff (e.g., those outside of our preferred bandwidth of h^*=2). We then predict the fitted values from this regression for all observations in the data, including those in our analytical window. This process creates a single linear combination of our covariates that we can use to test covariate balance.

We then fit our primary RD model, excluding these covariates, to assess whether this composite is smooth on either side of the cutoff. In other words, we replace our outcome in the RD model with this covariate composite. We find no evidence of any discontinuity in this composite. Our point estimates are 0.0016 (h=2) and 0.0018 (h=3), neither of which are statistically significant.