The Distributional Effects of Curricular Intensification

Andrew M. Penner

Thad Domina

Emily Penner

AnneMarie Conley

The Distributional Effects of Curricular Intensification

Abstract: Current educational policies in the United States attempt to boost student achievement and promote equality by intensifying the curriculum and exposing students to more advanced coursework. This paper investigates the relationship between one such effort -- California's push to enroll all 8th grade students in Algebra -- and the distribution of student achievement. We build on previous research that examines the average effects of curricular intensification policies by examining differences in inverse propensity score weighted distributions to investigate how these policies changed the distribution of student achievement more broadly. We find that California's attempt to intensify the curriculum did not raise test scores at the bottom of the distribution, but did lower scores at the top of the distribution. These results highlight the efficacy of inverse propensity score weighting approaches to provide unconditional differences across the distribution that nonetheless adjust for differences in observable characteristics between different groups, and provide a cautionary tale for curricular intensification efforts.

JEL classification: I24

1. Introduction

Education policies in the U.S. typically focus on two broad goals: (1) Increasing student learning and (2) Decreasing inequality between students. A wide range of contemporary educational policies pursue these dual goals by exposing all public school students to a standardized and rigorous academic curriculum. Policies such as academic standards, test-based accountability, and high-stakes requirements for grade promotion and graduation all attempt to raise student achievement levels by exposing students to higher levels of academic course content. At the same time, these policies attempt to narrow educational inequalities by insuring that a wide range of students have the opportunity to learn the same material. In this study, we evaluate the distributional consequences of California's ambitious effort to improve high school mathematics achievement and narrow achievement inequalities by standardizing middle school mathematics curricula. This effort culminated in 2008, when the state attempted to require all 8th graders to enroll in Algebra.

Educational research typically focuses on differences in the average test scores of students, examining, for example, the difference between the average test score of students who were exposed to higher curricular standards and students who were not. Such analyses are informative about the effects of such policies on the average level of student achievement. However, they potentially miss important information about policy effects on the distribution of student achievement more broadly. That is, we might imagine that enrolling all 8th grade students in Algebra could have countervailing effects on the top and bottom of the achievement distribution. Students at the bottom of the distribution might fare better, as they are now enrolled in Algebra instead of pre-Algebra, and have more opportunities to interact with higher achieving peers. By contrast, students at the top of the distribution might fare worse, as the Algebra class that they were in now has a greater number of low achieving students, and teachers in these classes may struggle to teach Algebra to a more heterogeneous class than they are used to (cf. Duflo, Dupas, and Kremer 2011). While this might mean that these policies have no effect on average, or even a negative average effect (if the negative effects at the top outweighed the positive effects at the bottom), such results would indicate that the policies were successful in decreasing inequality, even if they had no or negative effects on average.

To investigate the effects of curricular intensification on the distribution of student achievement more broadly, we calculate differences between quantiles of the inverse propensity score weighted distributions of scores of students who were and were not exposed to schools that had intensified their 8th grade Algebra policy.[1] Intuitively, this can be thought of as providing information about the difference between the p^{th} percentile score of students who were exposed to the policy and the p^{th} percentile score of students who were not exposed to the policy.

While distributional approaches have a relatively long history in economics (e.g. Koenker and Bassett 1978; Buchinsky 1994), they have only recently begun to be applied in the fields of sociology and education (e.g., Penner and Paret 2008; Grodsky, Warren, and Kalogrides 2009; Bitler, Domina, and Penner 2013). One explanation for the underutilization of distributional approaches lies in the difficulty in understanding how to interpret conditional and unconditional

[1] It is important to note that we use the term "effect" somewhat loosely, as we can only match on observable characteristics, so that our identification of true causal effects hinges on students in schools that have implemented these policies having similar unobservable characteristics as students in schools that have not. While we believe that this is plausible given our covariates, it is of course possible that this is not the case, and to the degree that there is selection on unobservable characteristics our results may not represent causal effects.

quantile effects. In an effort to address issues associated with non-random selection into treatment conditions, social scientists typically attempt to estimate the relationship between of educational interventions and educational outcomes, controlling for a wide range of student characteristics, including demographics and prior achievement measures. While this approach greatly increases the utility of observational data, it introduces interpretive challenges in the context of quantile regression, where substantial translation is necessary to get the unconditional quantile treatment effect from an estimate that is conditional on control variables.

This problem is summarized succinctly by Firpo, Fortin and Lemieux (2007), who note that "existing methods cannot be used to answer a question as simple as 'what is the impact on median earnings of increasing everybody's education by one year, holding everything else constant?'" (pg. 1). However, even many leading researchers often discuss their results on conditional quantiles in ways that might be interpreted as pertaining to unconditional quantiles, which likely adds to confusion around correct interpretation (cf. Budig and Hodges 2010; Grodsky, Warren, and Kalogrides 2009; Hao and Naiman 2007; Konstantopoulos and Li 2012; McGuiness and Bennett 2007; Penner 2008; Philips 2011). This distinction between conditional and unconditional quantiles is potentially important, as Firpo et al. (2007) show that the effect of union membership on log wages is positive at the conditional 90th percentile, but negative effect at the unconditional 90th percentile. To address this, Firpo et al. propose a recentered influence function that allows users to recover effects on the marginal distribution from conditional quantile regression models. This provides a much needed tool that allows researchers to condition on important covariates while retaining results that are interpretable from the perspective of the larger distribution.

Here we take a different approach, drawing on work by Firpo (2007) that highlights the promise of propensity score weighting to provide more easily interpretable estimates of how two marginal distributions differ while still accounting for underlying differences on other covariates. Propensity score based methods have grown increasingly popular in the social sciences as a means of accounting for selection on observables in non-experimental settings. Like regression-based approaches to causal effects estimation, propensity score approaches separate the relationship between outcomes and treatment variables from the potentially confounding relationship between other observable characteristics and treatment odds. However, these two approaches proceed in very different ways. Regression approaches condition estimates of the relationship between treatment and outcome across a population for observable covariates. On the other hand, propensity score weighting models the observable factors that predict treatment, and then focuses the analysis on cases with similar likelihoods of treatment participation. This approach has two primary advantages: First, inverse propensity score weighting minimizes the importance of cases outside of the area of common support, so that only cases that could plausibly be in either treatment or control influence estimates. Second, propensity weighting can yield estimates of the relationship between treatment and outcome that are unconditional. This aspect of propensity weighted estimates considerably eases their interpretation, particularly in the context of distributional analyses.

An additional advantage that a propensity score weighting approach has over a regression control approach is that we can easily use either the treatment or control distributions as our baseline. That is, in addition to using propensity score weights to estimate treatment effects in the population, we can also weight the control group to be similar to the treatment group (which provides information about how the 1st percentile treatment score differs from what the 1st percentile control group score would be if the control group was similar on observables to treatment). Alternatively, we can weight the treatment group to look like the control group's observed distribution (which allows us to see the effects using the control group as the basis for the percentiles). Substantively, this distinction is important as by using these two different distributions as the reference these approaches provide answers to related but analytically distinct questions, both of which are potentially of interest. In the context of policies designed to enroll more students in early Algebra, one can think of "Algebra for All" schools as the treatment group and schools that enroll some students in grade-level math and others in Algebra as a control. Using the distribution of students in Algebra for All schools as a baseline provides an estimate of how the implementation of Algebra for All policy affected the students in the schools that implemented the policy—that is, the effect of the treatment (being enrolled in an Algebra for All school) on the treated (the students who were in Algebra for All schools). However, it is also potentially interesting to estimate what the effect would have been if control schools had become Algebra for All schools (treatment on the untreated). Thus, in addition to highlighting the efficacy of distributional approaches for understanding how the effects of policies might vary at the top and bottom of the distribution, we also provide an example of how inverse propensity score weighting can be used to in conjunction with distributional analyses to obtain easily interpretable estimates of policy effects from observational data.

