HHS Public Access

Author manuscript

Soc Sci Res. Author manuscript; available in PMC 2016 July 01.

Published in final edited form as:

Soc Sci Res. 2015 July ; 52: 627–641. doi:10.1016/j.ssresearch.2015.03.008.

Curricular Policy as a Collective Effects Problem: A Distributional Approach

Andrew M. Penner,
Department of Sociology, University of California, Irvine

Thurston Domina,
School of Education, University of California, Irvine

Emily K. Penner, and
Center for Education Policy Analysis, Stanford University

AnneMarie Conley
School of Education, University of California, Irvine

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 suggest that this effort is an instance of a “collective effects” problem, where the population-level effects of a policy are different from its effects at the individual level. In such contexts, we argue that it is important to consider broader population effects as well as the difference between “treated” and “untreated” individuals. To do so, we present differences in inverse propensity score weighted distributions to investigate how this curricular policy 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 for estimating collective effects, and provide a cautionary tale for curricular intensification efforts and other policies with collective effects.

1. Introduction

In the effort to develop an empirical base for social policy-making, scholars often draw upon a medical research model to identify the anticipated effects of different policy interventions. In idealized form this model proceeds in three steps: (1) Based on basic research and observational data, policy-makers or other social actors develop an intervention to address a documented social problem; (2) Evaluators test this intervention on a small scale, typically by comparing outcomes for individuals who are exposed to the intervention (“treated”) with those who are not (“control”); (3) Having demonstrated desirable effects in this experimental setting, policymakers design policy to mandate or facilitate the intervention's

Direct correspondence to: Andrew Penner, University of California, Irvine, Irvine, CA 92697-5100; andrew.penner@uci.edu.

adoption at scale. While this design and validation model holds great promise for improving the evidence base of social policy, several scholars have noted that the effects of social policies implemented at scale are often very different from the effects observed for the same interventions in small-scale demonstration projects (Dodge 2011; Welsh, Sullivan and Olds 2010).

In this paper, we consider one such example: Based on evidence indicating that students benefit when they take advanced courses (c.f. Domina 2014; Heppen et al. 2012; Long et al. 2012), California public schools dramatically expanded 8th grade Algebra enrollments between 2005 and 2010. Our analyses, reported here and elsewhere (Domina, McEachin, Penner and Penner 2014; Domina, Penner, Penner, and Conley 2014) indicate this this policy effort was counter-productive. We introduce the concept of "collective effects" in an attempt to explain this disconnect. We argue that most evaluation research that informs policy-making focuses on the effects of interventions on individuals. But most social policies affect not just individuals, but also schools, neighborhoods, and societies. Put simply, collective effects arise when the effect of a policy on a given individual diverges from the effects of the policy on the population at large.

A simple illustration encapsulates this insight: Standing up at a baseball game is likely to improve any given spectator's view. However, if every spectator in the stadium stands up at the same time, nobody's view is likely to improve appreciably. In other words, the observation that standing improves views at the individual level is insufficient for estimating the effects on a policy requiring all spectators to stand up. Collective effects exist in many domains. Thus, while we often analyze social policies from a partial equilibrium perspective, holding everything in the model constant while shifting a single parameter, a general equilibrium model is likely to be more appropriate, since policies often lead to large-scale changes in the access to given interventions (cf. Lise, Seitz, and Smith 2004). Put differently, while ceteris paribus is a helpful concept for understanding individual effects, when policies are put into place at the population level many things change.

More technically, one can view collective effects as suggesting that for many social policy interventions, the stable unit treatment value assumption (SUTVA) for causal inference is unlikely to be met unless assignment to treatment occurs at the population level (e.g. schools instead of students; communities instead of individuals) so that the effect of the treatment is not affected by whether others were treated. However, we believe it is more helpful to think about individual and population level effects as being fundamentally different questions, and that it is only under certain conditions that they have the same answer. Since the effects of many social interventions spill-over across individuals, we suggest that estimates of an intervention's effect derived from settings in which a limited number of individuals are treated may be of limited value for understanding the intervention operating at scale.

To return to the baseball analogy, a stadium designer likely cares less about the view from each particular seat than the broader distribution of views. Likewise, when designing social policy, we argue that it makes sense to think about effects on the population broadly. Most analysts would argue in favor of adopting a policy that has desirable effects when implemented at the population level, even if complying with the policy had undesirable effects on an individual who complies with the policy in isolation. For example, in a world in which few drivers comply with traffic regulations, compliance might arguably be dangerous for any given driver. However, near-universal compliance with traffic regulations undoubtedly improves safety for all drivers, including the few who do not comply.

We further argue that in considering the population perspective, it is often helpful to think beyond average differences and consider the broader distribution of outcomes. Once again the baseball stadium analogy is useful. If the people who are most likely to stand when others at the stadium are sitting are the shortest (i.e., those who have the most to gain by standing), they are likely to lose the most if everyone stands, and their standing view may be substantially worse than if everyone (themselves included) were seated. However, one could imagine that the average view quality is the same regardless of whether people are sitting or standing, even though there is more inequality in views when people are standing. We thus argue that in thinking about these questions at the population levels, it is helpful to compare the entire distribution of outcomes, for example, by comparing each of the different percentiles of the distributions.[1]

In this paper, we develop the notion of collective effects as 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. Our analyses indicate that policy environment is a clear example of an instance in which individual effects and collective effect diverge, both at the average and across the distribution. In the discussion, we build on this insight to provide a preliminary typology of collective effects in educational and social policy settings.

2. Collective effects and curricular intensification

California's effort to universalize 8th grade Algebra culminated in 2008, when the state attempted to require all 8th graders to enroll in Algebra. This push to intensify the mathematics curriculum entails two major changes for schools. First, and most obviously, it involves exposing more students to relatively advanced Algebra concepts in the 8th grade. Second, the 8th grade Algebra push also 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 high-skilled students. In effect, therefore, this policy aims to detrack mathematics instruction in California middle schools. To understand this change and its potential implications, it is therefore useful to review the literature related to course-taking patterns in secondary school as well as the broader literature on school tracking.

