The Effects of Tracking with Supports on Instructional Climate and
Student Outcomes in High School Algebra

Takako Nomi
Elaine M. Allensworth

Consortium on Chicago School Research
University of Chicago

May 18, 2010

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

Abstract

Tracking has been criticized for impeding the academic progress of low performing students; however, eliminating tracking has also been shown to have negative consequences, particularly for high achieving students. This study examines the consequences of a policy which sorted ninth-grade algebra classes by students' abilities, but provided doubled instructional time and additional supports for low ability students and their teachers. Results show that low-ability students received more demanding instruction and better pedagogy; these benefits helped mitigate negative effects of concentrating students with behavioral problems together. High ability students received more demanding instruction and had fewer classmates with behavioral problems, and this led to improvements in their test scores, but higher failure rates.   We discuss implications for tracking practices.

There has been a fierce debate over whether tracking should be eliminated from U.S. high schools. Tracking places students in differential academic trajectories based on their initial academic skills, interests, and future occupational paths. Although historically tracking has been the norm in American comprehensive high schools, it has been widely criticized for impeding academic progress of low performing students (Gamoran, 1987; Oaks, 2005; Powell, Farrar, & Cohen, 1985; Rosenbaum, 1976).[1] For example, low-track classrooms are often characterized as having low-level content, low expectations, and poor instructional environment (e.g., Oakes, 2005; Gamoran & Mare, 1989; Lucas, 1999; Powell, Farrar, & Cohen, 1985; Rosenbaum, 1976). Also, teachers tend to spend more time drilling or dealing with behavioral problems in low-track classrooms, while spending more time on critical thinking in high-track classrooms (Oakes, 1985; Page, 1991; Rosenbaum, 1976).

To address problems of tracking, a number of schools and districts have eliminated curriculum tracks, mixing students of different ability levels into the same class. In principle, all students in detracked classes receive curriculum and instruction at the same level and rigor as those in college-preparatory classes (Oakes, 1985; Wheelock, 1992). However, schools' detracking efforts often encounter various challenges, both within and outside schools. For example, teachers in detracked schools often continue to believe that tracking is necessary to address variability in students' academic skills. Many teachers struggle with providing effective instruction in mixed-ability classrooms and continue to hold low expectations for students after classes are de-tracked. Also, schools often face resistance from middle-class parents because they believe tracking benefits their children if they are placed in high-track classrooms (Gamoran & Weinstein, 1998; Oakes, 1994; Wells & Oakes, 1996; Rubin, 2008). Some schools have successfully detracked classrooms and improved instruction for low-ability students, but

characteristics of such schools are exceptional—a shared belief in diversity among staff, successful professional development that led teachers to use inclusive pedagogical practices, and additional supports for struggling students (e.g., extra support courses) (Boaler & Staples, 2008; Oakes, 2005; Rubin, 2008).

Besides being difficult to implement, detracking may actually have some negative consequences on academic achievement. The most able students in detracked classrooms are more likely to become bored and disaffected than they had been in tracked classes as teachers typically lower instructional levels to accommodate lower ability students (Rosenbaum, 1999). In a study of detracking that used NELS data and made adjustments for student selection bias and teacher qualifications, Argys, Rees, and Brewer (1996) found that achievement of high-ability students declined when moved out of high-ability tracks. Also, in Massachusetts, detracking resulted in fewer students performing at "proficient" and "advanced" levels on state tests, compared to schools that kept tracking students. There may also be disadvantages to low-ability students from placement with high-ability students, such as negative effects on their self-esteem (Loveless, 1999). Thus, while tracking seems to lead to poor instructional climates for low-skill students, merely eliminating tracking is not clearly preferable.

Some writings suggest an alternative approach—keeping homogeneous classes for organizational reasons, but ensuring that low-track students receive more challenging coursework and better instruction than they would receive under traditional tracking (Hallinan, 1994; Loveless, 1999). In 2003, such an approach was put in place by the Chicago Public Schools (CPS) when they implemented a new algebra policy. The policy required ninth-grade students with incoming math scores below the national median to take two periods of algebra—regular algebra plus an algebra-support course, and provided curricular resources and

professional development to teachers of algebra support courses to help them effectively use two periods for algebra instruction. This policy led schools to separate students into algebra classes based on whether their incoming math scores were above or below the national median. As a result, the policy intensified tracking in algebra classes, but students in low-track classes received twice as much instruction in algebra, and their teachers received professional development and resources to help improve algebra instruction.

In an earlier study, we showed that this policy led to improved Algebra test scores for both below norm and above norm students, but no improvements in pass rates. We now examine why this occurred. This study shows how this approach affected the classroom learning environment, the pedagogy, academic demand, and the concentration of students with behavioral problems, and how these changes in classroom environment affected students' math achievement.

This research adds to the existing literature on tracking in several ways. First, most research on tracking and classroom learning environments comes from qualitative studies where the number of students is often very small. Here we use a large quantitative data to test relationships between classroom academic composition and learning environment. Second, while most research has focused on test scores as a measure of academic outcomes, this study examines both course failure and test scores, as they may have differential relationships with classroom academic composition. This is an important limitation in prior research because course grades and failure are shown to be much better predictors of high school graduation, college attendance, college graduation, and future earnings than test scores (Allensworth & Easton, 2007; Bottoms, 2008; Miller, 1998; Noble & Sawyer, 2002). Third, prior research on tracking is likely to suffer from selection bias at both school and student levels. School-level

selection bias could occur if schools that track differ from detracked schools in unmeasured ways which are also related to students' outcomes. Student-level selection may occur if students select into schools with a particular curriculum, or into particular tracks within schools for reasons not observed by researchers (e.g.,motivation, parental support) which are also correlated with students' outcomes. The exogenous changes in curriculum and academic composition brought by the policy allow us to address selection bias to determine the effects of tracking.

The Double-Dose Algebra Strategy in Chicago

The double-dose algebra policy[2] in Chicago Public Schools (CPS) 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 intended to provide all students with a curriculum that would prepare them for college, eliminating inequality in course enrollment (Authors, year). However, many CPS students entered high school with math skills well below grade level, and failure rates were particularly high in ninth-grade algebra (Roderick & Camburn, 1999). Thus, to improve high failure rates in algebra the district instituted a new policy in 2003, requiring first-time ninth graders with 8th-grade math scores below the national median on the Iowa Tests of Basic Skills (hereafter referred to as "below-norm students") to enroll in two periods of algebra—a support algebra course and a regular algebra course.

The policy also provided curricular resources and lesson plans for double-dose algebra teachers. Prior to the policy, algebra curricula varied considerably across schools in CPS. The double-dose Algebra policy attempted to improve coherence in algebra curricula across the district 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. Additionally, the district ran professional development workshops three times a year for double-dose algebra teachers where it provided suggestions on how to use the two periods for Algebra instruction.

In addition, the policy guidelines encouraged schools to offer the courses sequentially in the day, with the same teacher and the same students. To follow these guidelines, most schools sorted above-norm and below-norm students into separate algebra classes—a single period algebra for above-norm students and double-period algebra for below-norm students. As a result, peer ability levels for below-norm students declined considerably post-policy, while above-norm students were more likely to have higher ability peers post-policy than were pre-policy students with similar incoming test scores (see Table 1).

The double-dose algebra in Chicago separated algebra classes by students' skills for instructional purposes, providing additional instructional time and supports for low-skill (below-norm) students. Additional instructional time allowed more flexible time use for instruction in double-dose algebra classes, which made teachers more likely to try the new practices suggested in the professional development. According to CPS internal and external evaluations[3], teachers who taught support classes reported that they were able to focus on skills that students lack and cover materials in a different order than simply following the textbook (Starkel, Martinez, and Price, 2006; Wenzel, Lawal, Conway, Fendt, and Stoelinga, 2005). Teachers were also concerned that students who do not like mathematics would be disengaged from class if they were required to take two periods of math. To facilitate students' engagement teachers tried to minimize time for lectures and use instructional activities, such as working in a small group, asking probing and open-ended questions, and using board work. External observers also

reported that support course teachers used more time in these interactive activities than regular algebra teachers who tended to spend more time giving lectures and letting students work individually.

Our earlier study, evaluating the overall policy effects on students' outcomes, showed that the double-dose algebra policy benefited the learning of all students; both below-norm students who were required to enroll in double-dose algebra, and above-norm students who took single algebra, showed significantly higher test scores post-policy (Author, year). However, course grades improved very little for below-norm students, while course grades *declined* for above-norm students (Author, year). The current study builds on this earlier works to explain the pattern of results; we examine how the policy affected classroom instruction and learning environment, and how such changes in instruction and learning environment affected students' test scores and pass rates.

