Opening the Gates: Detracking and the International Baccalaureate

ALLISON ATTEBERRY

University of Colorado School of Education

SARAH E. LACOUR

University of Colorado School of Education

CAROL BURRIS

Network for Public Education

KEVIN WELNER

University of Colorado School of Education

JOHN MURPHY

South Side High School

Background/Context: There is broad agreement about the benefits of taking AP and/or IB courses in high school. Nonetheless, student access to such courses remains uneven and inequitable, due largely to the practice of tracking students by perceived "ability." These tracking practices are often defended based on the contention that detracking and mixed-ability classes are impractical or unworkable. This study's conclusions inform the policy debate on the efficacy of detracking as an instructional strategy and add to the emerging literature concerning the potential of providing a school's most challenging, highest quality curriculum to all students.

Research Questions: We study a reform that combines two basic elements: detracking in Grades 6 through 10 plus open IB enrollment in Grades 11 and 12. We answer four questions: (a) Did greater numbers (and proportions) of students enroll in IB courses as the district progressively detracked its math and ELA courses? (b) As detracking took place over time, did IB courses become accessible to a broader range of students with respect to prior achievement? (c) How did students who enrolled in IB courses after detracking perform on their end-of-course IB assessments, conditional on prior achievement? Here we particularly focus on whether high-performing students appeared to perform worse on the IB assessments as the IB classes

were composed of higher numbers of students with lower prior achievement. (d) Conditional on taking a math IB course, did students become less likely to take the more challenging Math SL IB course (relative to the less challenging Math Studies IB course)?

Intervention/Program/Practice: Policy to detrack access to high school IB courses.

Research Design: We use an interrupted time series approach to examine whether the onset of detracking coincided with (a) increased IB participation or (b) decreased IB scores. We also document whether low-, middle-, or high-prior skill groups of students perform less well during detracking. Finally, we explore whether racial achievement gaps on statewide assessments were exacerbated by detracking.

Recommendations: The results associated with this detracking reform challenged two widespread beliefs. First, the school's highest achievers continued to succeed in the more heterogeneous IB classes. Second, the average IB scores for the school's lower achievers were the same or higher after detracking began, even though many more such students enrolled in those courses. In short, this case study documents the potential for not rationing the enriched, world-class curriculum of the International Baccalaureate.

Recognizing the far-reaching advantages of studying in challenging high school classes, increasing numbers of students seek out the most rigorous college-preparatory courses available to them, most notably International Baccalaureate (IB) and Advanced Placement (AP) courses (Clemmitt, 2006; Sadler, Sonnert, Tai, & Klopfenstein, 2010). This corresponds with the demands of policymakers for greater academic outcomes and for schools to expect their students to take more demanding courses (Obama, 2012). Yet, despite broad agreement about the benefits of taking such courses in high school (e.g., Kyburg, Hertberg-Davis, & Callahan, 2007), student access remains uneven and inequitable (Mathews, 1998; Oakes et al., 2000; Schmidt, Burroughs, Zoido, & Houang, 2015; Solórzano & Ornealas, 2004; Tyson, 2013).

Most schools that offer AP or IB[1] courses offer them only on a limited basis due to factors such as a perceived lack of student interest, low enrollment, inadequate funding, or insufficiently trained teachers (Clemmitt, 2006; Santoli, 2002). Students are also often denied access to these classes through school selection policies and practices. These selection policies limiting course access are generally based on a combination of the following three beliefs: (a) that curriculum should be stratified to meet the needs of students of differing achievement; (b) that schools have both the right and the obligation to decide student placement; and (c) that the placement of lower achieving students in high-track classes, such as IB and AP classes, will result in their own frustration and will, in addition, harm the learning of their higher achieving classmates (Klopfenstein, 2003).

Educators and others who advocate for maintaining selective access to IB and AP classes point to a subset of studies finding that detracked,

heterogeneous classes are associated with decreases in the learning of higher achievers (Brewer, Rees, & Argys, 1995; Kulik, 1992). Pursuant to this line of reasoning, they believe that if schools were to open the restrictive gates and allow all students access to advanced curricula, the results would be watered-down content and a decline in the achievement of the highest achieving students (see Attewell, 2001; Oakes, Wells, Jones, & Datnow, 1997). However, there are other studies that report no such decline in the learning of high achievers in heterogeneously grouped classes (Burris, Heubert & Levin, 2006; Figlio & Page, 2002; Figlio & Page, 2002; Mosteller, Light, & Sachs, 1996; Rui, 2009).[2] Moreover, even in those studies that do show a tracking benefit for high achievers, researchers cannot disentangle the effects of tracking from the effects of curriculum and other factors associated with high-track classes, such as better instruction and more qualified teachers (Kerckhoff, 1986; Oakes, 1986; Slavin & Braddock, 1993).

Yet the rationale for locking the gates to IB and AP classes extends beyond such arguments commonly heard from supporters of tracking. In his study of "star" public high schools, Attewell (2001) describes how some schools limit access to classes like IB and AP in order to enhance the chances of their most competitive students gaining entry to Ivy League colleges. He asserts that schools use sorting and stratification to help some students "stand out from the rest" in the college admission process (p. 268). His study also suggests that these schools limit access in an attempt to increase average scores on external exams, thus presenting an image of school excellence. This is an example of what has been referred to as "opportunity hoarding," where parents with advantages work to maintain those advantages for their children without regard to the consequences for other children (Lewis & Diamond, 2015).

Over the past couple of decades, however, new perspectives on who should have access to IB and AP courses have emerged (Clemmitt, 2006). Concerned with unequal access to IB and AP courses, the National Research Council now recommends that schools develop a "coherent plan" to increase the numbers of students who are prepared to take such courses and that schools treat "all students as potential" AP and IB participants when they are in Grades 6 through 10 (National Research Council [NRC], 2002, p. 198). The NRC thus advises that if we are to prepare more students to take IB and AP courses, that path must begin with an earlier set of enriched, challenging courses for all. Middle- and high-school course offerings should not be designed to "weed out" those students not expected to succeed. As explained later in this article, this is precisely the approach taken in the district under study—an approach that links detracking in Grades 6 through 10 to later IB success.

The NRC advice is echoed by the Association of American Universities (AAU) and the Pew Charitable Trusts, which recommend that key IB principles be extended and applied school-wide (Conley, 2005). Indeed, pointing to the key role of high expectations and a challenging curriculum, educators and researchers have argued that treating all students as potential participants in IB and AP courses is one key way of improving the achievement of struggling, at-risk students (see Hoffman, 2003) and that the combination of heterogeneous grouping with enriched, accelerated curriculum will not diminish the learning of higher achievers (Burris, Heubert & Levin, 2006).

Recent studies have described successful implementation of detracking reforms that expand the heterogeneity of the most challenging classes without adversely affecting learning, thus allowing more students to enjoy the opportunity of studying the best curriculum that schools have to offer (Alvarez & Meehan, 2006; Boaler, 2006; Burris, Welner, Wiley & Murphy, 2007). This detracking research follows decades of work documenting the harms of tracking for students placed in lower tracks (Oakes, 2005; see also Schofield, 2010, describing international research on this issue).

In this article we describe how a diverse suburban high school implemented a policy of unrestricted access to, and universal preparation for, its IB program by creating a common ninth- and tenth-grade accelerated program for all students. That is, we are studying a reform that combines two basic elements: detracking in Grades 6 through 10 plus open IB enrollment in Grades 11 and 12. In fact, during the later years of the study, the reform included IB as the only course option in English. The treatment accordingly combines pre-high-school preparation with high-school access. We analyze data from an 18-year period during which the IB program expanded from serving a few, elite students to serving a majority of the school's students. Using an interrupted time series approach, we find that overall achievement on the IB exams did not decline as the gates were opened. To the contrary, for IB math, the detracking initiative led to an increase in the number of students taking IB classes, as well as the same or higher performance on math IB tests across the full range of the prior achievement distribution. For the English IB test, participation increased dramatically for all students, and while there are no clear improvements in mean IB scores, there is also no evidence that top-performing students were harmed by the increasing heterogeneity of these classes.

The current study utilizes a broad policy lens through which to examine the effects of this detracking initiative. It reports on one piece of a set of long-term studies of this district and its high school. In other work, we also give voice to the more qualitative dimensions of the rollout and experiences of this reform (Welner, 2001; Burris et al., 2006; Burris et al.,

2007; Burris, Welner, & Bezoza, 2009). This study's conclusions inform the policy debate on the efficacy of detracking as an instructional strategy and add to the emerging literature concerning the potential of providing a school's most challenging, highest quality curriculum to all students. Further, given that tracking practices are often defended based on the contention that detracking and mixed-ability classes are impractical or unworkable (Loveless, 1998), this study provides an important existence proof. It provides solid evidence that a well-executed detracking reform can substantially improve both equity and excellence.

RESEARCH QUESTIONS

Using data provided by the Rockville Centre School District, combined with publicly available data from Common Core of Data (CCD) and information from IB Statistical Bulletin documents, we analyze IB trends associated with this district's reform effort. In particular, we first confirm that the reform was in fact associated, for the IB courses, with greater overall enrollment as well as the increased enrollment of students with lower prior achievement (as measured by the PSAT). We then examine IB test score trends before, during, and after the reform was instituted, and also as a function of students' prior PSAT scores. In order to disentangle these outcome trends from demographic trends, we control for any time-varying cohort composition trends using cohort-level demographic control variables from CCD. Finally, since there are two options for IB math—a more challenging "Math SL" (Standard Level) course, formerly known as Math Methods, *versus* the IB "Math Studies" course—we also examine the changing probability of enrolling in the more advanced of the two math IB courses. We do so to speak to a concern that tracking took place even within IB offerings. Formally stated, our research questions are as follows:

1. Did greater numbers (and proportions) of students enroll in IB courses as the district progressively detracked its mathematics and English language arts (ELA) courses?

2. As detracking took place over time, did IB courses become accessible to a broader range of students with respect to prior achievement?

3. How did students who enrolled in IB courses after detracking perform on their end-of-course IB assessments, conditional on prior achievement? Here we particularly focus on whether high-performing students appeared to perform worse on the IB assessments as the IB classes were composed of higher numbers of students with lower prior achievement.

4. Conditional on taking a math IB course, did students become less likely to take the more challenging Math SL IB course (relative to the less challenging Math Studies IB course)?

THE CONTEXT OF THE STUDY

THE DISTRICT OF STUDY

The study site is a suburban community of 28,000 in Nassau County, on Long Island, New York. The Rockville Centre School District's student population is approximately 3,500. Nearly three quarters of students at the school are white, approximately 10% are African American, around 12% are Latino, and roughly 3% are Asian. Most of the district's African American students are eligible for free or reduced-price lunch and live in a HUD (Department of Housing and Urban Development) housing project; the majority of Latino families reside in Section 8 subsidized housing in the downtown area of the larger village (see Burris, 2014). By contrast, most of the district's white families earn upper-middle-class incomes. During the period covered by this study, approximately 14% of all high school students were eligible for free or reduced-price lunch. In Appendix A, we present demographic trend data for the study population across the 18 study cohorts in order to demonstrate that the demographics are relatively stable over the study period. This is important because it addresses any concerns that changes in overall IB performance might be due to a changing student composition. If, for example, the school had seen a decrease in the proportion of students on free or reduced priced lunch—a student demographic that has historically struggled on IB tests—then one might think that such a change was the driving force behind any positive changes in student performance. However, the absence of clear demographic shifts in the available data serves to negate such a concern.

THE INTERNATIONAL BACCALAUREATE DIPLOMA PROGRAM

The IB program began in 1967 in order to serve the educational needs of geographically mobile students, such as the children of military personnel, diplomats, and international executives. These students needed high-quality academic education in order to meet university entrance requirements in any of 119 different countries (Duevel, 1999).

Participating schools must be accredited by the International Baccalaureate Organization (IBO) and must make a substantial commitment to teacher training and development.[3] Due to the demanding nature of the IB curriculum and the assessments, colleges around the globe

recognize the challenging curriculum of the program, and thus students may earn college credit for IB coursework in a manner consistent with the earning of AP credit. Students may elect either of two IB options. The first option is to take individual courses and earn an IB course certificate. The second is to declare an intention to take the courses and complete the other requirements for a full IB diploma.[4] (This study focuses only on IB courses, not the diploma.) Student learning is measured by criterion-referenced assessments, which are designed to be consistent from year to year and applied equally across schools (International Baccalaureate Organization [IBO], 2005). International senior examiners grade student work and assign scores that are in accordance with sound psychometric practices. (For an extensive discussion of the grading practices of the IB program see IBO, 2004.)

The district's International Baccalaureate program, introduced in 1981, was originally intended to serve high school students identified as "gifted and talented."[5] Thus, it began as a highly exclusive program serving only a handful of students, and it stayed small and selective for over a decade. Although over 90% of students graduating from the district's only high school, South Side High School (SSHS), went on to attend college, only 20% took one or more IB or AP courses, the most challenging courses that the school had to offer in any given subject area. After reviewing these data, the district concluded that the primary reason for the static size of the IB program was its selectivity—school requirements preventing or discouraging students from participating. The district focused its attention on identifying the gates that closed off access: teacher recommendations, grades, and prerequisite courses.

DETRACKING THE HIGH SCHOOL

The primary focus of reform soon became South Side, the district's only high school. Encouraged by the success of its earlier detracking efforts (for a complete history of earlier efforts to detrack middle school, see Appendix B), the high school began in 1999 to introduce heterogeneous grouping in the ninth grade. English and social studies classes were detracked first, followed by science in 2000 and mathematics in 2001. Since foreign language was detracked in the 1980s, the ninth-grade class that entered the high school in 2001—which we refer to as Cohort 2001—was the first class to be heterogeneously grouped in all subjects.

The ultimate target for detracking, however, eventually became the preparation of greater numbers of students for the IB program.[6] Accordingly, the tenth grade was detracked over time while the curriculum was transformed to become more enriched and challenging.[7] The school's revised curriculum was designed to be taught to all students in heterogeneously grouped classes.