A great deal of research examines the consequences of course enrollment in middle and high school mathematics, where nearly all students are exposed to a sequence of courses that

[1]We can also think about the distribution of individual effects. In this analogy, this would amount to examining how much each individual's view would change if only that particular spectator stood up, and examining the distribution of these changes (ignoring how this would affect the views of those seated behind the spectator). While distributional approaches can provide information about effects at the individual or population levels (like mean differences), we suggest that thinking about effects at the population level lends itself to thinking about how the broader distribution of outcomes changes.

begins with Algebra I and concludes with Calculus. In many schools, 8th grade is the first point at which student trajectories through this math sequence diverge, with relatively advanced students taking 8th 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 and Hannigan 2000). Furthermore, mathematics course-taking is a strong predictor of mathematics learning and achievement, as well as postsecondary educational attainment (Schiller and Muller 2003; Attewell and Domina 2008; Long, Conger, and Iatorola 2012; Domina 2014). Consistent with these findings, Heppen et al. (2012) provide evidence from a recent 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. Taken together, this research tradition provides strong evidence to suggest that policy efforts to enroll more 8th graders in Algebra ought to have positive average effects on student achievement.

However, several studies that evaluate large-scale curricular intensification efforts return considerably less encouraging results (see Stein, Kaufman, Sherman, and Hillen 2011 for a review). In a series of instrumental variable analyses that take advantage of rapid curricular intensification in 10 North Carolina school districts, Clotfelter, Ladd and Vigdor (2012, 2015) find that 8th grade Algebra enrollment has a detrimental effect on student achievement, particularly for low- performing students who are placed into Algebra. Similarly, 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 and Allensworth 2009), Nomi (2012) finds that curricular intensification in Chicago had unintended negative effects for high-achieving students.

The available evidence regarding California's 8th grade Algebra-for-all effort is similarly discouraging. Using statewide district panel data, Domina, McEachin, Penner, and Penner (2014) find that student achievement growth rates are lower in districts with more students enrolled in 8th grade Algebra. Likewise, in a case study of one large California school district, Domina, Penner, Penner, and Conley (2014) demonstrate that student achievement growth slowed for 8th graders enrolled in both pre-Algebra and Algebra courses as the district increased 8th grade Algebra enrollment rates over a short period of time.

If exposure to advanced courses increases learning for a broad range of students, why are the effects of curricular intensification policies like California's 8th grade Algebra-for-all efforts often negative? We suggest that the collective effects framework can help reconcile this apparent paradox. 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, asking what would happen if a counterfactual person in an identical world was (or was not) exposed to an advanced course. As such, these analyses hold constant many factors that one might expect to change in the event of a larger-scale policy change; including classroom peer composition, the teacher and his or her level of preparation to teach the course, and the social meaning of the course. Given that policies are not implemented in this partial equilibrium, 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).

3. Analytical approach

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. Drawing upon the idea of collective effects, we argue that Algebra-for-all policies might have very different effects across the distribution of student achievement. For example, we might imagine that enrolling all 8th grade students in Algebra could have countervailing effects on the top and bottom of the achievement distribution.

Universalizing 8th grade Algebra might raise the bottom of the achievement distribution by insuring that all students are exposed to more rigorous coursework and higher achieving peers. By contrast, the same policy effort might have negative effects at the top of the distribution if negative peer effects also operate and teachers in reorganized classes focus their attention on teaching struggling students in heterogeneous settings (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.

Setting

In the analyses that follow, we explore the effects of curricular intensification in Towering Pines, a large, diverse public school 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, this Southern California district increased Algebra enrollments from 32 percent to 84 percent.[2] These 8th grade course placements increased students' odds of taking advanced mathematics courses throughout high school. 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

[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.

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.

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 who are frequently excluded from high-level courses. Descriptive statistics are presented in Table 1.

Methods

While much policy analysis focuses closely on estimating policy effects on individuals who are exposed to policy “treatments,” attending to collective effects underscores the important ways in which policies might have larger, unanticipated consequences across a population. Furthermore, as the above baseball stadium example makes clear, collective effects can change the distribution of policy-relevant outcomes in important ways. To investigate the effects of curricular intensification on the distribution of student achievement, we therefore 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.[3] 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, Penner, and Hoynes forthcoming). One explanation for the underutilization of distributional approaches lies in the difficulty in understanding how to interpret conditional and unconditional 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 student demographics and prior achievement. While this approach greatly increases the utility of observational data, it introduces interpretive challenges in the

[3]It is important to note that we use the term “effect” loosely in this empirical context, 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.

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).[4] However, even many leading researchers often discuss their results on conditional quantiles in ways that could be interpreted as pertaining to unconditional quantiles, which likely adds to confusion around correct interpretation (cf. Addo and Lichter 2013; Budig and Hodges 2010; Grodsky, Warren, and Kalogrides 2009; Konstantopoulos and Li 2012; McGuiness and Bennett 2007; Penner 2008; Philips 2011). This distinction between conditional and unconditional quantiles is potentially important. For example, Firpo et al. (2007) show that the effect of union membership on log wages is positive at the conditional 90th percentile, but negative at the unconditional 90th percentile (see also Killewald and Bearak (2014) on the motherhood wage penalty).

To address this, Firpo (2007) 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. Regression approaches condition estimates of the relationship between treatment and outcome across a population for observable covariates, while propensity score weighting models the observable factors that predict treatment, and then focus the analysis on cases with similar likelihoods of treatment participation. Propensity score approaches also minimize 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. In the context of estimating distributional effects, however, propensity score weighting is helpful in that it readily yields estimates of the relationship between treatment and outcome that are unconditional (given unconfoundedness), considerably easing their interpretation.[5]

An additional benefit of the propensity score weighting 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 the treatment group).