2. Curricular intensification policy background, theory and research

California's effort to intensify middle school mathematics curricula dates to at least 1987, when the California Superintendent of Public Schools argued that establishing a common set of standards in middle school math was central to raising student achievement. In 1992, the state board of education formalized goal of "heterogeneous grouping and detracking" and common standards; and in 1997 the state's revised standards called for all 8th graders to be enrolled in Algebra I. The standards had their intended effect, and the proportion of 8th graders enrolled in Algebra in California more than tripled between 1999 and 2008, rising from 16 percent to 51 percent (Rosin, Barondess, Leichty 2009). From this perspective California's 8th grade Algebra for all policy was a resounding success, exposing a larger proportion of students to higher level mathematics in middle school. Further, these gains put students on a trajectory of higher level mathematics coursetaking that persisted through high school (Domina, Penner, Penner and Conley 2013). In 2008, the state's Board of Education attempted to capitalize on these gains by making Algebra the "sole test of record" for the state's 8th graders. The state never fully implemented this mandate, opting instead to adopt the national Common Core state standards. However, California's revised middle school math standards continue to encourage schools to enroll all 8th graders in Algebra (Wurman & Evers 2011).

California's effort to universalize 8th grade Algebra entails two major changes for middle and high schools. First, and most obviously, it involves exposing more students to relatively advanced Algebra mathematics concepts in the 8th grade. Second, the 8th grade Algebra push also likely precipitates important changes in the skills-composition of 8th grade mathematics classrooms, moving low-performing students from pre-Algebra or less advanced 8th grade math courses to 8th grade Algebra courses that were once reserved exclusively for relatively highskilled students. In effect, therefore, this policy aims to detrack mathematics instruction in California middle schools. To understand this change and its potential implications, therefore it is useful to review the literature related to course-taking patterns in secondary school as well as the broader literature on school tracking.

We begin by reviewing research related to high school course-taking. Much of this research has focused on middle and high school mathematics, where nearly all high schools send students through a sequence of academic mathematics courses that begins with Algebra I and concludes with Calculus. In many schools, 8^{th} grade is the first point at which student trajectories through this math sequence diverge, with relatively advanced students taking 8^{th} grade Algebra and less advanced students taking pre-Algebra coursework. These early placement decisions have important consequences for students. Students who take Algebra early complete more – and more advanced – high school mathematics courses than their peers, even after controlling for a broad array of background characteristics (Gamoran & Hannigan 2000). Furthermore, mathematics course-taking is a strong predictor of mathematics learning and achievement, as well as postsecondary educational attainment (Schiller & Muller 2003; Attewell & Domina 2008; Long, Conger, & Iatorola 2012). Taken together, this research tradition provides strong evidence to suggest that policy efforts to enroll more 8^{th} graders in Algebra ought to have positive average effects on student achievement.

However, one major limitation in this correlational research literature involves selection bias. Students self-select and are placed into courses based on a wide variety of observable and unobservable characteristics. Many of the studies referenced above employ a rich set of covariates to separate the effects of these student characteristics from the effects of course placement. Nonetheless, it seems likely that unobserved selection biases these estimates of the effects of course-taking. We know of two studies that address these biases using experimental or quasi-experimental methods. Heppen et al. (2012) analyze data from a randomized controlled trial in which high-achieving 8th graders in 68 randomly-selected small, rural middle schools were offered access to an online Algebra course. In this case, access to online Algebra had a moderate positive effect on these high-achieving students' Algebra achievement as measured at the end of 8th grade (effect size=0.39), as well as their subsequent high school math course-taking. While encouraging, it is not clear whether or not these experimental findings generalize to settings in which 8th grade Algebra is traditionally administered.

On the other hand Clotfelter, Ladd, & Vigdor (2012a, 2012b) provide quasi-experimental evidence suggesting that accelerated Algebra has a negative effect on achievement. In a series of instrumental variable analyses that take advantage of rapid curricular intensification in 10 North Carolina school districts, Clotfelter, Ladd & Vigdor (2012a, 2012b) find that students who enroll in 8th grade Algebra are more likely to pass Algebra by the end of 10th grade, but are less likely to complete Geometry, and 8th grade Algebra enrollment depresses student performance on end-of-course Algebra I tests. Interestingly, their analyses indicate that these negative effects are most pronounced for low-performing students who are placed into Algebra, suggesting that students must have a sufficiently high level of baseline skills to benefit from Algebra instruction. These findings raise important questions about the effectiveness of policy efforts – like those undertaken in California – to universalize 8th grade Algebra. For our purpose, the evidence regarding pronounced negative effects of Algebra for low-achieving students is particularly troubling. This evidence indicates that 8th grade Algebra efforts fail to raise average achievement even as they exacerbate educational inequality.

A second line of research that is closely related to work on coursetaking considers the effects of tracking – the practice of assigning students to skills-homogenous ability groups for targeted instruction – on the distribution of student achievement. Policy efforts like California's 8th grade Algebra push do more than change a handful of students' course-enrollment patterns. Rather, by eliminating low-level courses and integrating students who once would have taken these courses into more advanced classrooms, these policies likely cause middle schools to assign a more skills-heterogeneous group of students to middle school mathematics courses. These changes in classroom composition may have independent consequences on student learning (Nomi 2012), as well as effects on teacher instructional content and methods (McPartland & Schneider, 1996; Rosenbaum 1999).

Despite continuing interest in the effects of tracking on the distribution of student achievement, relatively few studies have directly investigated this question. In a classic meta-analysis, Slavin (1990) focuses exclusively on studies that estimate the effect of enrolling in a course with a homogeneous group of students (i.e. tracking) compared to enrolling in a heterogeneously-grouped course. Slavin finds that tracking has no effect on achievement. Furthermore, he finds no evidence to suggest that the effects of tracking vary with student skills.

However, a handful of studies suggest that tracking may have uneven effects across the distribution of student achievement. Kerckhoff (1986) compares British secondary students enrolled in tracked and untracked schools. While his analysis suggests that tracking has no average effect on student achievement, he finds that attendance in a tracked school is associated with relatively rapid test score growth for high-achieving students and slower test score growth for low-achieving students. This differential effect is particularly pronounced in mathematics. Argys, Rees, & Brewer (1996) find evidence of a similar pattern among U.S. high school students. In this analysis, the authors categorize students' educational environments based on their teachers' characterization of the skills-composition of students in high school mathematics classes (above average, average, below average, or heterogeneous.) Argys, Rees, & Brewer's analyses indicate that students with relatively low levels of skill benefit from enrolling in skills-heterogeneous math classes. In contrast, they find that high-performing students learn less in skills-heterogeneous math classes than they might have in tracked math classes. While these two studies operationalize tracking somewhat differently, both indicate that tracking exacerbates educational inequalities.

Since policy efforts like California's 8th grade Algebra-for-all push change both student-level course enrollment patterns *and* course-level composition, their effects are somewhat unpredictable (Stein, Kaufman, Sherman, and Hillen 2011). Allensworth et al. (2009) find no evidence to suggest that a Chicago Public Schools effort to enroll all 9th graders in Algebra I and college prep English improved student achievement, graduation rates, or college-going. While difference-in-difference analyses suggest that the "double-dose" Algebra curriculum that Chicago implemented as a part of this effort was effective for low-achieving students (Nomi & Allensworth 2009); Nomi (2012) finds that curricular intensification in Chicago had unintended negative effects for high-achieving students. Furthermore, preliminary evidence from California is similarly discouraging. Descriptive analyses of state-wide data (Liang, Heckman, & Abedi, 2012) as well as district case-study data (Domina et al., 2012) indicate that California students placed in 8th grade Algebra learn no more than similarly-skilled students placed into pre-Algebra.

How should we understand the growing body of research from curricular intensification policy interventions that finds iatrogenic effects on student achievement in light of previous research suggesting that heterogeneous ability grouping and the exposure to more advanced material increase student achievement? We suggest that it is an important example of the difference between general and partial equilibria. Much of the work demonstrating the benefits of advanced course taking and curricular intensification does so in a context where the only thing that is changing is whether a given individual is placed into a higher level course. This framework approaches the question of course placement from a partial equilibrium perspective, so that everything else assumed to remain constant: the peers in the course, the teacher and their level of preparation to teach the course, the social meaning of the course, and other factors that could affect achievement do not change, allowing researchers to understand what would happen if a counterfactual person in an identical world was (or was not) exposed to the advanced course. We suggest that policies as implemented, however, rarely operate in this partial equilibrium, but rather that when a policy like California's 8th grade Algebra for All policy is implemented, it changes not just whether a given individual receives access to the advanced coursework, but also many other things, such as the teachers and peers that an individual is likely to encounter in Algebra. Put differently, we suspect that Algebra means something different in the Algebra for All schools than in the Baseline schools. As such, from a policy perspective we believe it is important to understand not just the effects of placing any given individual into Algebra ceteris paribus, but also the effects of implementing a broad-based Algebra-for-all policy on student achievement (or other outcomes of interest). This general equilibrium approach allows for the fact that curricular intensification policies change the broader dynamics of peer and teacher interactions.