English IB Coursework

Beginning with Cohort 2009, it became district policy that IB English was the only option for students in Grade 11. Likewise, beginning with Cohort 2011, IB English was the only option in Grade 12. Given this policy, one might expect to see in our results that 100% of students in the final cohorts participate in English IB. However, to foreshadow our findings, the highest observed participation rate we report for any given cohort is 82%. There are two explanations for this. First, a small number of students were not eligible to participate in English IB coursework even in these latter years (e.g., students in out-of-district placements, developmentally delayed students in a life-skills program, students in the district alternative school [fewer than 15], and students temporarily learning at home). All remaining students did participate in English IB coursework. However, the second reason that the participation rate appears lower than 100% is that a small group of students who enrolled in the course did not participate in the assessments, and we only observe students with IB scores. Together, these factors account for the discrepancy between a 100% participation policy and our reported 82% participation rate in the final cohorts. Our results are therefore an underestimate of actual participation.

Math IB Coursework

Members of Cohorts 1994 through 2000 experienced tracking in *both* ninth and tenth grade math courses. During this time, only students in upper tracks were prepared for IB coursework starting in Grade 11. Cohorts 2001 through 2003 experienced a detracked ninth grade, but a tracked tenth grade. Cohort 2004 was the first to experience fully detracked math curricula in ninth and tenth grade, which gave all students access to an enriched curriculum designed to prepare them for open enrollment IB math classes in Grade 11.

The International Baccalaureate Organization offers four options for high school math courses: Math Studies, Mathematics Standard Level (SL), Mathematics Higher Level (HL), and Further Mathematics HL. At SSHS, only the first two courses—Math Studies and Math SL—are

offered (typically taken in eleventh grade). Students can take either the less intensive Math Studies course or the more intensive Math SL course in Grade 11 (the latter of which includes an introduction to calculus), but not both. In lieu of IB Math HL for grade 12, SSHS instead offers AP Calculus AB, AP Calculus BC, and AP Statistics.[8] Typically students who take IB Math Studies in Grade 11 can go on to take AP Calculus AB or AP Statistics in Grade 12. Students who opt for the more rigorous IB Math SL course in Grade 11 are prepared to take AP Calculus BC in Grade 12. There could therefore be some concern that students are tracked even *within* the eleventh grade IB course offerings—a hypothesis we investigate below.

In summary, the detracking of South Side High School took place in three distinct phases: In Phase I, students were members of cohorts[9] that experienced tracking in *both* ninth and tenth grade—that is, before high school was detracked at all. In Phase II, cohorts experienced a detracked ninth grade, but a tracked tenth grade. In Phase III, cohorts experienced fully detracked math and ELA curricula in both ninth and tenth grade, which gave all students access to an enriched curriculum designed to prepare them for open enrollment, and in the last years, mandatory enrollment in IB classes in the final two years of high school.[10] A summary of the sequence of detracking is shown in Table 1.

Table 1 makes it clearer that detracking was complete in all grades leading up to IB course-taking (i.e., through grade 10) in South Side High School at different times for different subjects. Students enrolled in ninth grade in 2002–03—that is, Cohort 2002—were the first to experience a fully detracked English and social studies curriculum (through grade 10). Cohort 2004 was the first to experience a fully detracked math curriculum. Finally, Cohort 2006 was the first to experience a fully detracked science curriculum.

Table 1. District's Annual Progress Detracking its Middle School and High School

	Subjects That Were Detracked, by Grade Level				
Cohort	Experience in Grades 6–8	Experience in Grade 9	Experience in Grade 10	Experience in Grade 11	Experience in Grade 12
Cohort 1994	FL, ELA, SS	FL	FL		
Cohort 1995	FL, ELA, SS	FL	FL		
Cohort 1996	FL, ELA, SS	FL	FL		
Cohort 1997	FL, ELA, SS	FL	FL		
Cohort 1998	FL, ELA, SS, Mth	FL, ELA, SS*	FL		
Cohort 1999	FL, ELA, SS, Mth, Sci	FL, ELA, SS	FL		
Cohort 2000	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci	FL		
Cohort 2001	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL		
Cohort 2002	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS		
Cohort 2003	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS		
Cohort 2004	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth		
Cohort 2005	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth		
Cohort 2006	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci		
Cohort 2007	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci		
Cohort 2008	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci		
Cohort 2009	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci	ELA	
Cohort 2010	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci	ELA	
Cohort 2011	FL, ELA, SS, Mth, Sci	FL, ELA, SS, Sci, Mth	FL, ELA, SS, Mth, Sci	ELA	ELA

Note. Cohorts defined by Fall of 9th grade year (e.g., Cohort 1994 were in 9th grade in the 1994–95 school year). ELA = English, Mth= Math, Sci=Science, SS= Social Studies, FL = Foreign Language. *SS with honors by application.

DATA

The data for the current study come from two sources. First, information about each student's tenth grade PSAT score and IB exam score(s) is provided by the district itself. We identify each student as a member of one of 18 cohorts (1994 through 2011).[11] Each cohort is defined by the fall of its ninth-grade year (e.g., Cohort 1994 entered high school in the 1994–1995 academic year). On average, cohorts comprised approximately 270 students. Each student has one data record, including his or her verbal and math PSAT scores, as well as scores on any of the three IB tests taken. One limitation of the district data is that we only have records for students who ultimately took an IB exam.[12] We therefore cannot conduct a student-level analysis on the probability of taking IB courses over time. We therefore supplement the district data with information from the Common Core of Data (CCD) providing the overall size of each cohort and its demographics. This allows us to explore descriptively both the number *and* percentage of students who take IB exams over time, despite only having student records for IB takers. Below we discuss two variables of particular interest—tenth grade PSAT scores as a measure of prior achievement, and IB scores as the primary outcome.

To measure achievement or preparedness *prior to* IB enrollment, we use students' tenth-grade scores on the verbal and mathematics batteries of the PSAT[13]—a shortened version of the Scholastic Assessment Test (SAT)—to characterize students as initially low-, middle-, and high-performing as they enter into IB grades. Because the school district pays for all of its students to take the PSAT in the tenth grade, all students who take the IB have a PSAT score that precedes their enrollment in eleventh grade, as long as they were enrolled in the district in grade 10. In our analyses, we will use PSAT verbal scores when considering IB English test scores, and we will use PSAT math scores when examining IB Math test scores.

Our primary analyses in this study focuses on achievement on the International Baccalaureate subject-area assessments (a series of assessments taken during and at the end of IB courses, resulting in a single score). These are, we contend, outcomes that matter to the students and the school; the assessments are well constructed and measure important knowledge and skills (see IBO, 2004, for descriptions of the assessments). Moreover, almost all students in this school who are in IB courses take the assessments.[14] The IB assessments are scored on a scale of 1 to 7, wherein a score of '1' signifies work of the poorest quality; a '7' is considered exemplary.[15] Additional information about the IB English, IB Math Studies, and IB Math SL courses can be found in Appendix C.

METHODS AND ANALYSIS

We explore our four research questions by examining whether the relevant outcomes (IB participation, prior achievement score distributions of participants, IB scores, and IB Math test choice) change over time in a manner that corresponds to the three phases of detracking. We use an approach akin to an interrupted time series ("ITS") method. Due to data limitations (described above), we answer Research Questions 1 and 2 descriptively through data visualizations, while for Research Questions 3 and 4 we supplement visualizations with formal interrupted time series models that estimate changes in outcomes across the phases of detracking, separately for low-, middle-, and high-achieving students.

Given its centrality to this methodological approach, it is worth revisiting and formally defining our Phase I, Phase II, and Phase III indicator variables. Phase I occurs before the high school had begun the detracking reform, and members of these cohorts experienced tracking in *both* ninth and tenth grade. Eleventh- and twelfth-grade IB courses were therefore only open to students who had completed upper track coursework in the first two years of high school. In Phase II, cohorts experienced a detracked ninth grade, but a tracked tenth grade. In Phase III, cohorts experienced fully detracked math and ELA curricula in ninth and tenth grade, which gave all students access to an enriched curriculum designed to prepare them for open enrollment IB classes in the final two years of high school. Note that English courses were detracked earlier than math courses, and therefore the "phase" variables depend on subject. For English, Phase I includes Cohort 1994 to Cohort 1997, Phase II includes Cohort 1998–2001, and Phase III includes Cohort 2002–2011. For math, Phase I lasted longer (Cohort 1994–2000), Phase II includes Cohorts 2001–2003, and Phase III includes Cohorts 2004–2011.

For Research Question 1, which explores whether participation in IB courses increased as detracking occurred, we produce descriptive figures showing the percent of each cohort that participates over time. We look for evidence that relative to Phase I during which the high school was fully tracked, a greater proportion of students participate in IB courses in Phase II and Phase III. For Research Question 2, we produce histograms and descriptive statistics for the distribution of prior achievement (PSAT) scores among IB participants over time. Here, we are particularly interested in whether the average PSAT score of IB participants has decreased (which we would expect to happen as the access gates were opened), and thus whether the range of prior achievement is wider in the post-detracking phases. In this and subsequent analyses that look at subject-specific outcomes, we use matched subject prior achievement as a predictor (e.g.,

we use PSAT math scores to predict IB math outcomes and PSAT verbal scores to predict IB ELA outcomes).

For Research Questions 3 and 4, we use student-level data to estimate the impact of detracking on IB outcomes via an interrupted time series model. We compare outcomes across the three detracking phases, and we separately run these analyses for three groups of students classified as low-, middle-, or high-achieving (based on their performance on the tenth grade PSAT).[16] By conducting the analysis separately by prior achievement group, we can address what has arguably been the most contentious aspect of open-door policies to IB and AP classes: Does the inclusion of students with lower incoming achievement in IB classes attenuate the learning—and ultimately the IB exam scores—of students with higher incoming achievement? Below we present the basic interrupted time series model in Equation (1) estimated via ordinary least squares (however see Appendix D for all analyses with IB score outcomes also rerun using instead an ordered probit model[17]). The model below is run separately by prior achievement group and by IB subject (English, Math SL, and Math Studies):

$$IBScore_{ic} = \beta_0 + \beta_1(Cohort_{ic}) + \beta_2(Ph2_{ic}) + \beta_3(Ph3_{ic}) + \beta_4(Cohort_{ic} \times Ph2_{ic})$$
$$+ \beta_5(Cohort_{ic} \times Ph3_{ic}) + \beta_6(UnderP_{ic}) + (C_c)\gamma + \varepsilon_i \quad (1)$$

In Equation (1) we model IB scores for student i in cohort c ($IBScore_{ic}$) as a linear function of his/her ninth-grade cohort $Cohort_{ic}$, subject-specific detracking phase indicators, $Ph2_{ic}$ and $Ph3_{ic}$, interactions between cohort and phases, and a set of controls described below. In essence, this model allows us to estimate the mean level of IB scores and trends (increasing or decreasing) in IB scores in each of the three phases, and to do so separately for three categories of pre-achievement. If detracking led to lowered performance, then we would expect to see decreased means and/or downward trends in IB score performance across phases, particularly among the group with higher prior scores.

The coefficient on the variable, $Cohort_{ic}$, which is centered so that zero equals the last cohort that was part of Phase I, captures the linear trend in IB scores over time in Phase I. Due to the centering, the estimated coefficients on $Ph2_{ic}$ and $Ph3_{ic}$ provide us with an estimate of the difference in outcomes at the start of each period, relative to the end of Phase I. Thus, if high-achieving students did worse in Phase III than they did when the IB courses were more selective, we would anticipate a negative coefficient on $Ph3_{ic}$ when this model is run on the group of high achievers. We interact $Cohort_{ic}$ with subject-specific detracking phase indicators, $Ph2_{ic}$ and $Ph3_{ic}$. The coefficients on these interaction terms tell us how the linear trends over time changed across phases. Here, the concern is that the

students in the highest group on prior achievement would have negative slope in Phase III; that is, as IB classes are open to all students, that top students would see a decline in IB scores over time. We also control for a vector of grand-mean-centered cohort demographic characteristics (C_c) that include percentage Black, Latino/a, Asian, "Other or Unknown," and percentage female for each of Grades 9 through 12. We also include the only available student-level demographics, ($UnderP_{ic}$), a dummy variable indicating underprivileged status (also grand-mean centered).

For Research Question 4, we explore whether, conditional on participating in a math IB course, students are more likely to enroll in the Math SL or Math Studies course. As noted above, Math SL is a more advanced course than is Math Studies. Because the increased enrollment in IB courses likely draws disproportionately from the lowest group on prior achievement, we were curious to see whether these students were taking the more or less challenging of the math options. There was also some concern that in preparing all students for IB course work, even students from the highest group on prior achievement might become increasingly likely to take the less demanding of the courses, with their new colleagues. To examine this, we slightly modify Equation (1) to have a binary outcome (where 1 indicates Math SL rather than Studies (MLS_{ic}) for students who took one of the two math courses) and to use a logistic regression approach, while the right-hand side of the model remains the same. Here, again, the coefficients on $Ph2_{ic}$ and $Ph3_{ic}$ indicate the change in probability of taking the more strenuous course at the beginning of these phases, relative to the end of Phase I. The coefficients on the interacted variables indicate changes in the trend between phases.

While this time series approach provides us with information about trends over time for students in each of three groups based on prior achievement, we were also concerned that this obscured something important about the composition within those three broad groups and what might be happening in different phases to students with the same prior achievement. We thus took a second approach to answering Questions 3 and 4, looking specifically at how students with the same prior achievement (PSAT score) performed on each of the IB exams in each of the three phases.

$$IBScore_{ic} = \beta_0 + \beta_1(PSAT_{ic}) + \beta_2(Ph2_{ic}) + \beta_3(Ph3_{ic}) + \beta_4(PSAT_{ic} \times Ph2_{ic}) + \beta_5(PSAT_{ic} \times Ph3_{ic}) + \beta_6(UnderP_{ic}) + (C_c)\gamma + \varepsilon_i \tag{2}$$

In Equation (2) we model IB scores for student i in cohort c ($IBScore_{ic}$) as a function of prior achievement as measured by $PSAT_{ic}$—a continuous PSAT test score that has been centered at the middle of the range of

scores for a given achievement group.[18] We interact $PSAT_{ic}$ with subject-specific detracking phase indicators, $Ph2_{ic}$ and $Ph3_{ic}$, and we control for the same vector of demographics. Now, the coefficients on $Ph2_{ic}$ and $Ph3_{ic}$ represent the difference in IB scores for a student at the middle of each prior achievement group relative to Phase I. Whereas the ITS approach in Equation (1) follows a more traditional approach of looking for evidence that levels of, and trends in, IB scores change as the detracking policy changed, it is not entirely clear that we should expect *jumps* in IB scores or strong increases in IB score trends in later phases. The approach in Equation (2) instead highlights whether students across the distribution of prior achievement appear to be doing better, worse, or the same in later phases of the detracking policy.

