**Title:** The distributional consequences of high school exit exams

**Authors:** Marianne Bitler, UC Irvine and NBER; AnneMarie Conley, UC Irvine; Thurston Domina, UC Irvine; & Hilary Hoynes, UC Davis and NBER

**Background/context:**

Until recently, high school students in the US earned diplomas exclusively by accumulating course credits. Over the last three decades, however, a growing number of states have implemented high school exit exams and mandated that students pass these minimum-competency tests in order to earn their diplomas. Today, high school exit exams are in place in 22 U.S. states, and approximately two-thirds of US high school students must pass one of these exams in order to earn a high school diploma. California's high school exit exam (CAHSEE) was mandated by the state legislature in 1999 and has been administered in English to all students in the spring of their 10th grade year since 2002. The CAHSEE was implemented in two steps: Pursuant to a State Board of Education ruling, students who were 10th graders in California public schools in the spring of 2002 and 2003 but were allowed to graduate without passing the CAHSEE. However, all students in later cohorts must pass the exam in order to earn a high school diploma.

Several studies conducted in California and elsewhere strongly indicate that high school exit exams have negative effects on student odds of earning high school diplomas (Clark, 2009; Dee & Jacob, 2007; Martorell, 2004; Papay, Murnane & Willet, 2008; Warren & Edwards, 2005). In California, nearly one-third of students fail to pass the exit exam on their first attempt (Human Resources Research Organization, 2008) and regression discontinuity analyses indicate that students who barely fail the exam on their first attempt are considerably more likely to drop out of high school than they would otherwise have been had they barely passed (Reardon, Atteberry, Arshan, & Kurlaender, 2009). On average, Warren, Jenkins, & Kulik (2006) estimate that high school graduation rates fall by approximately 2 percentage points when states implement high school exit exams.

The effects that high school exit exams have on student learning are less clear. While some evidence points to a positive relationship between the existence of high school exit exams and student achievement (Bishop, 1997; Bishop, Moriarty, & Mane, 2001), the most rigorous available studies indicate that exit exams have no causal effect on achievement (Jacob, 2001). However, since they estimate only the average effects of high school exit exams, these studies neglect the possibility that high school exit exams have different effects on students with differing levels of academic skill.

A smaller number of studies have estimated the effects of exit exams on the distribution of student achievement. Reardon, Atteberry, Arshan, & Kurlaender's research in California (2009) provides some preliminary evidence to suggest that scores on the relatively low-stakes California Standards Exams for 11th graders at the bottom the test score distribution declined modestly after the CAHSEE was implemented as a high stakes exam. Furthermore, using quantile regression techniques and nationally-representative National Assessment of Educational Progress (NAEP) test score data, Grodsky, Warren, & Kalgorides (2009) demonstrate that the effects of high school exit exams are flat across the skill distribution. Both of these studies suggest that the distributional consequences of exit exams are minimal. However, we hypothesize below that the distributional effects of exit exams are likely to be sharply discontinuous. We argue that more fine-grained distributional analyses are necessary in order to capture the sharply uneven effects that high school exit exams have on student achievement.

**Objective:**

We contribute to the literature by analyzing the effects of the CAHSEE on student graduation odds using official enrollment and graduation data for several southern California districts. Our analyses pay particular attention to the high stakes exam's distributional consequences, estimating policy effects across the distribution of students' latent probability to earn a high school diploma.

In addition, our work supplements the uneven existing literature on the effects of exit exams on student achievement. By establishing a minimum competency threshold and giving students strong incentives to clear it, high school exit exams aim to raise the academic achievement of low-performing students. In this paper, we investigate the effects of California's exit exam policies across the academic achievement distribution.

Our hypothesis, illustrated in Figure 1 in Appendix B, is that exit exams have stronger positive effects on students near the bottom of the achievement distribution (near the cutoff for passing) than on the top end of the achievement distribution, and may also have negative effects for students at the very bottom of the distribution. By raising the stakes associated with minimum competency on the skills they assess, the CAHSEE may encourage students who expect to be close to the minimum competency threshold to redouble their academic efforts. On the other hand, we do not expect the CAHSEE to have a positive effect on the academic achievement of the large number of students who expect to score well above the minimum competency threshold. In fact, the exit exam may even have a modest negative effect on these relatively high-achieving students' academic achievement, if they are motivated to some extent by the desire to earn a high school diploma. Lastly, for students who expect to score well below the threshold, we also anticipate that there may be negative effects. Existing literature on high school exit exams and other accountability interventions fail to examine effects on the full distribution. Our work fills an important gap in this literature.

