

Abstract Title Page
Not included in page count.

Title:

The Price of Just Failing: Consequences of High School Exit Examinations for Urban Students in Massachusetts

Author(s):

John P. Papay, Harvard Graduate School of Education
Richard J. Murnane, Harvard Graduate School of Education
John B. Willett, Harvard Graduate School of Education

Abstract Body
Limit 5 pages single spaced.

Background/context:

Description of prior research and/or its intellectual context and/or its policy context.

As part of standards-based educational reforms introduced over the past two decades, many states have implemented exit examinations that students must pass in order to earn high school diplomas. Advocates argue that such examinations create incentives for students to work at learning important cognitive skills. By certifying that high school graduates have mastered the state-defined academic content standards, the examinations may also increase the economic value of a high school diploma (Evers & Walberg, 2002). Opponents of these tests suggest that they put unnecessary stress on students and encourage them to drop out of high school (Thomas, 2005; Jones, Jones, & Hargrove, 2003). They also argue that such tests place the greatest burden on the very groups, such as low-income students, that are already struggling in the educational system (Kornhaber & Orfield, 2001). Because high school graduation is associated with many positive life outcomes, the effects of high-stakes testing on high school completion are important for educational policymakers to understand.

Given the possible consequences of high-stakes testing, scholars have begun to explore whether exit examinations reduce graduation rates. This work has taken two main forms: some researchers have examined how imposing high-stakes testing policies affects aggregate student outcomes, while others have focused on the consequences of individual student performance on the test in states with high-stakes testing regimes.

Exploiting variation in exit examination policies across states and/or over time, some recent work provides at least tentative support that imposing exit examinations increases dropout rates (Reardon & Galindo, 2002; Warren, Jenkins, & Kulick, 2006; Nichols, Glass & Berliner, 2006). In contrast, Carnoy & Loeb (2002), Greene & Winters (2004), and Carnoy (2005) find no relationship between state accountability policies, such as exit examinations, and high school completion rates. Some recent work suggests that exploring aggregate patterns may obscure heterogeneity in effects for different groups of students (Jacob, 2001; Dee & Jacob, 2006).

Research that examines the relationship between individual student performance on exit examinations and high school completion remains much less common. Martorell (2005) finds no effect of failing the Texas exit examination on high school graduation for students who barely failed. This finding holds for every examination until the very last administration of a student's senior year.

Purpose/objective/research question/focus of study:

Description of what the research focused on and why.

Our identification strategy enables us to make causal claims about the effect of failing the exit examination – as opposed to passing it – for students of essentially equal ability near the pass/fail cutoff. We look at the effects of dividing a continuous measure of student proficiency into two categories – pass and fail – at an arbitrary cut score. Because setting a cut score is an inevitable feature of any exit examination policy, understanding its consequences is important.

We first examine the effect of failing the 10th grade mathematics examination. We then investigate students who fail their 10th grade test, examining their persistence and success on retests. Finally, we examine the effects of student performance on the 8th grade mathematics test.

In all cases, we pay particular attention to heterogeneous effects for students by family income and urban schooling. Specifically, we address three primary research questions:

- RQ1. Does failing the high school mathematics exit examination as a 10th grader make students on the margin of passing less likely to graduate from high school?
- RQ2. Do students who fail the 10th grade mathematics exit examination persist in retaking the examination and succeed in passing these retests?
- RQ3. Does failing the 8th grade mathematics examination cause students on the margin of passing to leave school before taking the 10th grade mathematics examination?

Setting:

Specific description of where the research took place.

We use data from the 2006 graduating cohort in Massachusetts. In the 15 years since the Massachusetts legislature passed the *Massachusetts Education Reform Act* of 1993, the state has invested more than one billion dollars per year in additional funding for K-12 public education. These investments appear to have borne considerable fruit. For example, a 2006 study by the Fordham Foundation praised the Massachusetts academic standards as the most rigorous in the country (Finn, Julian, & Petrilli, 2006) and a 2006 *Education Week* report concluded that the state-wide examinations were very well aligned with these standards (*Quality Counts*, 2006). Most importantly, Massachusetts students appear to be doing well. In 2007, Massachusetts 8th graders ranked first nationally in mathematics, tied for first in reading, and third in writing on the National Assessment of Educational Progress (NAEP) examinations (NCES, 2008). Furthermore, since the introduction of state testing under standards-based reform, the state's NAEP performance has improved rapidly. Thus, we examine the consequences for students of failing the MCAS examination in the context of a system that has brought about significant accomplishments.