RESULTS

Before addressing each of our four research questions in turn, we begin by examining how IB scores in this district compare to world mean scores over time. Figure 1 shows mean IB performance for each of the three exams in question for SSHS (solid line) and the world (dashed line). These graphs indicate that the mean SSHS performance on each test fared no worse—relative to the international average—following detracking than before the detracking reforms were implemented (where the dashed vertical line indicates when the content area was fully detracked). In fact, the trends in SSHS mean annual scores seem relatively flat, suggesting that fluctuations in class mean scores appear to be independent of the increases in IB enrollment and the broader range of incoming achievement. This provides some first-order evidence that the detracking initiative was not a driver of lowered overall IB performance; however, we explore this more concretely below.

RQ 1: IB PARTICIPATION RATES

As the high school progressively detracked Grades 9 and 10, enrollment in eleventh and twelfth-grade IB courses grew for the detracked cohorts. In Figure 2, we see that overall participation in both subject areas increased from a starting point of only about 20 to 30% of students taking the course and having an IB score to recent numbers in the 70 to 80% range.[19] Detracking, by providing all with a pre-IB curriculum, is associated with a dramatic increase in the numbers and proportions of students taking IB courses. Next, we consider whether the increased participation coincided with increased heterogeneity among participants with respect to prior achievement.
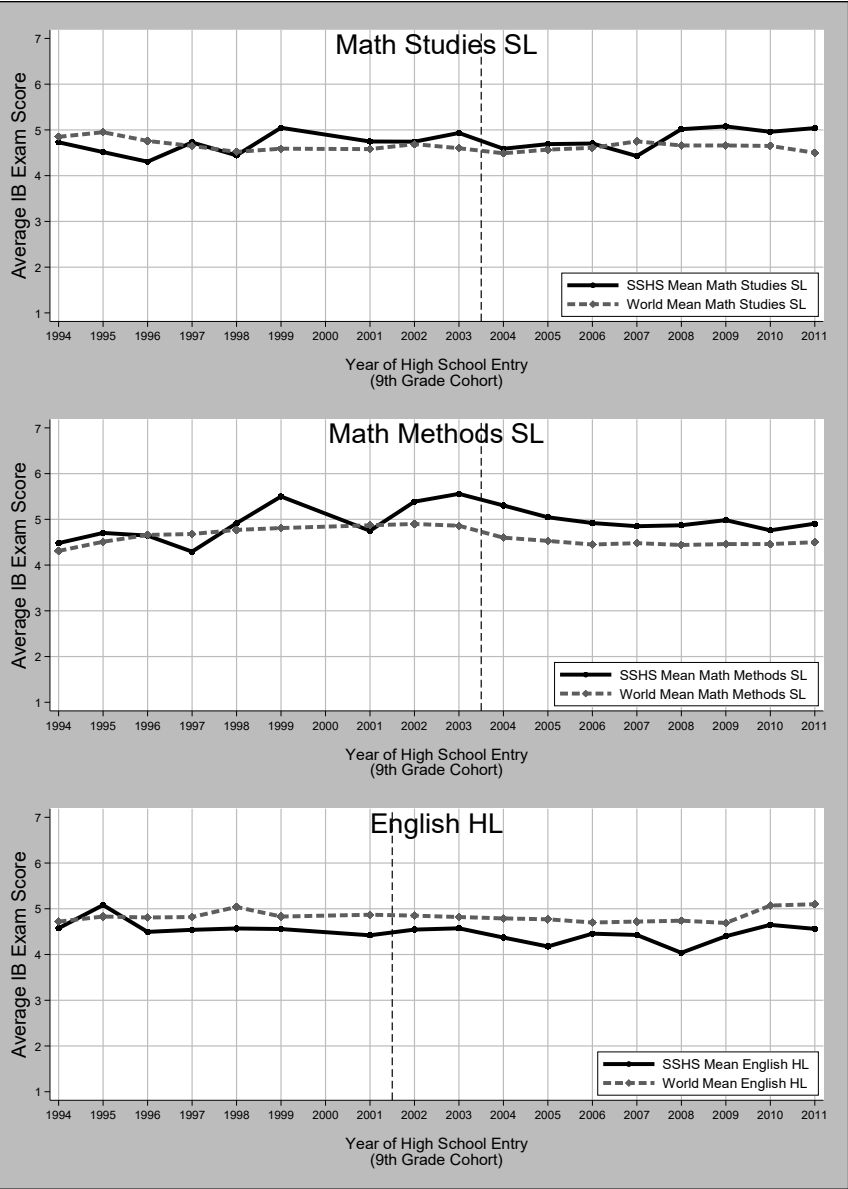
Figure 1. Study school (SSHS) and world mean IB scores over time in standard level (SL) and higher level (HL) courses
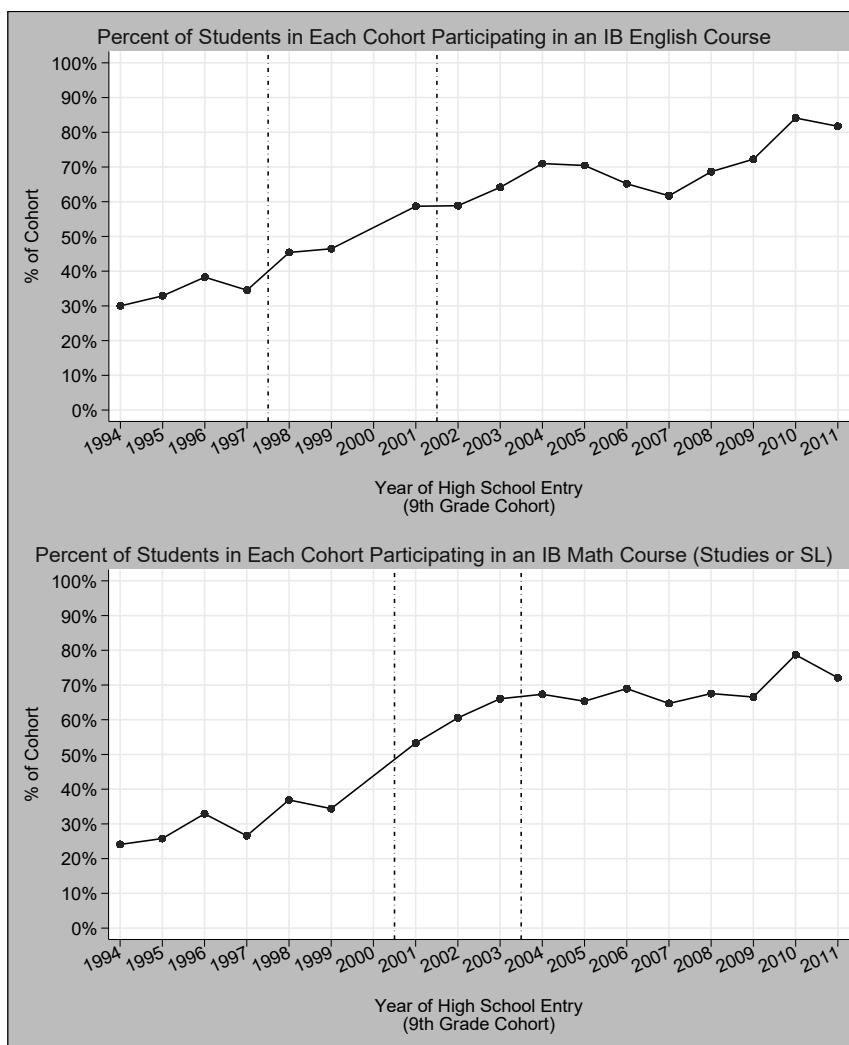
Figure 2. SSHS IB participation rates across cohorts, by subject

RQ 2: PRIOR ACHIEVEMENT DISTRIBUTION OF IB PARTICIPANTS

The increased participation in IB classes following detracking appears to have drawn from students with lower PSAT scores, those who generally have been denied access to IB credit. In Figure 3, the histograms depict the distribution of prior PSAT achievement for each of the three phases of detracking. For all three IB subjects, these distributions shift to the

left, suggesting that, as expected and intended, students with lower prior achievement in the same subject have greater access to IB courses in the later phases of detracking.

Although the distributions shift left, it is somewhat surprising that they do *not* systematically become wider over time (as documented by nondiminishing standard deviations across the phases). This means that while detracking attracted students from the lower groups on prior achievement, a lower number of participating students had PSAT scores at the very high level in Phase II and III.[20]

Assuming that this lower number did not occur by chance, there are at least two possible explanations why one might observe a decreased presence, in IB courses, of students who scored at the highest percentiles on the PSAT. One possible explanation is that because of changes in curriculum and/or tracking in lower grades, the highest achieving students no longer score as well on the PSAT. We considered this even though the SAT and PSAT are not achievement tests in the same sense as the standardized tests used by states to measure classroom learning—and are therefore less likely to reflect curricular changes.[21] This explanation runs counter to this school district's documented success with detracking, which earlier research has shown to be either beneficial or neutral (Burris, Heubert, & Levin, 2006; Burris, Wiley, Welner, & Murphy, 2008) for the district's high achievers.[22]

A second possibility, which we consider to be more likely, is that a small number of highly advantaged families left this district during the early years of the study (the difference between cohorts in the three phases is just a few students per year in these high percentile ranges). PSAT and SAT scores are correlated with wealth, so very high scorers would likely disappear if there were fewer families in the highest income brackets living in the district. With the available data, we cannot observe this latter phenomenon of wealthy migration, but we do know anecdotally that during this time period wealthy families from this and other districts moved to Long Island's "Gold Coast" on the north shore.

RQ 3: IB SCORES OVER TIME, BY PRIOR ACHIEVEMENT

In Tables 2–4, we examine the temporal relationship between cohort and IB score in Math Studies, Math SL, and English IB courses, respectively (see also visualizations of Tables 2–4 in Appendix C). These correspond to Equation (1). If detracking had the sort of negative consequences sometimes feared, then we should expect to see lower IB scores for high-PSAT students in Phases II and III than in Phase I.
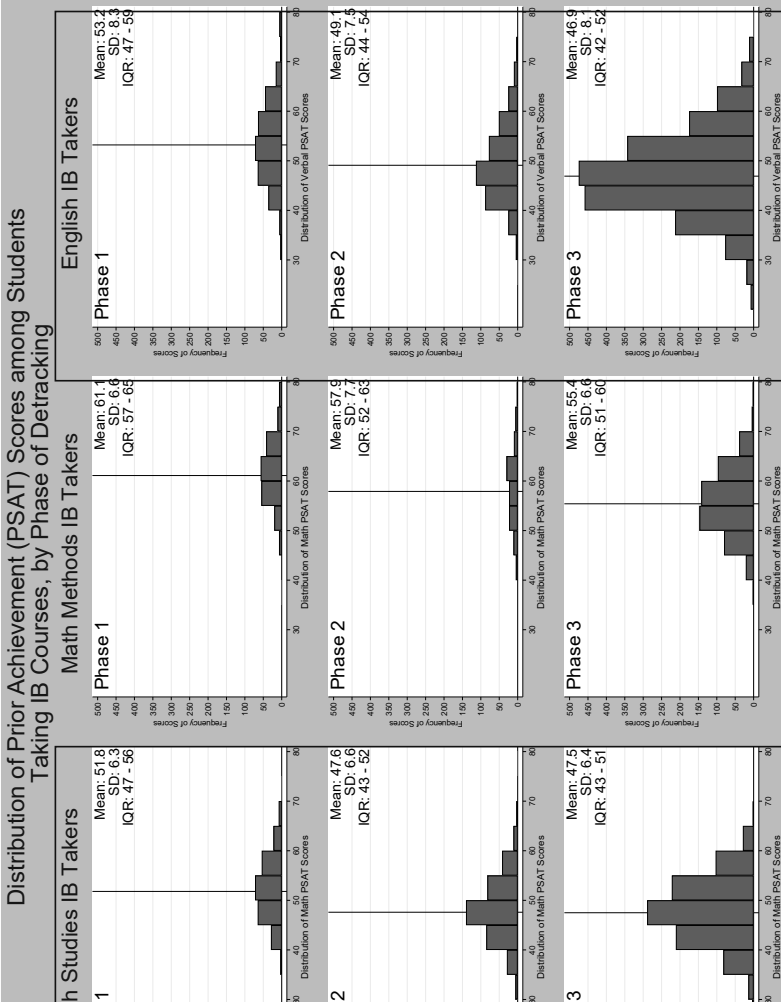
Distribution of Prior Achievement (PSAT) Scores among Students Taking IB Courses, by Phase of Detracking

Figure 3. Prior achievement (PSAT) distribution of students taking IB courses, by subject and phase

Table 2. Math Studies: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	4.647 ***	5.212 ***	5.503 ***	4.777 ***	5.149 ***	6.144 ***
	(0.197)	(0.204)	(0.336)	(0.920)	(0.635)	(1.032)
Phase 2	-0.210	-0.075	-0.307	0.387	-2.253	-5.804
	(0.270)	(0.339)	(0.743)	(2.793)	(3.750)	(7.317)
Phase 3	-0.648 **	-0.430	-0.205	-0.770	0.514	3.358
	(0.237)	(0.295)	(0.560)	(1.235)	(2.333)	(4.311)
Cohort	0.115	0.143 *	0.088	-0.123	0.255	0.230
	(0.063)	(0.056)	(0.089)	(0.402)	(0.168)	(0.393)
Cohort x Phase 2	-0.055	-0.070	0.198	-0.190	0.904	2.329
	(0.103)	(0.129)	(0.274)	(1.421)	(1.316)	(2.642)
Cohort x Phase 3	-0.042	-0.064	-0.029	0.162	-0.286	-0.758
	(0.065)	(0.062)	(0.106)	(0.368)	(0.504)	(0.989)
R-Squared	0.025	0.105	0.094	0.102	0.152	0.213
N	1,063	411	136	1,063	411	136
Covariates?	No	No	No	Yes	Yes	Yes