[4]We focus here on observables, in order to interpret these differences as causal effects one would need to assume unconfoundedness (cf. Firpo, Fortin and Lemieux 2007). A later version of this paper (that does not include this quote) was published in *Econometrica* (Firpo, Fortin and Lemieux 2009).

[5]It is possible to recover unconditional estimates from conditional quantile regression models, however this requires additional work (cf. Machado and Mata 2005; Firpo, Fortin and Lemieux 2009; Chernozhukov, Fernández-Val, and Melly 2013).

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). 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, what is the effect of being enrolled in an Algebra for All school for the students who were in Algebra for All schools (compared to similar students who were in Baseline schools). However, it is also potentially interesting to estimate what the effect would have been if students in Baseline schools had been in Algebra for All schools, which is potentially a different question.[6]

Analysis

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.

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.

[6]The distinction here is often referred to as the difference between estimates of the effects of treatment on the treated (TOT) and the effects of treatment on the untreated (TUT).

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.[7]

While the demographic characteristics of the students in Towering Pines did not change substantially as schools intensified their curricula and there are 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).[8] 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 achievement, 7th grade ELA achievement, demographic characteristics, and interactions between demographic characteristics and baseline achievement.[9] 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

[7]Given that our outcome measure is in 10th grade, there are many mechanisms through which being in a Baseline, Transition, or Algebra for All school in 8th grade might matter. Here we are not concerned with the mechanisms per se, but are rather interested in observing how the population of students who attended different types of schools in 8th grade fared. It is also important to note that since this district had 10 middle schools and 3 high schools, students would have been exposed to peers in high school who went to different middle schools. Thus, the results we present here should not necessarily be interpreted as reflecting how curricular intensification might affect the distribution of student achievement outside of this context.

[8]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.

[9]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.

treatment *t* (Baseline, Transition, and Algebra for All), we define our first inverse propensity score weight as:

$$W=1/\hat{p}_t \quad (1)$$

where $\hat{p_t}$ is the predicted probability that a student received the treatment he or she actually received. 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.

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

$$W=p_{t=\hat{A}fA}/\hat{p}_t \quad (2)$$

$$W=p_{t=\hat{B}}/\hat{p}_t \quad (3)$$

where $p_{t=\hat{A}fA}$ represents the predicted probability that a student was in an Algebra for All school and $p_{t=\hat{B}}$ the predicted probability that a student was in a Baseline school. Thus, for example, in equation (2), students from an Algebra for All school receive a weight of 1 (because for these students the numerator and denominator are identical), while the students from Baseline or Transition schools are weighted to more heavily if they have higher predicted probabilities of being from an Algebra for All School. In weighting the distribution of students from Baseline and Transition schools to look more like the observed distribution of students from 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, 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.

We 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 of students from the Baseline schools to the weighted first percentile score of students from 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 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 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 few significant differences between students in Baseline schools and students in Transition or Algebra for All schools.

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.[10] To account for the fact that our students are nested in schools and cohorts, we stratify on schools and cohorts and bootstrap 999 replicates for our 95 percent confidence intervals.

4. Results

Figures 5 and 6 present the differences between the 10th grade math achievement of students from Baseline schools and students from Transition schools (Figure 5) and Algebra for All schools (Figure 6). As in Figures 3 and 4, the x-axis represents the percentile at which the distributions are being compared, and the y-axis shows the difference between the quantiles of the two 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 of students from Transition schools is roughly .15 standard deviations lower than the median score of students from Baseline schools. As the confidence interval does not include 0, we conclude that 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 distributions of students from Transition and Baseline schools, but that around the 25th percentile a statistically significant gap of about .1 standard deviations emerges. The difference between the two distributions fluctuates

[10]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 roughly 15 percent of 7th graders 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.

somewhat, and is largest near the median, where we see that students from Transition schools are scoring nearly .2 standard deviations lower. 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 students from Transition and Baseline schools.

We see a very different pattern of results in Figure 6, where we compare students from Algebra for All schools and Baseline schools. While there is no difference between the very bottom of the Algebra for All and Baseline distributions, we find that students from Algebra for All schools score worse than students from Baseline schools throughout a large portion of the achievement distribution. This gap increases in a monotonic fashion up until about the 60th percentile. Students in the 60th to 85th percentile from Algebra for All schools are scoring about a third of a standard deviation lower than the students in the 60th to 85th percentile from 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 of students earn the maximum score possible on the CAHSEE test. Given this ceiling effect, we cannot estimate the effects of curricular intensification at the top of the distribution. However, supplemental analyses using logistic regression models to estimate the odds of earning the maximum score possible are 30 percent smaller among students from Algebra for All schools relative to those from Baseline schools, while the odds of students from Transition schools hitting the test ceiling were 18 percent lower than those from Baseline schools. Thus, while we cannot observe the test score differences for quantiles affected by the CAHSEE test ceiling, we find differences in the odds of achieving the highest score, particularly when comparing students from Baseline and Algebra for All schools. It is also 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 the bottom of the distribution of students from Transition and Algebra for All schools is lower than the bottom of the distribution of students from Baseline schools. Thus, while a priori we might have expected that the bottom of the achievement distribution would have been lifted among students from the Algebra for All schools, we find no evidence that this is the case. We do find evidence, however, that student achievement at the top of the distribution is lower among students from Transition and Algebra for All schools than among students from Baseline schools, and that for Algebra for All school students these differences grow increasingly large the towards the top of the distribution. 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 Transition and Algebra for All students to look like Baseline, while in Panel B we present results that weight Baseline and Transition students to look like Algebra for All students on their observable characteristics. Overall, we see that the results are largely similar, but that the results in Panel B (which weight towards Algebra for All school students) provide are somewhat more negative than those in Panel A (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 students from the Baseline schools or the 25th percentile of students from 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 should think carefully about whether they are interested in differences relative to the treated or untreated distributions.

