

“Double-dose” Algebra as an Alternative Strategy to Remediation:
Effects on Students’ Academic Outcomes

Takako Nomi
Elaine M. Allensworth

Consortium on Chicago School Research
University of Chicago

December 8, 2008

Running Head: Double-dose Algebra Policy

We gratefully acknowledge support for this work from the Institute for Educational Sciences, U.S. Department of Education, grant # R305R060059. However, the responsibility for the findings and conclusions reported in this manuscript is the authors', not the U.S. Department of Education. We thank Valerie Lee, Steve Raudenbush, Sean F. Reardon, Julia Gwynne, Chris Mazzeo, and Vanessa Coca for their helpful comments on earlier versions of this article. We also thank Nicole Tysvaer for her support in the production of this article. For information on this study, contact the first author.

Contact Information:

Takako Nomi
1313 E. 60th St., Chicago IL. 60637
Tel: (773) 834-1879
Fax: (773)702-2010
Email: nomit@ccsr.uchicago.edu

Elaine M. Allensworth
1313 E. 60th St., Chicago IL. 60637
Tel: (773) 834-3061
Fax: (773)702-2010
Email: elainea@ccsr.uchicago.edu

This is a preprint of an article whose final and definitive form has been published in the *Journal of Research on Educational Effectiveness* © 2009 copyright Society of Research on Educational Effectiveness; *Journal of Research on Educational Effectiveness* is available online at <http://www.informaworld.com/smpp/content~db=all?content=10.1080/19345740802676739>

Abstract

Expanded instructional time has become increasingly popular as a strategy to improve the academic outcomes of low-skilled students, particularly in the ninth grade. We evaluate the efficacy of a double-period algebra policy initiated in the Chicago Public Schools in 2003. This policy required all students with eighth-grade test scores below the national median to enroll in a support algebra course in addition to regular algebra in the ninth grade. Using regression discontinuity combined with interrupted time series designs, and instrumental variable models, we show the effects of the policy on students' grades, failure rates and test scores in 9th-grade algebra and 10th-grade geometry. Providing support courses improved algebra test scores for the target population, but only modestly affected grades and failure rates. Students with very low initial abilities benefited less than students close to the national median. The policy also led schools to track algebra classes by students' entering math skills. As a result, it affected academic outcomes among students not targeted by the policy; test scores among high-ability students improved while their grades declined.

There is currently a nationwide push to increase the rigor of high school coursework so that all students graduate with college-ready skills. Criticisms of low academic standards have grown since the early 1980s with the *Nation at Risk* report, which described U.S. public high schools as “a sea of mediocrity” (National Commission on Excellence in Education, 1983). More recently, changes in the U.S. economy have brought increasing criticism that high schools are not sufficiently preparing students for either college or workforce. Furthermore, there is growing consensus that success in the workforce requires the same skills that students need for college (ACT, 2006; American Diploma Project, 2004). In response, policy-makers have been calling on states and districts to abandon low-level coursework and require all students to complete college-preparatory course sequences (National Governor’s Association, 2005).

At the same time, there is also growing concern about high failure rates in ninth-grade courses, particularly in ninth-grade algebra (e.g. Horwitz & Snipes 2008; Herlihy & Kennelly, 2007). It is widely recognized that many students enter high school with weak math skills—on the 2005 NAEP 30 percent of eighth graders scored below basic level, while only 30 percent were at or above the proficient level (U.S. Department of Education, 2007). In Chicago, where we base this study, half of ninth-grade students fail at least one course, and pass rates are lowest in math (Roderick & Camburn, 1999; Allensworth & Easton, 2005). Concern about ninth-grade failure is amplified by findings that ninth-grade course failure is directly tied to dropping out in later grades, as failures prevent students from accumulating the credits that they need to graduate (Allensworth & Easton, 2007; Bottoms, 2008).

Thus, educators are caught in a dilemma. Requiring college preparatory coursework for all students makes sense in an age when the vast majority of students aspire to attain a four-year college degree.¹ However, raising requirements without sufficient support mechanisms could

lead more students to fail and drop out (Darling-Hammond, 2004). This presents a challenge for educators who are charged with teaching students entering high school with academic abilities that are well below grade level.

Historically, most high schools have offered a differentiated (tracked) curriculum that has provided separate courses based on students' academic skill levels and their interests (Angus & Mirel, 1999; Cremin, 1961; Tyack & Cuban, 1995). However, tracking has received considerable criticism for impeding the academic progress of students in lower tracks and exacerbating existing educational and social inequalities. A large volume of research in the 1990s documented strong links between students' academic and social backgrounds and their course taking (Lee, 2002; Oakes, 1985; Powell, Farrar, & Cohen, 1985).

In response to concerns about tracking, some school districts have actively "de-tracked" their high schools, creating mixed-ability classrooms and providing low-ability students with curriculum and instruction at the same level as those in college-preparatory classrooms. Existing studies have shown a number of problems and difficulties accompanied by schools' de-tracking efforts (Gamroan & Weinstein, 1998; Oakes, 1994; Rosenbaum, 1999-2000). For example, teachers experienced difficulty accommodating instruction and giving sufficient attention to meet the needs of both low and high ability students (Rosenbaum, 1999-2000).

Other school systems have attempted to improve instruction for low-performing students without mixing low- and high-skilled students in the same classrooms. In the late 1980's and early 1990's, California and New York introduced "transition" math courses for under-performing students aimed at bridging the gap between basic and college preparatory mathematics (White, Gamoran, Smithson, & Porter, 1996, p.285). California replaced lower-level math courses with Math A, which emphasized more complex cognitive tasks of

understanding, application of knowledge, and reasoning (White et al., 1996). In Rochester, New York, lower track math was replaced with the Stretch Regents sequence, which covered the same integrated material as the college-preparatory Regents courses, but taught at a slower pace (White et al., 1996). Buffalo, New York, introduced University of Chicago School Math Project (UCSMP) courses as an alternative to the Regents courses; ordinarily a six-year sequence from 7th to 12th grade, students in Buffalo could enroll in the first course (Transition Mathematics) in 9th grade (White et al., 1996). These transition courses seemed to be successful in terms of moving students from “dead-end”, low-level sequences to college-preparatory coursework (White et al., 1996). However, evidence was inconclusive as to whether these courses improved academic outcomes. Gamoran, Poter, Smithson, and White (1997) found that the learning gains of students in transition courses were better than those of students who took general math/pre-algebra although not at a level which reached statistical significance, and not as high as those who took regular algebra courses.

The Double-Dose Strategy

An increasingly popular approach is to offer extended instructional time in algebra, where low-performing students receive an extra period of instruction each day. At some schools this is achieved through a second “shadow” or “support” algebra course; other schools use blocked scheduling where algebra takes up two periods in the day instead of one (Chait, Muller, Goldware & Housman, 2007). We refer to both of these as a “double-dose” strategy--where low-skill students receive twice the algebra instruction that they otherwise would receive.

Historically, many Catholic schools enrolled students in two periods of English or math if they were unprepared for high school instruction (Bryk, Lee, & Holland, 1993). “Double-dose”

coursework was used as one of several strategies, including slower-paced coursework and tutoring activities, intended to help underperforming students succeed in a college-preparatory curriculum. The “double-dose” approach in Catholic schools seemed successful. Bryk et al. (1993) reported that in schools where students enrolled in additional coursework, the number of students in need of remediation halved by the eleventh grade and almost all students reached the college-preparatory level by their senior year.

More recently, a comprehensive high school reform model, the Talent Development Model (TDM), has documented some success using double-period programs, and their ideas have begun to spread nationally (Kemple, Herlihy, & Smith, 2005; Mac Iver, Balfanz, & Plank, 1998). The TDM includes a mandated college-prep curriculum with increased learning time (double-period courses), along with small learning communities, and teacher professional development. Kemple et al (2005) found that students in the Talent Development schools made modest gains in ninth-grade attendance, academic course credits earned, and promotion rates, compared to students in non-Talent Development schools that were similar in demographic and academic compositions.²

Yet, besides the TDM evaluation, little research exists evaluating the effectiveness of double-dose courses as a district or state reform strategy. The success seen in Catholic schools may not be applicable to public high schools as many differences besides double-dose coursework exist between the two contexts. Furthermore, the TDM evaluation is not applicable to districts that are implementing double-dose courses without the full TDM. It is simply not known whether a double-dose strategy by itself is effective. To address this knowledge gap, this study examines the effects of a systemwide “double-dose” algebra policy implemented in the Chicago Public Schools (CPS) in 2003.

Implementation of the “Double-Dose” Algebra Policy in Chicago Public Schools

The double-dose algebra policy³ in Chicago built on an earlier curricular policy initiated in 1997 which eliminated remedial math courses and required all first-time freshmen to enroll in algebra in ninth grade, followed by geometry and algebra II in the subsequent two years. The 1997 policy was intended to provide all students with a curriculum that would prepare them for college. However, many CPS students entered high school with math skills well below grade level, and failure rates in ninth-grade algebra were high (Roderick & Camburn, 1999). The 2003 double-period algebra policy was a response to those high failure rates.⁴ This policy required all first-time ninth-grade students who tested below the national median on the eighth-grade math test,⁵ referred to here as “below-norm students,” to enroll in two periods of algebra coursework. Students enrolled in a regular algebra class, plus an additional “algebra support” class, for a full academic year. All CPS high schools were subject to the “double-dose” algebra policy, including 60 neighborhood schools, 11 magnet schools, and 6 vocational schools.

Because CPS is fairly decentralized, algebra curricula had varied considerably across schools prior to the policy. The double-dose Algebra policy attempted to improve coherence in algebra curricula by providing resource materials to double-dose algebra teachers with two curricular options—Agile Mind and Cognitive tutor—and providing stand-alone lesson plans that teachers could use. The district also ran professional development workshops three times a year for double-dose algebra teachers where it provided suggestions on how to use the extra time for Algebra instruction. However, we do not know what teachers actually taught in the support courses or the extent to which the curriculum was consistent across schools. A limitation of this

study is that we do not address variation in the instructional content. Therefore, we are studying the average effects of doubling students' exposure to algebra instruction without regard for the particular content of that instruction.⁶

While CPS did not mandate a particular curriculum, the district did provide guidelines for the structure of the support courses. To try to ensure that schools offered coherent instruction for targeted students, the district strongly suggested that schools program their algebra support courses in three specific ways: 1) double-dose algebra students should have the same teacher in algebra and algebra support; 2) the courses should be offered sequentially; and 3) students should take their algebra support course with the same students that were in their regular algebra course. Although these structural guidelines were presented only as recommendations, and not requirements, most schools followed them in the initial year (Table 1). In 2003, 80 percent of “double-dose” students had the same teacher in their regular algebra and support course and 72 percent of students took the two courses sequentially. Also, in a typical algebra course taken by “double-dose” students, 92 percent of their classmates were also “double-dose” students.