Population/Participants/Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

To analyze the effect of failing the 10th grade examination (our first research question), we focus on the 66,347 students who first took the 10th grade mathematics MCAS examination as sophomores in 2004 and for whom it was a high-stakes test. For our third research question, we use the 69,127 students who took the 8th grade mathematics examination in 2002. This sample includes students who dropped out of school before 10th grade, but excludes any students who entered Massachusetts public schools after 2002.

Intervention/Program/Practice:

Specific description of the intervention, including what it was, how it was administered, and its duration.

Massachusetts began administering the MCAS mathematics and ELA examinations in 1998. For the class of 2003, the 10th grade tests became high-stakes exit examinations. Students must pass both tests in order to receive a high school diploma. We focus on the consequences of failing the mathematics examination because this test has proved to be a much greater barrier to high school graduation than the ELA examination. The state allows students to take the tests without time constraints and to retake them repeatedly if they fail, attempting explicitly to make the MCAS as minimal a barrier to graduation as possible.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

We use a quasi-experimental regression discontinuity design to answer our primary research questions. Conceptually, we would like to take students who scored identically, right at the pass/fail cut score, and randomly assign them to either a “pass” or a “fail” condition. This assignment process would render them equivalent in expectation on all observable and unobservable characteristics prior to treatment, allowing us to identify any differences in the ultimate outcome (high school graduation) as a causal effect of simply failing the examination, rather than of earning lower scores. Such an experiment is, of course, both politically impossible and unethical. However, we can take advantage of the state’s exogenous imposition of the minimum passing score to provide a natural experiment from which we can make equivalent causal conclusions. We examine students with similar MCAS performance, but just on either side of this exogenously-assigned cutoff, to determine the fitted probability of graduation for students who passed and who failed but scored right at the cutoff. Clearly, our sample includes no students who achieved the minimum passing score and failed; thus, we use the underlying relationship between MCAS score and graduation to predict the outcomes for these (hypothetical) students. We can then interpret any differences in the graduation outcomes between these students scoring at the cutoff who passed and who failed as the causal effect of failing the examination for these students “on the margin” of passing (Shadish, Cook, & Campbell, 2002). Any effect thus represents an unintended consequence of the exit examination because policymakers would hope that the mere label of “passing” or “failing” would not cause differences in outcomes for equally able students.

Data Collection and Analysis:

Description of plan for collecting and analyzing data, including description of data.

The Massachusetts Department of Education has compiled a comprehensive database that tracks students longitudinally throughout high school, allowing for clear description of student graduation outcomes. For the 2006 graduating cohort, the records contain each student’s MCAS mathematics and ELA test results, demographic characteristics, and status at cohort graduation, including whether the student graduated, dropped out, is still enrolled, transferred out, was expelled, or any of eleven possible outcomes. It tracks students who leave the system, including those who pursue GEDs, and enables us to identify students who graduate in five years and those who leave the system at one point but return to graduate. This dataset allows for much more precise estimation of the probability of high school completion than do most previous studies, and it permits investigation of the direct link between student performance on high-stakes tests and graduation outcomes at the individual level.

Using the regression discontinuity strategy first proposed by Thistlethwaite & Campbell (1960), we can analyze data from our natural experiment to make causal inferences for students at the margin of passing. Because the probability that a student passes the examination goes unequivocally from zero to one at the single cut score, the discontinuity is sharp. Because we do not know the exact functional form of the relationship between the probability of graduation and MCAS score, we display this relationship over the MCAS score range using a nonparametric smoothing approach. As standard nonparametric smoothing strategies have poor boundary properties, Hahn, Todd, & Van der Klaauw (2001) recommend estimating such limits using the method of local linear regression. Our implementation of this strategy follows the recommendations of Imbens and Lemieux (2008).