3. Analytical approach and setting

In the analyses that follow, we explore the effects of curricular intensification in Towering Pines, a large, ethnically and socio-economically diverse public school district located in an inner-ring Southern California suburb. California is at the forefront of the national movement to intensify the curriculum, and Towering Pines was a district that sought to fully implement the state's Algebra for all policy. From 2004 to 2008, as the state as a whole increased the proportion of 8th graders enrolled in Algebra from 38 to 56 percent, Towering Pines increased from 32 percent to 84 percent. Importantly, we see that these gains persist beyond 8th grade, so that it is not the case that the 8th graders pushed into higher level math fail yielding no benefits to subsequent coursetaking.[2] Further, the district was thoughtful in implementing Algebra for All, and sought to prepare students for the advanced mathematics courses they would be taking, as can be seen in rising test scores in 6th grade mathematics. They also allowed schools to vary the timing of this transition, rather than forcing all schools to make the transition at the same time. This suggests that Towering Pines is in many ways a best case scenario for evaluating what kinds of effects these policies will have when implemented by a school district.

[2] There is some evidence that some Towering Pines schools relabeled classes from pre-Algebra to the first year of a two year Algebra sequence. As students in these classes would not be on track to complete calculus by 12th grade, and as they did not count as being in Algebra according to the state's accountability system, we do not consider these students as being enrolled in 8th grade Algebra for the purposes of our analyses.

The 10 middle schools in Towering Pines together enroll approximately 4,000 eighth graders each year. More than fifty percent of the district's 8th graders are Latino, approximately 25 percent are Vietnamese, and approximately 15 percent are white. Most of the remaining students are Asian and 1 percent of the students in the district are African American. Over 60 percent of the students in the district were English-language learners when they enrolled in school, and while a large proportion of these students had been reclassified as English-proficient by the time they were 8th graders, more than a third of the sample remained classified as English Language Learners (ELLs) in their 8th grade year. This sample is clearly not representative of 8th graders nationwide or statewide, and it is difficult to know whether the results of the Towering Pines case study are generalizable. However, the district's ethnic, economic, and linguistic diversity makes it a rich research site, especially since students of color and English-language learners are frequently excluded from high-level courses. Descriptive statistics are presented in Table 1.

[Insert Table 1 about here.]

Our key independent variable is the degree to which a student was exposed to curricular intensification. Rather than conceptualize this as a continuous treatment (e.g. using the percent of students in a school who were in 8th grade Algebra) or a dichotomous treatment (intensified curriculum vs. not), given that Figure 1 reveals a trimodal distribution of the percent of students in a school who were enrolled in Algebra we examine how students were affected by being in schools falling into one of three treatment categories. The first category, which we refer to as the Baseline Schools, contains students in schools where less than 46 percent of students are in Algebra or higher; the second, which we label the Transition Schools, contains students in schools ranging from 46 percent to 74 percent in Algebra or higher; and the final group, which we call the Algebra for All Schools, contains greater than 74 percent of students in Algebra or higher.

[Insert Figure 1 about here.]

Table 2 provides information on the rate of the curricular intensification at the 10 different middle schools in the district from 2004 through 2008. In addition to listing the percent of students who were in Algebra or higher, we also shade each school according to whether it is a Baseline, Transition, or Algebra for All school. We see that both the initial rates of 8th grade Algebra placement and the rates at which placement intensified varied across the schools in Towering Pines. For example, we see that school 4 is a baseline school in 2005, a Transition school in 2006, and an Algebra for All school in 2007 and 2008, while school 5 remains a baseline school for an additional year. The starting points also vary: in 2005 school 1 has 25 percent of students in Algebra or higher, and eventually becomes a Transition school, while school 8 was already a Transition school (with 49 percent of students In Algebra or higher) in 2005. Overall, we see that all schools enrolled a higher percentage of their students in Algebra in 2008 than they did in 2004, and that there were no Algebra for All schools in 2004 and no Baseline schools in 2008.

[Insert Table 2 about here.]

We have detailed administrative data for all students who enrolled in 8th grade in Towering Pines between 2004-05 and 2007-08. Because students take different tests depending on the course that they are enrolled in, we cannot examine the gap in mathematics achievement between the 8th

graders who did and did not take Algebra, and instead examine the effects of the level of curricular intensification by looking at 10th grade achievement on the California state exit exam (CAHSEE) taken by all students. This exam is designed to test student mastery of basic mathematics skills, and is administered to all students for the first time in 10th grade. To ease interpretation, we create a z-score based on the CAHSEE, so that the differences observed can be interpreted in standard deviation units. As the purpose of the CAHSEE is to establish a basic level of competency, there are some ceiling effects which preclude an examination of the effect for the top 10 percentiles. Given that the coursetaking gains we find from placing students in higher level mathematics courses in 8th grade persist through 10th grade, examining achievement in 10th grade allows us to assess how the policy's success in changing student coursetaking trajectories affects their longer term achievement.

While the demographic characteristics of the students in Towering Pines did not change substantially over the study period as schools intensified their curricula and there few differences demographic differences between Baseline, Transition, and Algebra for all schools, there were marked gains in prior achievement in both mathematics and English Language Arts (ELA).[3] We account for differences between Baseline, Transition, and Algebra for All schools by creating inverse propensity score weights based on the likelihood of being in these three categories (Imbens 2000). To do so we first estimate a multinomial logistic regression model predicting student odds of enrollment in the three school categories for all Towering Pine 8th graders in the 2004-05, 2005-06, 2006-07, and 2007-08 school years, based on their 6th grade math

[3] Prior achievement is operationalized using 6th grade score for mathematics achievement, because not all students took the same test in 7th grade. For ELA, all students took the same test in 7th grade, and so we use 7th grade scores instead.

achievement, 7th grade ELA achievement demographic characteristics, and interactions between demographic characteristics and baseline achievement.[4] We then use this model to generate predicted probabilities of attending Baseline, Transition, and Algebra for All schools for each student. We use these predicted probabilities to generate inverse propensity score weights which, following Imbens (2000) we define as the inverse of the conditional probability of being in a particular treatment category given the pre-treatment variables. More concretely, we use three sets of weights. For students at each category of treatment *t* (Baseline, Transition, and Algebra for All), we define our first inverse propensity score weight as:

$$W = {}^{1}/_{\widehat{p_t}} \qquad (1)$$

where $\widehat{p_t}$ is the predicted probability that a student receives the treatment he or she actually receives. This inverse propensity score weighting scheme balances treatment groups on observable characteristics by up-weighting students who actually received a given treatment but were unlikely to do so based on observable characteristics (and, conversely down-weighting students who were highly likely to receive the treatment they received). These weights use the overall sample of respondents as the population that they weight towards, and we refer to them as the population weights.

[4] Demographic characteristics include gender, race, and English language status. We sort students in to three language status categories: ELLs are students who entered the district with limited English language skills and have not demonstrated English-language proficiency by their 8th grade year; Reclassified Fluent English Proficient describes students who had limited English skills when they entered the district but who demonstrated proficiency before 8th grade; all other students are in the third category which includes native English speakers and students who were bilingual in English and another language upon district entry. We opted not to use a measure of free and reduced lunch status, as only 25 to 30 percent of students in a given year do not receive free and reduced lunch, and it is unclear whether those who do not are from higher SES families, undocumented and reticent to use services, or some combination of both.