Table 1 about here

Because many schools complained about scheduling difficulties, CPS removed the first structural guideline—having the same teacher in algebra and support courses—in 2004. However, 54 percent of “double-dose” students continued to have the same teacher in the two classes. This change also brought declining adherence to the second guideline; in 2004, less than half of “double-dose” students took the two courses sequentially. However, double-dose students continued to attend algebra courses with other double-dose students in 2004; in a typical algebra

course taken by double-dose students, 87 percent of their classmates were also double-dose students.

As a result of the three structural guidelines, the “double-dose” algebra policy led to considerable changes in the ways schools programmed students into their freshman year math classes. The easiest way to adhere to these guidelines was by sorting above- and below-norm students into separate algebra courses, and this is what happened in the majority of schools. After the policy was implemented, math classes became more homogenous in terms of students’ incoming math abilities. This decline in classroom heterogeneity occurred among both low- and high-ability students, but was largest among students with incoming test scores just below the national median.⁷ The average ability level of classroom peers declined for students entering high school with math scores below the national median because most of them no longer had above-norm students in their algebra class. Peer ability levels increased for students with math scores above the national median because they no longer took algebra with below-norm students who were in separate double-period classes.⁸ It is important to recognize these changes in classroom structure in order to understand how the policy affected students’ learning experiences. For below-norm students who were targeted by the policy, the intervention was not just receiving extra instructional time, but receiving extra instructional time within homogenous low-ability classes. There were also possible policy effects on above-norm students as their algebra classes became more homogenous without below-norm students.

A Conceptual Framework for Studying the Effects of Double-dose Algebra Policy

The purpose of the double-dose algebra policy was to reduce course failure in algebra among low-skilled students by providing an extra period of algebra instruction. This strategy is consistent with substantial research showing instructional time to be important for student

learning (Anderson, 1984; Bloom, 1974; Millot, 1995). To further support double-dose algebra teachers, the district provided new resource materials and optional professional development. Thus, the policy could have affected students' academic outcomes through two intentional mechanisms: 1) expanded instructional time, and 2) changes in instructional content and pedagogy resulting from new curricular resources and professional development. These policy mechanisms would only affect students who received a double-dose of algebra.⁹

To maintain instructional coherence for double-dose students, the district also provided guidelines as to how schools should structure the support classes (same teacher, same peers, sequential courses). As a result of the guidelines, the policy induced tracking of algebra classes. Thus, tracking was an unintentional third mechanism through which the policy could have affected student outcomes. Shifts in classroom composition could themselves induce changes in instructional content and pedagogy, and also affect classroom learning climate. These unintended effects of the policy would not only affect targeted students, but also non-targeted (high-ability) students, although in opposite ways.

Literature on ability grouping suggests that classroom academic composition can affect content difficulty and instructional pace because instruction is typically geared toward average students in classrooms (Bar and Dreeben, 1983). In addition, in classrooms with high-ability students, teachers tend to spend more time on critical thinking rather than drilling or dealing with behavioral problems than in classrooms with low-ability students (Oakes, 1985; Page, 1991; Rosenbaum, 1976). This suggests that the double-dose algebra policy would unintentionally improve instruction and classroom climate for high-ability students. On the other hand, it could have deleterious effects on instruction and climate in algebra classes for low-ability students, decreasing rigor and increasing behavioral problems. These adverse effects might be offset by

the intentional effects of the policy--greater instructional time, instructional resources, and professional development. As we evaluate the consequences of policy for the target students, we cannot separate the intended effects from those that were not intended; we are examining the effects of double-dose algebra in homogenous low-ability classrooms.

When evaluating the policy, we must also ask which outcomes should be affected. The goal of the policy was to reduce failure rates in ninth-grade algebra. However, the mechanisms of more time and more consistent instruction were intended to boost students' grades indirectly—by improving their skills in algebra. This makes sense for an intervention based on low skills (test scores determine eligibility). But research on ninth-grade course failures in Chicago has shown that grades are much more strongly related to students' behaviors than their incoming academic skills (Allensworth & Easton, 2007).¹⁰ Teachers consider multiple factors when assigning a grade besides content knowledge, including attendance, assignment completion, and students' performance relative to other students. Therefore, we cannot necessarily expect improvements in grades and failure rates to accompany improvements in academic skills. In fact, even though tracking may improve test scores for above-norm students because they are taught more challenging materials in a better learning climate, their course grades may be lower than in classrooms with lower ability students due to harder assignments and teachers' perceptual changes in their abilities relative to classroom peers. In contrast, low-ability students could improve grades without improving algebra skills simply because expectations should fall when high-ability students are no longer in their classes. This suggests that we need to look at multiple outcomes—test scores, algebra grades, and algebra failure rates. In addition, school district personnel hoped that the policy would have long-term effects, with improvements in algebra performance leading to better performance in geometry in the

subsequent year. Yet, because the policy only brought changes in algebra course structure, there may have been little reason to believe that geometry performance would be greatly affected.

Lastly, we must also consider whether the policy effects were consistent across students with different academic skills. While the policy effects should differ between students entering high school with test scores above and below the national median (determining eligibility for double-dose algebra), we could further hypothesize that students' incoming math skills would moderate the effects of the policy on their algebra performance. By placing the cut-point for enrollment at the national median, the policy affected the math curriculum of more than half of the students entering CPS high school.¹¹ It could be that students close to the national median did not need much support to pass algebra. Alternatively, these students may have more easily responded to increased support and instruction than students with very low skill levels. Unintentional effects of classroom compositional changes may also vary by students' academic skills. For example, students who were just below the 50th percentile became the highest ability students in their algebra class, while very low-ability students continued to be the lowest-skilled students in their algebra class.

Thus, evaluating this policy calls for an examination of whether students' outcomes were affected, which outcomes were affected, whether the effects depended on students' initial skills, and the mechanisms through which the policy worked. This article focuses on all but the policy mechanisms. A subsequent article will address how the policy effects were mediated by classroom composition, instruction, and pedagogy.¹²

Research Questions

The policy was designed to help students entering high school with below-average math skills handle ninth-grade algebra. Thus, our primary research question: *(1) did putting lower performing students in double-dose algebra lead to significant improvement in their math achievement, measured by test scores, grades and pass rates?* There are two potential counterfactuals to enrolling in double-dose algebra, and we explore both in this analysis. One counterfactual to taking double-dose algebra post-policy is taking single-period algebra post-policy (counterfactual A). This comparison tells us what might happen, for example, if a below-norm student took single-period algebra because she was mis-assigned post-policy or the cut-off for double-dose changed slightly. A second counterfactual (B) to taking double-dose algebra post-policy is taking single algebra without a double-dose policy. This comparison tells us the effect of having the policy on targeted students and is the more pertinent comparison for evaluating the policy.

Because the “double-dose” algebra policy led to changes in the classroom organizational structure of all algebra courses, it also might have affected the outcomes of students not targeted by the policy. Therefore, we also ask: *(2) Did the policy affect the academic outcomes of students who were not targeted by the policy (i.e., those entering high school with math scores above the national median)?* Finally, we look for differences in policy effects based on students’ initial math skills. We ask: *(3) how did policy effects differ by students’ incoming math skills?*

Data and Methods

Data

The study uses data on the population of first-time ninth graders who entered CPS high schools between 2000 and 2004 (five ninth-grade cohorts). Approximately 85 percent of students were eligible for free/reduced lunch programs; the racial-ethnic composition was 54% African-American, 34% Latino, 9% white and 4% Asian. Seven magnet schools were excluded from the analyses because almost all of their students entered with math scores above the national median, so the policy did not affect them. Also, because our analyses compare changes in student outcomes from pre-policy to post-policy years, we restricted our analyses to schools to those existed in both time periods. The analytic population includes 92,432 students in 64 schools.

In general, the five cohorts had similar demographic compositions, while academic composition differed slightly (Table 2). The percentage of students entering high school with math scores below the national median varied from 59% in 2001 to 50% in 2002. The 2002 cohort had slightly higher initial math abilities than other cohorts.

Table 2 about here

This study uses multiple data sources provided by CPS. Administrative records provide demographic information, including student enrollment status, age, gender, race, and special education status. Indicators of students' socioeconomic status are derived from U.S. census data about the economic conditions in students' residential block groups. Semester-by-semester course transcript and grade data files contain detailed class information, such as teacher IDs, class periods, subject names, subject specific course codes, and course grades. These were used

to classify students' algebra courses. Elementary achievement test scores are based on the Iowa Test of Basic Skills, taken by all students from the third through eighth grade. High school achievement test scores come from the PLAN exam, a test that is part of the EPAS system developed by ACT, Inc. which CPS students take in the fall of the tenth grade.

Measurement

The outcome variables of this study include course grades and failure in algebra and geometry and algebra test scores in the fall of 10th grade. Grades are taken from students' primary algebra course (not their support course) in their first year of high school and from their geometry course in their second year of high school. (Almost all students take geometry in their second year, even if they failed algebra in their first year.) Grades are measured on a four-point scale, where A=4 regardless of level (e.g., Honors, regular). The average algebra grade in our analytic population was 1.63 with a standard deviation of 1.18; the average geometry grade was 1.61, with a standard deviation of 1.25. Across all students, algebra grades are fairly normally distributed. However, grades are not normally distributed if we just examine students entering high school with below-average test scores. Therefore, we substitute dummy variables indicating whether students received a grade of "B" or higher for the continuous measures of grades for analyses of student subgroups based on incoming test scores. Algebra course failure is a dummy variable indicating that a student received an "F" in their primary algebra course.

Algebra test scores come from a subset of the standardized math test (PLAN) developed by ACT, which was administered in October of 10th grade. The algebra subtest contains 22 multiple choice questions (with five response categories each); raw scores are converted to a scale score ranging from 1 to 16. The national average PLAN algebra score is 8.2, with a standard deviation of 3.5. The content of the exam is based on surveys conducted by ACT, Inc.

of high school teachers, and includes problems found in first-year high school algebra classes (ACT, 2007). In our analytic population, the average score on the subset was 6.0 with a standard deviation of 2.5. Test scores are not available for the 2000 cohort, therefore, analyses involving test scores use only two pre-policy cohorts (2001 and 2002).

Students' entering math abilities are measured in several ways. Using eighth-grade ITBS percentile scores, we created 2 sets of categorical variables: 1) a binary variable denoting whether students entered high school below the 50th percentile, which indicates double-dose algebra eligibility; 2) a set of dummy variables distinguishing percentile rank groups—below the 20th percentile, 20th to 35th percentile, 35th to 50th percentile, 50th to 65th percentile, 65th to 80th percentile, and above the 80th percentile. The extended set of dummy variables is used to examine differential policy effects for students with different incoming abilities (research question 3). In addition, we used a continuous measure of math ability as a control variable to adjust for cohort variation in students' ability levels upon entering high school. This indicator, which we call the latent ability score, is a much more precise and reliable indicator of students' true ability than the percentile score, as it uses information on students' learning trajectories over 6 years prior to high school with equated forms of the test.¹³ The latent scores were centered on the average score of students at the 50th percentile across 5 cohorts. We also created a set of continuous latent score variables for each percentile rank group, each centered on the average score of that group, to provide a consistent control for achievement over time in a way that was uncorrelated with the percentile rank dummy variables.