Table 3. Math SL: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	4.667 ***	5.074 ***	5.700 ***	5.080 **	6.178 ***	4.035 ***
	(0.769)	(0.301)	(0.214)	(1.744)	(1.301)	(0.782)
Phase 2	-1.199	-1.850 **	-0.622	-1.062	-6.662	-0.630
	(1.013)	(0.577)	(0.427)	(3.500)	(7.689)	(1.318)
Phase 3	-0.015	-0.314	-0.170	0.319	0.912	1.325
	(0.849)	(0.382)	(0.342)	(1.594)	(5.008)	(0.796)
Cohort	0.200	0.272 **	0.210 ***	0.050	0.582	0.318
	(0.281)	(0.102)	(0.055)	(0.678)	(0.311)	(0.226)
Cohort x Phase 2	0.257	0.468 *	0.128	0.127	1.853	0.982
	(0.400)	(0.234)	(0.182)	(0.530)	(2.692)	(0.717)
Cohort x Phase 3	-0.261	-0.264 *	-0.196 **	-0.244	-0.883	-0.227
	(0.284)	(0.106)	(0.065)	(0.714)	(1.055)	(0.181)
R-Squared	0.042	0.084	0.128	0.16	0.11	0.19
N	153	303	389	153	303	388
Covariates?	No	No	No	Yes	Yes	Yes

Table 4. English: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low V PSAT	Mid V PSAT	High V PSAT	Low V PSAT	Mid V PSAT	High V PSAT
Constant	4.092 ***	4.922 ***	4.979 ***	5.185 ***	7.681 ***	7.588 ***
	(0.100)	(0.124)	(0.164)	(0.482)	(0.600)	(0.804)
Phase 2	0.212	0.237	0.092	-1.755 *	-4.277 ***	-4.733 ***
	(0.140)	(0.212)	(0.321)	(0.751)	(0.965)	(1.320)
Phase 3	-0.037	-0.041	0.243	1.608 ***	1.168 *	2.743 **
	(0.121)	(0.180)	(0.292)	(0.371)	(0.557)	(0.873)
Cohort	-0.013	0.066	-0.128	0.927 **	1.424 ***	1.472 ***
	(0.065)	(0.072)	(0.084)	(0.306)	(0.343)	(0.386)
Cohort x Phase 2	-0.026	-0.184	0.206	-1.401 **	-1.909 ***	-1.925 ***
	(0.074)	(0.096)	(0.134)	(0.432)	(0.493)	(0.567)
Cohort x Phase 3	0.030	-0.074	0.131	-1.168 **	-1.820 ***	-1.960 ***
	(0.065)	(0.073)	(0.088)	(0.374)	(0.424)	(0.482)
R-Squared	0.006	0.01	0.009	0.066	0.092	0.111
N	1,705	574	351	1,704	573	351
Covariates?	No	No	No	Yes	Yes	Yes

For both of the math tests, we find that student performance, across achievement groups, was increasing modestly in Phases I and II, but appeared to even out in Phase III (although for Math SL, in Phase III the slopes for the highest and middle achievement groups appear slightly positive while the slope of the lowest group is slightly negative). The average IB score, nonetheless, does not appear to be significantly changed from the end of Phase I to the beginning of either Phase II or III. In corresponding regression models that include covariates (right-hand side of Tables 2 and 3), there are no statistically significant increases in mean IB scores from the end of Phase I to the start of Phase II or III, and no statistically significant differences in trends across the phases (recall that the coefficients on the interaction terms ($Cohort_{ic} \times Ph2_{ic}$) and ($Cohort_{ic} \times Ph3_{ic}$) capture how the linear trends over time changed across phases). In other words, detracking (along with the corresponding increase in participation in IB courses) is not associated with lower mean math IB performance for any group.

For English IB scores (Table 4), the story is slightly different. In Phase I and II, trends in IB scores are more varied across prior achievement groups. However during Phase III, the trends for all groups appear mostly flat. Overall, the findings for English IB scores support the assertion that students in Phase III perform similarly, on average, to their Phase I counterparts, when far fewer students were taking the challenging courses. Results in Table 4 generally exhibit no statistically significant differences in mean IB scores or trends in IB scores across the phases. When we control for student- and cohort-level covariates, the positive difference in IB scores in Phase III, relative to Phase I, is statistically significant for all three achievement groups. Taken as a whole, these results tend to support the idea that detracking does not harm students in the highest achievement groups. Still, we look next at IB achievement as a function of PSAT achievement (rather than as a function of time).

We now revisit Research Question 3 by examining whether students with the same PSAT scores appear to perform better if they were in Phase I, II, or III of detracking—a slightly different lens on the same question. Tables 5–7 contain the estimated coefficients that capture the relationship between PSAT achievement and IB score for Math Studies, Math SL, and English across different phases. Figures 4 through 6 represent correspond to these three tables.

Table 5. Math Studies: IB Scores as a Linear Function of Math PSAT Scores and Phase, Separately by Prior Achievement Groups

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	1.957 **	4.750 ***	4.577 ***	2.543 ***	5.534 ***	5.285 ***
	(0.633)	(0.096)	(0.386)	(0.651)	(0.307)	(0.654)
Phase 2	1.155	0.596 ***	1.717 **	0.436	-0.235	1.128
	(0.686)	(0.140)	(0.553)	(0.715)	(0.370)	(0.807)
Phase 3	0.650	0.662 ***	1.886 **	-0.017	-0.346	1.507
	(0.655)	(0.115)	(0.667)	(0.684)	(0.377)	(0.889)
PSAT Score	0.111 ***	0.047	-0.083	0.117 ***	0.015	-0.129 *
	(0.029)	(0.049)	(0.047)	(0.028)	(0.049)	(0.050)
PSAT x Phase 2	-0.036	0.058	0.133 *	-0.039	0.092	0.202 **
	(0.032)	(0.070)	(0.064)	(0.031)	(0.069)	(0.065)
PSAT x Phase 3	-0.010	0.010	0.165 *	-0.024	0.048	0.298 ***
	(0.031)	(0.058)	(0.076)	(0.030)	(0.057)	(0.083)
R-Squared	0.153	0.09	0.115	0.232	0.171	0.306
N	1,063	411	136	1,063	411	136
Covariates?	No	No	No	Yes	Yes	Yes

Table 6. Math SL: IB Scores as a Linear Function of Math PSAT Scores and Phase, Separately by Prior Achievement Groups

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	0.189	4.301 ***	5.354 ***	2.891	5.618 ***	5.636 ***
	(2.319)	(0.180)	(0.124)	(2.797)	(0.501)	(0.386)
Phase 2	1.784	0.543 *	0.914 ***	-4.498	-0.919	0.881
	(3.120)	(0.255)	(0.236)	(3.783)	(0.623)	(0.451)
Phase 3	1.411	0.519 **	1.004 ***	-1.076	-1.114	0.705
	(2.445)	(0.192)	(0.196)	(2.927)	(0.600)	(0.484)
PSAT Score	0.176	0.073	0.082 ***	0.074	0.093	0.101 ***
	(0.102)	(0.091)	(0.018)	(0.120)	(0.095)	(0.017)
PSAT x Phase 2	-0.064	-0.036	0.017	0.199	0.012	0.006
	(0.138)	(0.127)	(0.033)	(0.162)	(0.137)	(0.032)
PSAT x Phase 3	-0.059	0.018	0.028	0.028	0.013	0.015
	(0.108)	(0.096)	(0.027)	(0.125)	(0.100)	(0.026)
R-Squared	0.103	0.053	0.214	0.247	0.145	0.338
N	153	303	389	153	303	388
Covariates?	No	No	No	Yes	Yes	Yes

Table 7. English: IB Scores as a Linear Function of Verbal PSAT Scores and Phase, Separately by Prior Achievement Groups

|	Models Without Covariates			Models With Covariates		
	Low V PSAT	Mid V PSAT	High V PSAT	Low V PSAT	Mid V PSAT	High V PSAT
Constant	3.505 ***	4.852 ***	5.764 ***	3.173 ***	5.197 ***	5.446 ***
	(0.318)	(0.078)	(0.148)	(0.341)	(0.298)	(0.530)
Phase 2	-0.382	0.050	0.251	-0.162	-0.167	0.259
	(0.379)	(0.111)	(0.289)	(0.389)	(0.242)	(0.447)
Phase 3	-0.090	-0.032	-0.182	0.290	-0.435	0.212
	(0.324)	(0.088)	(0.201)	(0.359)	(0.354)	(0.634)
PSAT Score	0.030	0.050	0.093 ***	0.029	0.056	0.093 ***
	(0.016)	(0.039)	(0.019)	(0.015)	(0.038)	(0.020)
PSAT x Phase 2	0.026	0.049	0.013	0.025	0.038	0.011
	(0.019)	(0.056)	(0.036)	(0.018)	(0.055)	(0.036)
PSAT x Phase 3	0.015	-0.004	-0.044	0.016		-0.041
	(0.016)	(0.043)	(0.026)	(0.016)	(0.043)	(0.027)
R-Squared	0.105	0.024	0.116	0.158	0.079	0.156
N	1,705	574	351	1,704	573	351
Covariates?	No	No	No	Yes	Yes	Yes
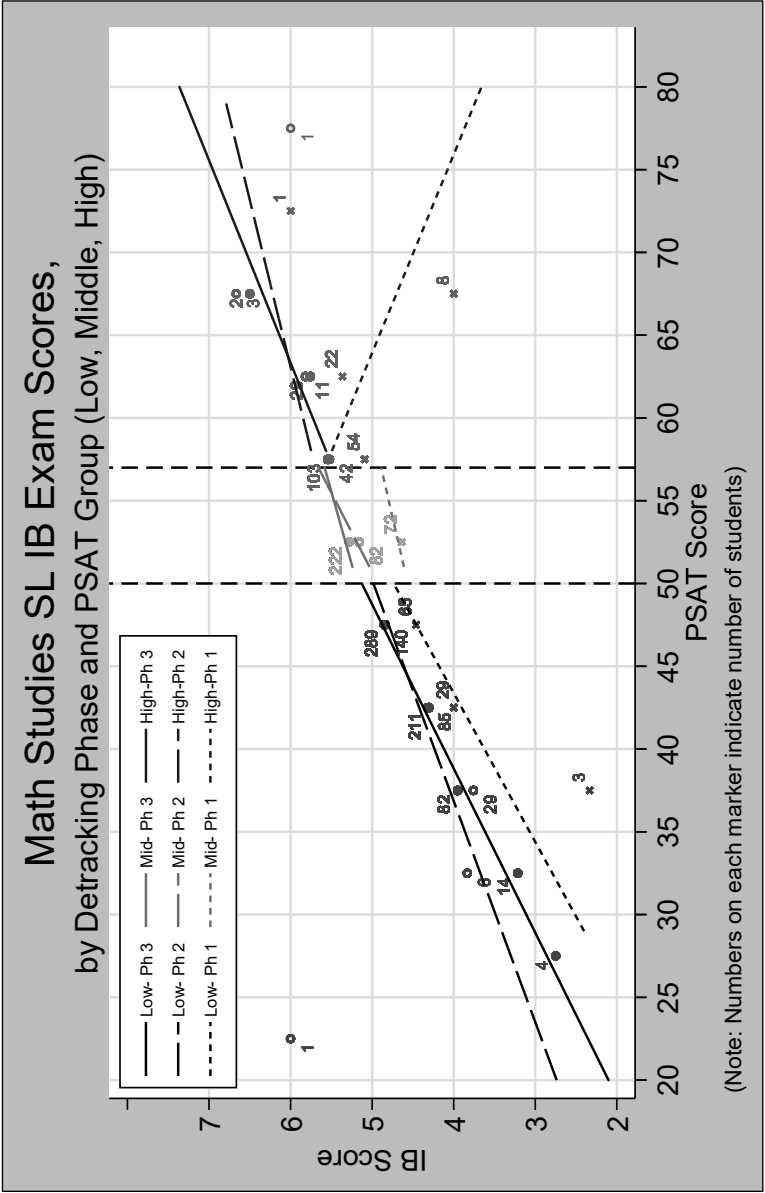
Figure 4. Math Studies SL: IB scores as a linear function of PSAT scores and phase, separately by prior math achievement groups
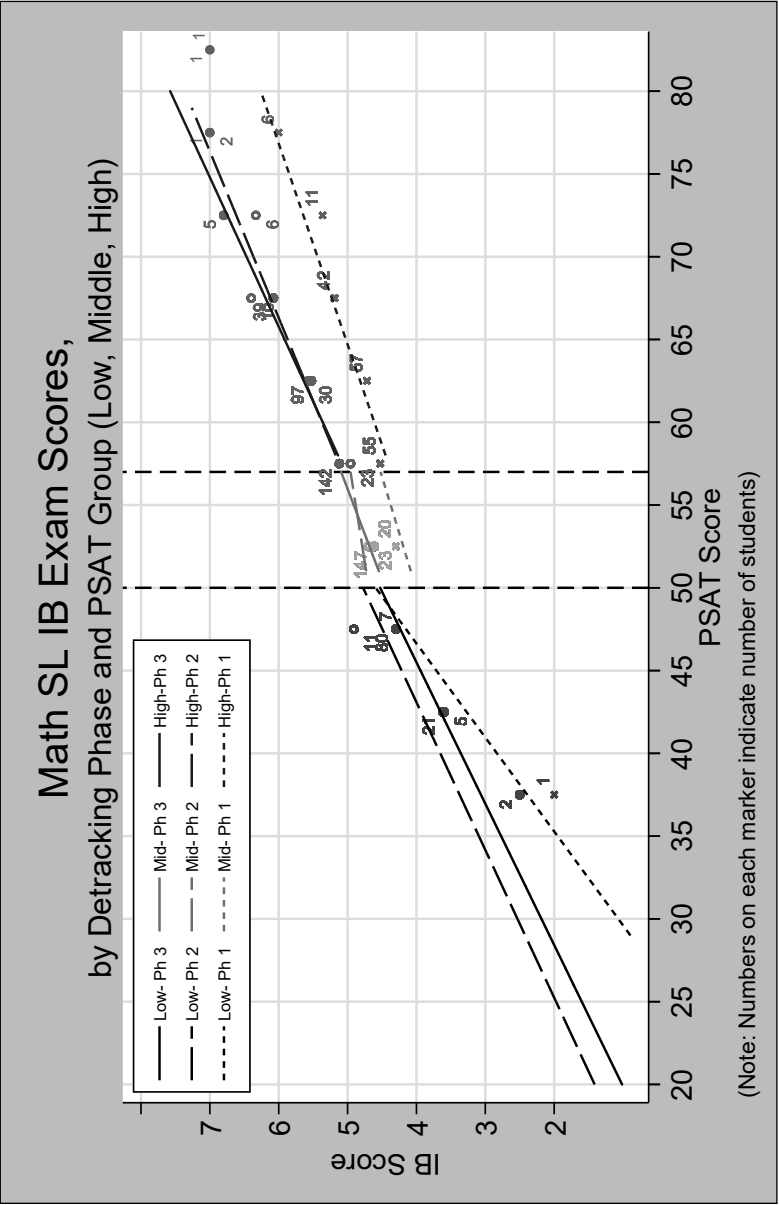
Math SL IB Exam Scores,
by Detracking Phase and PSAT Group (Low, Middle, High)