5. Discussion

Research in the social sciences often focuses on individuals as the unit of analysis, estimating how a given individual's outcome would be expected to differ if this individual was or was not exposed to some experience. While differences between individuals are informative, we argue that from a policy perspective it is often more valuable to understand effects on the broader population. This is particularly important because the effects of a policy at the individual and population levels are not necessarily congruent, as the presence of collective effects can lead to population-level effects that diverge from the sum of individual-level effects.

California's attempt to enroll more 8th graders in Algebra represents a clear example of collective effects. While there is strong evidence that individual students benefit when they take more advanced courses, our results indicate that the attempt to enroll all students in more challenging middle school mathematics courses had negative achievement consequences for a wide range of students. These findings are striking: the gap between the achievement of students from 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 short term costs attributable to institutional inertia. 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 initially do more harm than good if 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, analyses examining district-level longitudinal data in California indicate that the negative effects of increases in 8th grade Algebra enrollment occur both immediately after large shifts in enrollment patterns and after more gradual shifts (Domina, McEachin, Penner, and Penner 2014), suggesting that this is somewhat unlikely to be the case.

California's Algebra-for-all effort is typical of a broad range of policies aimed at intensifying curricula in U.S. public schools (e.g. raising high school graduation requirements, Common Core). We argue, accordingly, that more distributional research is needed to assess whether curricular intensification policies are having their desired effects of increasing student learning and decreasing inequality when they are implemented at scale. As research seeks to examine this and other questions where 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 provide intuitive results about the distribution of achievement while adjusting for differences in observable characteristics.

More generally, we argue that the collective effects framework can help clarify our understanding of unintended policy effects in many settings. Below we broaden our discussion beyond education to discuss three settings in which we might expect effects at the population level to diverge from effects at the individual level. While not exhaustive, we highlight three ways in which collective effects emerge: 1) spillover effects, where policies have externalities that affect the collective, 2) structural conditions, where interventions targeting individuals seek to address problems that are caused by community-level or structural considerations, and 3) policy drift, where the form and content of policies and interventions change as they move from implementation at the individual level to implementation at the population level.

Spillover effects

First, we might imagine collective effects arising when interventions affect not only treated individuals, but also produce effects that spill over to people who were not directly treated by the intervention. One example of this is the concept of community or herd immunity, where if a sufficiently high proportion of the population is immunized, then the community as a whole (including members who were not immunized) benefit from the treatment. In other settings, the externalities may be more local, as in the case of contact immunity, in which individuals who come into contact with immunized individuals can catch their immunization.

This latter model tends to dominate thinking in education, where spillover effects are conceptualized as being driven by direct contact with somebody who was treated (e.g. classmates helping each other with homework). However, the more diffuse spillover effects may also be important. For example, early-education interventions that improve all students' basic skills may pay greater than expected dividends if they allow teachers and curriculum developers to focus their instructional time and effort on more advanced material in later grades.

We suggest that the importance of these direct spillover effects is likely to be heightened when there is frequent contact between units targeted by the intervention and those that are not targeted. If a policy intervention targets some students in a classroom, its effects are likely to spill over to non-treated students. Similarly, an intervention that targets some companies in a given industry may affect other companies in the same or allied industries.

Structural Conditions

We would also expect individual and collective effects to diverge when individual-level interventions are applied to structural problems. This point is well-articulated by Dodge (2009: 198) in the context of antisocial behavior:

> Clinicians ‘work around’ or ‘work with’ community risk factors; they almost never work to change these factors. Going to scale with individual-level interventions may ignore cultural and community causes, leading to the perpetual replication of new cases with little net impact on community rates of problems. Removing one drug trafficker from the street corner may only lead a new trafficker to emerge; removing the class deviant from the middle school classroom may only grow a new student to fill this role.

In such cases it is easy to see how one could find large and robust effects of an individual-level treatment on individuals' problem behaviors, and yet not find any reduction in the incidence rates of problem behaviors when the same treatment is operationalized at the population level. Even if individual treatments influence which individuals fill particular structural positions, they may do little to change the overarching structure. In many cases larger social structural considerations or cultural factors likely play important roles in shaping the outcomes that individual-level interventions are seeking to change, leading to large and far reaching collective effects.

Policy drift

A third set of collective effects occur when the process of implementing an intervention at scale fundamentally changes the nature of the intervention. Several sources exist for policy drift, including the fundamental contextual nature of an intervention (Dodge 2011), poor implementation fidelity (Dodge 2009), resource scarcity (Stecher et al. 2001), and positional effects (cf. Raftery and Hout 1993).[11] Many interventions that are effective in closely-monitored trials prove to be much less effective when implemented in contexts where program designers are unable to insure that all aspects of an intervention are faithfully executed, or when they are purposively adapted to better fit in different communities (cf. Dodge 2011).

But even carefully implemented interventions can be subject to policy drift when they are implemented the population level. For example, experimental data from Tennessee and elsewhere provide strong evidence to suggest that class size reductions should boost student achievement and narrow educational inequalities (cf. Nye, Hedges, and Konstantopoulos 2000). However, limitations in the available supply of qualified teachers substantially limited the effectiveness of class size reduction policies when they were implemented at scale in California (Stecher et al. 2001). In this example, the resource scarcity changes the

[11]Research has also highlighted the importance of accounting for low enrollment and high attrition rates in individual-level studies, as well as sample selection processes targeting the small segment of the population likely to benefit the most from an intervention. These considerations also suggest that large effects in the study sample may not translate into effects at the population level (Welsh, Sullivan and Olds 2010; Daro, McCurdy, Falconnier, and Stojanovic 2003). As Dodge (2011) notes, while such studies can help inform our understanding of whether the intervention can change human behavior, these limitations make it difficult to assess whether the intervention is likely to be effective in achieving population-level changes. While such processes clearly lead to divergent effects at the individual and population levels, we view them as issues of external validity, and not instances of collective effects.

nature of the intervention, forcing schools to staff classrooms with under-prepared teachers in order to satisfy the class size mandate.