Other student control variables include a dummy variable on gender and special education status. Race/ethnicity is measured by a set of dummy variables distinguishing African American, Hispanic, White, and Asian students. Two measures of SES variables are constructed

using the block-level 2000 U.S. census data, linked to students' home addresses.¹⁴ They include neighborhood poverty and social status, which are standardized to have a mean of zero and standard deviation of one. Neighborhood poverty is a composite measure of the male unemployment rate and the percent families under the poverty line in the block group, and social status is a composite measure of average educational attainment and percentage of employed persons who are managers, executives or professionals in the block group. Residential mobility is measured by a set of dummy indicators distinguishing no moves (omitted category), moving once, and moving twice or more in the three years prior to entering high school. Age at entry into high school is measured by three variables--number of months old for entering high school, a dummy variable indicating if students are slightly old, and a dummy variable indicating if students are young for starting high school.

Analysis

We apply the conceptual framework of Rubin's causal model (Holland, 1986; Rubin 1978) to examine the effects of double-dose algebra enrollment on student outcomes. The key assumption of Rubin's causal model is that, using the language of experiment, subjects assigned to treatment and control conditions have potential outcomes in both states and the causal effect is defined as the difference in potential outcomes between these two states. We define the double-dose algebra enrollment effects as differences between two potential outcomes—one outcome when students received the treatment (i.e., double-dose algebra coursework) and the other when they did not receive the treatment (i.e., single algebra coursework).

One method of estimating the effects of double-dose algebra course enrollment exploits exogenous variations in treatment created by the policy using a regression discontinuity analysis.

If all students adhered to the policy, the treatment (double-dose course enrollment) should be perfectly correlated with an observed variable (i.e., the percentile score), and there should be no unobserved characteristics that are correlated with the treatment and outcomes. This allows us to estimate the average treatment effects by comparing outcomes between students just below and just above the cutoff for double-dose eligibility. In fact, system-wide adherence to the policy was relatively high. Figure 1 shows that the enrollment in double-dose algebra was strongly defined by the 50th percentile cutoff. Most students below the national median enrolled in double-dose algebra while most students above the national median did not, in both 2003 and 2004. Policy non-compliance was higher among low-ability special education students than among regular education students. Among regular education students, over 80 percent of students complied with the policy although compliance was slightly lower close to the 50th percentile cutoff score (Figure 2). This sharply-defined participation in double-dose algebra appears, on first glance, to be ideal for studying the effects of double-dose algebra with a regression-discontinuity design.

However, there are several limitations to using a standard regression discontinuity model for the analysis. First, a standard regression discontinuity design requires students not eligible for the treatment to not be affected by the policy. This is not the case with this policy which caused all algebra classrooms to become more homogenous in their ability composition. Thus, we cannot estimate policy effects with a traditional regression discontinuity, only the effects of enrolling in double-dose algebra post-policy versus single algebra post-policy (counterfactual A). Second, a regression discontinuity analysis is not informative about the policy effects for students who are far away from the 50th percentile.¹⁵ Third, while adherence to the policy was high, it was not complete, and this could affect our estimates of double-dose algebra effects. Thus, our analyses are modified to address these three limitations, as described below.

Figure 1 and Figure 2 about here

Addressing the first limitation: enrollment effects are not equivalent to policy effects. A traditional regression-discontinuity design would compare the outcomes of *post-policy* students just below the 50th percentile cutoff (who took double-dose algebra) to other *post-policy* students who were just above the cutoff, and did not take double-dose algebra. This comparison tells us what might happen to students if the cut-off score was lowered slightly (resulting in fewer students enrolled in double-dose algebra) or if a student near the cutoff score was assigned to single algebra when eligible for double-dose algebra. In other words, it tells us the trade-off for students to be assigned to one versus the other algebra class under the policy. While these are important issues, a more relevant policy question is what would happen to students if there were no policy. Post-policy comparisons cannot tell us this because the policy could have affected outcomes of above-norm students, due to classroom compositional changes that accompanied the policy. To understand the overall policy effects, we need a comparison with students entering high school below the 50th percentile cut-off before the policy was enacted—comparing pre- and post-policy cohorts with an interrupted time series design.

By combining the two designs—a regression discontinuity and a time series design we address limitations that are usually inherent in each. The weakness of cohort designs is that it is usually not possible to discern the extent to which post-policy changes are due to the policy versus other unrelated changes that may be occurring at the same time. The regression discontinuity strengthens the time series design by providing a within-cohort comparison; if outcomes are similar among students immediately below and above the 50th percentile pre-

policy, but diverge post-policy, this is good evidence that the observed changes were due to the policy. The cohort comparison strengthens the regression discontinuity by allowing us to test whether the policy affected outcomes among non-targeted students and to discern policy effects for students with entering test scores far from the national median. It also allows greater confidence that we are correctly modeling the underlying relationship between the outcome and the percentile scores from which the discontinuity is discerned by using information from pre-policy cohorts.

To discern the enrollment and policy effects using cohort comparisons combined with a regression-discontinuity we used multi-level models with students nested in cohorts nested in schools. The student-level model to estimate the outcome Y for student i in cohort j in school k is written as:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{BelowNorm})_{ijk} + \pi_{2jk}(\text{Percentile} - 50^{\text{th}} \text{percentile})_{ijk} + \pi_{3jk}(\text{Percentile_squared})_{ijk} + \sum_{p=1}^P \pi_{4+pjk}(X)_{ijk} + e_{ijk}, \quad (1)$$

where X is a vector of student-level control variables, including latent math abilities centered around the average of those at the 50th percentile, special education, age, gender, race, socio-economic characteristics, and mobility. These student-level variables are correlated with course grades and test scores and included to control for shifts in academic and demographic characteristics across the cohorts.

The intercept π_{0jk} represents the average outcome for students at the 50th percentile on the eighth-grade test in cohort j and school k , controlling for student background characteristics. The coefficient π_{1jk} is the discontinuity estimate, indicating the difference in the outcome between students who are just below the national median and students just above the national median in cohort j and school k . The coefficients π_{2jk} and π_{3jk} indicate relationships between the outcome and the percentile scores centered on the 50th percentile and its square term, respectively.

The cohort- and school-level models for π_{0jk} through π_{3jk} are specified as:

$$\begin{aligned}
 \pi_{0jk} &= B_{00k} + B_{01k}(Yr01)_{jk} + B_{02k}(Yr02)_{jk} + B_{03k}(Yr03)_{jk} + B_{04k}(Yr04)_{jk} + r_{0jk}, \\
 \pi_{1jk} &= B_{10k}(Yr00)_{jk} + B_{11k}(Yr01)_{jk} + B_{12k}(Yr02)_{jk} + B_{13k}(Yr03)_{jk} + B_{14k}(Yr04)_{jk} + r_{1jk} \\
 \pi_{2jk} &= B_{20k} + B_{21k}(Yr01)_{jk} + B_{22k}(Yr02)_{jk} + B_{23k}(Yr03)_{jk} + B_{24k}(Yr04)_{jk} + r_{2jk}, \\
 \pi_{3jk} &= B_{30k}, \\
 B_{00k} &= \gamma_{000} + u_{00k}
 \end{aligned} \tag{2}$$

We use year dummy variables to gauge trends in the outcome because there are only 3 pre-policy time points and the outcome trends may not be linear. The intercept B_{00k} for π_{0jk} represents the average outcome for students just at the cut-off in the base-line pre-policy year in school k , and the coefficients B_{01k} , B_{02k} , B_{03k} , and B_{04k} indicate differences in outcomes for later cohorts among students at the cut-off. The policy was initiated in 2003, thus, B_{03k} and B_{04k} represent post-policy changes in outcomes for students just at the cut-off (those not targeted by the policy). We use 2000 as the base year in the first set of models to show trends over time clearly. However, 2002 was the final pre-policy cohort and the closest comparison cohort for judging the policy. Thus, we report the degree to which the post-policy coefficients were significantly different from those in 2002, as well as from the base year.

The key parameter estimates for discerning effects of double-dose course enrollment on targeted students are B_{10k} , B_{11k} , B_{12k} , B_{13k} and B_{14k} ; these represent differences in the outcomes between just above- and below- norm students in each year. There is no intercept in this equation so that each year's discontinuity is captured in full by the coefficient, rather than as a deviation from the discontinuity in 2000. The coefficient B_{10k} shows whether the outcomes of students just below the cut-off differ from the outcomes of students just above the cut-off in the base-line year in school k ($H_0: B_{10k} = 0$, no difference in outcomes between students just above and below the cut-off in the base year). The coefficients B_{11k} , B_{12k} , and B_{13k} indicate whether cohort changes in the outcomes from the base-line year were different between students just below and just above the cutoff. Because pre-policy cohorts did not enroll in double-dose algebra, we expect no differences in the outcomes between students just below and above the cutoff pre-policy ($B_{10k} = B_{11k} = B_{12k} = 0$). However, if double-dose algebra enrollment improved student outcomes, we expect B_{13k} and B_{14k} to be greater than zero when predicting test scores and grades, and less than zero when predicting failures.

This analysis tells us post-policy double-dose algebra effects (counterfactual A) through coefficients B_{13} and B_{14} ($H_0: B_{13k} \neq 0, B_{14k} \neq 0$). These coefficients provide the discontinuity estimate among post-policy cohorts. The policy effects (i.e., the second counterfactual of experiencing the policy vs. no policy) can also be discerned from these analyses. Policy effects for targeted students (counterfactual B) can be discerned by calculating the extent to which outcomes for below-norm students changed in post-policy cohorts ($H_0: B_{03} + B_{13k} \neq 0, B_{04} + B_{14k} \neq 0$). Policy effects on non-targeted students can also be found by examining the extent to which outcomes shifted post-policy for students above norms ($H_0: B_{03} \neq 0, B_{04} \neq 0$).

For prediction of the coefficients on student control variables, π_{pjjk} , preliminary analyses determined whether cohort dummy variables and random effects were included at the cohort level. At the school level, a random intercept model was used for the intercept B_{00k} . For B_{01k} through B_{24k} , preliminary analyses determined whether they were allowed to vary. All the other B's for control variables are fixed at school levels without school predictors.¹⁶

Addressing the second limitation: Variations in policy effects by students' incoming abilities. The second methodological limitation to the regression discontinuity is that the estimated effects of double-dose algebra enrollment for below-norm students using the above design are applicable only to students in the narrow range around the 50th percentile. The comparison between taking double-dose Algebra and taking single-dose Algebra under the policy (i.e., counterfactual A) is only a logical comparison for students with test scores close to the national median.¹⁷ Furthermore, the regression-discontinuity analysis does not allow us to discern different policy effects by students' incoming abilities (research question 3). To address these limitations, we present an alternative analysis that uses pre-policy cohorts as the only comparison groups (i.e., no within-cohort comparisons). For this analysis, we create six percentile rank groups—three groups below and three above the 50th percentile. We then compare outcomes between post-policy students who were subjected to the policy and pre-policy students who did not experience the policy for each percentile group.