**Setting/Population/Participants/Subjects:**

This paper uses data from several southern California public school districts for 10th graders for the 2003–2004, 2004–2005, and 2005–2006 academic years. For one district (N=13,583), we have the universe of data for 10th graders (and their past and future test scores and measures of high school completion from this district). For the other two districts we have data on a subset of schools. These districts serve a student population with large shares of Hispanics (68%), non-Hispanic Vietnamese (14%), and non-Hispanic White students (12%), along with smaller shares of students from other ethnic groups. We focus on students in 10th grade in these particular cohorts because of the nature of the testing intervention.

**Intervention:**

The intervention that we analyze in this paper is the implementation of the CAHSEE as a high-stakes qualifying exam for a high school diploma. In the words of California state superintendent of instruction Jack O'Connell, the CAHSEE "measures absolutely the least our students must know as they move on to their next step in learning and earning." The exam consists of two parts: (1) a multiple choice mathematics portion, which is aligned with the California state math standards for 6th and 7th graders, as well as Algebra I, and (2) an English language arts (ELA) portion, which uses multiple choice questions and one essay to assess student mastery of 10th grade and earlier content standards. The test is untimed and administered to students over the course of two days. Each part is scored on a scale that ranges from 275 to 400. To pass the test and earn a high school diploma, students must score at least 350 on each part. The state allows students who fail either portion of the test to retake that portion in order to qualify for a diploma.

The original CAHSEE legislation stipulated that students in the high school class of 2004 and students in all subsequent graduating classes would be required to pass the exam in order to earn their diplomas. However, after independent evaluators argued that the original legislation did not give schools enough time to adjust their curricula to the exam's requirement, the State Board of Education voted in July of 2003 to defer the CAHSEE graduation requirement for two years. As a result, the 10th graders who took the test in 2004 were the first students to take the CAHSEE as a high-stakes exam. While nearly all 10th graders in 2002, and all 10th graders in 2003 took the CAHSEE, these students could earn high school diplomas even if they failed to pass the test.

**Research Design:**

Our analyses take advantage of the CAHSEE's staggered implementation to construct a quasi-experimental evaluation of the high-stakes test's distributional consequences. The treatment group in this study consists of students who were 10th graders in the spring of 2004 or the spring of 2005. These students were in the first two cohorts for which the CAHSEE was a high-stakes high school exit exam. The control group consists of students who were 10th graders in the spring of 2003. While these students were required by law to take the exam, they did not have to pass it in order to earn a high school diploma.

We estimate the distributional effects of the CAHSEE by comparing the distribution of student outcomes in the treatment and control groups. Our preliminary analyses have focused on the distribution of student scores on the CAHSEE exam itself. In future analyses we will investigate the effects of the CAHSEE on the distribution of student scores on low-stakes California Standards Tests (CST) in Mathematics and English language arts, as well as on the distribution of the latent probability of high school graduation.

**Data Collection and Analysis:**

Our first comparison will test whether our treatment and control groups look balanced according to pre-intervention test scores. In addition to CAHSEE data, we also have information on other test scores for years the students were in the districts–including both the CSTs (used for NCLB accountability purposes and thus high stakes for teachers and schools but not for students) and district-level tests (which count towards student grades and should be high stakes for students). We will compare the means and distributions of these scores for our students for the year before they were in 10th grade to assess balance.[1] Then we will compare the test score distributions for students in our districts who took the CAHSEE but did not need to pass to graduate (spring 2003 10th graders) to the test score distributions for students in adjacent cohorts (spring 2004 and spring 2005 10th graders) who did need to pass. Thus, unlike Reardon et al.

[1] If there is lack of balance across observable dimensions, we will use inverse propensity score weights based on predicting being in the treatment group to adjust for observable differences (see Firpo 2007 for a discussion of the appropriateness of this for uncovering quantile treatment effects).

(2009), who focus on local treatment effects for students right at the threshold of passing or failing or in the bottom of the pre-CAHSEE distribution, we will test for effects of the existence of this high stakes test across the full distribution of student CAHSEE outcomes.

We will calculate the effects of CAHSEE on the distribution using the quantile treatment effect estimator assuming that the 2003–2004 cohort is a valid control group for the two treated cohorts (we will also test this assumption). Quantile treatment effects (QTE) are the distributional analog to the mean differences and regression adjusted mean differences that are the workhorse of randomized control trials. In the experimental setting, QTE are produced by comparing various percentiles for the treatment and control groups. Thus the QTE for the $10^{th}$ percentile is the difference in the $10^{th}$ percentile of the treatment and control groups. Just as the mean difference is a consistent estimate of the average treatment effect, the QTE are consistent estimates of the treatment effect at specific points in the outcome distribution. Inference will be performed using bootstrapped confidence intervals.[2]

In order to insure that the results of our analyses of CAHSEE test scores are not simply artifacts of increasing familiarity among teachers and students with the test, we will estimate the distributional consequences of CAHSEE implementation on several additional student outcomes. We will make use of the $10^{th}$ grade year test scores on the district and CST tests as alternative comparison distributions for the treatment group's scores in the years that the CAHSEE exams are required. In addition, we will examine the smoothed binary quantiles (Kordas, 2006) for the distribution of the latent probabilities the dichotomous outcome of high school completion.