Table 1 about Here

Conceptual Framework

There are three key mechanisms by which a double algebra strategy—tracking with additional supports for struggling students— could have affected students' outcomes: 1) expanded instructional time; 2) improvements in instructional content and pedagogy resulting from curricular resources, professional development, and expanded instructional time for teachers; and 3) ability grouping into more homogeneous classes. The first two mechanisms only directly affected low-skill students who enrolled in double algebra. However, because some teachers of double-algebra also taught high-ability students in single-period algebra classes, it is

possible that the professional development they received for double-dose classes affected instruction in their other algebra classes. Furthermore, some students who took support (double) algebra took their main algebra class with mixed-ability peers. By taking support algebra, they may have been less likely to hold back the pacing of their main algebra class. This would be a second potential spillover effect on high-ability students. The third mechanism, ability grouping, affected both low- and high-skill students.

The importance of instructional time for student learning is discussed in a number of places (Anderson, 1984; Bloom, 1974; Millot, 1995). For low-skill students who took two-period algebra (i.e., below-norm students), extended time would allow greater content coverage, greater instructional time on the same content, more flexible time use for instruction, and more time for students to learn the material. Professional development and curricular resources with lessen plans were intended to help teachers improve instructional practices with extended time. In this context, grouping low-skill students in the same classroom might have allowed teachers to target instruction at the skill level of their students. Thus, even though tracking is often thought of as detrimental for low-skill students due to low-level content coverage, low expectations, poor instruction, and disciplinarily problems in low-track classrooms, such problems may be mitigated if low-skill students and their teachers are provided with additional time and supports.

For above-norm students, ability grouping was the only mechanism that could have directly affected their academic outcomes: they continued to enroll in single-period algebra and their teachers received no additional resources, but their academic composition improved considerably because fewer low-skill students attended their algebra classes post-policy. If teachers adjusted instruction to higher skill levels of students, post-policy above-norm students would receive more challenging instruction. In addition, grouping students by ability could have

led single-period classes to have less disruption and better overall attendance, given that one criticism of low-track classes is that they have a disproportionate number of students with behavior problems. This also could have lead to greater learning for above-norm students.

While the double-dose algebra policy improved learning, it was not accompanied by improvements in course grades or failures. Grades often reflect more than content mastery because teachers consider multiple factors when assigning them, such as attendance, assignment completion, engagement, and performance relative to other students. Prior studies have shown that students tend to receive lower course grades in classrooms with higher ability peers even though they are likely to have higher test scores in high-ability classrooms (Farkas, Sheehan, & Shuan, 1990; Kelly, 2008). To explain these relationships, researchers have often focused on a "frog pond effect"—the classroom contextual effect on teachers' perception of the student, suggesting that teachers assign higher grades to students who look better in their classes relative to their peers. Few studies, however, have examined how classroom academic composition is related to other instructional factors, such as content difficulty and course demand, and how this, in turn, affects students' course grades. For example, students could receive lower grades in high-ability classrooms because they are taught more challenging materials at a faster pace, and this may make them less likely to complete or do well on their assignments.

Research Questions

This study examines how a double-dose algebra strategy—tracking with support for low-skill students—affected classroom learning experiences for both high-skill and low-skill students, and how these experiences mediated the effects of the policy on students' achievement (grades and test scores). The first set of questions addresses the policy effects on classroom

learning environments for high- and low-skill students. We ask: for high-skill students, did classroom instruction and environment improve as a result of taking algebra with high-ability peers? For low-skill students, did classroom instruction and environment improve or decline as a result of taking two periods of algebra with low-ability peers? We then ask: for both high-ability and low-ability students, how did the changes in classroom academic composition and instruction/learning environment brought by the policy affect their test scores and failure rates?

Data and Methods

Data

This study uses data on two cohorts of first-time ninth graders in the Chicago Public Schools (CPS). Chicago has the third largest school district in the nation and serves predominantly low-income and minority students. Approximately 85 percent of students are eligible for free/reduced lunch programs. The racial-ethnic composition is 54% African-American, 34% Latino, 9% white and 4% Asian.

We use 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 conditions, including educational attainment, occupational levels, poverty and employment status of residents in students' residential block groups.

Semester-by-semester course transcript and grade data files contain detailed class information, including teacher IDs, class periods, subject names, subject specific course codes, and course grades. These were used to classify students' algebra courses and group them with their classmates. These files also provided information on the number of absences students had

in each of their classes. Elementary achievement test scores are based on the Iowa Test of Basic

Skills (ITBS), taken in third through eighth grades. Disciplinary files were used to calculate

students' disciplinary records, and to corroborate information on disciplinary problems gathered

through the surveys. High school achievement test scores come from the PLAN exam, a test that

is part of the EPAS system developed by ACT, Inc. which all CPS students take in the fall of the

tenth grade. Surveys of students conducted biannually by the Consortium on Chicago School

Research provide information about the climate and instruction in math classrooms, including

instructional activities, academic demand, and students' disciplinary problems.

Sample

Our analyses use two cohorts of first-time ninth-grade students—one pre-policy (2002-

03) cohort and one post-policy (2004-05) cohort of students who responded to questionnaires

about their math classrooms on the biannual survey. We restrict our analyses to students in

schools that were in existence in both time periods to make comparisons between two cohorts of

students in the same school. We exclude students who received special education services

because many of them were exempt from double-dose algebra and they often enrolled in self-

contained special education classrooms pre-policy, which would not be comparable to typical

pre-policy algebra classrooms attended by regular education students.

While the biannual survey was given to all CPS students in the spring semester,

questionnaires about Math classes were administered to a subset of students. In the spring 2003

survey, students were randomly selected to respond to either English or Math questionnaires. In

the 2005 survey, students were asked whether they had English or Math classes first on Monday,

and, then continued to answer questions on the marked class. Among ninth-grade regular-

education students, the overall survey response rates were 58 percent for the 2002-03 cohort and 67 percent for the 2004-05 cohort. Of survey respondents, 50 percent responded to math questionnaires in 2003 and 43 percent in 2005. We were concerned that differences in the survey response rates between the two cohorts might bias our results. Survey respondents were similar to the general ninth-grade population in academic and demographic characteristics, and the two cohorts of survey respondents had similar characteristics to each other (see Table 2). However, two cohorts of students may differ in unmeasured ways which also affect their outcomes. Therefore, we examined the potential for response bias by replicating the analyses of compositional effects with the population of CPS students (not just survey takers) to determine if the estimates were similar to those obtained when survey data were included.

Table 2 about here

One additional restriction was made for the main part of our analyses; this was to limit the analyses to students who adhered to the policy—below-norm students who enrolled in double-dose algebra and above-norm students who enrolled in single-period algebra. By making this restriction, we attempted to estimate the policy effects for policy-complying students.[4] Excluding students who did not take the required course made it easier to model relationships between classroom composition and students' outcomes.[5] However, this introduces selection bias if policy adherence was correlated with unmeasured characteristics of students in a way that was correlated with their outcomes. Thus, we also performed an instrumental variables analysis to estimate the unbiased treatment effect for policy-complying students and compared that estimate

to the estimate obtained by simply excluding students who did not take the required course. The final sample size for the analyses presented in the main text was 6779 students in 55 schools.

Measurements

Students' academic outcomes include algebra test scores and failure in algebra. 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). The average score on the subset for CPS sample was 6.0 with a standard deviation of 2.5. Course failure was a dichotomous variable where one indicated failing the primary algebra course (not the support course) in the first year of high school and zero indicates passing the primary algebra course.

Students' entering math abilities are based on students' national percentile rank scores on the 8th-grade ITBS math test; these are the scores used to determine double-dose algebra enrollment in CPS. However, because percentile scores are bounded between one and one hundred they are not a precise measure of ability for students with very low and high abilities. Therefore, we constructed an additional variable of achievement using a vector of students' test scores on the Iowa Tests of Basic Skills from third through eighth grade, standardized to have a mean of zero and standard deviation of one.[6] Latent ability indicators were included as control variables for very low ability students (those with below -2 SD) and high ability students (those with latent scores above 2 SD). All other students have a value of zero to avoid collinearity with percentile scores.

We created two measures of classroom academic composition. First, classroom average abilities were constructed by taking the average of students' 8th-grade latent math ability scores in their algebra classes. This variable captures the average initial skill levels of students in algebra classes upon entering high school. The second measure was a set of dummy variables, indicating the percent of students in regular algebra classes (not support classes) who were enrolled in support courses. Among below-norm students, classrooms were coded one if all of their students took support classes and zero otherwise. Among above-norm students, classrooms were coded one if their algebra classes contained any students taking double-dose algebra.[7] This variable attempted to capture time-varying compositional factors over the course of the year. We reason that teachers tailor instruction based on not only students' incoming skills, but also the progress students make during the course of the year. Thus, having students who take support courses in regular algebra classes may benefit the pace of the regular algebra class, controlling for the average incoming skill levels of students, because taking support coursework would facilitate students' learning, and teachers would modify instruction accordingly.