The following three-level statistical models are used with students nested within cohorts within schools. The student-level model is expressed as;

$$Y = \pi_{0jk}(\text{Rank1})_{ijk} + \pi_{1j}(\text{Rank2})_{ijk} + \pi_{2j}(\text{Rank3})_{ijk} + \pi_{3jk}(\text{Rank4})_{ijk} + \pi_{4jk}(\text{Rank5})_{ijk} + \pi_{5jk}(\text{Rank6})_{ijk} + \sum_{p=1}^P \pi_{5+pjk}(X)_{ijk} + e_{ijk}, \quad (3)$$

where X is a series of student-level control variables.

For each percentile rank group, the student outcome is a function of cohort, which is specified as;

$$\begin{aligned} \pi_{0j} &= B_{00k} + B_{01k}(\text{Yr01})_{jk} + B_{02k}(\text{Yr02})_{jk} + B_{03k}(\text{Yr03})_{jk} + B_{04k}(\text{Yr04})_{jk} + u_{0j}, \\ \pi_{1j} &= B_{10k} + B_{01k}(\text{Yr01})_{jk} + B_{12k}(\text{Yr02})_{jk} + B_{13k}(\text{Yr03})_{jk} + B_{14k}(\text{Yr04})_{jk} + u_{1j}, \\ &: \\ \pi_{5j} &= B_{50k} + B_{01k}(\text{Yr01})_{jk} + B_{52k}(\text{Yr02})_{jk} + B_{53k}(\text{Yr03})_{jk} + B_{54k}(\text{Yr04})_{jk} + u_{5j} \end{aligned} \quad (4)$$

The student-level models do not include an intercept so that coefficients π_{0jk} through π_{5jk} represent the average outcome of students for each percentile rank group in cohort j and school k .

The cohort-level equations include dummy variables for cohort year for each percentile group with the 2000 cohort as a reference category. The cohort-level intercept B_{00k} represents the average outcome for the lowest-ability students in 2000 in school j . For each percentile rank group, the coefficients for the cohort year dummy variables indicate differences in the outcome between 2000 and the given cohort year in school j .

The school level models for percentile rank variables do not include covariates and preliminary analyses determined whether B_{00k} through B_{54k} were allowed to vary. For student level control variables, preliminary analyses determined for which π to include cohort dummy variables and whether they are fixed at the cohort level. The school-level models for control variables were fixed at the school level without predictors.

If the policy had an effect, we should see that trends in math outcomes observed pre-policy change post-policy (Ho: $B_{03k} \neq B_{00k} \dots B_{53k} \neq B_{50k}$). The limitation of cohort comparisons is that the changes observed in post-policy cohorts might be due to other factors that occurred simultaneously without the policy. Therefore, we specifically looked for a divergence in trends for below- and above-norm groups at the year of policy implementation ($B_{03k}, B_{13k}, B_{23k}$ compared to $B_{33k}, B_{43k}, B_{53k}$). Such a divergence would suggest a policy effect, rather than a cohort effect, since the organizational effects of the policy were opposite for students below versus above the cutoff.¹⁸ While the earlier analyses with the regression-discontinuity design specifically tested this, the current analyses extend the comparison beyond the narrow range of students with scores close to the national median.¹⁹

Addressing the third limitation: Discrepancies in policy adherence. The third methodological limitation arises from less-than-perfect adherence to the policy. While double-dose enrollment was very strongly defined by the policy, about 20 percent of students did not follow the policy guidelines. Thus, by comparing students based on eligibility for double-dose algebra (i.e., the cutoff scores), we are estimating intent-to-treat effects, but not the effects of *enrollment in double-dose algebra*. To deal with the problem of policy compliance, we apply instrumental variable (IV) methods and conduct two analyses. The first analysis, referred to here as a “within-cohort analysis,” uses the cutoff score as an instrument to estimate the effect of enrolling in double-dose algebra vs. not enrolling under the policy (counterfactual A).

The second analysis uses policy year as an instrument to examine the enrollment effect among the target students (i.e., below-norm students) in comparison with their pre-policy counterparts who did not take double-dose algebra because the policy was not yet implemented (counterfactual B). We refer to this as a “cross-cohort analysis.” The IV estimators identify the local average treatment effects (LATE), or the average treatment effects among compliers, which is defined as the average treatment effect for subjects who are induced to take the treatment (i.e., double-dose coursework) on the account of variation in the instrument (Imbens and Angrist, 1994). Because the level of policy compliance differs between 2003 and 2004, the enrollment effects are separately estimated for each post-policy cohort in both analyses.

The two-stage least squares models for the within-cohort analysis are:

$$T_{ij} = \gamma_0 + \gamma_1 Z_{ij} + \gamma_2 X_{ij} + \gamma_3 W_j + \upsilon_i \quad (5)$$

$$Y_{ij} = \beta_0 + \beta_1 \hat{T}_{ij} + \beta_2 X_{ij} + \beta_3 W_j + \varepsilon_i \quad (6)$$

For student i in school j , T indicates treatment participation (i.e., enrollment in double-dose algebra), Z is the instrument (i.e., a dummy variable indicating whether students scored below the cutoff), X is a vector of student covariates, including ITBS math percentile scores centered around the 50th percentile, its square term, special education status, and demographic characteristics (age, race, and SES), and W is a vector of school fixed effects. The school fixed effects and a special education indicator, together with the instrument, allow for very good prediction of policy adherence, as discrepancies are largely defined by the ways that particular schools enrolled their students in double-dose Algebra and students’ special education status. The treatment effect β_1 represents the within-school average effect of double-dose enrollment on the outcome Y among compliers in a given post-policy year.²⁰

The cross-cohort analysis compares the outcomes of post-policy below-norm students to the outcomes of pre-policy below-norm students (i.e., policy vs. no policy for the targeted students). To examine the degree to which double-dose algebra effects differed by student incoming abilities, we estimate the treatment effect separately for each of the three below-norm percentile groups. For the first-stage equation, we use the equation 5, but the instrument (Z) here is the post-policy year dummy variable. In addition, because double-dose course enrollment depends on which schools students attended and their special education status, we include a vector of interaction terms between the post-policy and school dummy indicators, an interaction between the post-policy dummy variable (Z) and special education status, and a vector of three-way interaction terms among the post-policy dummy, special education status, and school dummy variables.

In each percentile rank group, we specify the outcome Y for student i in school j as;

$$Y_{ij} = \beta_0 + \beta_1 \hat{T}_{ij} + \beta_2 C_{ij} + \beta_3 X_{ij} + \beta_4 W_j + \varepsilon_i, \quad (7)$$

where C is a set of dummy variables distinguishing pre-policy cohort years, X is a vector of student covariates, including latent ability scores centered around the average score in that percentile group, special education status, and demographic characteristics (age, race, and SES), and W is a vector of school fixed effects. Analyses presented in earlier stages use 2000 as the reference category to show outcome trends as differences from the base-line year. However, doing so produces coefficients that compare post-policy cohorts to cohorts from two years prior to the policy. Thus, this analysis uses 2002 as the base year for estimating policy effects, as this is the last pre-policy cohort and we do not show pre-policy trends with this analysis.

Additional Causal Assumptions

Rubin's causal model makes several assumptions which must hold for making valid causal inferences. The stable unit treatment value assumption (SUTVA) states that potential outcomes associated with ones' treatment are not affected by the treatment status of other subjects. However, this study could violate SUTVA for two reasons. First, because the policy recommended schools teach double-dose students in the same algebra classrooms, the level of adherence to this guideline affects all algebra classroom compositions. Second, algebra classroom compositions are also affected by the extent to which schools enrolled their non-targeted students in double-dose algebra courses. For example, if schools enrolled many above-norm students in double-dose algebra, and they took the same algebra class with below-norm students, such classrooms would be more heterogeneous and have higher average abilities than classrooms that are composed of all below-norm students. This suggests that the potential outcome of a given student depends not only on whether he/she enrolled in double-dose algebra, but also on the enrollment status of other students and the extent to which schools followed the structural guideline of the policy. Because policy compliance varies by schools and policy year, our estimates of the enrollment and policy effects are specific to a policy year with a given level of policy compliance. In other words, the results are generalizable to schools, holding policy compliance levels constant. We also make the "intact schools" assumption where the average treatment effects are generalizable to schools, holding current student membership constant (Hong & Raudenbush, 2007).

Results

Intent-to-treat Effects among Students near the Cutoff Score

Table 3 displays the coefficients described in equations 1 and 2 from which we can discern the intent-to-treat effects. The coefficients in the top half of the table show the average math outcome in the base cohort (2000 cohort for all but the test scores), and the deviation from the base year for students who scored just above the 50th percentile on the eighth-grade test. The coefficients in the bottom half of the table show the degree to which outcomes were different for students who were just below the 50th percentile on the eighth-grade test. If there was a policy effect, we should see that math outcomes were similar pre-policy between students just above and just below the 50th percentile, but diverged in 2003 when the policy was implemented. In fact, this is exactly what we see in terms of ninth-grade (algebra) outcomes.

Table 3 about here

Algebra grades (GPA), failure rates, and test scores were similar between students just above and below the national median among pre-policy cohorts, but students just below the cutoff performed better post-policy than students just above the cutoff, particularly in 2003. For example, students who were just below the national median had similar algebra grades to students just above the national median in all of the pre-policy cohorts (the coefficients of -.05, .00 and -.03 were not significantly different from zero); but in the 2003 cohort students just below the national median received higher algebra grades by .20 GPA points ($p < .01$) and were less likely to fail by .18 logits (4 percentage points, $p < .10$) in 2003, compared to students just above the national median. These differences were somewhat smaller in 2004.

At first glance, the discontinuity that appears in 2003 may make it appear that the policy had beneficial effects on the grades of students entering high school with below-norm math skills. However, this is not the case. Post-policy differences in GPAs and failure between just below- and above-norm students largely occurred because the policy had adverse effects on above-norm students. Algebra GPAs had been improving and failure rates had been declining during pre-policy years for both just above- and below-norm students. The upward trends in algebra GPAs continued and algebra failure rates leveled off for students just below the cutoff post-policy. In contrast, above-norm students received *lower* GPAs and *higher* rates of course failure post-policy compared to 2002. The divergence in algebra grades between students just below and just above the national median can be more clearly seen in the top two panels of Figure 3.