(Note: Numbers on each marker indicate number of students)

Figure 5. Math SL: IB scores as a linear function of PSAT scores and phase, separately by prior math achievement groups
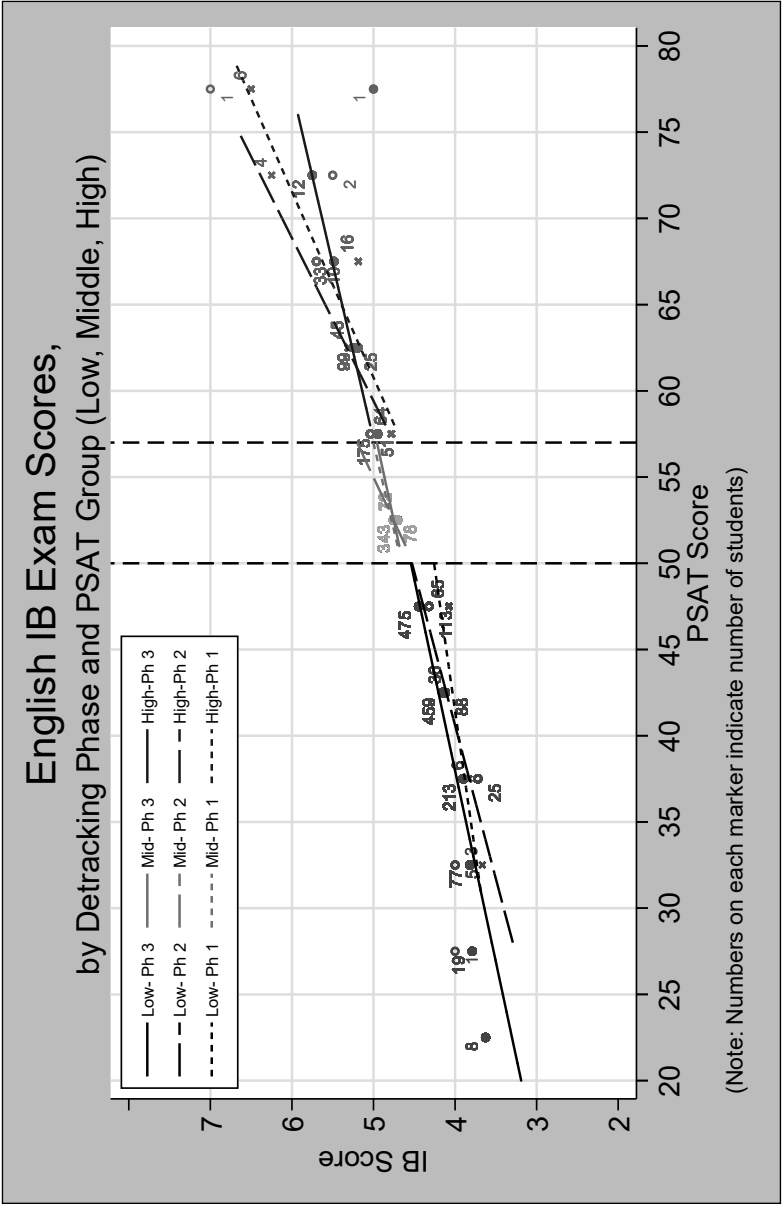
Figure 6. English: IB scores as a linear function of PSAT scores and phase, separately by prior verbal achievement groups

If students of the same prior achievement are doing better in later phases of detracking, we should expect to see that the lines (representing phases) are higher for the dashed (Phase II) and solid (Phase III) lines than for the dotted (Phase I) line. In Figures 4 and 5—Math Studies and Math SL—we see that across the distribution of prior achievement, students are obtaining higher scores on the IB exam in Phase II and III than were similar students in Phase I. Importantly, this means that students in the top tier of PSAT takers perform slightly *better* in the final phase of detracking than they did prior to detracking. We note that these improvements are statistically significant for both the middle and top tiers of students when estimated without covariates. However, when we control for cohort racial/ethnic demographics, the estimated coefficients on phase are no longer statistically significant.

For English IB scores, shown in Figure 6 (and Table 7), patterns are less distinct across the phases. While there appear to be no differences in the IB score performance of students with low PSAT scores, there is some evidence that students at the very highest end of the PSAT score distribution obtain slightly lower mean IB scores in Phase III. We note, however, that these differences are not statistically significant and largely reflect the noise associated with having so few students who fall into this highest category of prior achievement.

Many of the estimated differences in outcomes across phases are not statistically significant, and in the current study that null finding is of great interest. It suggests that during a time when participation in IB courses among students with low PSAT scores increased substantially, performance on IB assessments did not change. Indeed, when it comes to math IB scores, there is even some evidence that students with the highest PSAT scores performed slightly *better* in Phase III. In Appendix D, we present nonparametric versions of Figures 4, 5, and 6, which provide some assurance that our findings are not driven by our choice of functional form or cut scores for low, middle, and high PSAT scores.

RQ 4: CHALLENGE OF SELECTED COURSE

In response to our final research question, we find that the students were not substituting the less challenging Math Studies course for the more challenging Math SL course during the final phase of detracking (Table 8). The likelihood of choosing Math SL rather than Math Studies appears higher, on average, in the third phase for students from all three categories of prior achievement, although it appeared to decrease somewhat over time for those students in the highest achievement group during Phase II. As we have done for all analyses, we also model

this outcome as a function of prior PSAT scores, separately by phase (see Figure 7 and Table 9). Here we can more clearly see that though the probability of taking the more difficult Math SL course is much lower among students with low prior PSAT scores, the propensity to take the harder course does not decrease across the distribution of prior achievement. This suggests to us that students in all achievement groups were equally confident in their preparation to meet the challenges of the harder course and assessment in every phase of detracking.

In sum, the district's detracking reform is associated with three key findings: First, detracking opened IB coursework to students who were historically denied access based on lower achievement. Moreover, these newly incorporated students succeeded in the more challenging curriculum without lowering the average school scores on IB assessments. Second, the onset of the reform is associated with higher scores on the math IB exams across all three levels of prior achievement. Third, detracking was not associated with a decreased likelihood of taking the more challenging mathematics IB course, across all levels of prior achievement.

LIMITATIONS

The study described here documents a dramatic increase across 18 cohorts in IB participation alongside sustained IB test score performance across the distribution of prior achievement. Yet it is important to acknowledge that the aggregated, quantitative data used herein offers only a certain kind of insight into the detracking experience. In other work, we rely on more qualitative evidence to illuminate the individual experiences of detracking in Rockville Centre School District (see Burris & Welner, 2007; Burris, 2014). As discussed in the data section, our quantitative analysis is limited somewhat by the fact that we only have records for students who ultimately completed an IB assessment, rather than the full population of SSHS students in every cohort. This dataset allows us to examine changes among IB test-takers but does not allow us to compare them to students in those cohorts who opted not to take IB assessments.

Table 8. Probability of Taking Math SL Rather Than Math Studies, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	0.115	0.478 ***	0.731 ***	0.088	0.453	0.864 **
	(0.059)	(0.086)	(0.073)	(0.285)	(0.291)	(0.271)
Phase 2	-0.057	-0.241	0.275	-0.577	-0.640	0.512
	(0.081)	(0.152)	(0.149)	(0.880)	(1.711)	(0.461)
Phase 3	-0.085	-0.085	0.048	-0.082	0.441	-0.025
	(0.071)	(0.117)	(0.119)	(0.381)	(1.087)	(0.272)
Cohort	0.008	0.052 *	-0.010	0.082	0.042	-0.066
	(0.019)	(0.025)	(0.019)	(0.128)	(0.073)	(0.080)
Cohort x Phase 2	-0.005	-0.040	-0.148 *	0.186	0.146	-0.289
	(0.031)	(0.059)	(0.060)	(0.444)	(0.600)	(0.250)
Cohort x Phase 3	0.008	-0.038	0.006	-0.055	-0.093	0.045
	(0.020)	(0.027)	(0.022)	(0.117)	(0.231)	(0.064)
R-Squared	0.024	0.047	0.021	0.033	0.07	0.053
N	1,214	708	523	1,214	708	522
Covariates?	No	No	No	Yes	Yes	Yes

Table 9. Probability of Taking Math SL Rather Than Math Studies as Function of Math PSAT Scores and Phase, Separately by Prior Achievement Groups

|	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Constant	-0.133	0.298 ***	0.894 ***	0.017	0.313 *	0.810 **
	(0.199)	(0.042)	(0.048)	(0.210)	(0.129)	(0.138)
Phase 2	0.018	-0.027	0.029	-0.154	-0.075	0.072
	(0.217)	(0.061)	(0.097)	(0.232)	(0.158)	(0.172)
Phase 3	-0.064	0.206 ***	0.111	-0.241	0.204	0.213
	(0.206)	(0.048)	(0.076)	(0.221)	(0.157)	(0.174)
PSAT Score	0.011	0.061 **	0.024 ***	0.010	0.060 **	0.024 ***
	(0.009)	(0.022)	(0.007)	(0.009)	(0.022)	(0.007)
PSAT x Phase 2	-0.001	-0.036	0.015	-0.001	-0.030	0.008
	(0.010)	(0.031)	(0.013)	(0.010)	(0.031)	(0.013)
PSAT x Phase 3	0.007	-0.029	0.012	0.008	-0.028	0.011
	(0.010)	(0.024)	(0.010)	(0.010)	(0.024)	(0.010)
R-Squared	0.055	0.064	0.090	0.070	0.093	0.125
N	1,214	708	523	1,214	708	522
Covariates?	No	No	No	Yes	Yes	Yes
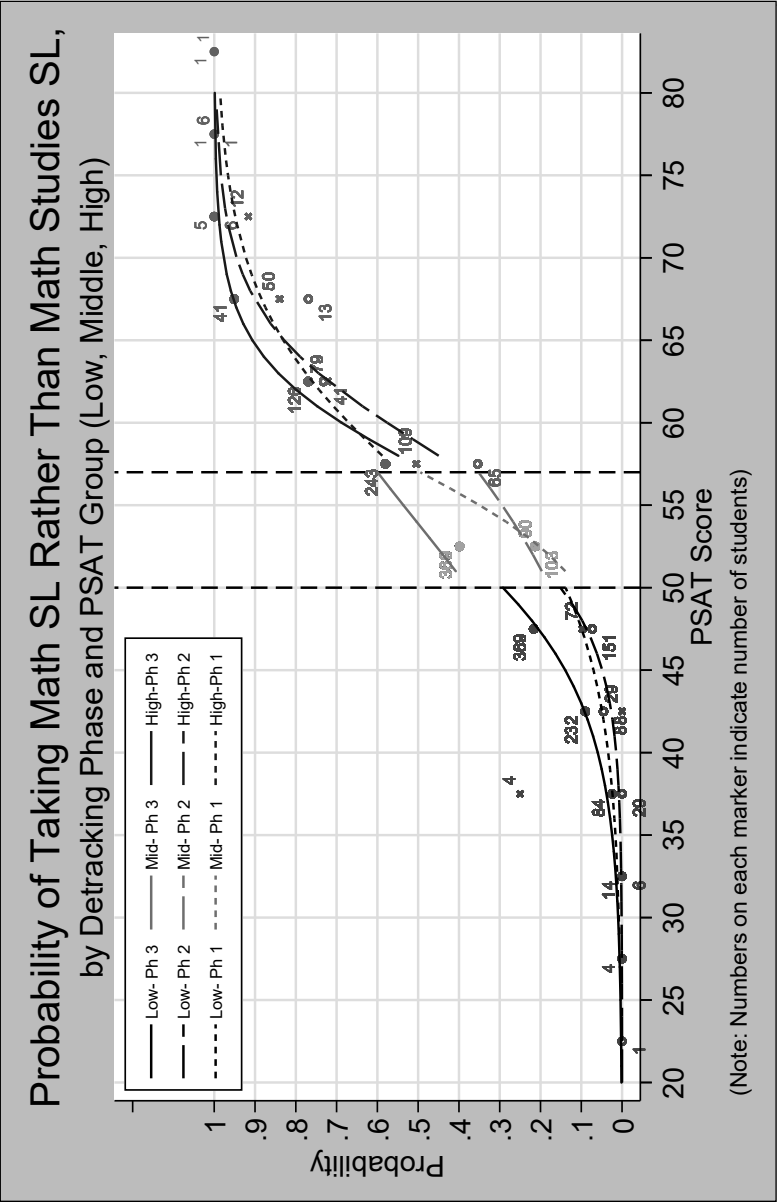
Figure 7. Probability of taking Math SL rather than Math Studies SL as function of math PSAT scores and phase, separately by prior achievement groups

We examine PSAT scores of participants, participation rates, and IB scores across three distinct phases of detracking in this school district. The modified interrupted time series approach we use is relatively straightforward, with the main counterfactual being the same district prior to the onset of detracking. The change in IB participation rates corresponds to the phases of detracking that occurred in this district. Nationwide IB trends did not change in ways that mirror the changes observed in SSHS, making it unlikely that our results merely reflect secular trends in IB participation and performance. From an internal validity perspective, it seems likely that the changes in participation rates we estimate are caused by the detracking initiative. That said, we do not have data on how IB participation or scores in other districts were or were not correlated with tracking or detracking policies—including any changes in nearby districts that continued to track during the same period.