We also calculate weights that weight respondents to look either like the Algebra for All or Baseline students by calculating the weights

$$W = \widehat{p_{t=AfA}} \Big/ \widehat{p_t} \qquad (2)$$

$$W = \widehat{p_{t=B}} \Big/ \widehat{p_t} \qquad (3)$$

where $\widehat{p_{t=AfA}}$ represents the predicted probability that a student was in an Algebra for All school and $\widehat{p_{t=B}}$ the predicted probability that a student was in a Baseline school. Thus, for example, in the equation (2), students in an Algebra for All school receive a weight of 1 (because for these students the numerator and denominator are identical), while the students in Baseline or Transition schools are weighted to more heavily if they have higher predicted probabilities of being in an Algebra for All. In weighting the students in Baseline and Transition schools to look more like the students in Algebra for All schools, we are using the Algebra for All distribution as our standard. This approach provides an estimate of the effect of treatment on the treated, as it tells us what the effect of Algebra for All was at different points in the Algebra for All distribution, if we weight our Baseline students to be similar to our Algebra for All students on observables. Likewise, using the weight from equation (3) uses the Baseline schools as the underlying standard, and follows the logic of estimates of treatment on the untreated. Intuitively, it can be helpful to think of this from a matching perspective; equation (1) is akin to using the area of common support, equation (2) is akin to matching Baseline and Transition students to Algebra for All students (i.e., finding control students who look like treatment students), and equation (3) is akin to matching Transition and Algebra for All students to students in Baseline schools (or matching treatment to control).

Figure 2 depicts the 6th grade mathematics achievement of students in Baseline, Transition, and Algebra for All schools unweighted (Figure 2a) and with the inverse propensity score weights weighting students towards the Algebra for All distribution (Figure 2b). We see that while the unweighted distributions vary considerably, applying the weights results in distributions that converge on the Algebra for All distribution.

[Insert Figure 2 about here.]

We can also check that the distributions are balanced by estimating quantiles of the inverse propensity score weighted test score distributions and comparing the 6th grade achievement of (1) students in Baseline schools to students in Transition schools, and (2) students in Baseline schools to students in Algebra for All schools. Our final estimates make the same set of comparisons for students' 10th grade mathematics scores. By estimating the differences at each percentile, we are comparing the value of the weighted first percentile score in the Baseline schools to the weighted first percentile score in the Algebra for All (or Transition) schools, and similarly for all other percentiles. Figures 3 and 4 present the differences between students in the Baseline schools and (1) students in the Transition schools (Figure 3) or (2) students in the Algebra for All schools (Figure 4), using inverse propensity score weights that weight respondents to look like the overall sample (Eq. 1). The x-axis represents the percentile at which the distributions are being compared, and the y-axis shows the difference between the two distributions of 6th grade test scores for the given point of comparison. The solid black line represents the point estimates, while the dashed grey lines represent the upper and lower bounds from bootstrapped confidence intervals. In both Figures 3 and 4 we find that the confidence intervals almost always include 0, so that when we use the inverse propensity score weights there are very minimal differences students in Baseline schools versus students in Transition or Algebra for All schools.

[Insert Figure 3 about here.]

[Insert Figure 4 about here.]

Since we lack achievement scores and course enrollment data for students who are not enrolled in Towering Pines schools, we examine only students who were enrolled in Towering Pines' schools for 6th grade and 10th grade. [5] To account for the fact that our students are nested in classrooms, we stratify on classrooms and bootstrap 999 replicates for our 95 percent confidence intervals.[6]

4. Results

Figures 5 and 6 present the differences between the 10th grade math achievement of students in Baseline schools and students in Transition schools (Figure 5) and Algebra for All schools (Figure 6). The x-axis represents the percentile at which the distributions are being compared, and the y-axis shows the difference between the two quantiles of distributions for the given point of comparison. The solid black line represents the point estimates, while the dashed grey lines represent the upper and lower bounds from bootstrapped confidence intervals. Figure 5, for example, shows that the median (50th percentile) score in Transition schools is roughly .2

[5] While the mathematics CSTs administered to 8th-12th graders are course-specific, all students in the 6th grade take the same grade-specific mathematics CST, as do most of the 7th graders. Because a small fraction of the 7th graders (roughly 15 percent) do take the Algebra I test, we control for 6th grade mathematics scores rather than 7th grade scores to ensure test uniformity. However, we use 7th grade ELA test score controls as all students take the same ELA test in every grade.

[6] Supplemental analyses using the 50 school-cohort units as strata yield similar results.

standard deviations lower than the median score in Baseline schools. As the confidence interval does not include 0, we conclude that the quantiles of the distributions of student achievement are statistically significantly different at this point. Overall, the pattern in Figure 5 suggests that there is no difference between the very bottom of the Transition and Baseline schools distributions, but that around the 35th percentile a statistically significant gap of about .1 standard deviations emerges. The difference between the two distributions fluctuates somewhat, and is largest near the median, where we see that students are scoring .2 standard deviations lower in Transition schools. The gap shrinks as we compare percentiles above the median, and we see that by the 70th percentile, there are no longer statistically significant differences between the achievement distributions of Transition and Baseline schools.

Turning next to Figure 6, where we compare Algebra for All schools to the Baseline Schools, we see a very different pattern of results. While there is no difference between the very bottom of the distribution in Algebra for All schools and Baseline schools, we find a surprisingly monotonic increase in the difference between the two distributions up until about the 60th percentile, so that students between the 60th and 85th percentiles in Algebra for All schools are scoring about a third of a standard deviation lower than students at the 60th to 85th percentiles in Baseline schools.

The lack of a gap at the very top of the distribution is driven by the ceiling effects on the test, as overall 12 percent earn the maximum score possible on the CAHSEE test. Supplemental analyses using logistic regression models to estimate the odds of earning the maximum score possible are 30 percent smaller among students in Algebra for All schools relative to those in Baseline schools, while the odds of students in Transition schools hitting the test ceiling were 18 percent lower than those in Baseline schools. This suggests that the results here may be conservative, and that in the absence of the CAHSEE test ceiling we might find larger differences at the top of the distribution, particularly when comparing Baseline and Algebra for All schools. However, it is important to note that one benefit of comparing the respective quantiles of the two distributions is that aside from the quantiles that are at the ceiling, the differences at the other percentiles are not affected by the test ceiling.

Overall, these results suggest that curricular intensification does not boost achievement at the bottom of the distribution, and that if anything students at the bottom of the distribution in Transition and Algebra for All schools score lower than the students at the bottom of the distribution in the Baseline schools. We do find evidence, however, that student achievement at the top of the distribution is lower in Transition and Algebra for All schools than in Baseline schools, and that in Algebra for All schools these differences grow increasingly large the towards the top of the distribution. Thus, while a priori we might have expected that the bottom of the achievement distribution would have been lifted in the Algebra for All schools, we find no evidence that this is the case. To the degree that Algebra for all schools are more equitable, it is precisely because they are less efficient; that is, there are no gains at the bottom that might offset the losses at the top of the distribution, so that if the distribution of student achievement is tighter, this is occurring solely through lowering achievement at the top of the distribution.

Figures 7 and 8 build on Figures 5 and 6 by reporting the same results using different weighting schemes. In Panel A of Figures 7 and 8 we present the differences from inverse propensity score weights that weight Baseline and Transition to look like the Algebra for All schools, while in Panel B we present results that weight Transition and Algebra for All schools to look like Baseline school on their observable characteristics. Overall, we see that the results are largely similar, but that the results in Panel A (which weight towards Algebra for All schools) provide are somewhat more negative than those in Panel B (which weight towards Baseline). The fact that the pattern of results does not vary substantially based on whether we are thinking about differences at the 25th percentile of the Baseline schools or the 25th percentile of the Algebra for All schools is reassuring. However, it is also important to note that the these two approaches do not provide identical answers, suggesting that researchers using this approach need to think carefully about whether they are interested in differences relative to the treated or untreated distributions.

5. Discussion

California is at the forefront of a national movement to enroll more students in more rigorous mathematics courses throughout secondary school, a movement in which enrolling 8th graders in Algebra plays a central role. While this effort has increased Algebra enrollments in California and around the nation, it is less clear if it has improved student achievement. Given that Algebra policy is not just about raising achievement across the distribution, but in particular is about changing the shape of the mathematics achievement distribution to make it more equitable, we argue that it is important to examine the effects of Algebra for All policies across the distribution of student achievement more broadly. However, unlike policies where large scale randomized control trials that allow for the estimation of Quantile Treatment Effects without adjusting for covariates, and as is often the case in social science research using observational data, we find important differences in observable characteristics at baseline that need to be accounted for in order to understand any potential effects that the policies might have had. As such, we also provide an empirical demonstration of how inverse propensity score weighting can be used to account for baseline differences and provide results that are more intuitively understandable than conditional quantiles (cf. Firpo 2007) .