Figure 3 about here

The third panel of Figure 3 shows the trends in test scores, and these look very different than the trends in grades. While algebra test scores were also similar between just below- and above-norm students among pre-policy cohorts, *both* groups of students had higher test scores post-policy than their pre-policy counterparts. The post-policy improvements were greater among students just below the cutoff, suggesting a double-dose benefit for those targeted by the policy. Students just below the cutoff scored .69 points higher in 2003 than their counterparts in 2002 (effect size=.26) while the pre- and post-policy difference among students just above the cutoff was .47 (effect size =.15).²¹

The evidence on policy effects on 10th-grade student outcomes is not consistent. Geometry course grades and failure rates were similar between students just below and above the national median pre-policy, but they diverged in 2003. While geometry GPAs were unchanged and course failure *increased* for students just above the national median in 2003, students just below the national median showed slightly higher GPAs in 2003 and their failure rates did not increase. There were no differences, however, in 10th-grade course outcomes in 2004.

Intent-to-treat Effects: Differential Effects by Students' Initial Abilities

The next set of analyses discerns the effects of the double-dose algebra policy by students' incoming abilities. Here, we compare the outcomes of post-policy students to those of pre-policy students by their percentile ranking on the eighth-grade test (see equations 3 and 4). We expect pre-policy trends to diverge for below- and above-norm groups at the year of policy implementation, if the policy had an effect.

As expected, before the policy was implemented, trends in outcomes look generally similar across all students, regardless of incoming abilities (see Table 4). Algebra GPAs slightly improved and course failure rates declined from 2000 to 2002. This pre-policy trend shifted with the first cohort subject to the policy. Among students entering high school with test scores above norms (high-ability students), algebra grades fell and algebra failure rates *increased* post-policy compared to the 2002 cohort. At the same time, algebra grades *improved* and failure rates did not change for below norm students—with the exception of the students in the very lowest ability group. Across the three above-norm (high-ability) groups, grades dropped by .10 to .18 points from 2002 to 2003 (the coefficients changed from .16, .12, and .19 to .06, .02 and .01, respectively), and failure rates increased by four percentage points (the differences from 2000

changed from $-.05$ and $-.04$ to $-.01$ and $.00$). Among all but the very lowest-ability below-norm groups, algebra grades increased by $.05$ to $.06$ points (the coefficients changed from $.16$ and $.17$ to $.21$ and $.23$, respectively), and failure rates were unchanged. However, the very lowest-ability students showed declines in algebra grades (by $.08$ points—from a coefficient of $.14$ to $.06$) and no significant changes in failure. The net effect, across all students was that GPAs declined very slightly by $.05$ points with the policy from 2002 to 2003, and failure rates increased slightly (an increase of 3%). For students targeted by the policy, including students far from the cut-off, there were only negligible changes in grades or failure rates with the policy (GPAs increased by $.01$ points and failure rates increased by two percentage points). Among above-norm students, GPAs declined by $.12$ points, on average, while failure rates increased by four percentage points.²²

The policy effects on students' grades and failure rates do not seem to support the aim of the policy. However, the results are very different when we look at students' test scores. Among students of all ability levels, algebra test scores were similar between the 2001 and 2002 (pre-policy) cohorts, but test scores improved among post-policy cohorts across all six ability groups. Students between the 20th and 50th percentiles made the largest improvements; in 2003 their scores improved by $.66$ to $.76$ of a point, compared to the 2000 cohort. This may seem like a small improvement; however, this is an artifact of the scale. The algebra subset of the exam has a standard deviation of 2.45, and on average, CPS students' math scores on the ACT exam sequence from ninth to tenth grade improve by only 1.87 points in a full year. Thus, a boost of $.76$ of a point is sizable. The net change in algebra scores across all six groups from 2002 to 2003 was $.56$, with an average increase of $.64$ for students targeted by the policy (below-norm students), and $.49$ for above-norm students.

Although the policy improved student achievement on the standardized test in algebra, 10th-grade course outcomes did not seem to improve as a result of the policy. Geometry grades were relatively stable across the cohorts, although students between the 35th and 50th percentiles showed a slight improvement in the first post-policy cohort ($p < .10$). Geometry course failure rates were stable during pre-policy years across all percentile groups. However, they increased among above-norm students between the 50th and 80th percentiles with the first post-policy cohort, and were higher among all students below the 65th percentile in the second post-policy year. The increase in geometry failure rates in the first post-policy cohort among above-norm students may be related to the increase in algebra failure rates among these same students. However, because the 2004 increase in failure rates is seen across both below and above norm students, we are hesitant to attribute it to the policy. Instead, it may reflect a change in school policies or emphases that occurred in this year.

Table 4 about here

The Effects of Double-dose Algebra Course Enrollment on Students who Complied with the Policy: Instrumental Variable Estimates

Our analyses so far did not account for incomplete compliance with the policy; some below-norm students did not enroll in double-dose algebra while some above-norm students did enroll in these courses. The next set of analyses examined the effect of the double-dose algebra enrollment itself on student outcomes among those who complied with the policy using instrumental variable methods. The first analysis examined the effect of enrolling vs. not enrolling in double-dose algebra post-policy (counterfactual A) by comparing students just

below and above the national median post-policy. The second analysis examined the effect of enrolling in double algebra versus not having double algebra classes available (counterfactual B) by comparing post-policy below-norm students who enrolled in double-dose algebra and pre-policy below-norm students who did not experience the policy. Note that the instrumental variables analysis provides our best estimates about the effects of enrolling in double-dose algebra among those who complied with the policy as opposed to simply being eligible for double-dose algebra (which is the estimate obtained from the cohort comparison and regression discontinuity). Furthermore, the instrumental variable analysis estimates the effects of enrolling in double-dose algebra as it was implemented in CPS schools, not the effects of 100% policy compliance.²³

Within-cohort analysis: Enrolling vs. not enrolling double-dose Algebra post-policy. The effect of double-dose algebra enrollment post-policy was estimated using two-stage-least squares models expressed in equations 5 and 6 for each post-policy year. Table 5 shows the treatment effect β_1 for each post-policy year as well as the intent-to-treat effect for a comparison. In this analysis, we re-estimated the ITT effects using the same model specification as the equation 6, but using the treatment assignment status Z instead of the predicted treatment status \hat{T} to provide an appropriate comparison.

Table 5 about here

As expected, the IV estimates are similar to, or stronger than, the intent-to-treat effects for each outcome. Students below the cutoff who enrolled in double-dose algebra had higher

GPA and lower failure rates in algebra and higher algebra test scores than students above the cutoff who did not take double-dose algebra. The IV estimates of double-dose effects on algebra grades and test scores were particularly stronger than the ITT estimates, suggesting even greater benefits from enrolling in double-dose algebra. Students who took double-dose algebra had grades that were a quarter of a point higher, on average, than similar students who took single-algebra, and their test scores were half a point higher (20% of a standard deviation). In the 2003 cohort, students who took double-dose algebra also had somewhat higher geometry grades and lower geometry failure rates than students who did not take double-dose algebra, but these differences were not seen in 2004. These results suggest that, post-policy, students eligible for double-dose algebra were better off taking it than taking regular single-period algebra in terms of their math performance and learning in their freshman year.

Cross-cohort analysis: Policy vs. no policy for the target students. The last set of analyses compared outcomes of post-policy below-norm students who took double-dose algebra to the outcomes of their pre-policy below-norm counterparts. We estimated equation 5 and 7 separately for each below-norm percentile group as defined earlier. For two outcome variables—algebra and geometry course grades—we used dummy coded variables for receiving a “B” or higher instead of continuous variables because course grades are not normally distributed among below-norm students.

Table 6 shows the differences in student outcomes between the 2002 cohort and post-policy cohorts among all students (i.e., ITT estimates) and policy-complying students (i.e., IV estimates). In general, the IV estimates are similar to the ITT estimates, reflecting the high compliance with the policy. With the exception of the lowest ability students (those below the

20th percentile), students who took double-dose algebra were more likely to get As and Bs in algebra and geometry, and had higher test scores, than similar students who entered high school before there were double-dose algebra courses. The size of the effects estimated from the instrumental variables models is larger than the intent-to-treat estimates, with students 2-4% more likely to receive a “B” or better in algebra and geometry if they took double-dose algebra, and averaging test scores that were about 0.8 points higher (an effect size of .3).

The instrumental variable models show very few positive effects of double-dose coursework for the lowest-ability students. Their test scores are higher, but the improvements are smaller than observed among students with average math abilities, and are modest in the second year of the policy. Very low-ability students were no more likely to get high grades in algebra or geometry as a result of double-dose algebra coursework, and were slightly more likely to fail. While double-dose coursework is often seen as a strategy to help students with very weak skills, these analyses suggest that the policy does not have substantial benefits for the lowest-ability students.

Table 6 about here

Discussion

Policy Implications

This policy was designed to address a clear problem in Chicago’s high schools—high failure rates in ninth-grade algebra. Research had shown a strong link between ninth-grade failure and eventual dropout, and district officials felt it was crucial to fix this problem. Yet, the strategy they developed—double-dose algebra—did not reduce failure rates. Failure rates remained the same for targeted students, and the policy unintentionally increased failure rates in

ninth-grade algebra for students who were not targeted by the policy. Many people in the district viewed the policy as a failure, and conversations with district personnel suggest that the policy was widely criticized. Why didn't this extended instruction in algebra have the expected outcomes?

One reason seems to be that math failure was the wrong outcome on which to judge this policy. This was a policy which sought to improve failure rates by improving students' algebra skills through more and better algebra instruction. It succeeded at doing this. Algebra test scores improved substantially—by almost a third of a standard deviation—among students who received double algebra classes.

However, this study shows that improvements in skills are not necessarily accompanied by improvements in grades and failure rates. Even though, based on their test scores, students entering high school post-policy learned more algebra than similar students in prior cohorts, their grades did not improve. Grades even declined among students with test scores above the national median. Although this seems counter-intuitive, there are several explanations for this pattern.

For high-ability students who did not enroll in double-dose algebra, the policy could only have affected their outcomes through the unintentional increase in tracking induced by the structural recommendations of the policy. High-ability students were in classrooms with higher-average ability levels post-policy. Therefore, individual students' abilities were relatively lower compared to the classroom average. This could have affected teachers' perceptions of the quality of students' work. It also could have affected students' self-perceptions, leading them to exert less effort. In addition, instructional content might have become more difficult if teachers perceived students better able to handle more complex material. Perceptual changes, and changes in course content and instruction, could explain why course grades declined among high-ability

students post-policy, even though their test scores improved. We are investigating these explanations in further work.

However, changes in classroom composition do not explain the lack of improvement in failure rates among low-ability students. Failure rates should have declined post-policy for below-average students, not only because their test scores increased, but also because their ability levels relative to classroom peers improved--especially for those near the 50th percentile. We suggest two potential explanations. First, even though classroom ability levels declined for below-norm students, content might have become more challenging. Their algebra teachers received new materials and professional development around algebra which may have led them to teach in a more demanding way. Their teachers may also have expected more of them because they were receiving twice as much instruction. Second, the policy may not have improved failure rates for the target students because it did not address some key reasons for course failure. Other work on Chicago schools has shown that course failures are related much more strongly to students' behaviors than to their skills, and course attendance is particularly important (Allensworth & Easton, 2007). Any reform efforts that address low skills will likely have limited effects on grades unless they address problems of low engagement and participation.