Findings/Results:

Description of main findings with specific details.

Passing the 10th grade MCAS mathematics examination increases the probability that a low-income, urban student on the margin of passing will graduate from high school on-time by eight percentage points ($p=0.015$). Given that 26% of low-income, urban students who just pass the exam do not graduate on time, this effect is quite substantial. Similarly, the effect on a student's probability of graduating in five years is seven percentage points ($p=0.027$). In Figure 1, we present the fitted nonparametrically smoothed relationship between on-time graduation and MCAS mathematics score for low-income urban students (please insert figure 1 here). For students at the margin, passing the examination substantially increases their subsequent probability of graduation. Visually, this effect appears as an interruption in the underlying smooth relationship between the probability of graduation and the MCAS mathematics score at the cut-off. For perspective, we have included the sample mean probabilities of on-time graduation at each MCAS score level.

Importantly, we find no such effects for wealthier urban students or for suburban students, regardless of family income. Thus, it is the interaction of low family income and an urban environment that appears to render students, on average, susceptible to the effects of failing. In Table 1, we present the effect of passing the examination, as opposed to failing it, for students in each of the four groups on both on-time and five-year graduation, with and without time-invariant student demographic controls (please insert table 1 here). In no cases are the effects for wealthier urban or suburban students statistically significant, and the effects for low-income urban students are statistically distinct from all three other groups.

We also find interesting descriptive patterns concerning student retesting behavior. Overall, the 8,269 students who fail the mathematics MCAS on their first try in 2004 show remarkable persistence in retaking the examination. Nearly 89% take the examination at least one more time and, of these students, 68% go on to pass the test at some point in high school. Among students with the same MCAS scores on the initial tests, low-income urban students are no less likely than suburban students to retake the mathematics examination. However, low-income urban students are nearly ten percentage points less likely to pass this retest than suburban students with the same initial scores ($p<0.001$). Finally, we find that, for urban students on the margin of passing, failing the 8th grade mathematics examination reduces the probability of persisting to 10th grade by three percentage points ($p=0.036$). Importantly, we find no effects of failing the 8th grade examination for suburban students.

Conclusions:

Description of conclusions and recommendations of author(s) based on findings and over study. (To support the theme of 2009 conference, authors are asked to describe how their conclusions and recommendations might inform one or more of the above noted decisions—curriculum, teaching and teaching quality, school organization, and education policy.)

We see several complementary explanations for the finding that failing the 10th grade mathematics examination reduces the likelihood of graduation for urban students from low-income families, but not for more affluent or suburban students. Importantly, we cannot distinguish whether just failing the examination causes these students to drop out or whether just passing it causes them to remain in school. Low-income urban students who pass may feel encouraged that they are doing well in school and may decide to persist to graduation. Similarly, schools or teachers may respond differently to students who pass, leading to improved

graduation outcomes.

On the other hand, low-income urban students who fail the examination may become discouraged or subject to institutional responses that reduce their likelihood of graduating on time. Families of low-income urban students may lack the resources to help them overcome the hurdle posed by failing the examination. Low-income urban students typically attend high schools in which many students have failed the 10th grade MCAS examinations. These schools are struggling to figure out how, with very limited resources, to respond to this problem. They may respond by retaining students in grade or placing them in remedial classes, which in turn may cause students to lose motivation and leave school. Finally, the interaction between school and home contexts may produce these effects.

Regardless of the mechanism at play, we find that low-income, urban students with essentially the same proficiency on the state test have substantially different graduation outcomes simply because they are categorized as “passing” or “failing” the examination. Thus, the inevitable necessity of choosing a cut point does produce effects for this potentially vulnerable group of students. These effects represent unanticipated consequences of efforts to prepare all students to meet the demands of 21st century life. These consequences are important and need to be at the center of efforts to make standards-based reforms work for all Massachusetts students in the years ahead.