The results that we find are quite striking: the gap between achievement in Algebra for All schools and Baseline schools is not favorable for Algebra for All schools at any point in the achievement distribution, and is increasingly unfavorable at the middle and top of the achievement distribution. The most optimistic interpretation of these findings that is warranted in our view is that these results evince a short term costs caused by institutional factors. That is, from an institutional perspective, we might expect that even changes that are beneficial in the long term might have iatrogenic short term effects, as the educational system that was in place is disrupted. Stigler (2009), for example, notes that educational structures in the United States facilitate suboptimal pedagogical practices, so that efforts to improve may do more harm than good when the broader system is not also changed to support the improvements. If such processes were at work here, we might expect that after the schools in Towering Pines have had a chance to adjust to the intensified curriculum, students may indeed fare better.

However, while this may be the case, the challenges associated with such changes should not be underestimated. Other research suggests that the negative effects of curricular intensification might be partially offset by attending to peer effects, increasing teacher preparation, and providing additional support to students who are struggling with the advanced material (Domina et al. 2012; Nomi 2012). Given the impending adoption of the Common Core throughout the United States, we suggest that more research on how the distribution of student achievement is affected by these policies is needed to assess whether curricular intensification policies are having their desired effects of increasing student learning while decreasing inequality. As research seeks to examine this and other questions in which it is important to understand not simply how average levels of achievement were affected, but how the broader distribution of achievement (or other outcomes) might have changed, we suggest that inverse propensity score weighted differences between quantiles offer a useful tool that allows researchers to examine the broader distribution of achievement, adjust for differences in observable characteristics, and provide intuitively interpretable results.

Works Cited

Allensworth, E., Nomi, T., Montgomery, N. & Lee, V. E. (2009). College- preparatory curriculum for all: The consequences of raising mathematics graduation requirements on students' course taking and outcomes in Chicago. *Educational Evaluation and Policy Analysis, 31*, 2, 367-391.

Argys, Laura M., Daniel I. Rees, Dominic J. Brewer. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management*, 15, 623–645.

Attewell, P. & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, 30, 1, 51-71.

Bitler, M., Domina, T., Penner, E. K., & Hoynes, H. (2013). Distributional Effects of a School Voucher Program: Evidence from New York City. NBER Working Paper No. 19271.

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica*, 405-458.

Budig, M.J. & Hodges, M. Differences in Disadvantage: How the Wage Penalty for Motherhood Varies Across Women's Earnings Distribution." *American Sociological Review* 75(5):705-28.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2012a). The aftermath of accelerating algebra: Evidence from a district policy initiative. *National Center for Analysis of Longitudinal Data in Education Research, Working Paper 69*, 1-47. Retrieved July 16, 2012, from www.caldercenter.org.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2012b). *Algebra for 8th Graders: Evidence on its Effects from 10 North Carolina Districts* (No. w18649). National Bureau of Economic Research.

Domina, T., Penner, A. M., Penner, E. K., & Conley, A. M. Forthcoming. "Algebra for all: California's 8th grade algebra initiative as constrained curricula." *Teachers College Record*.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of

Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review*, 101(5), 1739-74.

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

Firpo, S., Fortin, N.M., & Lemieux, T. (2007). Unconditional Quantile Regressions. NBER Technical Working Paper No. 339.

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, *22*(3), 241-254.

Grodsky, E., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971-2004. *Educational Policy*, 23(4), 589-614.

Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. Thousand Oaks, CA: SAGE Publications.

Heppen, J.B., Walters, K., Clements, M., Faria, A., Tobey, C., Sorensen, N., and Culp, K. (2012). *Access to Algebra I: The Effects of Online Mathematics for Grade 8 Students.* (NCEE 2012–4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706-710.

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review*, 842-858.

Koenker, R., & Bassett Jr., G., 1978. Regression quantiles. *Econometrica* 46, 33–50.

Konstantopoulos, S., & Li, W. (2012). Modeling Class Size Effects Across the Achievement Distribution. *International Journal of Sociology of Education*, 1(1), 5-26.

Liang, J., Heckman, P. E., & Abedi, J. (2012). What Do the California Standards Test Results

Reveal About the Movement Toward Eighth-Grade Algebra for All? *Educational Evaluation and Policy Analysis*.

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of High School Course-Taking on Secondary and Postsecondary Success. *American Educational Research Journal, 49*, 2, 285-322.

McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. *Economics of Education Review*, 26(5), 521-531.

McPartland, J. M., & Schneider, B. (1996). Opportunities to learn and student diversity: Prospects and pitfalls of a common core curriculum. *Sociology of education*, 66-81.

Nomi, T., & Allensworth, E. (2009). “Double-Dose” Algebra as an Alternative Strategy to Remediation: Effects on Students' Academic Outcomes. *Journal of Research on Educational Effectiveness*, *2*(2), 111-148.

Nomi, T., (2012). The unintended consequences of an algebra-for-all policy on high-skill students: Effects on instructional organization and students’ academic outcomes. *Educational Evaluation and Policy Analysis*. Published online before print, July 31, 2012.

Penner, Andrew M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology* 114: S138-S170.

Penner, Andrew M. and Marcel Paret. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research* 37: 239-253.

Phillips, D. J. (2011). Jazz and the Disconnected: City Structural Disconnectedness and the Emergence of a Jazz Canon, 1897–19331. *American Journal of Sociology*, 117(2), 420-483.

Rosenbaum, J.E. (1999). If Tracking Is Bad, Is Detracking Better? *American Educator*, 23, 24-29.

Rosin, M. S., Barondess, H., & Leichty, J. (2009). *Algebra policy in California: Great* *expectations and serious challenges*. Mountain View, CA: EdSource.

Schiller, K. S. & Muller, C. (2003). “Raising the bar and equity? Effects of high school graduation requirements and accountability practices on students’ mathematics course taking.” *Education Evaluation and Policy Analysis* 25: 299–318.

Slavin, Robert E. (1990). “Achievement effects of ability grouping in secondary schools: A bestevidence synthesis.” *Review of Educational Research* 60: 471-499.

Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research, 81*, 4, 453-492.

Stigler, J. W., & Hiebert, J. (2009). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: Free Press.

Wurman, Z., & Evers, W. M. (February 09, 2011). New Education Dashboard: Less Rigorous, Less Meaningful. *Education Week, 30*, 20.

Table 1: Descriptive statistics on 8th graders enrolled at Towering Pines Unified School District

	2004-2005	2005-2006	2006-2007	2007-2008
Gen Math in 8th grade (n)	2,414	1,867	988	612
(%)	64	50.79	25.48	15.65
Algebra in 8th grade (n)	1,216	1,501	2,438	2,773
(%)	32.24	40.83	62.87	70.9
Geometry in 8th grade (n)	142	308	452	526
(%)	3.76	8.38	11.66	13.45
ELL in 8th grade (n)	1,469	1,301	1,289	1,325
(%)	38.93	35.39	33.1	33.84
RFEP in 8th grade (n)	884	988	1,155	1,205
(%)	23.43	26.88	29.66	30.78
Eng only/FEP in 8th grade (n)	1,420	1,387	1,450	1,385
(%)	37.64	37.73	37.24	35.38
Hispanic (n)	1,939	1,852	2,075	2,111
(%)	51.39	50.38	53.29	53.92
Vietnamese (n)	834	867	921	963
(%)	22.1	23.59	23.65	24.6
White (n)	692	624	584	514
(%)	18.34	16.97	15	13.13
Other (n)	308	333	314	327
(%)	8.16	9.06	8.06	8.35
6th grade math score (mean)	-0.135	0.002	0.057	0.070
(sd)	(0.952)	(0.959)	(1.042)	(1.027)
7th grade ELA score (mean)	-0.146	-0.021	0.053	0.104
(sd)	(0.976)	(0.969)	(1.016)	(1.018)

Note: 6th grade math and 7th grade ELA scores are standardized across cohorts. Sample includes only students who were present in the 6th and and 10th grades.