Implications for Understanding the Effects of Tracking

While this study began as an evaluation of double-dose algebra, it has implications for understanding the effects of tracking. Although tracking has been widely criticized for impeding academic progress of low-ability students, our study showed that students learned more in homogenous low-ability classrooms when they were provided with additional coursework and their teachers received new curricular resources. This study did not tell us whether tracking alone

would have adverse effects on low-ability students, but that any potentially adverse effects from tracking were offset by increased instructional time and teacher resources.

For high-ability students, the way in which this policy was implemented provided a natural experiment for observing the effects of tracking free of selection bias. Sorting students into homogenous high-ability classes did seem to be beneficial for high-ability students' test scores, which confirms the findings of many prior studies. However, few studies on tracking have examined the effects on grades, and here we see the opposite effect—tracking led high-ability students to have lower grades. Grades are at least as important as test scores for students' future success—they determine eligibility for sports, programs, scholarships and college enrollment. The trade-off in learning is not free of cost for students.

Rethinking the Policy for Students with the Weakest Math Abilities

The double-dose algebra policy was least effective for students with the weakest math abilities. These students showed no improvements in grades or failure rates. They did not benefit from changes in classroom composition—prior to the policy they were the lowest-skilled students in their classes, and they continued to be the lowest-skilled students post-policy. Test scores also improved less among very low-skilled students than among other targeted students, particularly in the second year of the policy.²⁴ Ironically, this is the group that is seen as most in need of academic support to be able to handle algebra.

Most students entering high school below the 20th percentile received special education services, and these students may have been particularly ill-served by this policy²⁵. Large-scale curricular reforms that are targeted to low-ability students need to consider their effects on students with learning disabilities, as they comprise a large proportion of low-ability students in

many schools. Prior to the policy, special education students often took algebra in small, self-contained special education classrooms. When the policy was implemented, many of the self-contained special education algebra classes disappeared, and most special education students took double-dose algebra instead, together with other below-norm students who were not eligible for special education services. It is possible that very low-performing special education students were not yet ready for algebra, and that their pre-policy self-contained classrooms spent more time on remediation. Or, double-dose algebra teachers might have lacked strategies to reach special education students in larger and more heterogeneous classrooms. Conversations with CPS personnel suggest that many teachers of double-dose algebra felt unprepared to teach classes with large percentages of special education students.

Methodological contributions

This study adds to the existing methodological literature on policy evaluation by showing how two quasi-experimental designs can be combined to provide multiple comparisons, thus allowing for greater confidence in results. An education policy could affect all students, even those not targeted by a policy, and using a regression discontinuity alone is particularly problematic in such a case. If we had not used the cohort comparison, we would have overestimated the effects of double-dose algebra on grades and failure, and underestimated the effects on test scores for the students targeted by the policy. Cohort comparisons without the regression discontinuity would not have been able to discern the policy effects from other potential cohort effects, or to show the effects of complying with the policy (counterfactual A). Either method alone would have yielded an incomplete understanding of the effects of the double-dose Algebra policy on student outcomes. However, the current research design was

possible only because we had the following two conditions: 1) the availability of longitudinal student-level data; and 2) a policy design with explicit eligibility requirements and nearly universal implementation that allowed for unambiguous contrasts.

This study also shows that evaluation of curricular reforms must recognize the multiple ways in which they affect students' outcomes, beyond simply exposing students to different content. Curricular reforms may bring about changes in how classes are programmed. The double-dose policy led schools to reorganize all math classrooms, and as a result, classroom compositions changed considerably for all students. The effects of changing classroom composition are not small. In fact, their effects on algebra scores for above-norm students appear to be almost as large as the effects of doubling instructional time for below-norm students. Furthermore, the policy had larger consequences on algebra grades and failure for above-norm students who were not targeted by the policy than for below-norm students. The details of how curricular policies are implemented and programmed may matter as much as the changes in instructional time and content.

Future work

In future work we will explore the mechanisms through which the double-dose algebra policy affected students' outcomes—the ways in which classroom composition affected academic outcomes, and the changes that occurred in students' reports of classroom climate and instructional practices. We will also look to see whether changes in the math curriculum had broader effects on students' performance in school—for example, whether the loss of an elective period led students to become disengaged from school (which was a concern expressed by CPS school staff),²⁶ or whether students' overall grades declined because they were taking two credits

in a subject in which they were particularly weak. Doubling students' algebra instruction and providing resources to their teachers was a successful strategy for improving students' algebra skills, but it brought many other structural changes to the curriculum of students targeted and not-targeted by the policy which we have just begun to understand.

References

- ACT. (2006). *Ready for College and Ready for Work: Same or Different?* Iowa City, Iowa.
- ACT, Inc. (2007). The PLAN Technical Manual. ACT, Iowa City, IA.
<http://www.act.org/plan/pdf/PlanTechnicalManual.pdf>
- Allensworth, E., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago, IL: Consortium on Chicago School Research.
- Allensworth, E., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago Public Schools*. Chicago, IL: the Consortium on Chicago School Research.
- American Diploma Project. (2004). *Ready or not: Creating a high school diploma that counts*. Washington, DC: Achieve, Inc.
- Anderson, L. W. (Ed.). (1984). *Time and school learning*. London: Croom Helm.
- Angus, D., & M. J. (1999). *The failed promise of the american high school, 1890-1995*. New York: Teachers college Press.
- Barr, Rebecca and Robert Dreeben. 1983. *How Schools Work*. Chicago: University of Chicago Press.
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29, 682-688.
- Bottoms, G. (2008). *Redesigning the ninth-grade experience: Reduce failure, improve achievement and increase high school graduation rates*. Atlanta, Georgia.: Southern Regional Education Board.
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.

- Chait, R., Muller, R. D., Goldware, S., & Housman, N. G. (2007). *Academic interventions to help students meet rigorous standards: State policy options*. Washington, DC.: The National High School Alliance at the Institute for Educational Leadership.
- Cremin, L. A. (1961). *The transformation of the school: Progressivism in American education*. New York: Knopf.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*, 1047-1085.
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Education Evaluation and Policy Analysis, 19*, 325-338.
- Gamoran, A., & Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education, 106*, 385-415.
- Herlihy, C., & Kennelly, L. (2007). *State and district-level support for successful transitions into high school*. Washington, D.C.: National High School Center.
- Horwitz, Amanda and Jason Snipes. 2008. *Supporting Successful Transitions to High School*. The Council of the Great City Schools, Washington, D.C.
http://www.cgcs.org/publications/CGCS_SuccessfulTransitions.pdf
- Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association, 81*, 945-970.
- Hong, G., & Raudenbush, S. W. (2007). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*(475), 901-910.

- Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467-475.
- Kemple, J. J., Herlihy, C. M., & Smith, T. J. (2005). *Making progress toward graduation: Evidence from the talent development high school model*. New York: MDRC.
- Lee, V.E. (2002). *Restructuring high schools for equity and excellence: What works*. New York: Teachers College Press.
- Mac Iver, D. J., Balfanz, R., & Plank, S. B. (1998). An 'elective replacement' approach to providing extra help in math: The talent development middle schools' computer- and team-assisted mathematics acceleration (catama) program. *Research in Middle Level Education Quarterly*, 22(2), 1-23.
- Millot, B. (1995). Economics of educational time and learning. In M. Carnoy (Ed.), *International encyclopedia of economics of education* (pp. 353-358). Oxford: Pergamon-Elsevier.
- National Center for Education Statistics. (2006). *The condition of education 2006*. Washington, D.C.: U. S. Department of Education.
- National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Governors Association Center for Best Practices. (2005) *Getting it done: Ten steps to a state action agenda*. Retrieved November 08, 2005 from <http://www.nga.org/Files/pdf/05warnerguide.pdf>.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.
- Oakes, J. (1994). More than misapplied technology: A normative and political response to Hallinan on tracking. *Sociology of Education*, 67, 84-89.

- Page, R. N. (1991). *Lower track classroom: A curricular and cultural perspective*. New York: Teachers College Press.
- Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.
- Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of high school. *American Educational Research Journal*, 36, 303-343.
- Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York: Wiley.
- (1999-2000). If tracking is bad, is detracking better. *American Federation of Teachers* (Winter 1999-2000), 1-7.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6, 34-58.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). 2007. *The Condition of Education 2000-2007*.
<http://nces.ed.gov/programs/coe/>
- White, P. A., Gamoran, A., Smithson, J., & Porter, A. C. (1996). Upgrading the high school math curriculum: Math course-taking patterns in seven high schools in California and New York. *Education Evaluation and Policy Analysis*, 18, 285-307.

Figures

Figure 1. Percent of students taking double-dose math by 8th grade ITBS math percentile
All students

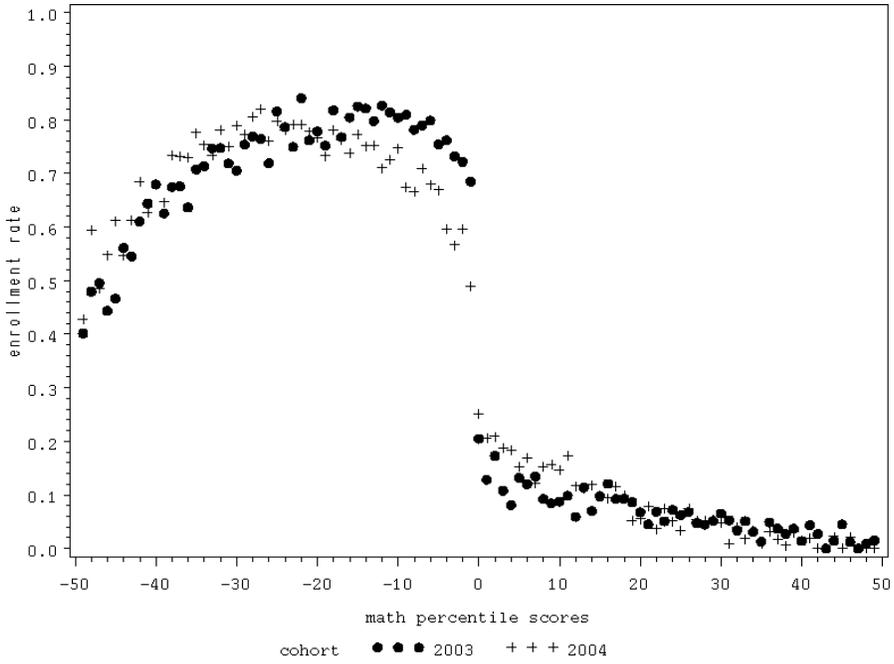


Figure 2. Percent of students taking double-dose math by 8th grade ITBS math percentile
Regular education students only