**Findings/Results**

We have begun analysis of the exit exam data from the first of our several school districts for our analysis of the effects of the California requirement that students pass the CAHSEE. We have calculated quantile treatment effects estimates of the effect of being in a cohort required to pass the CAHSEE to graduate (treatment group) compared to being in a cohort for whom the test was low stakes (passing the exam was not required to graduate, control group). Figure 2 in Appendix B shows the inverse cumulative distribution functions (CDF) for scaled math scores on the CAHSEE (exit exam) for students in the treatment group ($10^{th}$ graders in the 2004–05, and 2005–06 school years) and control group ($10^{th}$ graders in the 2003–04 school year). The solid line is the inverse CDF for the treatment group and the dashed line the inverse CDF for the control group. Then, as noted before the 0.50 quantile treatment effect is given by the vertical difference in these inverse CDFs at the median and so on.

Figure 3 in Appendix B presents the estimated QTE for various points in the scaled mathematics CAHSEE score distribution along with point wise 90% confidence intervals. These QTE were estimated at each of 9 points 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 (the deciles), along with 90% bootstrapped point wise percentile confidence intervals for each estimate. We find a suggestion that the exam has lead to a statistically significant increase in scores at the bottom of the distribution of math tests, but little or no effect at the top of the score distribution.

---

[2] Quantile treatment effects can also be calculated as the result of a quantile regression (Koenker and Bassett, 1982) of the outcome on a binary treatment indicator. Heckman, Smith, and Clements (1997), Koenker and Bilias (2001), and Bitler, Gelbach and Hoynes (2006) are recent applications of QTE in a setting with experimental data and Angrist, Lang, and Oreopoulos (2009) apply this to an education experiment. Reback (2007) and Neal and Schanzenbach (forthcoming) are two recent studies looking at distributional effects of other interventions.

We have also estimated QTE for several subgroups of students by race and ethnicity (Hispanics, Vietnamese, Whites, and others). While all of these QTE within subgroups show something similar (mostly positive effects, larger in the middle of the distribution), there are some interesting differences as well. For example, all of the QTE for Hispanics are statistically significantly different from 0, while for both Whites and Vietnamese, those at the very top of the distribution are negative and significant. Further work will include adding data from the other districts, examining balance in the pre-10$^{th}$ grade scores, and using the alternate comparison strategies.

**Conclusions**

Our preliminary findings suggest that in the case of our first school district, facing a mandatory exit exam may have had positive effects for some share of the bottom of the test score distribution but had little effect at the top. These findings are largely consistent with our hypothesis, and they suggest that looking across the distribution may reveal further effects of CAHSEE implementation that are obscured by in mean effects or regression discontinuity analyses.

Appendix A. References

Angrist, J., Lang, D., & Oreopoulos, P. (2009), ‘Incentives and services for college achievement: Evidence from a randomized trial,’ American Economic Journal: Applied Economics 1:1, 136–163.

Bishop, J.H. (1997). ‘Do curriculum-based external exit exams enhance student achievement?’ Consortium for Policy Research in Education Research Report, University of Pennsylvania. Retrieved August 22, 2009 from http://www.cpre.org/images/stories/cpre_pdfs/rr40.pdf.

Bishop, J.H., Moriarty, J.Y. & Mane, F.P. (2001). ‘Diplomas for learning, not for seat time: the impacts of the New York Regents examinations,’ Economics of Education Review 19:4, 333–349.

Bitler, M., Gelbach, J., and Hoynes, H. (2006). ‘What mean impacts miss: Distributional effects of welfare reform experiments,’ American Economic Review 96:4, 988–1012.

Clark, D., Martorell, P., & Yang, M. (2009). ‘The Effects of High School Exit Exam Performance: New Evidence from Florida,’ Society for Research on Educational Effectiveness.

Dee, T.S. & Jacob, B. (2007). ‘Do high school exit exams influence educational attainment or labor market performance?’ in, A. Gamoran, ed., On Standards-Based Reform and the Poverty Gap: Lessons for No Child Left Behind, Washington, DC: Brookings Institution Press.

Firpo, S. (2007), ‘Efficient semiparametric estimation of quantile treatment effects,’ Econometrica 75:1, 259–276.

Grodsky, E., Warren, J. & Kalogrides, D. (2009), ‘State High School Exit Examinations and NAEP Long-Term Trends in Reading and Mathematics, 1971–2004.’ Educational Policy 23: 589–614.