Table 2. Percent of students in the 10 Towering Pines middle schools in Algebra or higher, by year

School	1	2	3	4	5	6	7	8	9	10	Total
2005	25	35	57	29	31	32	26	49	43	40	36
2006	39	39	58	51	40	52	60	60	54	46	49
2007	53	94	64	94	66	83	92	86	68	52	75
2008	71	94	68	95	88	79	81	80	81	100	84

Note: Unshaded cell indicates Baseline, light gray shading indicates Transition, and dark gray shading represents Algebra for All

Figure 1. Kernel density estimate of the proportion of students enrolled in algebra or higher in the students' school

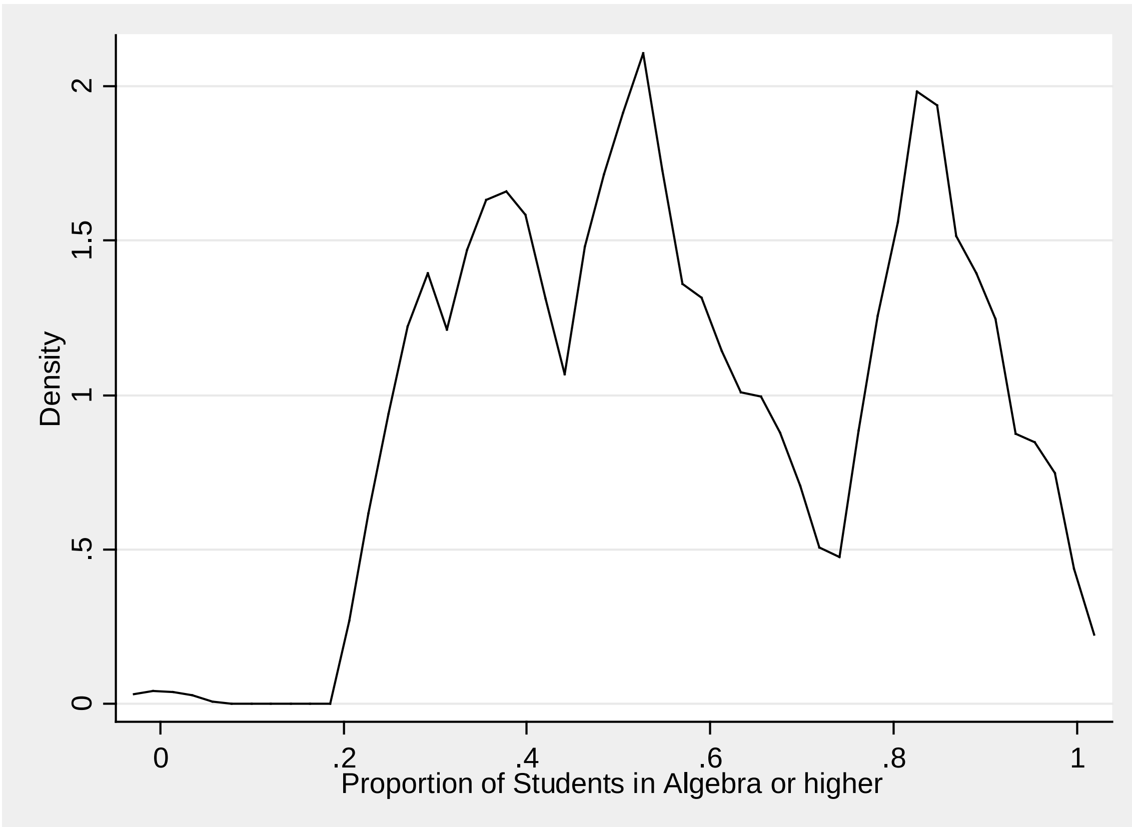
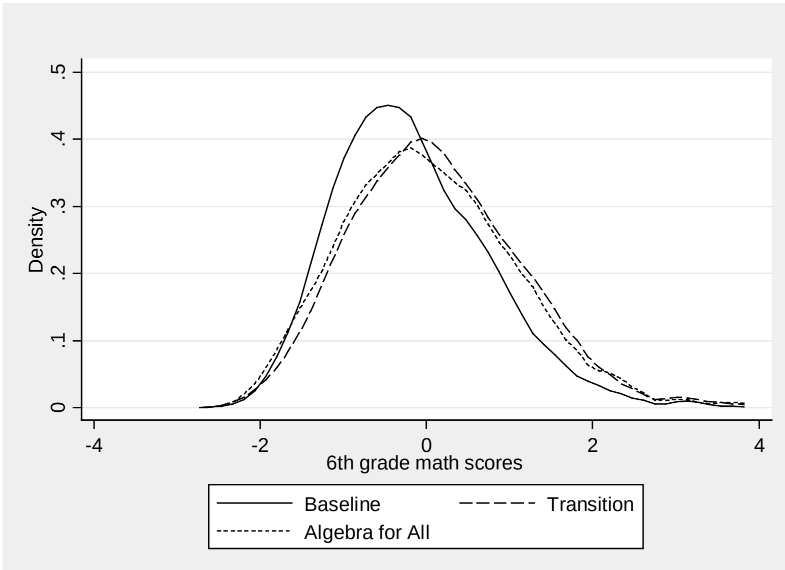


Figure 2. Distribution of 6th grade mathematics achievement in Baseline, Transition, and Algebra for All schools, unweighted and weighted towards Algebra for All

Panel A. Unweighted distributions



Panel B. Weighted towards Algebra for All distribution

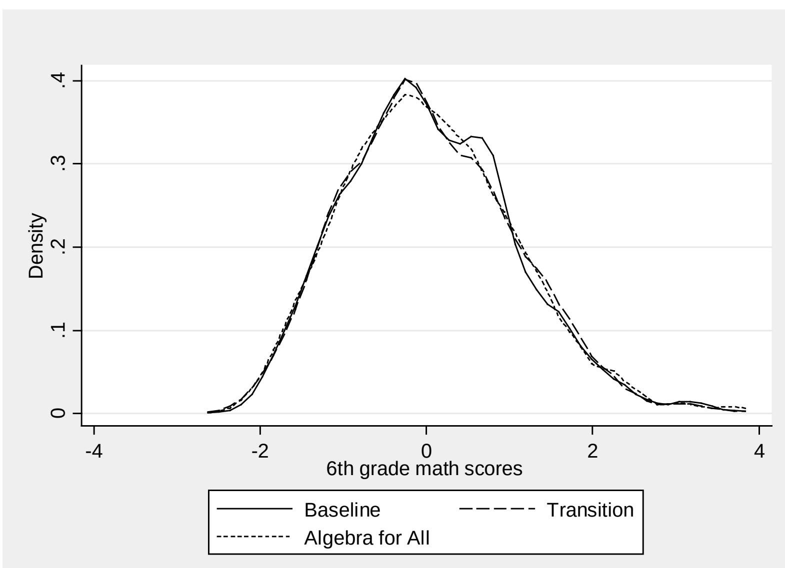


Figure 3. Differences between Transition schools and Baseline schools in 6th grade mathematics scores, using population inverse propensity score weights