Measures of classroom instruction were constructed using students' responses to survey questions about their math classes. They were created through Rasch analyses using the survey items described in Appendix A and taking the classroom average of the student measures. There are two measures of classroom instruction: 1) academic demand captures how difficult/challenging students find their math class; 2) interactive pedagogy captures the extent to which students are involved in interactive instructional activities, such as explaining and discussing how to solve a math problem to the class and writing math problems for other students to solve, as compared to listening to a lecture. Survey items on pedagogy are designed to be consistent with the process standards of the National Council of Teachers of Mathematics.

To capture classroom behavioral climate, we created a measure of the concentration of students with disciplinary problems and absentee problems. Classroom disciplinary problems come from students' survey on incidence of disciplinary actions (see Appendix A). The student measure was first created through Racsh analyses using survey items. We then took the classroom average of the student measure by algebra classes. Supplemental analyses that used discipline records provided by CPS provided confirmation of the degree of disciplinary incidents reported by students. Classroom absence was constructed by first calculating the total number of absent days per semester in ninth grade year for each student, across all of their classes, then averaging the total absent days among all students in the class.[8]

Other student control variables include a dummy variable on gender and a set of dummy variables on race/ethnicity distinguishing African American, Hispanic, White, and Asian students. Two measures of SES variables were constructed using the block-level 2000 U.S. census data, linked to students' home addresses.[9] 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.

We also created a cohort average ability variable by taking the average of student latent ability scores for each school each year, which is used to control for the ability level of incoming cohorts over time.

Analysis

To examine the effects of a double-dose algebra policy on classroom learning environment—academic demand, interactive pedagogy, and clustering of students with absentee problems—for high-skill and low-skill students, we use a regression discontinuity design combined with cross-cohort comparisons. A regression discontinuity analysis takes advantage of the fact that double-dose algebra enrollment was strongly defined at the ITBS 50th percentile score (see Figure 1).

A regression discontinuity analysis allows us to estimate the effects of enrolling in double-dose algebra on students' classroom environment and academic outcomes, compared to enrolling in single-period algebra for students near the cutoff scores *under the policy*. Among pre-policy cohorts, there should not be a discontinuous relationship between ITBS scores and outcomes at the cutoff scores because the policy was not yet enacted and no below-norm students enrolled in double-dose algebra. An advantage of using pre-policy cohort is to increase our confidence that we are correctly specifying underlying relationships between the ITBS percentile scores and the outcomes.

Figure 1 about here

However, comparing outcomes among post-policy students by the double-dose algebra eligibility does not tell us the policy effects—how students' outcomes are different because they experienced the policy vs. having no policy, because the policy affected the academic composition of all algebra classrooms. The policy effects for below-norm students came from enrolling in algebra in tracked (low-ability) classrooms with extra time and supports, compared to enrolling in single-period algebra with higher ability peers without support (i.e., pre-policy single-period algebra). For high-skill students, the policy effects come from taking single-period algebra with higher ability peers vs. lower ability peers. Thus, to understand the effects of the policy for low- and high-skill students, we need to compare outcomes between pre-policy cohorts who did not experience the policy and post-policy cohorts who did experience the policy. The regression discontinuity adds a confirmation that differences between the cohorts should not be attributed to other changes that might have occurred simultaneously.

By combining a regression discontinuity design with cross-cohort comparisons, we compare differences in post-policy outcome *changes* between students who are just below the 50th-percentile cutoff scores and those who are just above the cutoff scores. This would indicate the extent to which the policy had differential effects for low-skill (i.e., below-norm) students who enrolled in double-doles algebra and high-skill (i.e., above-norm) students who took regular algebra. If tracking was the only mechanism of the policy affecting students' outcomes, we expect the outcome changes to be opposite for below- and above-norm students because the direction of classroom compositional changes was opposite; classroom ability levels declined for below-norm students, but it improved for above-norm students. However, if double-dose algebra coursework mitigated potential negative effects of tracking for low-ability students, we may not observe such a discontinuity at the cut-off scores.

To further strengthen our analyses, we begin by estimating relationships between classroom academic composition and outcomes in the absence of the policy using only pre-policy cohorts. Pre-policy analysis suggests which outcomes should be affected by the policy and due to changes in academic composition. For high-skill students, the direction of post-policy outcome changes should be predicted by the pre-policy relationships between classroom composition and outcomes.

For below-norm students who enrolled in double algebra, the policy brought about simultaneous changes in instructional time, instructional supports—curricular resources and professional development for their teachers, and classroom composition. It is difficult to disentangle the effects of these various components on students' achievement. However, we can make inferences about their policy effects in several ways. For example, if post-policy outcome changes occurred only to below-norm students and if these outcomes had no relationship with classroom academic composition in the absence of the policy, this would be evidence that the supports and resources accompanying the policy induced such outcome changes. Also, if pre-policy relationships showed negative relationships between classroom academic composition and outcomes (e.g., if classrooms with lower ability peers were related to lower academic demand and lower test scores), but if we did not observe expected negative changes post-policy (e.g., post-policy changes in academic demand and test scores were similar to, or greater than, changes made by above-norm students), this would also suggest that potential negative effects of tracking was mitigated by taking double-dose algebra.

The final analyses looked for mediating effects of the policy outcomes through classroom learning environment. For high-skill students, we examined the extent to which post-policy changes in the academic outcomes were explained by improvements in peer ability levels and

other instructional changes. For low skill students, we can only show how classroom learning environment and academic composition were related to students' academic outcomes; a full set of mediating factors cannot be done for low skill students as we cannot separate the effects of instructional time, different instructional resources and increased professional development. Statistical models for our analyses are provided in Appendix B.

Results

Pre-policy Relationships

Before estimating the effects of the policy we examined how classroom academic composition was associated with classroom learning environment (academic demand, interactive pedagogy, and clustering of students with discipline and absentee problems) and students' academic outcomes (Algebra test scores and Algebra course failure) in the absence of the double-dose algebra policy, using pre-policy cohorts. While these relationships might be affected by selection bias, they demonstrate the normally-observed relationships of classroom composition to elements of the learning climate, providing a contrast with the relationships observed with tracking induced by the policy. We also examined whether the relationships varied by students' incoming skill levels; this was done by introducing interaction terms between classroom academic composition and students' incoming abilities. We report these results only when significant interactions were found.

The top panel in Table 3 presents analyses of classroom experiences—academic demand, interactive pedagogy, disciplinary problems, and absenteeism. Coefficients for class average ability showed that in the absence of the double-dose algebra policy, classrooms with higher average ability had greater academic demand and fewer students with disciplinary and absentee

problems, consistent with prior tracking research. In addition higher ability students were especially likely to experience greater academic demand if they were in higher ability classrooms, indicated by the positive interaction term between students' own ability and classroom average ability. In other words, high-ability students were more apt to perceive increased demand than low achieving students. Also, lower ability students were particularly likely to have peers with high absence in lower ability classrooms, as noted by the positive interaction term between students' own ability and classroom average ability. Interactive pedagogy was not related to classroom academic composition. Teachers' approaches to teaching algebra were not systematically different in high- versus low-ability classes.

The bottom panel in Table 3 shows relationships between classroom academic composition and students' academic outcomes. In the absence of the double-dose algebra policy, classroom average ability levels were associated with both students' algebra test scores and course failure rates, but in opposite ways. Students were likely to have higher test scores in algebra classes with higher ability peers. Once again, high-ability students were more affected by class ability levels than low-ability students, showing particularly high test scores in high ability classes. For example, students with average incoming ability have average algebra scores that are .43 points higher if in a high-ability classroom, compared to a low-ability classroom, but students with entering achievement one standard deviation above the mean would benefit by .89 points.[10] Students with incoming ability one standard deviation below the mean would not benefit from being in a high-ability classroom, compared to a low-ability classroom, in terms of their test scores. At the same time, all students were more likely to fail in classes with more high-ability peers.

Table 3 about here

These results suggest that, post-policy, high-skill students should have experienced greater academic demand and had classmates with fewer disciplinary and absentee problems than pre-policy students because the policy resulted in improvements in peer ability levels in their algebra classes. However, pedagogy should not have changed post-policy for high-skill students because it is unrelated to classroom academic composition. Also, high-skill students should have improved their test scores post-policy, but been more likely to fail due to improvements in peer ability levels. In contrast, declines in peer ability levels among low-ability students post-policy should have led to less academic demand, more students with disciplinary and absentee problems in their classes, lower test scores, and lower failure rates. However, potential negative effects of increased ability-grouping might have been mitigated by the additional time for learning, professional development and curricular resources for teachers.

Policy effects on classroom learning environment

The next analyses examined the extent to which the double-dose algebra policy affected classroom learning environments—academic demand, interactive pedagogy, and clustering of students with disciplinary and absentee problems—for high- and low-skill students. As discussed earlier, the policy effects were defined as differences in the outcomes between students who experienced the policy and students who did not. Policy effects were estimated by comparing the outcomes between pre- and post-policy cohorts. Differences in policy effects for high-skill versus low-skill students were estimated by comparing students just above the cut-off for double-dose eligibility (the 50th percentile) to students just below the cut-off (at the 49th

percentile); the variable representing entering achievement is centered at the 50th percentile to capture this discontinuity. We then introduced a variable representing the classroom average ability level to see the extent to which changes in the learning climate were attributable to changes in classroom academic composition.