Heckman, J. J., Smith, J. & Clements, N. (1997), 'Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts,’ Review of Economic Studies 64, 487–535.

Human Resources Research Organization (2008), Independent evaluation of the California High School Exit Examination (CAHSEE) fourth biennial report.

Jacob, B., (2001), ‘Getting tough? The impact of high school graduation exams,’ Educational

Evaluation and Policy Analysis 23: 2, 99–121.

Koenker, R. & Bilias, Y. (2001). ‘Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments,’ Empirical Economics 26:1, 199–220.

Koenker, R. & Bassett, G. (1978). ‘Regression quantiles,’ Econometrica 46, 33–50.

Kordas, G. (2006), ‘Smoothed binary regression quantiles,’ Journal of Applied Econometrics 21:3, 387–407.

Martorell, F., (2004), ‘Do High School Graduation Exams Matter? A Regression Discontinuity Approach.’ Working paper, University of California, Berkeley.

Neal, D. & Schanzenbach, D. (forthcoming), ‘Left behind by design: Proficiency counts and test-based accountability,’ Review of Economics and Statistics.

Papay, J.P., Murnane, R.H. & Willet, J.B., (2008), ‘The consequences of high school exit examinations for struggling low-income students: Evidence from Massachusetts.’ National Bureau of Economic Research Working Paper No. W14186.

Reardon, S., Atteberry, A., Arshan, N., & Kurlaender, M. (2009). Effects of the CaliforniaHigh School Exit Exam on Student Persistence and Graduation. Institute for Research on Educational Policy and Practice Working Paper, Stanford University. Retrieved August 22, 2009 from http://www.stanford.edu/group/irepp/cgi-bin/joomla/docman/reardon-et-al-cahsee-study/download.html.

Reback, R. (2008), ‘Teaching to the rating: School accountability and the distribution of student achievement,’ Journal of Public Economics 92, 1394–1415.

Warren, J.R. & Edwards, M.R., (2005), ‘High school exit examinations and high school completion: Evidence from the early 1990s.’ Educational Evaluation and Policy Analysis 27:1, 53–74.

Warren, J. R., Jenkins, K. N., & Kulick, R. B., (2006), ‘High school exit examinations and state-level completion and GED rates, 1975–2002.’ Educational Evaluation and Policy Analysis 28:2, 131–152.

Weiher, G.R. & Tedin, K. (2002). ‘Does Choice Lead to Racially Stratified Schools? Charter Schools and Household Preferences.’ Journal of Policy Analysis and Management, 21:1, 79–92.

Appendix B. Figures

Figure 1 Hypothesized distributional effect of CAHSEE

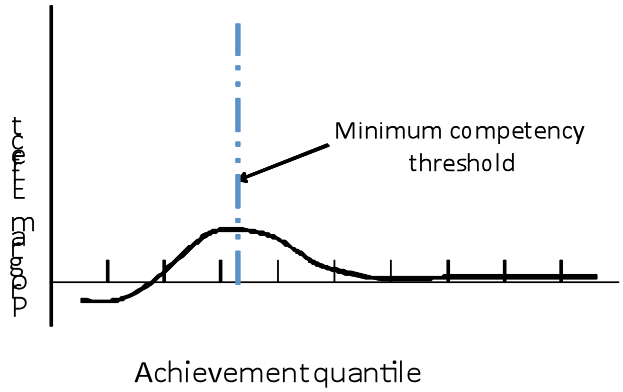


Figure 2 Inverse CDFs for scaled mathematics scores

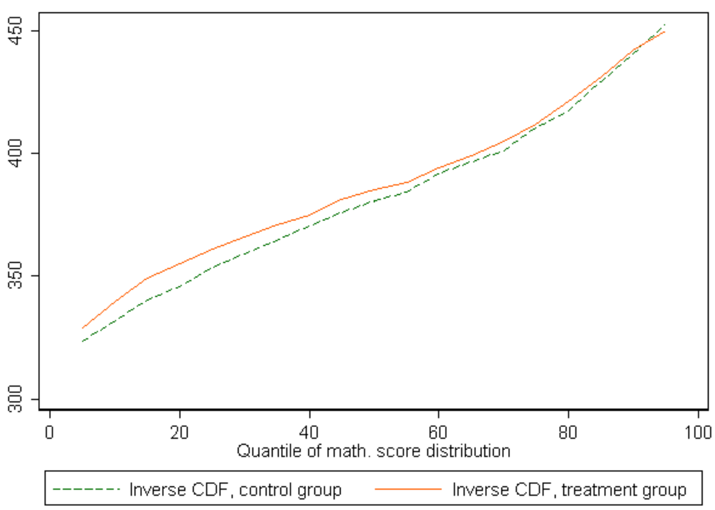

Figure 3 QTE for scaled mathematics CAHSEE scores