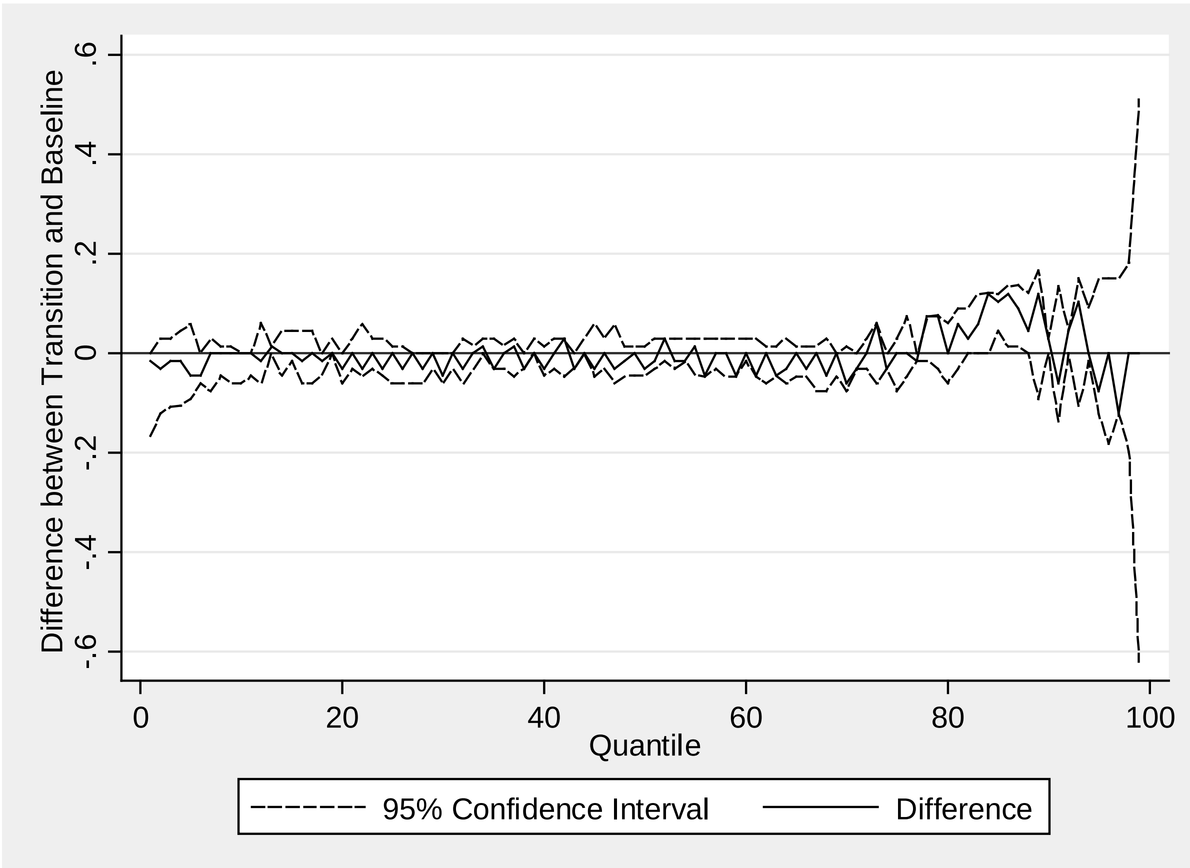


Figure 4. Differences between Algebra for All schools and Baseline schools in 6th grade mathematics scores, using population inverse propensity score weights

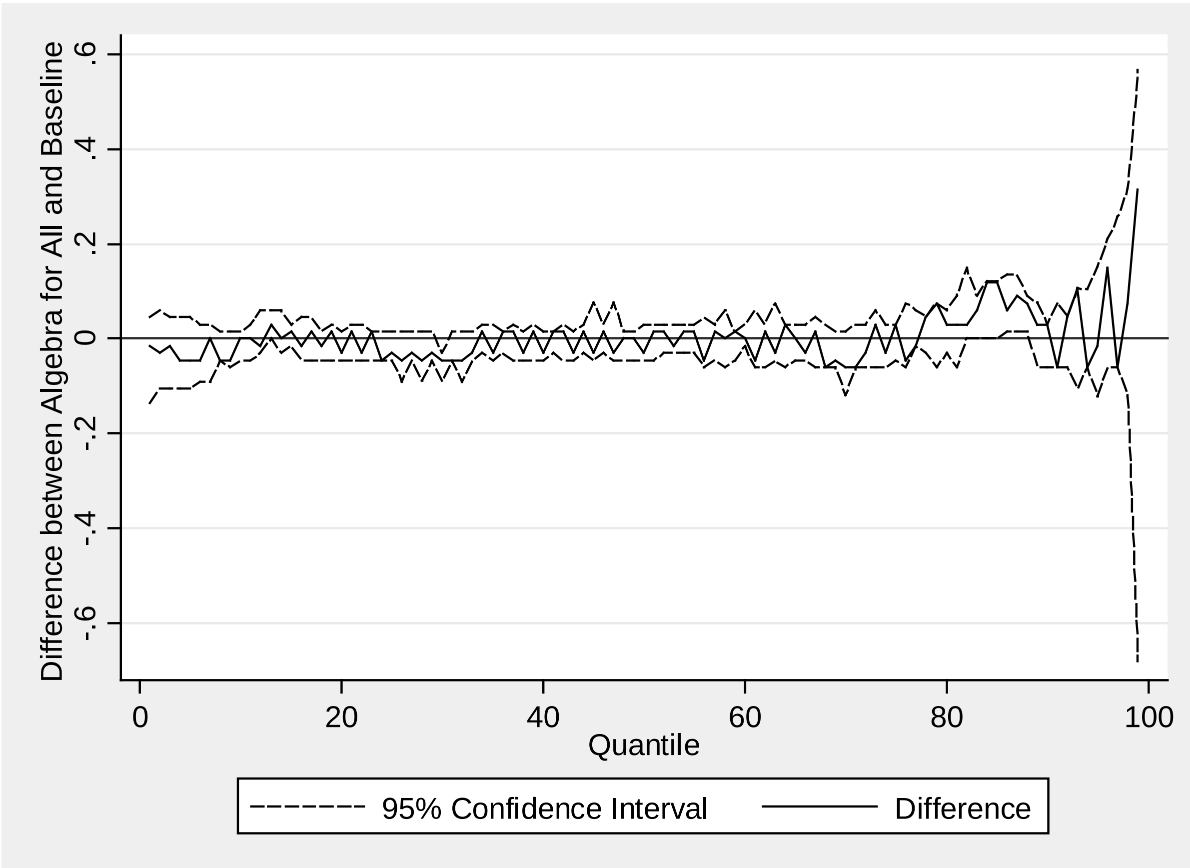


Figure 5. Differences between Transition schools and Baseline schools in 10th grade mathematics scores

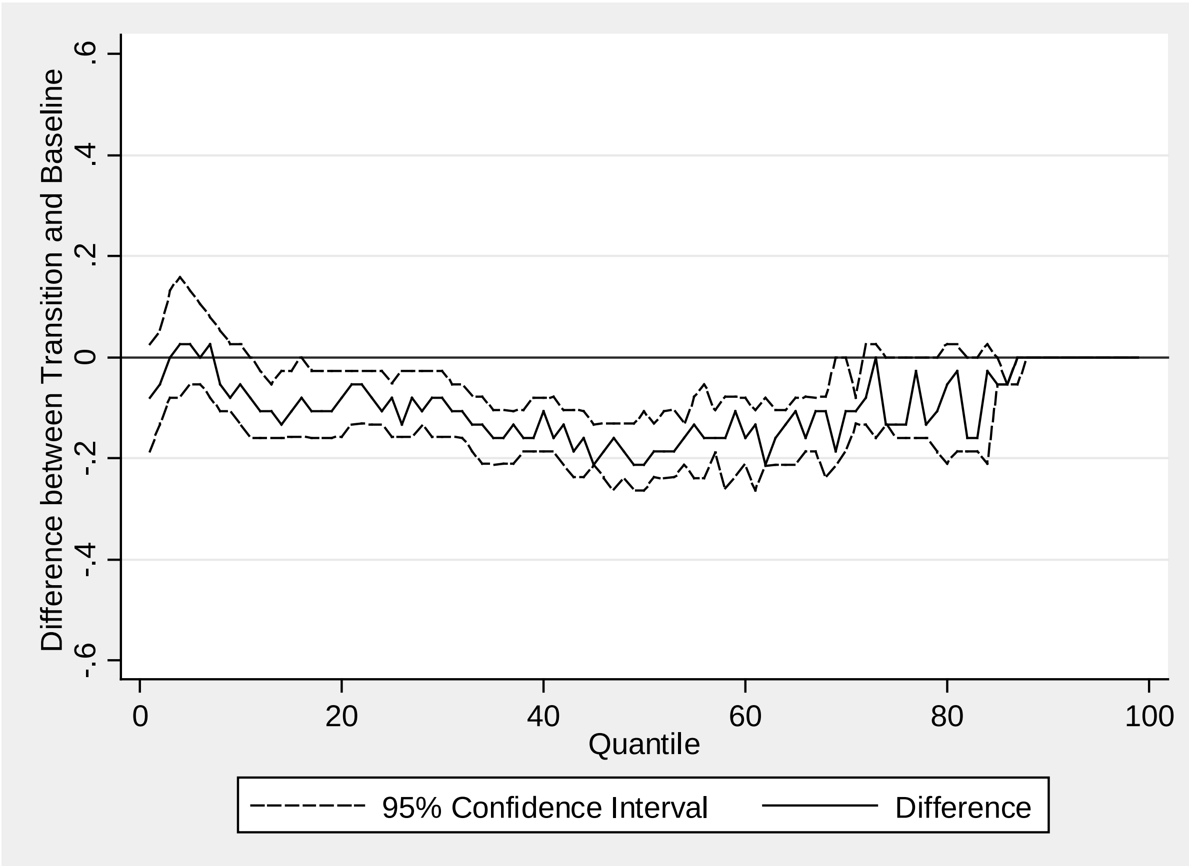


Figure 6. Differences between Algebra for All schools and Baseline schools in 10th grade mathematics scores