From an external validity perspective, it is important to keep in mind that these results pertain to one relatively well-resourced school district in New York and may not generalize to other contexts. There may be characteristics of SSHS in terms of its student population, families, culture, teacher workforce, leadership, or resources that may have enabled a successful detracking reform. Nonetheless we think our findings offer an important *proof of concept*: The apparent success of this detracking initiative refutes claims by skeptics of detracking who argue that higher achieving students will be harmed by mixed-ability classes, or that lower achieving students cannot be successful with a more accelerated curriculum.

Finally, the current study focuses on a particular curricular experience—the IB Programme—which is less common in the U.S. than the Advanced Placement program. Both IB and AP courses are designed to correspond to college-level courses, are externally prepared and scored, and result in the awarding of college credit if the score is deemed to be passing or above by most universities. However, since the particulars of AP and IB curriculum and assessments substantially differ, findings in an IB program context may not generalize to findings in an AP program context.

SUMMARY OF FINDINGS

By any criterion, the results associated with the SSHS detracking reform are desirable. Instead of the world-class IB curriculum being reserved for a relatively small number of elite students, the gates were opened and 82% of the school's students ended up completing an IB English assessment, while 72% completed an IB math assessment. Recall that it was district policy in the final cohorts that *all* students take IB English courses in eleventh and twelfth grade, and though we report that 82% of the final cohort

possess IB scores, the district reports that an even greater percentage of students participated in the course—although a subgroup did not complete the assessment.[23] As more students took these challenging courses, they correspondingly challenged two widespread beliefs. First, the school's highest achievers continued to succeed in the more heterogeneous IB classes. Second, the average IB scores for the school's lower achievers were
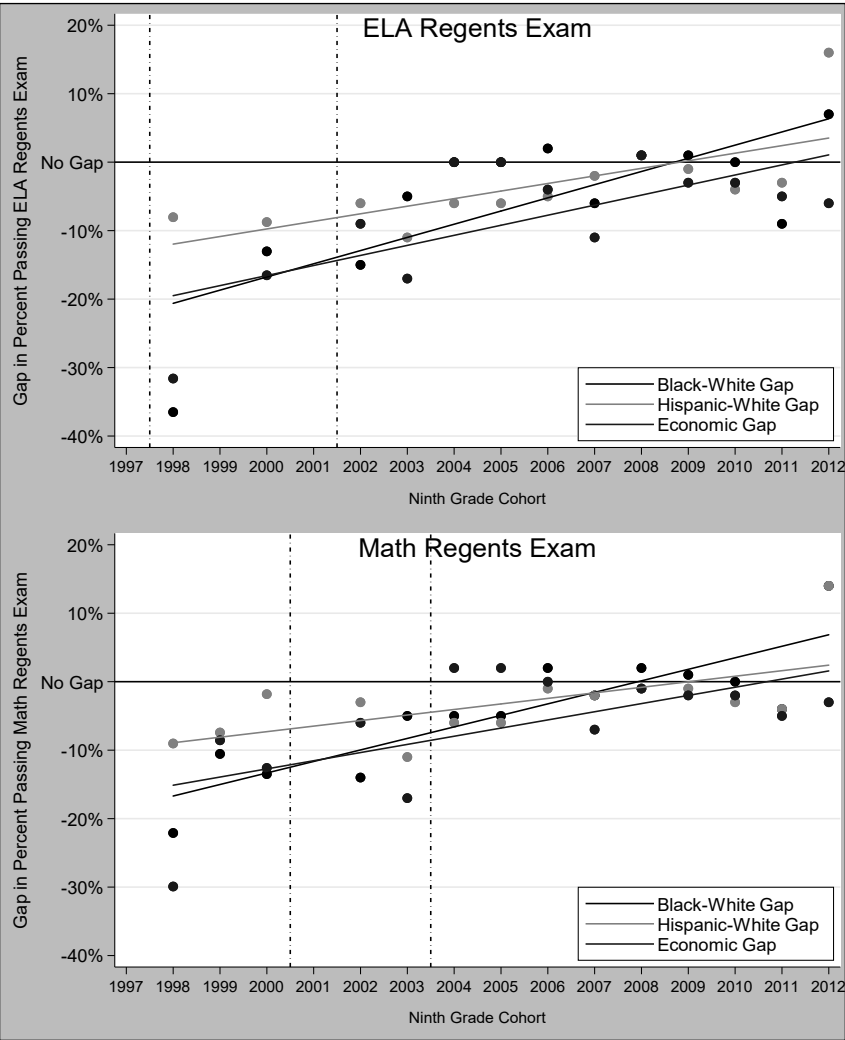

Figure 8. Gaps in Regents Exam passing rates in SSHS across cohorts, by subject

the same or higher after detracking began, even though many more such students enrolled in those courses. In short, this case study documents the potential for not rationing the enriched, world-class curriculum of the International Baccalaureate.

One may well ask whether the greater participation in IB courses ultimately narrowed racial and/or socioeconomic achievement gaps in SSHS. In particular, since we find that students with lower prior achievement are now exposed to the more rigorous IB curriculum and also that high-achieving students were not harmed by the broader inclusion, one might expect to see a narrowing of racial and socioeconomic outcome gaps on statewide assessments as a result of detracking. We cannot explore this directly with the current data because we lack student-level information on race, socioeconomic status, or other statewide assessment scores. However, we can look to aggregated school-level data to see if there is any evidence that achievement gaps closed during the detracking initiative. In Figure 8, we present SSHS's Regents Exam passing rate gaps along race and class dimensions, by cohort. It is clear that the Black-White gap, the Hispanic-White gap, and the gap among economically advantaged and disadvantaged students (as defined by the New York State Department of Education) have all closed from cohort 1998 to 2012 in SSHS. This is particularly compelling given that this district has remained relatively stable in terms of SES over this period and, if anything, was becoming more racially diverse. While we cannot definitively link individual students' exposure to IB courses to patterns in achievement, we think this finding is suggestive that detracking may have narrowed—and perhaps even closed—high school exam achievement gaps in this district.

DISCUSSION

Some past studies of tracking have found that the learning of high achievers can decrease in heterogeneous classes, but these studies fail to control for the effects of varying curricula as well as other factors associated with high-track classes. Kerckhoff (1986), reflecting on the results of his own study, states: "While the evidence presented here does strongly support the divergence hypothesis that tracking differentially effects [sic] performances of high and low ability groups, it does not provide an explanation of that effect" (p. 856). He goes on to suggest that the high-track advantage he saw may be the result of differentiated curriculum, better teachers in high-track classes, or classroom culture.

What sets this case study of Rockville Centre apart is the continuity of the IB curriculum, along with its external assessments of student learning. As the number of students taking IB courses in the school of study

increased, as students enrolled in the more challenging courses and as the program became more inclusive, the assessments of the IB program provided an objective, external, summative assessment of student learning that prevented the "watering down" of standards and instruction. The assessments of the IB gave teachers an external standard by which to monitor their own instruction as well as student learning. As described in Burris and Welner (2007) and Burris et al. (2008), traditional formal assessments, which are graded by trained IB examiners, when combined with internal assessments graded by the classroom teacher, were powerful tools available to the school as it maintained those standards even as the program expanded. As a result, standards were in fact maintained from year to year through the feedback teachers received from the externally assessed and moderated components, while differentiated instruction was implemented through personalized, internally assessed work.

Governors from across the nation met a dozen years ago at the National Education Summit on High Schools. Their concluding report, entitled "An Action Agenda for Improving America's High Schools," states the following:

> American high schools typically track some students into a rigorous college-preparatory program, others into vocational programs with less-rigorous curriculum and still others into a general track. Today, all students need to learn the rigorous content usually reserved for college-bound students, particularly in math and English. (National Education Summit on High Schools, 2005, p. 11)

Tracking, or as it is sometimes called, ability grouping, prevents some students from gaining access to the very curriculum that the governors recommend. What we as a research community are now learning is that well-executed detracking reforms can have remarkably positive results. Our study of a diverse suburban high school that has combined detracking with rigorous curricula for all students demonstrates that it is possible to meet the goal established by the National Educational Summit on High Schools. In this school, students of all levels of achievement and all socioeconomic backgrounds study an enriched, pre-IB curriculum together. Despite program growth that resulted from greater numbers of previously low-achieving students in IB classes, student performance on the assessments remained strong and competitive in the English and mathematics courses studied. When all students were provided by the school with IB access and preparation, the vast majority were found to be ready and willing to embark on college-level studies before high school graduation, even choosing the more difficult choice from among these rigorous options.

Is this research relevant to high schools that do not have an IB program? We believe that it is. The key instructional principles that are embedded in the IB program include the development of the following skills: critical thinking, oral and written communication, research, and problem-solving. These are the very skills that are needed for later college success (Conley, 2005). The underlying strategy was heterogeneous grouping combined with accelerated curriculum, support for struggling learners, and high expectations for all students. This strategy successfully prepared students in their middle- and high-school years for the rigors of IB classes in the eleventh and twelfth grades (see also Burris & Welner, 2005; Burris & Welner, 2007; Burris et al. (2008). There is every reason to believe that this strategy would prepare more students in other schools for Advanced Placement or other classes intended to prepare students for success in college.

To which schools, then, would the results of this case study be most generalizable? Prior research on this topic indicates that the key components of an extensive detracking reform are stability, values, and commitment (see Burris, Welner, & Bezoza, 2009). Stability of district leadership is needed because the careful dismantling of a tracking system takes time. Those who take on such a reform will likely meet with stiff resistance, especially from the parents of high-track students (Oakes, Quartz, Ryan, & Lipton, 2000; Oakes et al., 1997). Detracking is far more than a technical reform. Some obstacles are political and normative, while others are structural and curricular (Welner, 2001). Teacher support and training, student support, and community outreach are critical. As described in the beginning of this article, the forces that sustain elite programs in public high schools are strong and embedded in school culture.

Accordingly, a successful reform such as the one in Rockville Centre requires school leaders to take on longstanding practices and assumptions. These leaders should be guided by the belief, supported by this research, that all students deserve to study the best curriculum that a school has to offer, and that students will meet with success if they have the proper curriculum, teaching, and support. The results of this study will thus be most generalizable to districts that can muster the sort of sustained and complete effort that generated the reform we report on here. They must share a commitment to the belief that equity and excellence are not mutually exclusive. School leaders must have the stamina and tolerance needed to work through opposition. Finally, schools must be ready to back up that commitment with support classes and resources for struggling students, as well as supports for teachers (Burris, 2014).

Current policies that combine high standards, testing, and sanctions attempt to pressure schools to transform and improve, thereby creating high schools that prepare all students for college-level work. Faced with

this pressure, however, schools often turn to policies like grade retention and tracking that research has shown to be harmful (Oakes, 1985, Watanabe, 2008). This study demonstrates that high schools have another option. They can become places that provide all students access to programs based on world-class standards if they are willing to abandon sorting practices and instead provide all students with the opportunity and support needed to excel.

NOTES

1. The combined discussion here of IB and AP programs focuses on something they share in common: challenging, college-preparatory curriculum and instruction. However, they are fundamentally different programs, with different strengths and weaknesses and different admirers and critics. The empirical study discussed here concerns only an IB program, and readers should recognize that the outcomes may have been different had the school's advanced courses used the AP program.

2. The studies discussed here compare heterogeneous classes to tracked classes. A separate set of research compares higher and lower track placement within tracked systems (see, e.g., Card & Giuliano, 2016).

3. The IBO, which was established in 1968, is a nonprofit foundation that serves the needs of 1,468 member schools who offer one or more of its three courses of study known as the primary years, middle years, and diploma program (IBO, 2005).

4. The IB Diploma Program, which is offered in the final two years of secondary school, is a rigorous course of study that encompasses six areas of curriculum: (1) language A1 (the student's first language), (2) second languages, (3) individuals and society, (4) experimental sciences, (5) mathematics and computer science, and (6) the arts.

5. These descriptions of the school district are based on Burris & Weiner (2005), Burris et al. (2006), Burris & Weiner (2007), Burris et al. (2008), and Burris (2014).

6. Schools with IB need not be detracked. However, the structure and philosophy underlying the IB Diploma Program was an ideal fit for this school district's detracking reform initiatives. Both the school district and the International Baccalaureate Program profess a belief that "student capability … is not a static, invariant quality, such as a student's height would be, but is something more dynamic and variable in nature" (IBO, 2004). The educational practices of both institutions therefore presume that given the opportunity to study enriched, challenging curriculum that develops higher order thinking skills, student capacity to learn and think can grow and expand.

7. In addition to providing instructional supports, there has been a strong focus on ensuring that the curriculum and instruction put in place were demanding, engaging and linked to the later IB classes. In tenth-grade English classes, teachers integrated the use of the "IB Commentary" (a detailed, coherent literary

interpretation of a brief passage or poem), and in the social studies classes, they integrated the beginnings of the "IB Historical Investigation" (an internally assessed component of the course whereby students create an annotated bibliography based on a research question of their own making).

Teachers also developed process and product rubrics to assess individual student growth. Writing portfolios, combined with individual conferences, became part of practice in both English and social studies. Tenth-grade English support classes were transformed from a remedial model into classes focused on the accelerated content learned by all students, using pre-teaching, focused teaching, alternative teaching, and targeted review. Any student who wanted to take this every-other-day support class was encouraged to do so, and approximately 15% of all students requested it.

8. When the IB program was established at SSHS, the math department felt that the IB Math HL courses did not contain sufficient calculus, which they viewed as problematic given that many American universities favor calculus in the admissions process and prefer students to have a full year of calculus in high school.