That the vast majority of students who fail the 10th grade mathematics examination retake it and that low-income urban students retake the test at similar rates as their wealthier urban or suburban peers are encouraging. These findings suggest that these students are receiving the message that they should persist and retake the test. As a result, schools have time to work with these students and prepare them to meet the graduation requirements. However, low-income urban students are much less likely to pass this retest, even when comparing students with the same initial examination performance. Finding the explanation for this pattern is an important topic for research, with critical implications for improving equality of educational opportunity.

Our findings raise several questions for researchers, educators, and policymakers in Massachusetts and other states. First, the absence of effects of exit examination performance on high school completion for suburban students – including those from low-income families – suggests that it is possible to overcome the initial disappointment associated with failing a high-stakes examination (or to instill in students who fail the same sense of encouragement that equally able students who just pass have). Learning more about the initiatives that improve student retention could be helpful for districts struggling to support many failing students. A related question that we intend to pursue in future work is whether some urban districts or schools are more successful than others in supporting students who fail the 10th grade mathematics examination. If that is the case, then understanding the successful efforts of some urban districts might help others to improve their support to struggling mathematics learners.

This research argues strongly for the importance of examining heterogeneous effects. In future work, we hope to explore more fully the effects of failing on different groups of students, including students of different races and those with limited English proficiency. It also raises the question of whether the types of differential impacts we observe in Massachusetts may also be present in other states, especially those that use relatively demanding exit examinations. Finally, we wonder why the effect for urban students varies by income. Do wealthier students attend different schools, or do they receive additional support outside of school?

Appendixes
Not included in page count.

Appendix A. References

References are to be in APA format. (See APA style examples at the end of the document.)

- Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice*, Winter, 19-31.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Dee, T.S., & Jacob, B.A. (2006). Do high school exit exams influence educational attainment or labor market performance? Cambridge, MA: NBER Working Paper 12199. Retrieved June 26, 2008, from <http://www.nber.org/papers/w12199>.
- Evers, W.M., & Walberg, H.J. (Eds.). (2002). *School accountability*. Stanford, CA: Hoover Institution Press.
- Finn, C.E., L. Julian, & Petrilli, M.J. (2006). *The state of state standards*. Washington, D.C.: The Fordham Foundation. Retrieved March 26, 2008 from <http://www.edexcellence.net/foundation/publication/publication.cfm?id=358>.
- Greene, J.P. & Winters, M.A. (2004). *Pushed out or pulled up? Exit examinations and dropout rates in public high schools*. New York: Center for Civic Innovation at the Manhattan Institute for Policy Research.
- Hahn, J., P. Todd, & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-35.
- Jacob, B.A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99-121.
- Jones, M.G., B.D. Jones, & Hargrove, T.Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Kornhaber, M.L., & Orfield, G. (2001). High-stakes testing policies: Examining their assumptions and consequences. In G. Orfield & M.L. Kornhaber, (Eds.). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.
- Martorell, F. (2005). Does failing a high school graduation exam matter? Unpublished working paper: Author.

- National Center for Education Statistics. (2008). *State comparisons: National Assessment of Educational Progress (NAEP)*. Washington, DC: U.S. Department of Education. Retrieved April 5, 2008 from <http://nces.ed.gov/nationsreportcard/nde/statecomp/>
- Nichols, S.L., Glass, G.V., & Berliner, D.C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved June 26, 2008, from <http://epaa.asu.edu/epaa/v14n1/>.
- Quality Counts. (2006). Quality Counts at 10: A decade of standards-based education. *Education Week*, 25(17), 74.
- Reardon, S.F., & Galindo, C. (2002, April). Do high-stakes tests affect students' decisions to drop out of school? Evidence from NELS. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Shadish, W.R., T.D. Cook, & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
- Thistlethwaite, D.L., & Campbell, D.T. (1960). Regression-discontinuity analysis: An alternative to the *ex post facto* experiment. *The Journal of Education Psychology*, 51(6), 309-317.
- Thomas, R.M. (2005). *High stakes testing: Coping with collateral damage*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Warren, J.W., K.N. Jenkins, & Kulick, R.B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis*, 28(2), 131-152.