In other instances, the positional nature of an intervention may produce collective effects when the intervention is implemented at scale. Interventions such as honors track placement, selective university admissions, or judicial clerkships work at least in part as gate-keepers, identifying elites and conveying advantages to these elites at least in part because others were excluded. While interventions based on relative position can have important impacts on individuals, interventions making these opportunities available to all may undermine their effectiveness. Thus we would not expect to observe the same effects when implemented at the population and individual levels.[12]

Conclusion

Collective effects are widespread, important, and not currently well understood. In this paper, we use California's effort to universalize 8th grade Algebra as a context to develop the idea of collective effects and strategies for studying them. Despite strong evidence to suggest that students learn more when they are exposed to advanced courses, our analyses indicate that many students experience lower levels of achievement when their schools move from a course placement model in which 8th grade Algebra is reserved for relatively advanced students to a model in which nearly all students take the course. Taking advantage of propensity score weighting and distributional methods, these analyses document changes in the distribution of student achievement associated with school-level placement policy changes, net of potentially confounding observable student characteristics.

In doing so, we suggest that the prevailing emphasis on understanding individual-level effects in social science research, while important and useful in many domains, often fails to provide the information needed to understand the impacts of policies on the broader population. Further, as researchers begin to consider the effects of policies on populations more broadly (cf. Dodge et al 2004), we believe that it will be important to examine not just how the average outcome of a population is likely to be affected, but also to understand how the distribution of the outcome is affected more broadly. We highlight one fruitful strategy for undertaking such analyses, and argue that similar approaches will help us better understand not only educational policies, but also help to produce a deeper body of policy research across a wide range of topics. Future research should build on these approaches to better understand collective effects in education, health care, labor markets, crime prevention, and many other policy realms.

Acknowledgments

Research reported in this article was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award numbers P01HD065704 and K01HD073319, by the Institute for Education Sciences, and by the Spencer Foundation under award number 201400180. The

---

[12]While each of the above examples of policy drift involve instances in which interventions have positive effects for individuals, but null or negative effects at the population level, it is possible that a policy implemented locally could have negative effects for individuals but positive effects when implemented broadly. Exit exams, for example, could have negative effects for individuals not receiving a diploma when implemented at the local level. However, when implemented widely, such a policy could have positive effects on earnings by increasing human capital and improving the signaling power of and economic returns to high school diplomas.

content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Institute for Education Sciences, or the Spencer Foundation. The authors are grateful to Marianne Bitler, Ken Dodge, Greg Duncan, Sasha Killewald, and Tony Tam for useful comments and discussions, and for the support and hospitality of Stanford University's Institute for Research in the Social Sciences.

Works Cited

Addo, Fenaba R.; Lichter, Daniel T. Marriage, marital history, and black–white wealth differentials among older women. Journal of Marriage and Family. 2013; 75(2):342–362.

Allensworth E, Nomi T, Montgomery N, Lee VE. 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. 2009; 31(2):367–391.

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

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

Bitler M, Domina T, Penner EK, Hoynes H. Distributional Effects of a School Voucher Program: Evidence from New York City. Journal of Research on Educational Effectiveness. Forthcoming.

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

Budig MJ, Hodges M. Differences in Disadvantage: How the Wage Penalty for Motherhood Varies Across Women's Earnings Distribution. American Sociological Review. 2010; 75:705–28.

Chernozhukov, Victor; Fernández-Val, Iván; Melly, Blaise. Inference on counterfactual distributions. Econometrica. 2013; 81(6):2205–2268.

Clotfelter CT, Ladd HF, Vigdor JL. The aftermath of accelerating algebra: Evidence from a district policy initiative. Journal of Human Resources. 2015; 50(1):159–188.

Clotfelter, CT.; Ladd, HF.; Vigdor, JL. Algebra for 8th Graders: Evidence on its Effects from 10 North Carolina Districts (No w18649). National Bureau of Economic Research; 2012b.

Daro D, McCurdy K, Falconnier L, Stojanovic D. Sustaining new parents in home visitation services: Key participant and program factors. Child Abuse and Neglect. 2003; 27:1101–25. [PubMed: 14602094]

Dodge KA. Community intervention and public policy in the prevention of antisocial behavior. Journal of Child Psychology and Psychiatry. 2009; 50:194–200. [PubMed: 19220602]

Dodge KA. Context matters in child and family policy. Child Development. 2011; 82:433–442. [PubMed: 21291450]

Dodge KA, Berlin LJ, Epstein M, Spitz-Roth A, O'Donnell K, Kaufman M, Amaya-Jackson L, Rosch J, Christopoulos C. The Durham Family Initiative: A preventive system of care. Child Welfare. 2004; 83:109–128. [PubMed: 15068214]

Domina T, Penner AM, Penner EK, Conley AM. Algebra for all: California's 8th grade algebra initiative as constrained curricula. Teachers College Record. Forthcoming.

Domina T, McEachin A, Penner AM, Penner EK. Aiming High and Falling Short: California's Eighth-Grade Algebra-for-All Effort. Educational Evaluation and Policy Analysis. Forthcoming.

Duflo E, Dupas P, Kremer M. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. The American Economic Review. 2011; 101(5):1739–74.

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

Firpo S, Fortin NM, Lemieux T. Unconditional Quantile Regressions. NBER Technical Working Paper No 339. 2007

Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. Econometrica. 2009; 77(3):953–973.

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

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

Heppen, JB.; Walters, K.; Clements, M.; Faria, A.; Tobey, C.; Sorensen, N.; Culp, K. 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; 2012.

Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika. 2000; 87(3):706–710.

Kerckhoff AC. Effects of ability grouping in British secondary schools. American Sociological Review. 1986:842–858.

Killewald A, Bearak J. Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression. American Sociological Review. 2014; 79(2):350–357.