9. Each cohort is defined by the fall of their ninth-grade year (e.g., Cohort 1994 entered high school in the 1994–95 academic year).

10. Note that English courses were detracked earlier than math courses, and therefore the timing of each phase depends on subject: For English, Phase I includes Cohort 1994–Cohort 1997, Phase II includes Cohort 1998–2001, and Phase III includes Cohort 2002–2011. For math, Phase I lasted longer (Cohort 1994–2000). Phase II includes Cohorts 2001–2003, and Phase III includes Cohorts 2004–2011.

11. We do not have data for Cohort 2000 due to challenges with data collection.

12. Imagine, for instance, that there is a hypothetical outcome variable coded as 0 or 1 for students who do not and do enroll in an IB course. We only have data for students who have a "1" on that variable. These leads to an imbalance in the size of the data panel as greater numbers of students enrolled in IB courses over time. For instance, in the first cohort (1994) the dataset only contains 26 records, while in the last cohort, the dataset contains 296 records.

13. Research has established a strong relationship between the PSAT and student scores on the Advanced Placement AB Calculus exam and the Advanced Placement English Literature exam (Camara & Millsap, 1998).

14. This stands in sharp contrast to exam-taking rates among students in AP courses: Nationally, slightly over half of AP enrolled students sit for the exam (Jeong, 2009).

15. There is some available evidence about the quality of IB exam scores as a valid measure of student achievement. All assessments are created with the specific instructional objectives of the course in mind in order to maintain the construct validity of the course work. The same test is administered across schools internationally in either English, French, or Spanish. Examinations are sent to

examiners around the globe for scoring. Each assessment passes through no fewer than three discrete scoring processes. An assistant examiner is the first scorer of final examination papers. A team leader then moderates his or her work to ensure that standards of the first examiner are appropriate. These results are then sent to a chief examiner for approval. None of the above examiners works in the students' schools; they are employees of the International Baccalaureate Organization (IBO).

16. We chose cut scores that corresponded, roughly, with the 50th and 75th percentiles of national test-takers. Thus, even if students in the district were improving, on average, over time, student groupings are based on the more stable national score distribution. In other analyses, we do not rely on these cut scores to ensure that our results are not driven by these choices.

17. When the IB scores are the outcome of interest in our analyses, it would be appropriate to use, for instance, an ordered probit model to analyze them instead of ordinary least squares (OLS), which formally is used for continuous outcome models. In certain cases, the use of OLS regression modeling on an ordered categorical outcome is likely to yield quite similar results: When there are more than five categories and the scores are somewhat normally distributed—both of which are generally true in the case of IB scores—results will tend to be quite similar. We certainly find this to be the case in the current analysis in which choice of modeling approach does not change the overall takeaways of the paper, which suggests that those conditions are met. In comparing the results of the two methods, we note that for more than 95% of coefficients estimated, there was no change in whether or not the coefficient was statistically significant. For the <5% of cases in which there was a difference, the results were significant in the ordered probit, but not in the OLS. For a discussion of the use of OLS for ordered categorical outcomes see Angrist & Pischke (2009), chapter 3. In Appendix D we present all analyses in the paper with IB scores as outcomes rerun using an ordered probit model. Given that the ordered probit and OLS models produce consistent findings, we opt to present the more readily interpretable OLS model results in the narrative.

18. Thus, for example, where the lower group included those with PSAT scores from 20–50, we centered the PSAT score on 35.)

19. The quantitative dataset does not include information about IB diploma candidates; however, district personnel report that the number of diploma candidates also grew over time, which enriches the evidence of increased participation. The number of students pursuing the full IB diploma increased as well. Only 27% of students who entered the ninth grade at South Side High School in 1997 (the year prior to universal acceleration in mathematics) were IB diploma candidates. In contrast, 44% of students who began South Side in 2003 graduated as IB diploma candidates. In each subsequent year, that proportion has been exceeded or met. This includes an increased participation by the district's majority students as well as its minority students. While student-level demographic information was not available for our statistical analysis, reports from school personnel indicated that

13% of the school's African American or Latino students who began ninth grade in 2000 were IB diploma candidates—a figure that improved to 38% among the minority students who entered just two years later, in 2002.

20. There does appear to be a modest decrease over time in the number of IB takers in each cohort who have very high scores (i.e., above 60, which is about the 93rd percentile of PSAT performance). However, the decrease in the number of IB takers with PSAT scores above 60 is relatively small from Phase I to Phase II to Phase III. For instance, among students taking the IB Math Studies test, on average 6.6 students per cohort scored above a 60 in Phase I, 5.0 students per cohort in Phase II, and 3.2 students per cohort in Phase III. These differences across detracking phases at the top end of the PSAT score distribution are not large in absolute numbers, but it accounts for the slightly smaller standard deviations of PSAT scores among IB test-takers in Phase III, as reported in Figure 3.

21. Some research suggests that the SAT, used primarily for college admission purposes, is indicative of academic aptitude commonly referred to as "g" (Frey & Detterman, 2004), and is therefore a proxy measurement of "ability" or "aptitude" as opposed to "achievement." Our assumption here is that the PSAT scores are measuring both aptitude and achievement, and are therefore less likely to be impacted by curricular changes than would be a test designed to measure achievement.

22. The ELA tracking experience for students in Phase I, where the very high PSAT scores showed up, was also not much different from that of the later phases. During Phase I, middle school ELA classes in the district had already been detracked. The students then took tracked ELA classes in their Grade 9 year. During Phases II and III, the ninth grade ELA classes were also detracked. The PSAT is taken in October of the students' tenth grade year.

23. As explained in an earlier section, there are other reasons why participation rates in the final cohorts do not reach 100% even though the district's policy aimed for 100% participation. Our denominators for twelfth-grade enrollment are provided by the Common Core of Data, and these data are collected at the start of each school year. Enrollment likely fluctuates down slightly during the twelfth-grade school year. We also suspect that some members of each cohort are not eligible for IB course-taking for other reasons, including being in a specialized program and being educated offsite.

REFERENCES

Alvarez, D., & Meehan, H. (2006). Whole school detracking: A strategy for equity and excellence. *Theory into Practice, 45*(1), 82–89.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Attewell, P. (2001). The winner-take-all high school: Organizational adaptations to educational stratification. *Sociology of Education, 74,* 267–295.

Boaler, J. (2006). How a detracked mathematics approach promoted respect, responsibility and high achievement. *Theory into Practice, 45*(1), 45–46.

Brewer, D. J., Rees, D. I., & Argys, L. M. (1995). Detracking America's schools: The reform without cost? *Phi Delta Kappan, 77*(3), 210–214.

Card, D., & Giuliano, L. (2016, March). *Can tracking raise the test scores of high-ability minority students?* (NBER Working Paper No. 22104). Cambridge, MA: National Bureau of Economic Research.

Clemmitt, M. (2006). AP and IB programs: Can they raise high school achievement? *Congressional Quarterly Researcher, 16*(9), 193–216.

Conley, D. T. (2005). *College knowledge: What it really takes for students to succeed and what we can do to get them ready.* San Francisco, CA: Jossey-Bass.

Duevel, L. M. (2000). *The International Baccalaureate experience: University perseverance, attainment, and perspectives on the process* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (ProQuest Document ID 304548024)

Figlio, D., & Page, M. (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics, 51*(3), 497–514.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or *g*? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15*(6), 373–378.

Garrity, D. (2004). Detracking with vigilance. *School Administrator, 61*(7), 24–27.

Hoffman, N. (2003). College credit in high school: Increasing college attainment rates for underrepresented students. *Change, 35*(4), 42–48.

International Baccalaureate Organization. (2004). *Diploma Programme assessments: Principles and practice.* Retrieved from http://web3.ibo.org/ibis/documents/dp/d_x_dpyyy_ass_0409_1_e.pdf

International Baccalaureate Organization. (2005). *Education for life.* Retrieved from http://www.ibo.org/ibo/index.cfm

Jeong, D. W. (2009). Student participation and performance on Advanced Placement exams: Do state-sponsored incentives make a difference? *Educational Evaluation and Policy Analysis, 31*(4), 346–366.

Klopfenstein, K. (2003). Recommendations for maintaining the quality of Advanced Placement programs. *American Secondary Education, 32*(1), 39–48.

Kyburg, R. M., Hertberg-Davis, H., & Callahan, C. M. (2007). Advanced Placement and International Baccalaureate programs: Optimal learning environments for talented minorities? *Journal of Advanced Academics, 18,* 172–215.

Loveless, T. (1998, August). *The tracking and ability grouping debate.* Washington, DC: Thomas B. Fordham Foundation. Retrieved from https://fordhaminstitute.org/national/research/tracking-and-ability-grouping-debate

Mathews, J. (1998). *Class struggle: What's wrong (and right) in America's best public high schools.* New York, NY: Random House.

Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Education Review, 66*(4), 797–842.

National Research Council. (2002). *Learning and understanding: Improving advanced study of mathematics and science in U. S. high schools* (J. Gollub, M. Bertenthal, J. Labov, & P. Curtis, Eds.). Washington, DC: National Academies Press.

Oakes, J., Quartz, K. H., Ryan, S., & Lipton, M. (2000). *Becoming good American schools: The struggle for civic virtue in school reform.* San Francisco, CA: Jossey-Bass.

Oakes, J., Rogers, J., McDonough, P. Solórzano, D., Mehan, H., & Noguera, P. (2000). *Remedying un-equal opportunities for successful participation in Advanced Placement courses in California high schools: A proposed action plan.* An expert report submitted on behalf of the plaintiffs in the case of *Daniel v. the State of California.*

Oakes, J., Wells, A. S., Jones, M., & Datnow, S. (1997). Tracking: The social construction of ability, cultural politics and resistance to reform. *Teachers College Record, 98,* 482–510.

Obama, B. H. (2011, May). *Remarks by the President at Booker T. Washington High School Commencement in Memphis, Tennessee.* Retrieved from http://www.whitehouse.gov/the-press-office/2011/05/16/remarks-president-booker-t-washington-high-school-commencement

Rui, N. (2009). Four decades of research on the effects of detracking reform: Where do we stand?—A systematic review of the evidence. *Journal of Evidence Based Medicine, 2*(3), 164–183.

Sadler, P. M., Sonnert, G., Tai, R. H., & Klopfenstein, K. (2010). *AP: A critical examination of the Advanced Placement program.* Cambridge, MA: Harvard Education Press.

Santoli, S. P. (2002). Is there an Advanced Placement advantage? *American Secondary Education, 30*(3), 22–35.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher, 44*(7), 371–386.

Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record, 112*(5), 1492–1528.

Slavin, R. E., & Braddock, J. H., III. (1993). Ability grouping: On the wrong track. *College Board Review, 168,* 11–17.

Solórzano, D. G., & Ornealas, A. (2004). A critical race analysis of Latina/o and African American Advanced Placement enrollment in public high schools. *High School Journal, 87*(3), 15–26.

Tyson, K. (2013). Tracking, segregation, and the opportunity gap. In P. Carter and K. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 169–180). Oxford, UK: Oxford University Press.

Watanabe, M. (2008). Tracking in the era of high-stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record, 110*(3), 489–534.

APPENDIX A

Demographics in All Years

Appendix A provides summary demographic information for the school for a subset of four of the eighteen study cohorts. We first examine whether the school has been experiencing notable demographic trends across the study cohorts. As shown by the table, the demographic composition of the district's high school has been relatively stable. For example, ninth graders in Cohorts 1994, 1999, 2004, and 2009 are between 73.9 to 79.6% White, a relatively small change in composition. Likewise, percentage Latino/a, Asian, "Other/Unknown," and female are consistent across cohorts, exhibiting only fluctuations within only 1 to 3 percentage points. The only exception to this stability is the modest increase in percentage of students who are Black, which has risen between 1 and 6 percentage points across cohorts, depending on the grade. In addition, we can see that within any given cohort, there is little fluctuation in demographics from grade to grade as the cohort moves through school. While CCD does not provide free and reduced-price lunch (FRPL) eligibility at the grade-year level, we can examine whether this socioeconomic indicator exhibits a clear trend at the school level during the study panel. In all years between 1998 and 2015, the percentage of students who are FRPL eligible is between 10 and 15%, with the exception of 2008 when it drops down for one year to 8% (U.S. Department of Education, National Center for Education Statistics, Common Core of Data [CCD] "Public Elementary/Secondary School Universe Survey"). While there is some year-to-year fluctuation, there is no evidence of an upward or downward trend in poverty measures; indeed, the FRPL-eligible percentage is about the same in both 1998 and 2015.