For high-skill students, did classroom learning environment improve as a result of taking algebra with only high-ability peers? We first examined the policy effects for high-skill students. The intercept in Model 1 in Table 4 represents pre-policy outcomes for high-skill students—those who scored above the 50th percentile cutoff scores, while the coefficient for the 2004 cohort in the second row represents changes in the outcomes compared to the 2002 cohort, indicating the estimated policy effects for high-skill students.

Post-policy above-norm students reported greater academic demand, by 0.12 standard deviations (p<0.001), compared to their pre-policy counterparts. This post-policy change is consistent with the pre-policy analysis, showing that classroom average ability levels are positively related to academic demand. Also consistent with the pre-policy analyses, the use of interactive pedagogy did not change with the policy for above-norm students. In addition, post-policy above-norm students had fewer peers with disciplinary and absentee problems in their algebra classes. Disciplinary problems declined by 0.05 standard deviations for above-norm students and their peers were less likely to absent from school by 0.4 days post-policy than pre-policy.

We then introduced the classroom average ability in the models to see the extent to which post-policy differences were explained by changes in peer ability levels (see Model 2 in Table 4). Interaction terms were also introduced to pick up any deviations in compositional effects for: 1) post-policy above-norm students; 2) pre-policy below-norm students; and 3) post-policy below-

norm students, as compared compositional effects for pre-policy above-norm students. For above-norm (high-skill) students, improvements in peer ability levels explained most of the post-policy increases in academic demand; the coefficient on academic demand for post-policy cohort was reduced from 0.12 (Model 1) to 0.05 (Model 2) and it was no longer statistically significant (P.>05). Also, improvements in peer ability levels explained post-policy declines in disciplinary and absentee problems.

For low-skill students, did classroom instruction and learning environment improve or decline as a result of taking double algebra with low ability peers? In Table 4, the coefficient for the 2002 cohort under *below norm deviation* indicates the degree to which the outcome differed in 2002, for students just below the double algebra cut off (at the 49th percentile), compared to students just above the cutoff (at the 50th percentile). We did not expect to see differences in pre-policy outcomes between students below the cut-off compared to those just above the cut-off because the policy was not yet enacted, and we do not. None of the indicators of classroom instructional climate are significantly different for below-norm students, compared to above-norm students, pre-policy (with coefficients of -0.03, -1.0,-0.06, and -0.02 for academic demand, interactive pedagogy, absenteeism and disciplinary problems, respectively).

The coefficients for the 2004 cohort under below-norm deviation indicate the extent to which post-policy outcome changes differed for below-norm students, compared to above-norm students. Because the policy intensified tracking, it should have had opposite effects on academic demand and the concentration on students with disciplinary and absentee problems for below-norm students compared to above-norm students. However, if providing expanded instructional time and additional supports mitigated negative effects of tracking, we would not observe such discontinuities in the outcome changes.

Although tracking literature and the pre-policy relationships suggest course difficulty would decline due to declines in peer ability levels among below-norm students, academic demand actually increased among below-norm students as much as among above-norm students with the policy; the post-policy difference compared to above-norm students was -0.02 SD and not statistically significant (p>0.1). Unlike with above-norm students, the increase in academic demand was not due to changes in academic composition; academic demand improved by 0.12 SD (0.05+0.07) after adjusting for classroom peer ability. Increases in academic demand for below-norm students likely occurred because of other aspects of the policy, such as increased instructional time and pedagogical practices described below.

The most dramatic change in the classroom instructional environment was the increased use of interactive pedagogy for low-skill students who enrolled in double-dose algebra. Post-policy, below-norm students reported much more frequent use of interactive pedagogy than below-norm students pre-policy, the difference in the post-policy changes between below- and above-norm students was .55 SD (P<.001). This is consistent with the professional development and curricular resources given to teachers as well as the additional instructional time in double-algebra classes that allowed teachers more flexibility for instructional innovation. Improvements in pedagogy were also related to the improvements in academic demand; additional analyses (not presented here) showed that students in classrooms with more frequent use of interactive pedagogy reported greater academic demand both pre- and post- policy.

Classroom compositional changes did not explain improvements in the use of interactive pedagogy—the coefficients changed little after controlling for classroom compositional changes. This is not surprising as pre-policy analyses showed no relationship between pedagogy and peer ability levels. However, there was a change in the relationship between pedagogy and classroom

ability levels post-policy only among below-norm students. Post-policy, there was a negative relationship between classroom average ability level and the use of interactive pedagogy among below-norm students (a coefficient of -0.42 with p<0.05). This suggests that teachers were more likely to use interactive pedagogy the more that students in their double-algebra classes had very low abilities. In other words, teachers were most likely to attempt to change their pedagogical practices if they were teaching particularly struggling students.

While students in double-dose algebra received better instruction (i.e., better challenging and better pedagogical practice), declines in peer ability levels created greater concentration of students with behavioral problems. Peers in double-dose algebra were more likely to be absent from school by one day (-.38+1.37) and have disciplinary problem by 0.07 SD (-0.05+0.12) than classmates in pre-policy algebra classes. Consistent with prior studies on tracking, declines in peer ability levels explained post-policy increases in the concentration of disciplinary and absentee problems in double-dose algebra classes.

Table 4 about here

How did the changes in classroom academic composition and instruction/learning environment affect math achievement? Lastly, we present a series of analyses that predict algebra scores (Table 5)and algebra failure (Table 6) with sequential controls for 1) instructional climate, 2) classroom ability level and 3) the percentage of students in the primary algebra course who are also taking support algebra. These analyses discern the degree to which the changes in test score and algebra course failures observed post-policy can be attributed to changes in the classroom instructional environment, peer ability, and spillover effects from support algebra

classes. As with the prior models, post-policy effects are modeled off pre-policy main effects, while deviations for low-ability students are modeled as deviations off high-ability students. Model 1 shows policy effects without classroom covariates; for example, above-norm students had higher algebra scores by .56 points post-policy compared to similar pre-policy students. Below-norm students improved their test scores by additional 0.32 points post-policy, totaling 0.88 points higher than similar low-ability students pre-policy.

Model 2a through Model 2d includes variables on instruction/classroom learning environments, each entered separately. Model 3 shows these classroom variables together. These models show that students have higher test scores in classrooms with higher academic demand, more frequent use of interactive pedagogy and fewer students with absentee problems. Other analyses, not presented here, showed that these relationships did not differ by cohort or by student ability level. These changes explain a small portion of the improvements in test scores for both above and below-norm students, about 5 to 10 percent of the improvements.[11] Because having peers with absenteeism and disciplinary problems have negative relationships with test scores , these variables do not explain improvements in test scores for below-norm students; test scores improved more than would be expected, given that there were more students with high absenteeism and disciplinary problems in their classes.

Model 4 shows that high ability students' scores improved not only because of the changes in classroom instructional climate, but also because of the changes in peer ability levels; adding peer ability to the model decreases the post-policy improvement by an additional 23 percent, to 0.38. Thus, something about having high-ability peers was beneficial to their scores, beyond the measured effects in instructional climate. Yet, while classroom ability was related to test scores for low-ability students pre-policy, it did not matter post-policy; the relationship

between classroom average ability and algebra score for these students was -0.28 (which is 0.72-0.33+0.19-0.86), and was not significantly different from zero (p>0.1), after controlling for changes in other classroom variables.

Our last model (Model 5) examined the degree to which students benefited from having classmates who took support courses, after controlling for the classroom average incoming ability and other classroom variables. We reason that teachers not only tailored instruction based on students' incoming skills, but they also adjusted instruction as students made progress. Thus, if taking support coursework facilitates learning and teachers adjust instruction accordingly, students would benefit from having classmates who took an algebra support courses even though support course takers would have lower initial skill levels when they began high school. In fact, Model 5 shows that above-norm students did have higher test score gains of by about 0.6 points when their algebra classes had students who enrolled in support courses than above-norm students who did not have any such classmates (p<.01). That is, controlling for the incoming ability levels of classroom peers, students' scores improved more if their low-ability classmates were getting an additional period of algebra instruction. Adding this variable for spillover effects further explains the post-policy rise in test scores for above-norm students so that it is no longer significant.