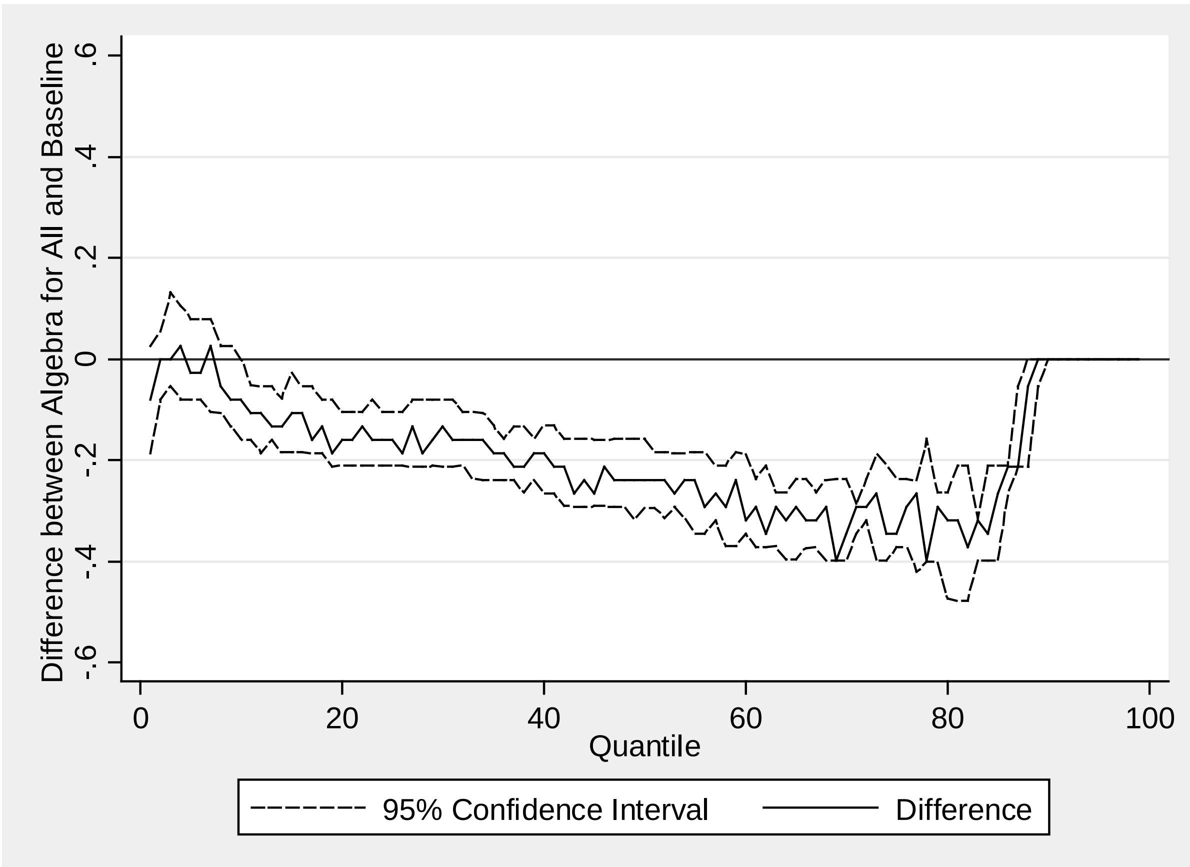
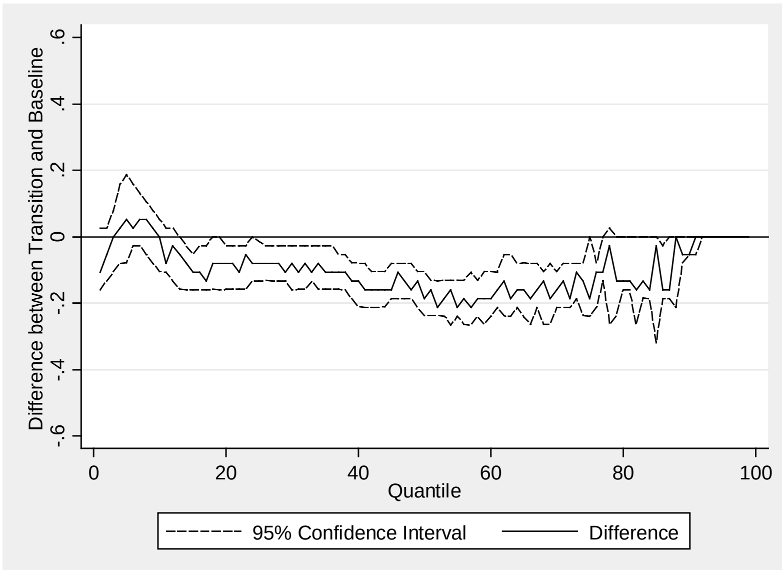


Figure 7. Differences between Transition schools and Baseline schools in 10th grade mathematics scores, weighted either towards Algebra for All or towards Baseline schools

Panel A. Weighted towards Baseline school distribution



Panel B. Weighted towards Algebra for All school distribution

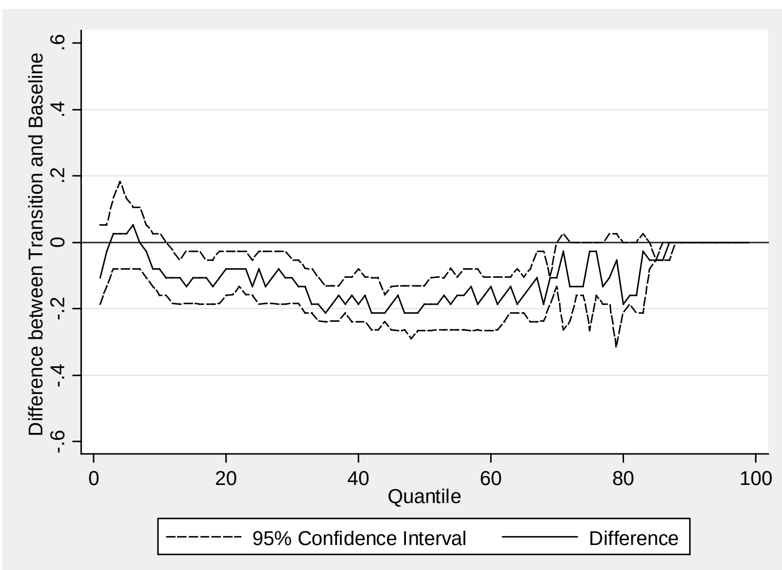
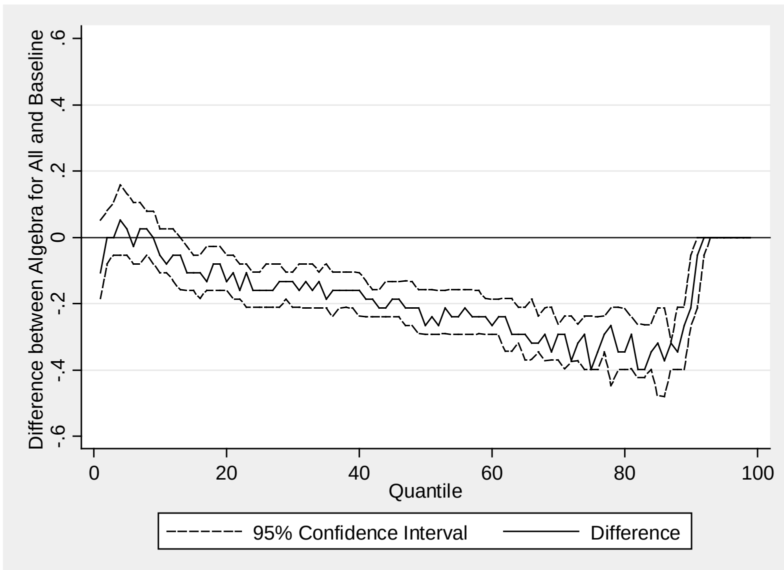


Figure 8. Differences between Algebra for All schools and Baseline schools in 10th grade mathematics scores, weighted either towards Algebra for All or towards Baseline schools

Panel A. Weighted towards Baseline school distribution



Panel B. Weighted towards Algebra for All school distribution