Koenker R, Bassett G Jr. Regression quantiles. Econometrica. 1978; 46:33–50.

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

Liang J, Heckman PE, Abedi J. What Do the California Standards Test Results Reveal About the Movement Toward Eighth-Grade Algebra for All? Educational Evaluation and Policy Analysis. 2012

Lise, J.; Seitz, S.; Smith, J. Equilibrium policy experiments and the evaluation of social programs (No w10283). National Bureau of Economic Research; 2004.

Long MC, Conger D, Iatarola P. Effects of High School Course-Taking on Secondary and Postsecondary Success. American Educational Research Journal. 2012; 49(2):285–322.

Machado, José AF.; Mata, José. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics. 2005; 20(4):445–465.

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

McPartland JM, Schneider B. Opportunities to learn and student diversity: Prospects and pitfalls of a common core curriculum. Sociology of education. 1996:66–81.

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

Nomi T. 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. 2012 Published online before print, July 31, 2012.

Nye B, Hedges LV, Konstantopoulos S. The effects of small classes on academic achievement: The results of the Tennessee class size experiment. American Educational Research Journal. 2000; 37(1):123–151.

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

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

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

Raftery AE, Hout M. Maximally maintained inequality: Expansion, reform, and opportunity in Irish education, 1921-75. Sociology of Education. 1993:41–62.

Rosenbaum JE. If Tracking Is Bad, Is Detracking Better? American Educator. 1999; 23:24–29.

Rosin, MS.; Barondess, H.; Leichty, J. Algebra policy in California: Great expectations and serious challenges. Mountain View, CA: EdSource; 2009.

Schiller KS, Muller C. Raising the bar and equity? Effects of high school graduation requirements and accountability practices on students' mathematics course taking. Education Evaluation and Policy Analysis. 2003; 25:299–318.

Slavin, Robert E. Achievement effects of ability grouping in secondary schools: A bestevidence synthesis. Review of Educational Research. 1990; 60:471–499.

Stecher B, Bohrnstedt G, Kirst M, McRobbie J, Williams T. Class-Size Reduction in California: A Story of Hope, Promise, and Unintended Consequences. Phi Delta Kappan. 2001; 82(9):670–74.

Stein MK, Kaufman JH, Sherman M, Hillen AF. Algebra: A challenge at the crossroads of policy and practice. Review of Educational Research. 2011; 81(4):453–492.

Stigler, JW.; Hiebert, J. The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom. New York: Free Press; 2009.

Wurman Z, Evers WM. New Education Dashboard: Less Rigorous, Less Meaningful. Education Week. Feb 09.2011 30:20.

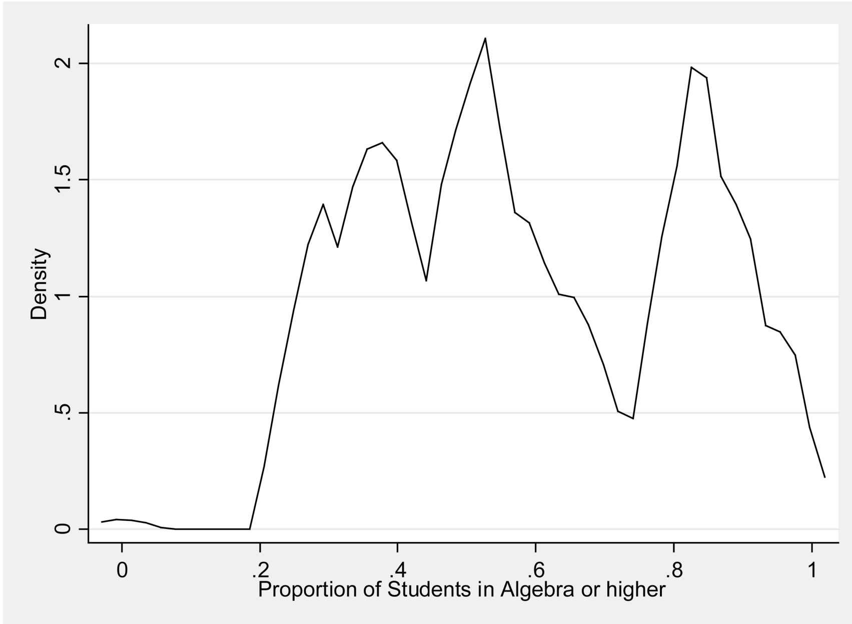
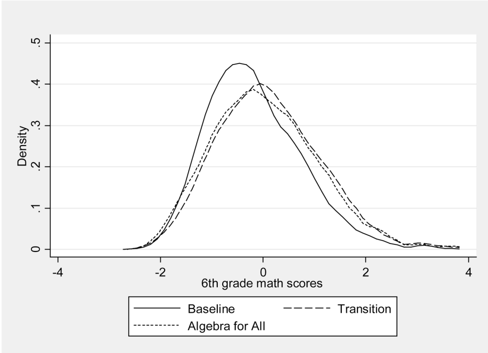


Figure 1. Kernel density estimate of the proportion of students enrolled in algebra or higher in the students' school

Panel A. Unweighted distributions



Panel B. Weighted towards Algebra for All distribution

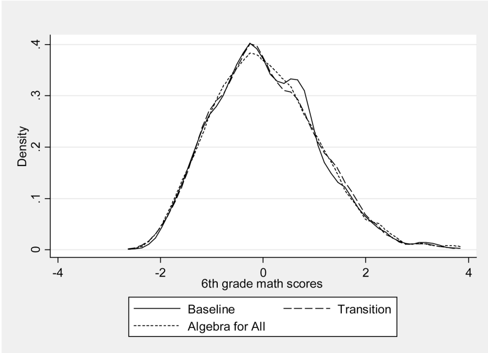


Figure 2. Distribution of 6th grade mathematics achievement in Baseline, Transition, and Algebra for All schools, unweighted and weighted towards Algebra for All