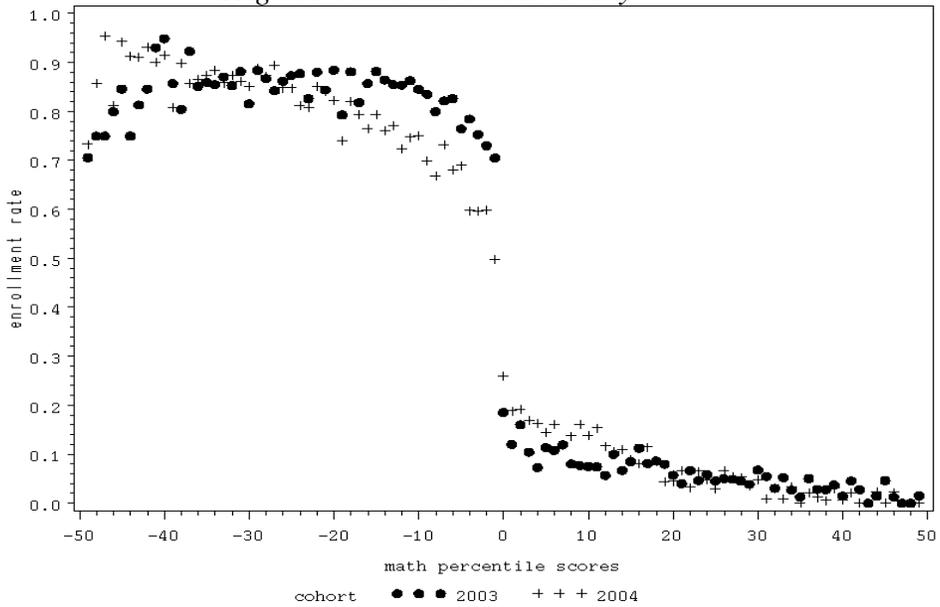


Figure 3.1. Cohort changes in algebra GPA for students just below and above the national median

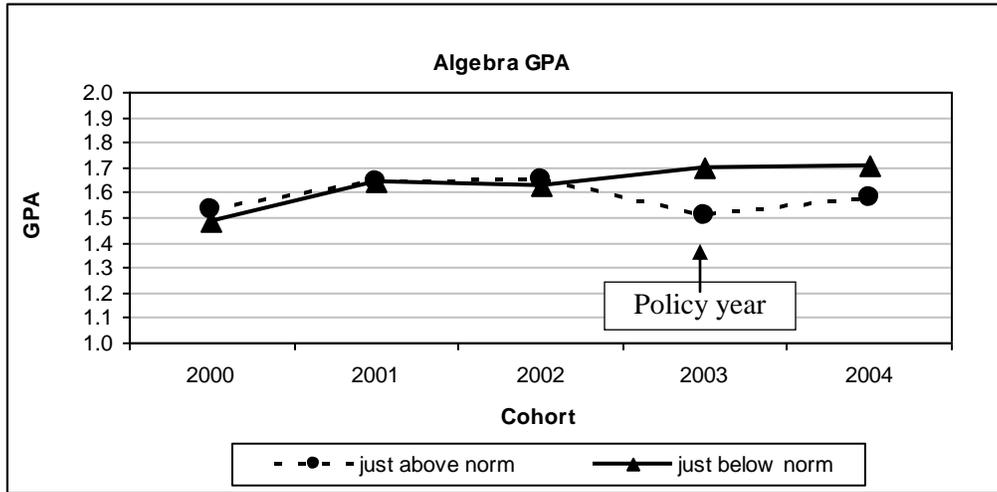


Figure 3.2. Cohort changes in algebra course failure for students just below and above the national median

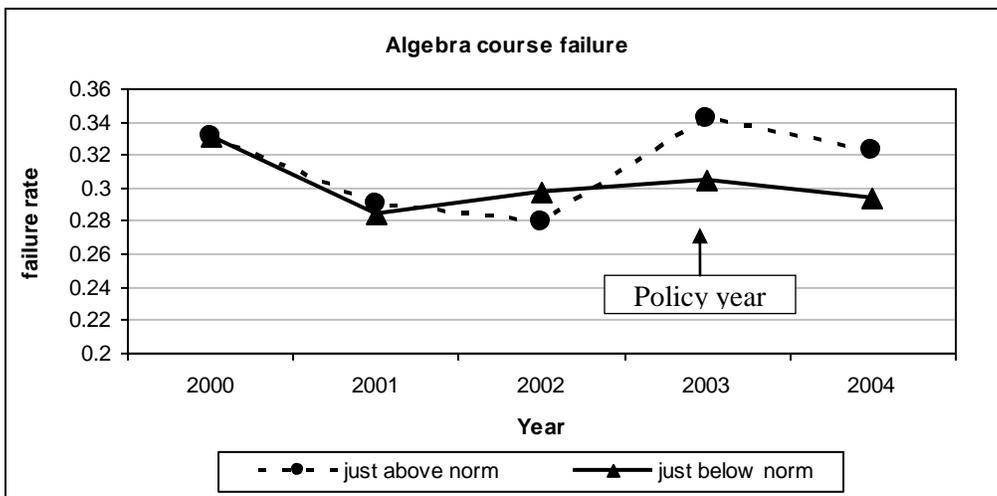
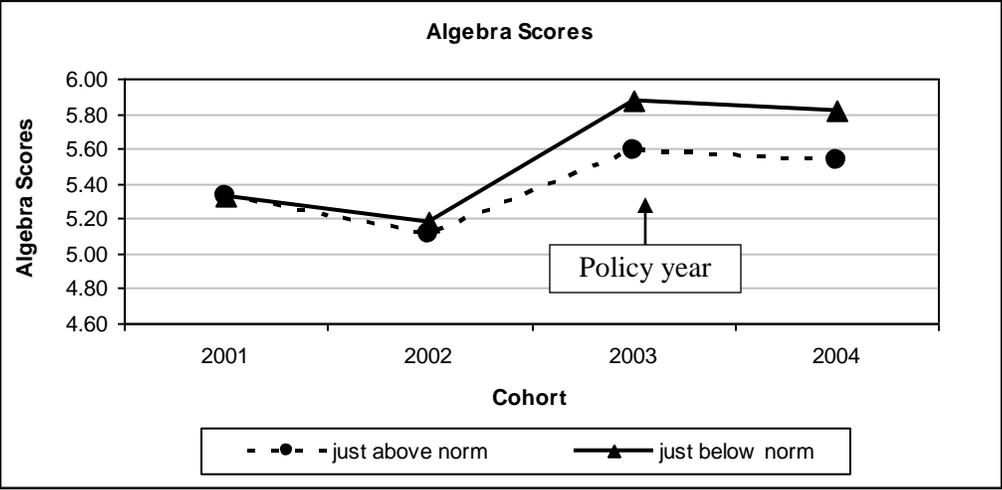


Figure 3.3. Cohort changes in Algebra scores for students just below and above the national median



Tables

Table 1. Compliance with three policy requirements by cohort year

Cohort	Percentage with the same teacher ¹	Percentage taking sequential courses ²	Percentage of double-dose students in the regular algebra class ³
2003	80.0	72.3	92.0
2004	54.0	48.0	86.6

¹ The percentage of double-dose algebra students taking support algebra and regular algebra with the same teacher

² The percentage of double-dose algebra students taking support and regular algebra sequentially

³ Among students taking double-dose algebra, the average percentage of students in their regular algebra class that are also taking algebra support

Table 2. Selected student characteristics by cohort

Variables	Cohort									
	2000		2001		2002		2003		2004	
	Mean	SD								
Below norm	0.58	0.49	0.59	0.49	0.50	0.50	0.53	0.50	0.54	0.50
Math ability (Z-score)	-0.09	0.87	-0.06	0.92	-0.04	0.95	-0.07	0.90	-0.09	0.92
Special Education	0.19	0.39	0.20	0.40	0.19	0.39	0.18	0.39	0.19	0.39
White	0.08	0.27	0.08	0.27	0.08	0.27	0.07	0.26	0.07	0.25
Hispanic	0.34	0.47	0.34	0.47	0.35	0.48	0.32	0.47	0.34	0.47
Asian	0.03	0.16	0.03	0.16	0.03	0.16	0.03	0.16	0.02	0.16
Poverty (Z-score)	0.01	0.99	-0.00	1.00	-0.03	1.07	0.03	1.00	-0.01	1.00
Number of students	17,276		17,681		17,623		19,512		20,055	

Table 3. Intent-to-treat effects on math outcomes (Regression discontinuity with cross-cohort comparisons)

	9 th -grade outcomes		10 th -grade outcomes		
	Algebra GPA	Algebra Failure (log odds)	Algebra Test Scores	Geometry GPA	Geometry Failure (log odds)
Intercept	1.53***	-0.71**	5.33***	1.54***	-1.20***
2001 Cohort	.11*	-0.19*		.04	-0.16~
2002 Cohort	.12*	-0.24**	-.22**	.01	-0.10
2003 Cohort (policy)	-.03	0.05	.25**	-.04	0.19*
2004 Cohort (policy)	.05	-0.03	.21**	-0.02	0.16~
<u>Below norm Deviation</u>					
2000 Cohort	-.05	0.00		-.05	-.04
2001 Cohort	.00	-0.03	-.01	.03	-.07
2002 Cohort	-.03	0.09	.07	-.03	.03
2003 Cohort (policy)	.20**	-0.18*	.29***	.08**	-.25*
2004 Cohort (policy)	.13***	-0.14~	.28**	-0.01	.02

~p<.10, * p<.05, ** p<.01, ***p<.001.

The intercept represents the average math outcome for students in the base year cohort (2000 for all outcomes but test scores, 2001 for test scores) who entered high school with test scores just above the cutoff (the 50th percentile). The coefficients for each cohort under the intercept show the change in outcomes from the base year for students just above the cutoff. The below-norm coefficient represents the difference in outcomes for students just below the cut-off (the 49th percentile). The coefficients for each cohort under the intercept show how the difference in outcomes between students just below and just above the cutoff changed, compared to the base year cohort (providing a difference in difference estimate).