Table 6 shows policy effects on algebra failure in log odds; the bottom of Table 6 shows the average failure rates for each year in percentage points for ease of interpretation. For above-norm students, failure rates increased post-policy by about 3 percentage points despite improvements in test scores (Model 1). In comparison, students just below the policy cut off score had lower failure rates post-policy, by 4 percentage points, compared to their pre-policy counterparts

Consistent with pre-policy relationships, Model 2 shows that below-norm and above-norm students were more likely to fail in classes with greater academic demand and especially when in classrooms with more students with disciplinary problems. Also, students were less likely to fail in classes with more frequent use of interactive pedagogy. While interactive pedagogy does not have a significant relationship with failure rates when entered alone, once we control for academic demand, greater use of interactive pedagogy is related to lower failure rates. In other words, classes that use more interactive pedagogy are likely to be more academically demanding, suppressing the relationship between pedagogy and failure rates. Additionally, students were more likely to fail classes with higher average ability, which is partly related to greater academic demand (Model 4). Algebra failure was not related to the percent of students who take support courses for below-norm or above-norm students (Model 5).

For above-norm students, most of the post-policy increase in failure rates was explained by changes in classroom environments and peer ability levels; differences between pre- and post-policy above norm students were reduced from three percentage points (Model 1) to one percentage point (Model 4) once we included all classroom variables and the post-policy change was no longer significant (0.08 logits, p>0.1). In comparison, for below-norm students, declines in failure rates were even greater once we controlled for increases in classroom behavioral climates (Model 2c and Model 2d). In other words, their failure rates declined by five to six percentage points more than we would have expected, given the greater concentration of students with high absence or disciplinary problems in their classes and the higher levels of academic demand.

Addressing selection bias problems

There are two potential sources of bias in this study. First, excluding students who did not adhere to the policy would result in selection bias if their unmeasured characteristics differed from the characteristics of students who adhere to the policy in a way that was related to their outcomes. Second, bias may result from differential survey response rates: survey respondents might be systematically different between pre- and post-policy cohorts in a way that was not measured by the researchers. To examine the extent of each selection bias problem, we conducted the following analyses: 1) estimating the effects of double-dose algebra using the instrumental variable method; and 2) examining students' academic outcomes using the population of ninth-grade regular education students instead of just the sample of survey takers.

An instrumental variable (IV) analysis provides an unbiased estimate of the treatment effect for complying students (i.e., those who took the assigned treatment) in a "fuzzy" regression discontinuity design when compliance is imperfect. This analysis uses the assignment variable (i.e., the cutoff scores) as an instrument to estimate the outcomes using two-stage least squares.[12] Although this strategy only allows us to estimate the effect of enrolling in double algebra vs. single algebra under the policy, we are able to see the extent of bias due to excluding miss-assigned students by comparing the results from the two analyses (also, see Author, year).

Our second analysis attempted to address cross-cohort selection bias due to differential response rates between the two cohorts. Here, we examine students' academic outcomes among all ninth-grade regular-education students without using the survey data, and compare these results to those based on survey takers.

In general, the two sensitivity analyses do not change the conclusions from our earlier analyses. Table 7 compares the IV results to the analyses based on only students who took the

assigned courses for students with survey data and the population of ninth-grade students when the data are available. Both analyses show that double-dose classes and single algebra classes had similar levels of academic demand under the policy, while the use of interactive pedagogy was greater and behavioral problems were greater in double-algebra classes than single algebra classes. Also, students in double-dose algebra had higher test scores and lower failure rates than students in single algebra post-policy. The results are similar regardless of whether we only include students who participated in the survey in the models versus the entire population of ninth graders.

Table 8 provides a further comparison of the survey sample to the population, with analyses that use both pre- and post-policy cohorts. In general, the coefficients predicting test scores are similar between the survey sample and the population. There was, however, some slight evidence of selection bias in estimating algebra failure rates. Survey respondents had lower failure rates than the ninth-grade population; the average failure rates among pre-policy students who scored at the national median on the 8th grade ITBS test was 25 percent (a coefficient of -1.1) among survey respondents as compared to 30 percent of the ninth-grade population (a coefficient of -0.81). Also, while both survey sample and population results showed post-policy increases in failure rates among above-norm students, the increase was larger for the population by about four percentage points. Differences in post-policy changes in failure rates for below-norm students were -6 percentage points for survey respondents and -9 percentage points for the ninth-grade population.

Yet, while survey respondents differed somewhat from the ninth-grade population with regard to course failure rates, the overall conclusions of this study did not change: algebra failure rates increased post-policy among high-ability students, while they declined somewhat for

below-norm students. Also, students were more likely to fail algebra in classes with higher average abilities.[13] The policy effects on failure rates simply become larger when the population is examined, rather than the sample.

Conclusions

Tracking has been one of the most controversial issues in education. The underlying concern is that tracking decreases the opportunity to learn among low-ability students, despite advantages for high-ability students. But this study has demonstrated that the issue of classroom ability-grouping is not so clear-cut; there are a number of important nuances in the ways in which classroom composition affects student achievement, beyond benefits for high-ability students and detriments for students with low-ability.

Consistent with prior literature, this study shows that grouping students into classes by incoming skills can lead to detriments for low-ability students. The more low-ability students are concentrated together, the lower the academic demands of instruction, and the lower students' learning gains. The most striking problem that arises from concentrating low-ability students together, however, is the concentration of students with disciplinary and absentee problems. The more that classroom peers have disciplinary and absentee problems, the more likely all students in the class are to fail and this relationship is quite strong. Classrooms with concentrations of students with discipline issues and absentee problems likely present difficult conditions for teaching; there may also be peer effects on students' behaviors and engagement that make them more likely to fail. Thus, an important issue to address in ability-grouping is support for behavioral issues in the classroom. Teachers in low-ability classrooms often struggle with

classroom management and attendance problems, and these struggles prevent them from being able to teach effectively (Sporte, Correa, Hart & Wechsler, 2009).

Prior literature has also found that high-ability students benefit from tracking, and this study supports some aspects of that conclusion—their learning gains are higher, classes are more demanding, and classroom peers are less likely to bring attendance and behavioral problems. Furthermore, high-ability students' learning benefits more from tracking than low-ability students' achievement does from de-tracking. While both high-and low-ability students learn more in classrooms with more high-ability peers, the relationships between classroom academic composition and test scores are stronger for high-ability students. This makes sense—a high ability student may be more likely to recognize differences between a highly difficult class and a moderately difficult class, while a low ability student might struggle in equally in either class. Thus, mixing students of varying abilities together has substantial negative effects on learning among high-ability students, while only modestly improving the learning of low-ability students.

Furthermore, this study highlights other costs for low-ability students when mixed into heterogeneous classrooms; their course grades suffer from taking classes with higher-ability peers, even though they show somewhat greater learning. Increases in failure rates are disturbing because each course that a student fails in high school increases the probability of dropping out by about 15 percentage points (Allensworth & Easton, 2007). Since low-ability students are already at high risk for dropout, increasing this risk is a sizable concern.

This finding about the effects of tracking on course grades adds a complication to what was previously known about ability-grouping. Tracking has the opposite effects on students' course grades as on their test scores. High-ability students received more academically demanding instruction post-policy due to intensified tracking, and higher academic demand

increased their failure rates, even though they learned more in the more demanding classes. Thus, tracking has some costs for students in high-track classes, as well as benefits; their grades suffer even though they learn more.

These findings stand in opposition to arguments that the elimination of curricular tracks creates greater equality without compromising excellence. Other research has discerned many difficulties that accompany schools' detracking efforts (Wells & Oakes, 1996; Rubin, 2008), and this study adds further evidence to this body of work. To be certain, there are cases of successful detracking, where low-ability students learn more in heterogeneous classrooms without hurting the learning of high-ability students (Boaler & Staples, 2008; Oakes, 2005; Rubin, 2008). However, schools that have successfully detracked classes have done so carefully; they allocated considerable recourses to low-ability students, including time and professional development for their teachers, with strong principal leadership and support from teachers. The results of this study are consistent with such approaches—the few heterogeneous classes that existed under the policy seemed to be more effective post-policy because the low-ability students in those classes were receiving extra instructional support. However, unless accompanied by considerable supports, simply mixing students in heterogeneous classrooms seems to produce greater costs for high-ability students than benefits for low-ability students.

In the end, this study supports the argument that the learning growth of both low- and high-ability students can benefit from tracking with additional instructional support for low-ability classrooms. Chicago's double-period algebra strategy intensified tracking. However, unlike the traditional form of tracking which includes low-level coursework (e.g., general mathematics and pre-algebra), all students took algebra, and intensified tracking was accompanied by doubled instructional time, professional development, and instructional

resources for double-period algebra teachers. Teachers in double-period classes used more interactive pedagogy, and this helped boost test scores along with doubling instructional time. Improvements in the rate of learning among students taking a support class also had positive spillover effects into regular algebra classrooms, so that all students benefited when low-ability students received supplemental algebra instruction. Thus, this strategy differs considerably from traditional tracking where low-ability students receive a different curriculum from high-ability students and often have less qualified teachers. It requires that low-ability students receive *more* resources than high-ability students--more instructional time, more support for their teachers--to offset the detriments of negative peer effects. While an inequitable distribution of resources may seem unfair, in the end, all students benefit. This strategy was effective for boosting the algebra scores of both low- and high-skill students, more than making up for potential disadvantages that might arise from grouping low-skill students together.