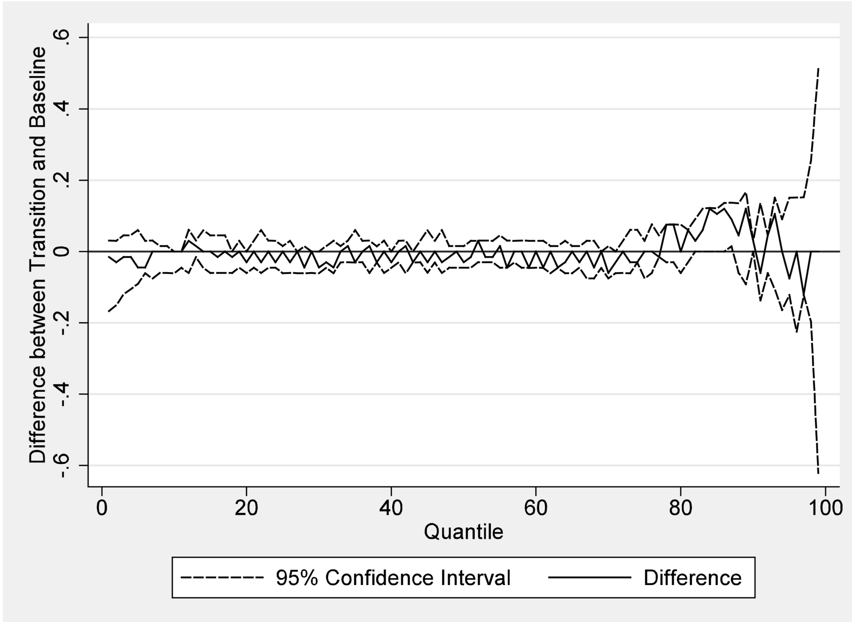


Figure 3. Differences between Transition schools and Baseline schools in 6th grade mathematics scores, using population inverse propensity score weights

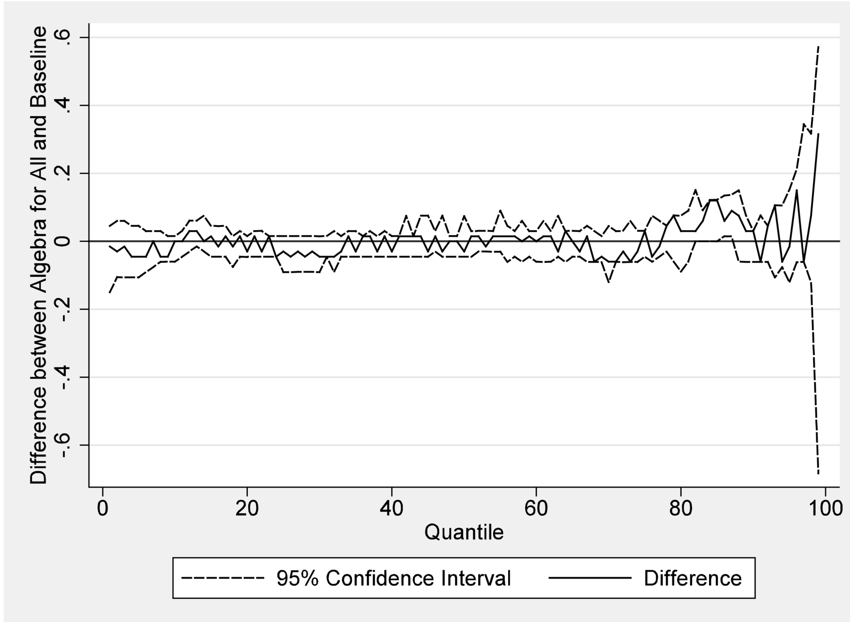


Figure 4. Differences between Algebra for All schools and Baseline schools in 6th grade mathematics scores, using population inverse propensity score weights

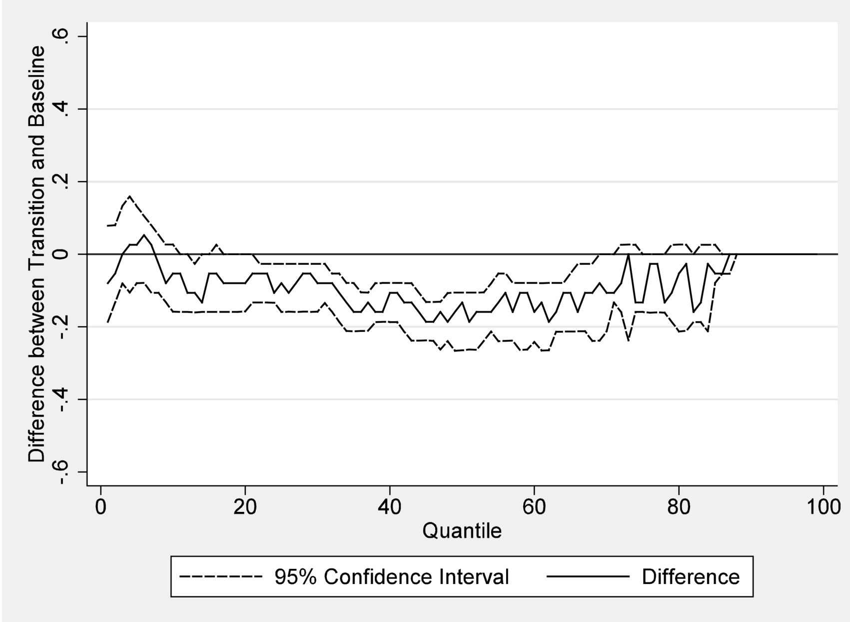


Figure 5. Differences between Transition schools and Baseline schools in 10th grade mathematics scores