Table 4. Intent-to-treat effects on math outcomes by 8th-grade math ability (Cross-cohort comparisons)

	Algebra GPA					
<i>8th Grade Percentile:</i>	<20 th	20 th -35 th	35 th -50 th	50 th -65 th	65 th -80 th	>80 th
2000 Cohort	1.11***	1.18***	1.40***	1.66***	1.95***	2.40
2001 Cohort	.09~	.10	.04	-.01	.05	.12
2002 Cohort (intercept)	.14*	.16*	.17*	.16*	.12~	.19**
2003 Cohort (policy)	.06_a	.21**_a	.23***_a	.06_a	.02	.01_a
2004 Cohort (policy)	.08_a	.25***_a	.27***_a	.13~_a	.10	.08_a
	Algebra Failure (probability)					
<i>8th Grade Percentile:</i>	<20 th	20 th -35 th	35 th -50 th	50 th -65 th	65 th -80 th	>80 th
2000 Cohort	.45*	.44**	.36***	.30***	.24***	.14***
2001 Cohort	-.03	-.04	-.02	-.01	-.03*	-.02
2002 Cohort (intercept)	-.05	-.05~	-.04~	-.05**	-.04***	-.04***
2003 Cohort (policy)	.01	-.05~	-.04	-.01_a	.00_a	.00_a
2004 Cohort (policy)	-.01	-.07*_a	-.05*	-.03_a	-.01_a	.01_a
	Algebra Test Scores					
<i>8th Grade Percentile:</i>	<20 th	20 th -35 th	35 th -50 th	50 th -65 th	65 th -80 th	>80 th
2001 Cohort	3.83***	4.30***	4.97***	5.68***	6.60***	8.57***
2002 Cohort (intercept)	.02	.07	-.05	-.05	-.11	-.17
2003 Cohort (policy)	.51***_a	.76***_a	.66***_a	.49***_a	.38**_a	.25~_a
2004 Cohort (policy)	.18*_a	.60***_a	.56***_a	.48***_a	.50***_a	.35*_a
	Geometry GPA					
<i>8th Grade Percentile:</i>	<20 th	20 th -35 th	35 th -50 th	50 th -65 th	65 th -80 th	>80 th
2000 Cohort	1.17***	1.25***	1.40***	1.68***	1.95***	2.38***
2001 Cohort	.02	.03	.05	-.02	.01	.10
2002 Cohort (intercept)	.05	.00	.06	.01	.01	.07
2003 Cohort (policy)	.02	.06	.13*	.04	.01	.02
2004 Cohort (policy)	-.09	.00	.04	.00	.04	.06
	Geometry Failure (probability)					
<i>8th Grade Percentile:</i>	<20 th	20 th -35 th	35 th -50 th	50 th -65 th	65 th -80 th	>80 th
2000 Cohort	.32***	.31***	.25***	.20***	.15***	.08***
2001 Cohort	.01	.00	-.01	-.01	-.02	-.01
2002 Cohort (intercept)	-.01	-.01	-.03	-.03	-.02	-.01
2003 Cohort (policy)	.03	.00	-.01	.02_a	.01_a	.00
2004 Cohort (policy)	.07***_a	.03*	.03_a	.02_a	.01	.01

Note. ~p<.10, * p<.05, ** p<.01, ***p<.001.

'a' indicates difference from 2002 cohort at p<05

Table 5. The Effects of enrolling vs. not enrolling in double-dose Algebra coursework post-policy (Counterfactual A)
Instrumental variables models (IV) compared to the intent-to-treat effects

Cohort	9 th -grade outcomes				10 th -grade outcomes					
	Algebra GPA		Algebra Failure (probability)		Algebra Test Scores		Geometry GPA		Geometry Failure (probability)	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV	ITT	IV
2003	.18***	.29***	-.03*	-.03**	.26***	.40***	.07*	.11*	-.04*	-.04*
2004	.11**	.25**	-.03+	-.03	.24***	.55***	-.01	-.02	-.01	-.01

Note. ~p<.10, * p<.05, ** p<.01, ***p<.001.

Table 6. The Effects of policy vs. no policy on students entering high school with below-norm math skills (Counterfactual B)
Instrumental variables models (IV) compared to the intent-to-treat effects

Cohort	9 th -grade outcomes				10 th -grade outcomes					
	Alg. GPA>B (probability)		Algebra Failure (probability)		Algebra Test Scores		Geo. GPA>B (probability)		Geometry Failure (probability)	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV	ITT	IV
<i><20th percentile</i>										
2003	.00	-.01	.01	.03*	.54***	.53***	.01	.02	.04+	.02
2004	-.01	-.01	.00	.02**	.17*	.14+	.00	-.01*	.06**	.04***
<i>20th-35th percentile</i>										
2003	.02**	.03**	.00	.01	.69***	.82***	.02+	.03**	.02	.02+
2004	.02**	.03**	.00	.00	.51***	.61***	.03*	.04***	.04**	.03**
<i>35th-50th percentile</i>										
2003	.03**	.03*	.00	.01	.70***	.85***	.03+	.03**	.01	.00
2004	.02*	.03**	.00	.00	.60***	.83***	.01	.02*	.05***	.04**

Note. ~p<.10, * p<.05, ** p<.01, ***p<.001

¹ Seventy-percent of twelfth-graders in the U.S. wanted to obtain a Bachelor's degree in 2006 (National Center for Education Statistics, 2006).

² Kemple et al (2005) used the comparative interrupted time series design where the Talent Development schools were matched with sets of comparison schools that are similar in several dimensions including racial compositions, ninth-grade promotion rates, mean test scores and attendance rates.

³ We refer to the Chicago policy as the "double-dose" policy to differentiate it from double-period (blocked) algebra, and from the mathematical term double-algebra. However, the district did not use the term "double dose." CPS staff members refer to the policy as "double-algebra."

⁴ We spoke with CPS staff members who were in charge of the double-dose algebra program in 2003 to learn the goals of the policy and the rationale for its structure. We were told that the district initiated the policy after learning about the high failure rates in the freshman year in the Roderick & Camburn (1999) study.

⁵ The Iowa Tests of Basic Skills

⁶ In a future study, we will look into the ways in which classroom instruction changed with this policy, using data from surveys of students on their perceptions of course difficulty, teacher personalism, and classroom assignments. However, while we have data on students' perceptions of their classroom environments, we do not know which curricular programs individual teachers used, how they organized the two periods for instruction, or the content that they covered in their classes.

⁷ The average classroom standard deviation of ITBS percentile scores fell from 18 to 14 for students with entering test scores just below the national median. Details of these changes are provided in another article (insert citation after review).

⁸ Average classroom peer ability levels declined by about .16 standard deviations for students testing just below the national median, and increased by about .12 standard deviations for students testing just above.

⁹ The district made new curricula and professional development available only to teachers teaching double-dose algebra. There was a possibility of spillover effects for teachers in regular algebra. However, the professional development was geared towards helping teachers structure two periods of instruction. In addition, CPS officials told us that, based on their observations of classrooms, they found that even teachers who taught both single-period and double-dose algebra tended to differentiate their instruction between the two types of classes. They tended to use new practices with the double-period class, but continued to use traditional methods with the single-period class. Teachers told them that they did not feel they needed to change methods with the advanced students, and that they were hesitant to try new practices that may be more time-consuming with just a single period. The double period allowed them to feel like they had the time to try new practices (e.g., cooperative groups).

¹⁰ In fact, absences and self-reported studying behaviors are eight times more predictive of failure than are incoming test scores (Allensworth & Easton, 2007).

¹¹ We asked administrators at CPS why this cut-off was chosen, and they replied that it was arbitrary; they looked to see if there was a point at which achievement levels were associated with a much higher failure rate, but found the relationship to be fairly linear.

¹² We are currently examining the ways in which classroom academic compositions affected students' perceptions of academic press, teacher personalism, and instructional tasks, and how those changes may have mediated policy effects on students' academic outcomes.

¹³ Percentile scores are not a consistent measure of achievement across cohorts of students taking different forms of the exam. They are not always properly equated to adjust for differences in the difficulty levels across test administrations. This inconsistency causes fluctuations in the relationship of incoming achievement with student outcomes across cohorts. Furthermore, one test score taken at the end of eighth grade is not the best measure of true ability. Students can have a bad day on the test (e.g., from illness), or get a few questions right or wrong out of luck. To better measure students' true math ability, we first equated students' raw scores on the ITBS through Rasch analysis. This removed form and level effects across different versions of the test. Then, to adjust for abnormally high or low performance on the eighth-grade test, we modeled students' growth trajectories from grades three through eight using the equated ITBS scores, with years nested within students. The 8th-grade latent scores were the estimated score from this model. Preliminary analyses showed inconsistent relationships between incoming achievement and students' later outcomes using the percentile scores, but these inconsistencies were not seen when latent scores were used. The latent scores allow for a consistent and precise measure of incoming math abilities across cohorts.

¹⁴ These SES measures provide a different value for students who live in different census block groups. While some students live in the same census block, these variables are much better at distinguishing economic status among students than the commonly-used indicator of whether students qualify for free and reduced lunch. Over 80% of CPS students qualify for free/reduced lunch so this variable provides little information. On the other hand, the variables based on census block show vast differences in the economic conditions of students, even among those who qualify for free/reduced lunch. Our SES indicators are also strongly related to student outcomes, more so than free/reduced lunch eligibility. There were 2450 census block groups represented among CPS students in 2004.

¹⁵ Particularly with this policy, extrapolating results beyond a narrow range around the cut-off score is problematic because who receives or does not receive double-dose algebra coursework influences classroom compositions, and this in turn affects student outcomes. While we can use the regression-discontinuity analysis to estimate the effects of a few students changing their algebra enrollment, the effects are not generalizable to a case of many students changing their enrollment because that would affect classroom compositions and, thus, the compositional effects of the policy would be different than those estimated here.

¹⁶ To assure that we controlled for all time-invariant unobserved school factors, the preliminary analysis used school fixed effects. All results were similar to the results from random effect models.

¹⁷ Counterfactual A tells us what would happen if a student enrolled in single-period algebra, when eligible for double-dose algebra, or if the cut-off were changed slightly. Students far from the cut-off are unlikely to be affected by these conditions. Furthermore, if many students far from the cut-off failed to follow the policy guidelines the changes in classroom composition would be different from those observed here, and this would change the estimates of policy and enrollment effects.

¹⁸ The policy brought a decline in average peer ability levels for students entering high school with below-norm skills, while it increased the average peer ability levels for students entering high school with above-norm skills.

¹⁹ The disadvantage, in comparison to the regression discontinuity, is that the below- and above-median groups are less comparable than students immediately above and below the national median.

²⁰ Equation 6 is essentially the same as the equation 1, except for that we use \hat{T} instead of Z (i.e., *BelowNorm* in the equation 1).

²¹ Pooled standard deviations are 1.69 for below norm students and 2.37 for above norm students.

²² The net effects across all students and targeted/non-targeted students are calculated as the weighted average effects, comparing outcomes in 2003 to outcomes in 2002, across the groups represented in Table 4.

²³ If schools had 100% compliance to the policy, it would create greater levels of tracking than observed. This would cause different classroom compositional effects than observed, and, therefore, different overall policy effects. Instead, the IV analysis shows the effect of taking double-dose algebra under the policy as it was actually implemented.

²⁴ It is possible students in the lowest-ability group were not measured well by the test, and that their ability levels were so low that the test was insufficiently sensitive to any improvements in their skills. Even if this is the case, it suggests that the double instruction in algebra was not successful at bringing their achievement up to a level that would be minimally expected at the end of ninth-grade algebra. This is not strong support for the policy for very low-skill students.

²⁵ We conducted an additional analysis to see whether removing special education students changed our conclusions. However, removing these students is the same as removing students below the 20th percentile score. Thus, it did not change the pattern of our results.

²⁶ When the policy was implemented, staff at many schools expressed concern that students would have less motivation to attend school if not able to take elective classes that they enjoyed.