---

[1] The form of tracking has changed over time. Historically, U.S. high schools used rigid curricular tracking (e.g., academic, regular, and vocational tracks), but this has changed to more seemingly flexible curricular choice (see Lucas, 1999).

[2] 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."

[3] These evaluations used classroom observations and focus groups based on a small sample of schools (teachers from 12-15 schools).

[4] Shadish, Cook, and Cambell (2002) suggest excluding missassigned cases if such cases are less than five percent of the total population. Instrumental variable methods (IV) can also be used to estimate the treatment effects for compliers. In our study, the percent of missassigned students exceeded five percent and we also conducted sensitivity analyses using IV methods, as discussed.

[5] Changes in academic composition were opposite for students who adhered to the policy and those who did not. For example, above-norm students who took double-dose algebra, even though they were not required to do so, experienced declines in peer-ability levels, and despite declines in peer-ability levels, their test scores were likely to improve because of additional instructional time and supports. For these students, classroom compositional changes would not explain post-policy improvements in test scores. Thus, including non-adhering students in the analyses on

above-norm students would reduce explanatory power of classroom academic composition on their test scores. It would also suppress relationships between classroom academic composition and test scores

[6] A two-level HLM, nesting years within students, modeled each student's learning trajectory; level 1 included variables for grade and grade-squared which were allowed to vary across students. There was also a dummy variable representing a repeated year in the same grade, to adjust for learning that occurred the second time in a grade, and a different dummy variable for repeating the eighth grade year so that additional learning that occurred when eighth grade was repeated could be added into a student's latent score. Before modeling students' growth trajectories, students' test scores were equated through Rasch analysis to remove form and level effects.

[7] Typical algebra classes taken by below-norm students had approximately 90 percent of their classmates also in support algebra classes. In comparison, the majority of above-norm students (78 percent) did not have any classmates taking support coursework in their algebra classes.

[8] Data on absence comes from a semester grade file provided by CPS. By calculating students' absences across all of their classes we attempted to capture whether students generally had attendance problems, not just whether they missed algebra. This was to avoid potentially confounding instructional effects of the algebra class on attendance with the effects of concentrating students with attendance problems together on the instructional environment of the algebra class.

[9] 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.

[10] The effect of being in a high-ability classroom compared to a low-ability classroom is estimated by comparing a classroom with average achievement one standard deviation above the mean to classroom with average ability one standard deviation below the mean. The standard deviation of classroom average ability (.528) is multiplied by two and by the coefficient on "class average ability" to produce the effect of classroom ability on a student with average achievement. The effect for a student with high entering achievement is calculated by adding to that the interactive effect: the standard deviation of incoming achievement (21.7) times the standard deviation of classroom average entering achievement (.528) times two for a 2-standard deviation difference, times the coefficient for the interaction term (0.02).

[11] For below-norm students, interactive pedagogy was the only variable that explained 13 % of post-policy test score improvements.

[12] The first-stage equation estimates double-algebra course enrollment using the cutoff score as an instrument with other students' control variables and school fixed effects. The second-stage equation estimates the outcomes using the expected value of the course enrollment as a predictor with the same control variables.

[13] Our prior work (Author, year), which evaluated the overall policy, showed that failure rates did not change post-policy for below-norm students, while the current study showed declines in failure rates for this population. The discrepancy between the two studies is likely to be due to differences in the sample included in these studies. While our earlier work used all students, the current study excluded students with disabilities who were, on average, more likely to fail as a result of taking double-dose algebra than regular-education students.

References

ACT, Inc. 2007. Aligning Postsecondary Expectations and High School Practice: The Gap

Defined, Policy Implications of the ACT National Curriculum Survey Results, 2005-

2006. ACT, Iowa City, IA.

Allensworth, E. & J. Easton. 2007. *What Matters for Staying On-Track and Graduating in the*

Chicago Public Schools. Chicago: Consortium on Chicago School Research.

Anderson, L. W. (Ed.). (1984). *Time and school learning*. London: Croom Helm.

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking america's schools: Equity at zero

costs? *Journal of Policy Analysis and Management, 15*, 623-645.

Boaler, J. & Staples, M. (2008). Creating mathematical futures through an equitable teaching

approach: The case of Railside School. *Teachers College Record 110*(3), 608-645.

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.

Farkas, G., Sheehan, D., & Grobe, R. P. (1990). Coursework Mastery and School Success:

Gender, Ethnicity, Poverty Groups Within an Urban School District.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of*

Education, 60, 135-155.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality:

Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*, 1146-

1183.

Gamoran, A., & Weinstein, M. (1998). Differentiation and Opportunity in restructured Schools. *American Journal of Education, 106*, 385-415.

Hallinan, M. (1994). Tracking: From theory to practice. *Sociology of Education, 67*(2), 79-84.

Kelly, S. (2008). What Types of Students' Effort Are Rewarded with High Marks? *Sociology of Education, 81*, 32-52.

Lucas, S. R. (1999). *Tracking inequality: Stratification and mobility in American high school*. New York: Teachers College Press.

Loveless, T. (1999). *The tracking wars: State reform meets school policy.* Washington, DC: Brooking Institution.

Miller, S. R. 1998. High School Grades as a Signal of Human Capital. *EEPA*, 20(4): 299-311.

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.

Noble, J. and R. Sawyer. 2002. *Predicting different levels of academic success in college using high school GPA and ACT composite score*. Iowa City, Iowa: ACT, Inc.

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.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven: Yale University Press.

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.

Rosenbaum, J.E. (1999). If tracking is bad, is detracking better? *American Educator, 23*(4), 24-29. 47.

Rubin, B.C. (2008). Detracking in context: How local constructions of ability complicate equity. *Teachers College Record, 110*(3), 646-699.

Sporte, Susan E., Macarena Correa, Holly M. Hart and Marjorie E. Wechsler. 2009. High School Reform in Chicago Public Schools: Instructional Development Systems. SRI International.
(http://ccsr.uchicago.edu/publications/other/Instructional_Development_Systems.pdf)

Starkel, R., Martinez, J., and Price, K. (2006). Two-period Algebra in the 05-06 school year: Implementation Report. Retrieved on January 15, 2010 from https://research.cps.k12.il.us/resweb/DownLoaderAdv?dir=program_evaluation&file=Two_Period_Algebra_final_report1.pdf

Wells, A. S., & Oakes, J. (1996). Potential pitfalls of systemic reform: Early Lessens from Detracking Research. *Sociology of Education, 69*, E135-E143.

Wenzel, S., Lawal, K., Conway, B., Fendt, C., and Stoelinga, S. (2005) *Algebra problem solving: Teachers talk about their experiences, December 2004.* Retrieved January 15, 2010 from

https://research.cps.k12.il.us/resweb/DownLoaderAdv?dir=program_evaluation&file=AP

S_external_evaluation_focus_group_report_010405.pdf

Wheelock, A. (1992). *Crossing the Tracks: How "Untracking" Can Save America's Schools*.

New York: New Press.

Appendix A

Academic Press (response category: strongly disagree, disagree, agree, and strongly agree), reliability=.76

In your math class, how often:
- Do you find the work difficult?
- Are you challenged?
- Does the teacher ask difficult questions on tests?
- Does the teacher ask difficult questions in class?
- Do you have to work hard to do well?

Interactive math instruction (response category: never, once in a while, once a week, almost everyday, everyday), reliability=.70

In your class this year, how often do you do the following?
- Write a few sentences to explain how you solved a math problem
- Explain how you solved a problem to the class
- Write math problems for other students to solve
- Discuss possible solutions to problems with other students
- Copy notes or problems off the board
- Apply math to situations in life outside of school

Disciplinary Problems (response category: never, 1-2 times, 3-5 times, and more than 5 times), reliability=.72

How many times this school year have you:
- Gotten into trouble at school?
- Gotten into a physical fight with another student at school?
- Been sent to the office for getting into trouble?
- Been put on in-school suspension?
- Been suspended from school?

Appendix B

We begin our analysis by examining how classroom academic composition is related to students' classroom experiences—academic demand, clustering of students with disciplinary and absentee problems, and interactive pedagogy—as well as their academic outcomes, in the absence of the policy. These analyses are based on students entering high school in the pre-policy (2002) cohort.

To estimate the outcome Y for student *i* in classroom *j* in school *k*, we estimated the following model using pre-policy students;

$$Y_{ij} = \beta_0 + \beta_1(ClassAverageAbility)_j + \beta_3(X)_{ij} + \beta_4(W)_k + \varepsilon_{ij} + u_j, \qquad (1)$$

where Y is an outcome (instruction/classroom learning environment and academic outcomes), X is a vector of student control variables, including latent incoming abilities, socio-demographic variables (age, race, gender, SES, and residential mobility in the three years prior to high school), W is a vector of school fixed effects, and ε_{ij} and u_j represent student and classroom error terms. Classroom-level error terms are estimated only when appropriate. The coefficient of interest is β_1, which represents the average within-school relationships between classroom average ability and the outcome, controlling for student background characteristics and characteristics of schools attended by students. In addition, we introduce an interaction variable between student incoming ability and classroom average ability to examine whether the relationship between classroom average ability and the outcome differs by students' incoming ability levels.