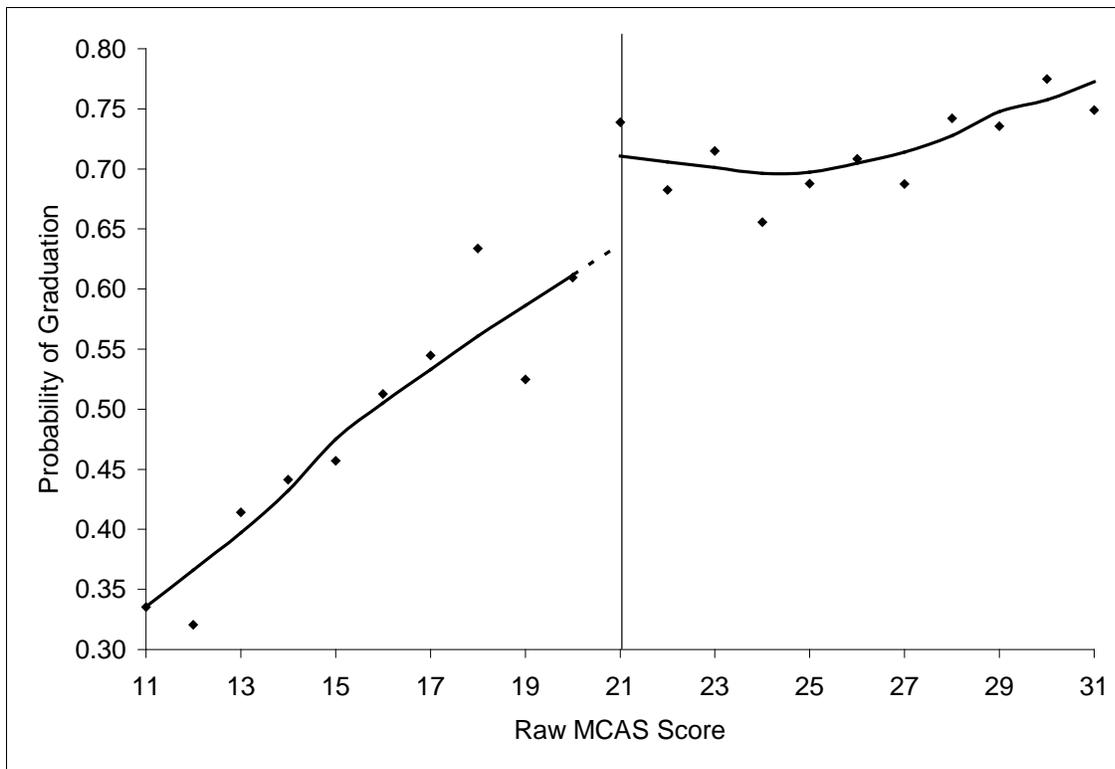
Appendix B. Tables and Figures
Not included in page count.

Table 1. Estimated causal effects of passing the 10th grade mathematics examination, as opposed to failing it, on the probability of on-time and five-year graduation for students at the margin of passing, with and without student-level covariates for different subgroups. Standard errors are in parentheses. (Bandwidth $h^* = 6$).

Group	On-time graduation		Five-year graduation		N
	No Covariates	Covariates	No Covariates	Covariates	
Urban, Low Income	0.074* (0.033)	0.080* (0.033)	0.064* (0.032)	0.071* (0.032)	3,469
Urban, Not Low Income	-0.045 (0.054)	-0.052 (0.054)	-0.048 (0.053)	-0.057 (0.053)	1,371
Suburban, Low Income	0.020 (0.042)	0.023 (0.042)	0.014 (0.040)	0.017 (0.040)	2,172
Suburban, Not Low Income	-0.011 (0.027)	-0.014 (0.027)	-0.019 (0.026)	-0.022 (0.026)	4,857

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Figure 1. Fitted smoothed nonparametric relationship (bandwidth=6) between the probability of on-time graduation and 10th grade mathematics score for low-income urban students, with the sample mean probabilities of graduation overlaid.



Note. We plot the nonparametric regression fit without student-level covariates.