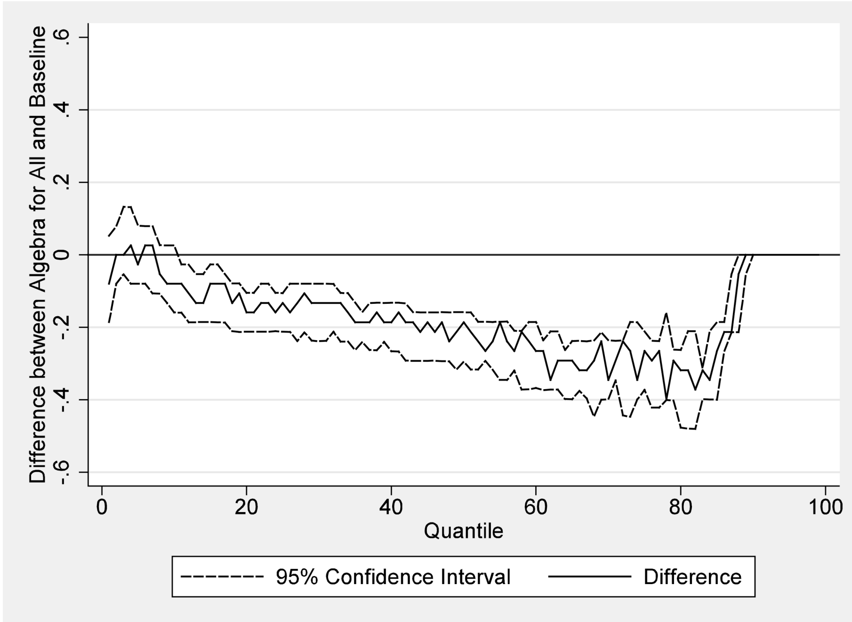
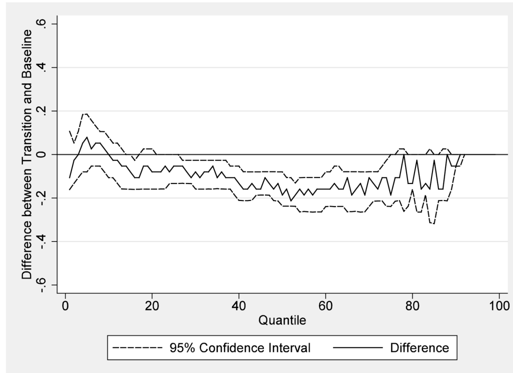


Figure 6. Differences between Algebra for All schools and Baseline schools in 10th grade mathematics scores

Panel A. Weighted towards Baseline school distribution



Panel B. Weighted towards Algebra for All school distribution

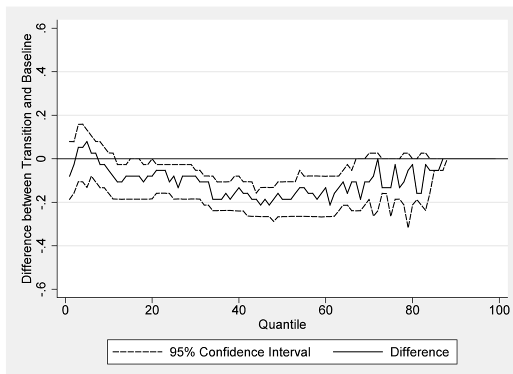
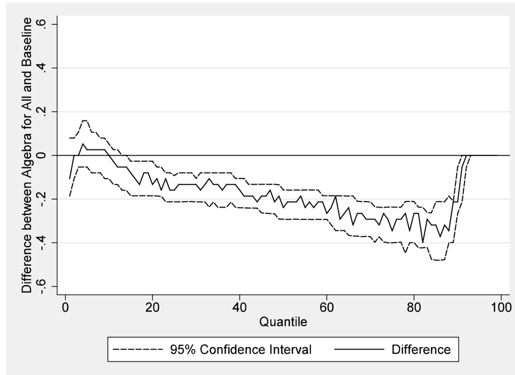


Figure 7. Differences between Transition schools and Baseline schools in 10th grade mathematics scores, weighted either towards Algebra for All or towards Baseline schools

Panel A. Weighted towards Baseline school distribution



Panel B. Weighted towards Algebra for All school distribution

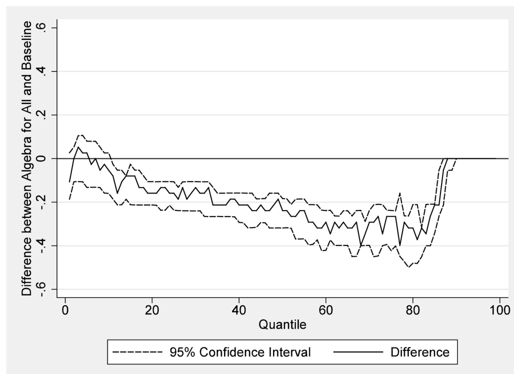


Figure 8. Differences between Algebra for All schools and Baseline schools in 10th grade mathematics scores, weighted either towards Algebra for All or towards Baseline schools

Table 1

Descriptive statistics on analytic sample by cohort

	2004-2005	2005-2006	2006-2007	2007-2008
Gen Math in 8th grade (%)	57	42	20	12
Algebra in 8th grade (%)	39	47	65	71
Geometry in 8th grade (%)	5	11	15	17
Attended Baseline school in 8th grade (%)	81	43	0	0
Attended Transition school in 8th grade (%)	19	57	51	21
Attended Algebra for All school in 8th grade (%)	0	0	49	79
ELL in 8th grade (%)	34	31	30	31
RFEP in 8th grade (%)	30	34	36	37
Eng only/FEP in 8th grade (%)	36	35	34	32
Hispanic (%)	49	46	51	52
Vietnamese (%)	25	28	26	27
White (%)	18	17	15	13
Other (%)	8	9	8	8
N	2,392	2,470	2,773	2,768

Note: 6th grade math, 7th grade ELA, and 10th grade math scores are standardized across all Towering Pines 8th grade students in these four cohorts.

Author Manuscript
Author Manuscript
Author Manuscript

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