The next analysis examined how the double-dose algebra policy affected instruction/learning environment—academic demand, clustering of students with disciplinary

problems, and interactive pedagogy—for below-norm and above-norm students. We estimated the following statistical models;

$$Y_{ijk} = \beta_0 + \beta_1(Yr04)_i + \beta_2(BelowNorm)_{ij} + \beta_3(BelowNorm*Yr04)_{ij} + \beta_4(ITBS\_Percentile)_{ij}$$
$$+ \beta_5(X)_{ij} + \beta_4(W)_k + \varepsilon_{ij} + u_j ,$$

$$(2)$$

where Y is an outcome (i.e., academic demand, interactive pedagogy, and clustering of students with absentee and disciplinary problems) for student i in classroom j in school k. *Yr04* is a dummy variable indicating the 2004 cohort (post-policy students). *BelowNorm* indicates whether students scored below the 50th percentile cutoff score, while *BelowNorm*Yr04*, indicates below-norm students in 2004. *ITBS_Percentile* indicates students' 8th-grade math ITBS percentile scores, centered on the 50th percentile. X is a vector of student control variables, W is a vector of school fixed effects, and ε_{ij} and u_j represent student and classroom error terms. Including school fixed effects controls for all time-invariant characteristics of schools, allowing us to estimate the average within-school changes in the outcomes. The ITBS percentile scores were centered on the 50th percentile so that the intercept β_0 represents the average pre-policy outcome for students at the 50th percentile. The functional forms for ITBS percentile scores were determined in preliminary analysis.

The policy effects for above-norm students were indicated by the coefficient for *Yr04* (β_1), which represents the average change in the outcome for above-norm students from pre- to post-policy. The coefficient for *BelowNorm*Yr04* (β_3) indicates the degree to which post-policy outcome changes were different for below-norm students, as compared to post-policy outcome changes made by above-norm students. If the policy had differential effects for below-norm students, the coefficient, β_3, would differ from zero. However, in the absence of the policy, we

do not expect the pre-policy outcomes to differ between below-norm and above-norm students who are just below or above the cut-off score, respectively $(\beta_2 = 0)$.

Our next analysis introduced the variable on classroom average ability. This analysis also examined the extent to which changes in classroom academic composition explained the policy effects on instruction/classroom learning environment. For above-norm students who enrolled in single-period algebra, their post-policy outcome changes would be largely explained by improvements in peer ability levels. For below-norm students, the classroom composition variable should not explain policy effects if other changes that occurred with the policy (e.g., additional time and support) also affected their outcomes.

Our last analysis examined how the changes in instruction/learning environment and classroom academic composition brought by the policy affected algebra test scores and algebra course failure. We first estimated the policy effects on students' outcomes using Equation 2. To estimate algebra course failure, we used logistic regression. Subsequently, we introduce variables on instruction/classroom learning environments and classroom academic composition variables to see the relationships between these classroom variables and the outcomes. Also, we looked to see the extent to which classroom variables explain post-policy changes in students' academic outcomes.

Figure 1.
Algebra enrollment rates by 8th-grade ITBS percentile scores
(Percentile scores centered around the national median)
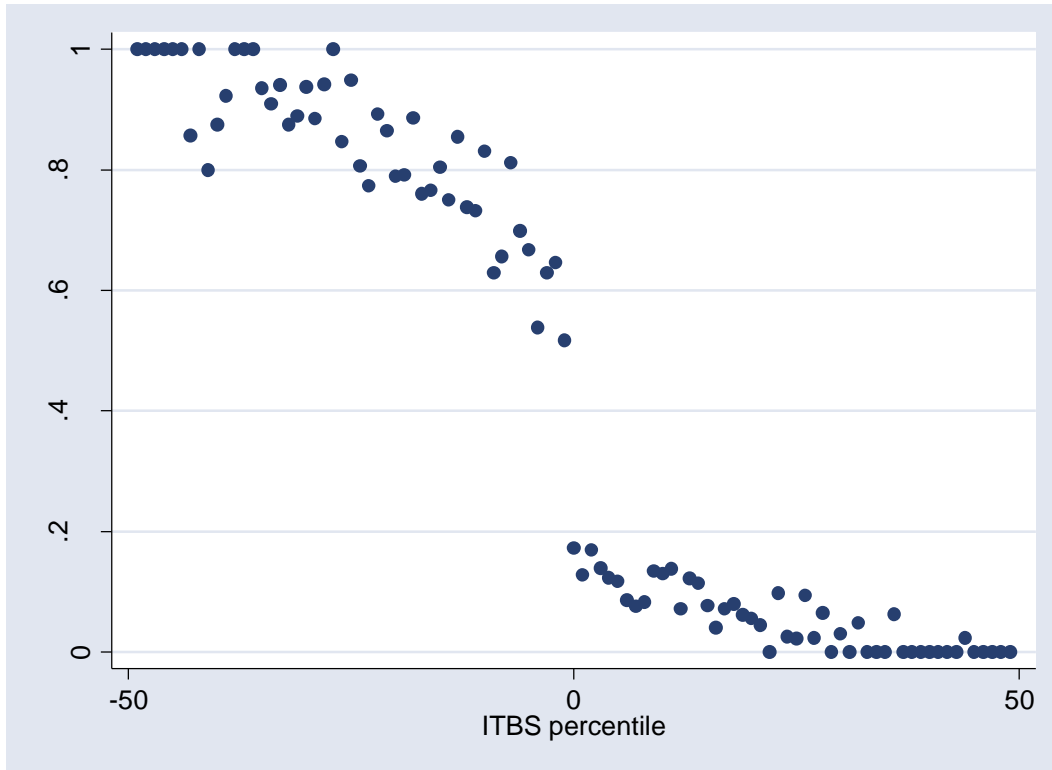
Tables

Table 1.
Average 8th grade math scores of classroom peers (standardized)

Year	Below-norm students	Above-norm students
2000	0.04	0.24
2001	0.04	0.26
2002	0.01	0.27
2003(policy)	-0.13	0.36
2004(policy)	-0.11	0.35

Note. "Below-norm" refers to students with eighth-grade test scores below the 50th percentile who were required to take two periods of algebra. Above-norm students' eighth-grade scores were at or above the 50th percentile and they were not required to take two periods of algebra.

Table 2.
Descriptive statistics of survey respondents on selected student characteristics by cohort

	Survey takers				Ninth-grade population			
	2002-03 cohort (pre-policy)		2004-05 cohort (post-policy)		2002-03 cohort (pre-policy)		2004-05 cohort (post-policy)	
Variables	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Below norm (%)	0.39	0.49	0.38	0.49	0.40	0.49	0.45	0.50
Incoming ability(Z-score)	.25	.69	.22	0.73	0.23	0.72	0.15	0.73
White (%)	0.07	0.26	0.07	0.25	0.08	0.27	0.07	0.26
Hispanic (%)	0.42	0.49	0.38	0.48	0.36	0.48	0.35	0.48
Asian (%)	0.03	0.16	0.03	0.18	0.03	0.18	0.03	0.17
Poverty (Z-score)	0.04	0.99	0.00	1.02	0.03	1.04	0.04	1.03
Student N	3278		3501		13979		15263	

Note. See the data section for variable descriptions.

Table 3.
Pre-policy relationships of classroom academic composition with classroom environment and academic outcomes

Outcomes		Est.	SE
Instruction			
Academic demand	Intercept	-0.03	0.04
	Class Ave. ability	**0.15***	0.06
	Student's own ability x Class Ave. ability	0.01**	0.00
Interactive pedagogy	Intercept	-0.11*	0.04
	Class Ave. ability	**-0.03**	0.06
Classroom behavior			
Disciplinary problems	Intercept	-0.11*	0.04
	Class Ave. ability	**-0.25*****	0.02
Absenteeism	Intercept	8.22***	0.05
	Class Ave. ability	**-2.89*****	0.10
	Student's own ability x Class Ave. ability	0.01***	0.00
Students' academic outcomes			
Algebra scores	Intercept	5.32***	0.19
	Class Ave. ability	**0.41****	0.13
	Student's own ability x Class Ave. ability	0.02***	0.00
Course failure (in logits)	Intercept	-1.11***	0.09
	Class Ave. ability	**0.31***	0.14

* p<.05, ** p<.01, ***p<.001. Models include controls for student backgrounds and incoming skills, as described in Appendix B.

Table 4. Pre- and post-policy changes in instruction/classroom environment,
Students just above and below the double-dose algebra eligibility cutoff score

Instructional outcomes	Academic Demand				Interactive pedagogy			
	Model 1		Model 2		Model 1		Model 2	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept (above-norm pre-policy)	-0.05	0.04	-0.05	0.03	-0.10*	0.06	-0.10*	0.06
2004 cohort deviation	**0.12*****	0.03	**0.05**	0.04	**-0.03**	0.08	**-0.05**	0.08
Below norm deviation	-0.03	0.05	-0.01	0.05	-0.10	0.07	-0.10	0.07
Below norm x 2004 cohort	**-0.02**	0.05	**0.07**	0.08	**0.55*****	0.10	**0.43*****	0.11
Class average ability			0.36***	0.05			0.04	0.06
Class ability x 2004 cohort			0.08	0.07			-0.03	0.09
Class ability x below norm			-0.23*	0.10			0.05	0.11
Class ability x below norm x 2004 cohort			-0.14	0.17			-0.42*	0.20

Classroom behavioral outcomes	Absenteeism (in days)				Disciplinary problem			
	Model 1		Model 2		Model 1		Model 2	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept (above-norm pre-policy)	8.32***	0.08	8.31***	0.07	-0.08***	0.02	-0.07***	0.02
2004 cohort deviation	**-0.38*****	0.11	**0.00**	0.10	**-0.05***	0.02	**-0.04**	0.02
Below norm deviation	-0.02	0.13	-0.04	0.13	-0.06	0.03	-0.06	0.03
Below norm x 2004 cohort	**1.37*****	0.19	**-0.07**	0.21	**0.12***	0.04	**0.02**	0.05
Class average ability			-2.63***	0.11			-0.24***	0.02
Class ability x 2004 cohort			-0.16	0.15			0.10**	0.03
Class ability x below norm			-0.41*	0.19			0.06	0.04
Class ability x below norm x 2004			-0.16	0.33			-0.12*	0.07

Outcomes are measured in standard deviations, with the exception of absences which is measured in days. *p<.05, **p<.01, ***p<.001.

Table 5. Pre- and post-policy Algebra test scores at the cutoff for double-algebra eligibility

	Algebra scores							
	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 3	Model 4	Model 5
	Est.	Est.	Est.	Est.	Est.	Est.	Est.	Est.
Intercept (above-norm pre-policy)	5.28***	5.29***	5.30***	5.17***	5.25***	5.18***	5.16***	5.17***
2004 cohort deviation	**0.56*****	**0.53*****	**0.56*****	**0.54*****	**0.55*****	**0.52*****	**0.38*****	**0.18**
Below norm deviation	0.07	0.06	0.07	0.06	0.05	0.06	0.04	0.04
Below norm x 2004 cohort	**0.32**	**0.32**	**0.28**	**0.43***	**0.35**	**0.40***	**0.38**	**0.65****
Instruction/classroom behavior								
Academic demand		0.27***				0.18**	0.12	0.11
Interactive Pedagogy			0.26***			0.18*	0.18*	0.20*
Absenteeism				-0.07***		-0.07***	-0.04***	-0.05***
Disciplinary problems					-0.19*	-0.06	-0	-0.01
Class average ability							0.72***	0.71***
Class ability x 2004 cohort							0.19	0.46**
Class ability x below norm							-0.33	-0.32
Class ability x below norm x 2004							-0.86*	-1.02*
Support course students in above-norm student's algebra class[1]								0.59***

* p<.05, **p<.01, ***p<.001. Models include controls for student backgrounds and incoming skills, as described in Appendix B.
[1] Classrooms are coded one if algebra classes taken by above-norm students had students taking support coursework and zero otherwise.

Table 6. Pre- and post-policy Algebra failure at the cutoff for double-algebra eligibility

	Algebra failure (in logits)							
	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 3	Model 4	Model 5
	Est.	Est.	Est.	Est.	Est.	Est.	Est.	Est.
Intercept (above-norm pre-policy)	-1.10***	-1.12***	-1.12***	-1.09***	-1.00***	-1.01***	-1.05***	-1.03***
2004 cohort deviation	**0.17***	**0.15**	**0.17***	**0.15***	**0.19***	**0.11**	**0.06**	**0.08**
Below norm deviation	-0.04	-0.05	-0.12	-0.11	-0. 1	-0.09	-0.09	-0.10
Below norm x 2004 cohort	**-0.40****	**-0.40****	**-0.39****	**-0.53*****	**-0.45*****	**-0.50*****	**-0.33***	**-0.36***
Instruction/classroom behavior								
Academic demand		0.23**				0.34***	0.31***	0.31***
Interactive Pedagogy			-0.12			-0.19*	-0.17	-0.17*
Absenteeism				0.06***		0.04**	0.06***	0.06***
Disciplinary problems					0.78***	0.72***	0.77***	0.77***
Class average ability							0.58***	0.55***
Support course students in above-norm student's algebra class[1]								-0.12

	Average algebra course failure rates by cohorts: logits converted to percentages							
	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 3	Model 4	Model 5
Above-norm students								
2002 cohort	25%	25%	25%	25%	27%	27%	26%	26%
2004 cohort	28%	28%	28%	28%	31%	28%	27%	27%
Below-norm students								
2002 cohort	24%	24%	23%	23%	25%	25%	24%	24%
2004 cohort	20%	20%	19%	17%	20%	18%	19%	19%

p<.10,* p<.05, **p<.01, ***p<.001. Models include controls for student backgrounds and incoming skills, as described in Appendix B.
[1] Classrooms are coded one if algebra classes taken by above-norm students had students taking support coursework and zero otherwise.

Table 7.

Sensitivity analysis: Estimates of the Effects of enrolling vs. not enrolling in double-dose Algebra coursework post-policy with different samples/methods

Comparison 1: Estimates from the original models that only use students who took the assigned courses (original) compared to instrumental variables models (IV)

Comparison 2: Estimates from models that only use survey takers to models based on the population of ninth graders

Differences in classroom environment/instruction between double- and single- algebra classes post-policy									
Academic demand		Interactive pedagogy		Absenteeism				Disciplinary problem	
Survey takers		Survey takers		Survey takers		Population		Survey takers	
Original	IV	Original	IV	Original	IV	Original	IV	Original	IV
-.05	-.01	.41***	.54***	1.4***	1.3***	1.3***	1.1***	.06*	.09

Differences in achievement outcomes between double- and single-algebra classes post-policy							
Algebra scores				Algebra failure (probability)			
Survey takers		Population		Survey takers		Population	
Original	IV	Original	IV	Original	IV	Original	IV
.41**	.83***	.25***	.58***	-0.08**	-0.06	-0.06***	-0.07*

Note. p<.05, ** p<.01, ***p<.001. Only students with survey data are included in these analyses.

Table 8. Sensitivity analysis: Estimates of post-policy changes in Algebra test scores and failure at the double-dose algebra eligibility cutoff scores for survey respondents and the ninth-grade population

	Algebra scores					
	Model 1		Model 2		Model 3	
	Survey	Pop	Survey	Pop	Survey	Pop
Intercept (above-norm pre-policy)	5.28***	5.36***	5.15***	5.08***	5.15***	5.08***
2004 cohort deviation	0.56***	0.56***	0.39***	0.47***	0.19	0.33***
Below norm deviation	0.07	0.02	0.04	0.12	0.03	0.12
Below norm x 2004 cohort	0.32	0.22*	0.40	0.43***	0.62*	0.59***
Peer absenteeism			-0.05***	-0.05***	-0.05***	-0.05***
Class average ability			0.75***	0.93***	0.74***	.97***
Class ability x 2004 cohort			0.19	-0.10	0.45*	0.04
Class ability x below norm			-0.37	-0.61***	-0.36*	-0.62***
Class ability x below norm x 2004			-0.94*	-0.24	-1.01*	-0.35
Support course students in above-norm student's algebra class [1]					0.56**	0.30**

[1]Classrooms are coded one if algebra classes taken by above-norm students had students taking support coursework and zero otherwise.

Table 8 (continued).
Pre- and post-policy changes in Algebra test scores and failure at the double-dose algebra eligibility cutoff scores

| | Algebra course failure (in logits) | | | | | |
| | Model 1 | | Model 2 | | Model 3 | |
	Survey	Pop	Survey	Pop	Survey	Pop
Intercept (above-norm pre-policy)	-1.10***	-0.81***	-1.16**	-0.90***	-1.13***	-0.96***
2004 cohort deviation	0.17*	0.24***	0.12	0.13	0.15	0.14
Below norm deviation	-0.04	0.01	-0.12	0.04	-0.10	0.04
Below norm x 2004 cohort	-0.40**	-0.28*	-0.36*	-0.23	-0.39*	-0.21
Peer absenteeism			0.10***	0.12***	0.10***	0.13***
Class average ability			0.57***	0.59***	0.54*	0.57***
Support course students in above-norm student's algebra class [1]					-0.14	-0.02

p<.10,* p<.05, **p<.01, ***p<.001. Analyses include all students who took the assigned courses.
[1]Classrooms are coded one if algebra classes taken by above-norm students had students taking support coursework and zero otherwise.