Table A1. Racial/Ethnic and Gender Characteristics across Grades, by Study Cohort

	White				Black			
Cohort	**9th**	**10th**	**11th**	**12th**	**9th**	**10th**	**11th**	**12th**
1994	79.6%	80.5%	78.7%	83.1%	7.0%	7.4%	6.2%	3.2%
1995	78.8%	80.0%	78.5%	82.4%	6.8%	7.2%	6.4%	3.6%
1996	79.2%	79.8%	78.3%	82.7%	7.2%	7.4%	6.5%	4.5%
1997	78.7%	79.5%	78.1%	76.6%	7.1%	7.5%	8.6%	4.0%
1998	79.6%	78.2%	75.7%	80.1%	7.7%	10.7%	2.9%	8.5%
1999	76.8%	78.9%	75.8%	75.2%	6.6%	6.7%	9.0%	8.2%

	White				Black			
Cohort	9th	10th	11th	12th	9th	10th	11th	12th
2000	80.8%	82.5%	82.0%	82.5%	6.4%	5.2%	5.5%	6.0%
2001	79.5%	81.1%	81.5%	81.5%	6.7%	6.7%	5.9%	5.4%
2002	77.0%	74.4%	71.9%	70.9%	7.4%	8.4%	9.4%	9.7%
2003	77.5%	77.4%	78.4%	77.0%	10.9%	10.2%	9.0%	9.7%
2004	77.3%	77.3%	78.7%	75.3%	8.7%	7.9%	7.3%	7.3%
2005	76.1%	75.6%	74.1%	73.0%	7.1%	8.0%	8.6%	9.5%
2006	78.7%	77.3%	77.2%	75.5%	7.3%	7.3%	8.0%	8.3%
2007	78.9%	78.7%	79.0%	77.0%	8.3%	9.9%	9.0%	8.9%
2008	74.6%	73.6%	73.5%	72.0%	8.1%	9.4%	9.0%	9.7%
2009	73.9%	71.4%	71.1%	70.3%	9.5%	10.3%	8.7%	9.1%
2010	77.7%	77.9%	77.2%	76.2%	8.1%	7.8%	8.5%	7.9%
2011	78.1%	78.1%	77.4%	71.4%	6.5%	6.9%	6.7%	10.7%

	Latino/a				Asian			
Cohort	9th	10th	11th	12th	9th	10th	11th	12th
1994	11.6%	10.5%	10.2%	9.3%	2.1%	1.9%	5.2%	4.4%
1995	12.7%	11.3%	10.3%	9.6%	2.0%	1.6%	5.0%	4.3%
1996	11.6%	11.0%	10.5%	9.9%	2.3%	1.9%	4.9%	2.9%
1997	12.2%	10.9%	9.8%	12.3%	2.2%	2.1%	3.5%	6.7%
1998	10.2%	8.0%	10.1%	8.5%	2.7%	3.1%	11.2%	3.0%
1999	13.9%	11.5%	11.9%	13.5%	2.7%	3.0%	3.2%	3.2%
2000	10.2%	9.7%	9.9%	9.0%	2.6%	2.6%	2.6%	2.6%
2001	11.8%	10.2%	9.1%	9.7%	2.0%	2.0%	3.5%	3.5%
2002	12.4%	14.0%	14.4%	15.4%	3.2%	3.2%	4.3%	4.0%
2003	8.7%	8.9%	9.4%	10.4%	2.9%	3.2%	2.9%	2.5%
2004	11.2%	11.3%	11.0%	14.3%	2.8%	3.4%	3.0%	3.0%
2005	14.5%	13.7%	13.9%	13.9%	2.4%	2.7%	3.4%	3.3%
2006	10.5%	11.2%	10.7%	12.1%	3.1%	4.2%	3.8%	3.8%
2007	10.2%	9.1%	9.4%	10.8%	2.6%	2.3%	2.6%	2.6%
2008	14.2%	14.0%	14.2%	15.3%	2.7%	2.6%	3.0%	2.6%
2009	12.6%	13.0%	15.6%	16.3%	3.6%	3.4%	3.4%	2.7%
2010	10.2%	10.0%	10.3%	10.8%	2.8%	3.9%	3.6%	3.2%
2011	11.6%	10.8%	11.1%	14.3%	3.9%	3.9%	4.0%	2.5%

	Other/ Unknown Race				Female			
Cohort	**9th**	**10th**	**11th**	**12th**	**9th**	**10th**	**11th**	**12th**
1994	0.0%	0.0%	0.0%	0.0%	51.6%	51.5%	50.6%	49.7%
1995	0.0%	0.0%	0.0%	0.0%	51.5%	51.3%	50.5%	49.6%
1996	0.0%	0.0%	0.0%	0.0%	51.2%	51.1%	50.4%	46.9%
1997	0.0%	0.0%	0.0%	0.4%	51.0%	50.9%	49.2%	48.8%
1998	0.0%	0.0%	0.0%	0.0%	50.6%	51.6%	52.5%	52.4%
1999	0.0%	0.0%	0.0%	0.0%	49.0%	47.8%	48.4%	48.2%
2000	0.0%	0.0%	0.0%	0.0%	52.3%	53.5%	52.6%	53.7%
2001	0.0%	0.0%	0.0%	0.0%	50.8%	50.8%	48.8%	49.4%
2002	0.0%	0.0%	0.0%	0.0%	52.1%	51.6%	50.8%	49.8%
2003	0.0%	0.3%	0.3%	0.3%	46.3%	47.5%	48.1%	47.8%
2004	0.0%	0.0%	0.0%	0.0%	51.4%	49.1%	50.5%	49.7%
2005	0.0%	0.0%	0.0%	0.4%	51.0%	51.5%	53.4%	52.2%
2006	0.3%	0.0%	0.3%	0.3%	47.0%	46.9%	47.4%	46.2%
2007	0.0%	0.0%	0.0%	0.7%	45.7%	45.6%	44.9%	45.0%
2008	0.4%	0.4%	0.4%	0.4%	50.8%	49.4%	48.9%	49.6%
2009	0.4%	1.9%	1.1%	1.5%	45.1%	45.4%	46.0%	45.6%
2010	1.1%	0.4%	0.4%	1.8%	52.3%	52.7%	51.2%	51.3%
2011	0.0%	0.3%	0.7%	1.0%	48.4%	49.3%	49.2%	48.4%

Note: Grades of high school are shown across the columns, while cohorts are shown down the rows (e.g., 1994 represents the cohort of students who began high school in fall of 1994). Cells contain the percentage students of each demographic group in each grade-cohort. "The other/unknown" race group includes children who indicated they were of two or more races, Hawaiian Native, Pacific Islander, American Indian, Alaska Native, or unknown race. The "two or more races" category has only been collected since 1998. Where data is missing, we use a linear imputation strategy. However, it is clear from the complete, most recent years that very few students in this school fall into this category.

APPENDIX B

Detracking Efforts Before High School

EARLY EFFORTS TO DETRACK

In order to prepare more students for the IB program and to increase the achievement of all students, the district gradually decreased its tracks, moving from three or more tiers to a two-track system that allowed any ninth or tenth grade student to opt into honors and then, in eleventh and twelfth grades, into AP and IB courses (Garrity, 2004). The ninth- and tenth-grade honors courses were the prerequisite courses for later AP/IB study. According to district leaders, these initial efforts resulted in a small, corresponding increase in the proportion of students enrolling in IB courses as well as the proportion of students pursuing the full IB diploma.

In theory, any student was allowed to enroll in IB courses in Grades 11 and 12. Yet decisions to do so were effectively predetermined very early in students' schooling. Teacher and counselor recommendations, for instance, sometimes discouraged students from enrolling in more challenging courses. Likewise, student choices in the ninth and tenth grade to study honors-track English, social studies, math, and science courses set the stage for later IB enrollment. However, pathways toward IB course taking formed even prior to entering high school.

DETRACKING MIDDLE SCHOOL

Perhaps the most important factor determining future IB access was the decision made by teachers, students, and their families three years before high school—in Grade 6—about whether to study accelerated mathematics. Since the 1980s, the New York State Board of Regents has required all districts to accelerate some students in math, meaning that the district would offer these students an algebra-based course (then called Sequential I Mathematics) prior to the ninth grade. This acceleration mandate was intended to allow select students either to graduate early or to earn college math credit while in high school (Maggio, 1988). Thus, accelerated math refers to a program of math study with two significant characteristics: (1) it teaches the usual sixth-, seventh-, and eighth-grade curricula in two years rather than three; and (2) it teaches the usual ninth-grade curriculum, an algebra-based course called Sequential I Mathematics, in the eighth grade.

District leaders came to understand that unequal access to the accelerated mathematics program might lead to *de facto* tracking throughout the

rest of the schooling years. Therefore in the late 1990s, as an early reform element, the district opened the gates to this accelerated math program, giving students access regardless of whether they had met the former criteria for entrance. Encouraged by the success of those students who did choose to accelerate in the early years of the middle-school math reform, administrators subsequently concluded that it was in the best interest of all students to eliminate stratified grouping for instruction and to provide math acceleration for all. The first step of this scaled-up reform was the development of a multiyear plan to eliminate tracking in middle-school mathematics. All tracking in mathematics would end with the sixth-grade class entering in the fall of 1995, and all future sixth-graders would study accelerated math in heterogeneously grouped classes. In addition, the school changed teaching and learning conditions in ways that school leaders believed would help all students succeed. Specifically, the superintendent and the middle-school leadership team concluded that a combination of three elements would enable all learners to be successful without reducing the achievement of the most proficient math students: (a) heterogeneous grouping, (b) high-track curriculum, and (c) pre- and post-teaching in alternate-day math workshops. In these workshops, students who needed extra support would receive it through pre-teaching, focused teaching (the expansion of a difficult concept), alternative teaching (using a different teaching strategy), and targeted review. As a result, all students who entered sixth grade in or after the 1995–1996 school year participated in accelerated math in heterogeneously grouped classes. This reform was seen by district educators as yielding prompt and demonstrable successes, which set the stage for the detracking of high school coursework.

APPENDIX C

Additional Detail on IB, by Subject

IB ENGLISH

All students who take the IB English HL (Higher Level) two-part assessment do so after studying the IB English curriculum for two years. 50% of the final IB score is based on two formal examinations (known as papers) given during May of the senior year of high school. The other half of the awarded score is determined by performance on two World Literature papers (totaling 20% of the score) as well as two oral presentations (totaling 30%).

IB MATH SL AND MATH STUDIES

The IB Math SL and IB Math Studies courses are Standard Level courses, meaning that they are 1-year courses. At this particular high school, they are usually taken by students in the junior year. The final score in each course is overwhelmingly (80%) based on a student's scores on two examinations (also referred to as papers) administered at the end of that course. Although students may take both courses, only one course is counted toward earning the IB diploma. Therefore, students generally select only one of these two.

Of the two courses, Math SL is the more difficult and it is usually chosen by more confident students—those who intend to study mathematics or science in college. Students in this IB Math SL class are required to study six core topics and one option. The core topics include numbers and algebra, functions and equations, circular functions and trigonometry, vector geometry, statistics and probability, and calculus. Options include statistical methods, further calculus, and further geometry. The remaining 20% of the score (that portion not based on the two papers) arises from an assessment of a portfolio consisting of three pieces of work completed by the student during the course. This work must show mathematical investigation, extended closed problem-solving, and mathematical modeling.

Math Studies is designed for those students who do not intend to pursue a math-related course of study in college. Here, the final 20% of the score is derived from the teacher assessment (reviewed by IB auditors) of a project requiring students to collect, measure and evaluate mathematical data.

VISUALIZATIONS THAT CORRESPOND TO INTERRUPTED TIME

SERIES ANALYSES RESULTS PRESENTED IN TABLES 2, 3, 4, AND 8

Appendix Figures C1–C3 illustrate the findings in Tables 2–4 and represent the temporal relationship between cohort and IB score in Math Studies, Math SL, and English IB courses, respectively. We designate prior achievement (PSAT) groups with color. The lowest-performing group is indicated by the red lines; the middle, the orange lines; the highest, the green lines. We use line pattern to indicate phase. Phase I is indicated with a dotted line; Phase II, a dashed line; Phase III, a solid line. If detracking had the sort of negative consequences sometimes feared, then we should expect to see lower IB scores for high-PSAT students in Phases II and III than in Phase I. Likewise, Appendix Figure C4 illustrates the findings in Table 8, which estimates the probability of taking Math SL rather than Math Studies, separately by prior math achievement groups.
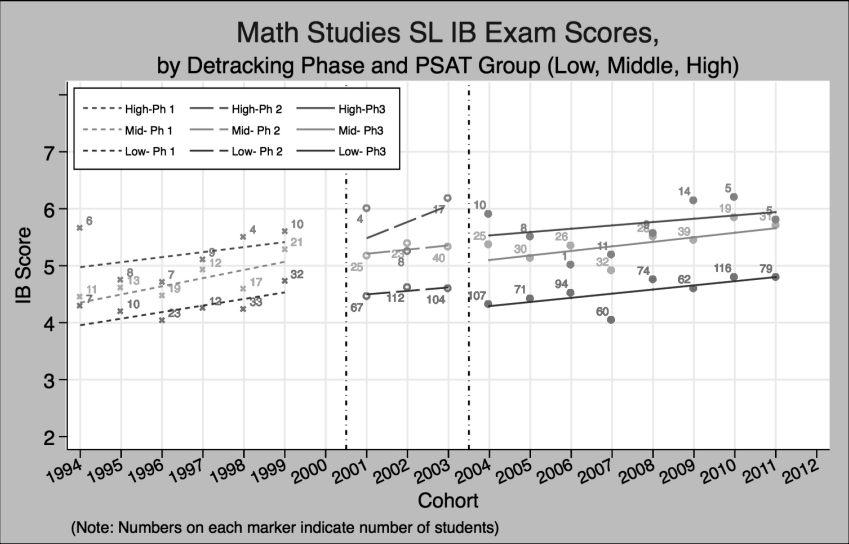

Figure C1. Math Studies: IB scores as a function of cohort and phase, separately by prior math achievement groups (interrupted time series analysis)
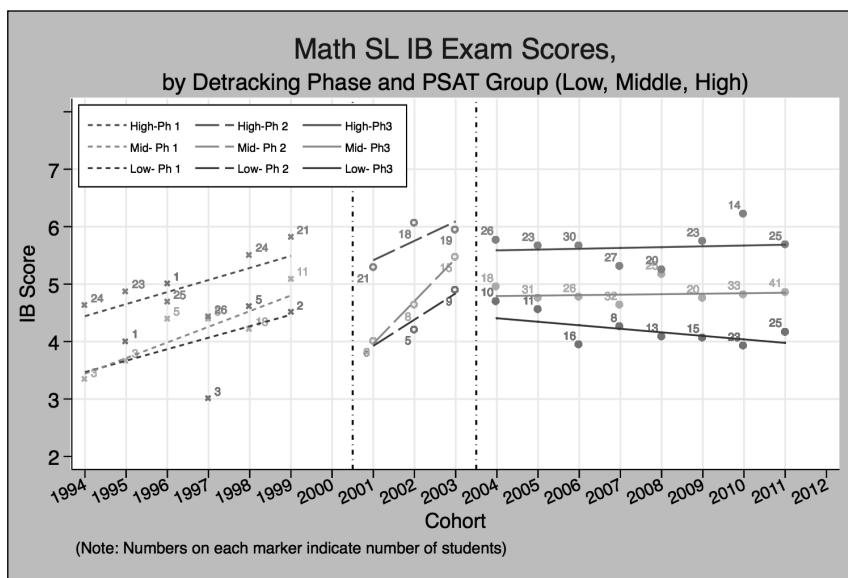
Figure C2. Math SL: IB scores as a function of cohort and phase, separately by prior math achievement groups (interrupted time series analysis)
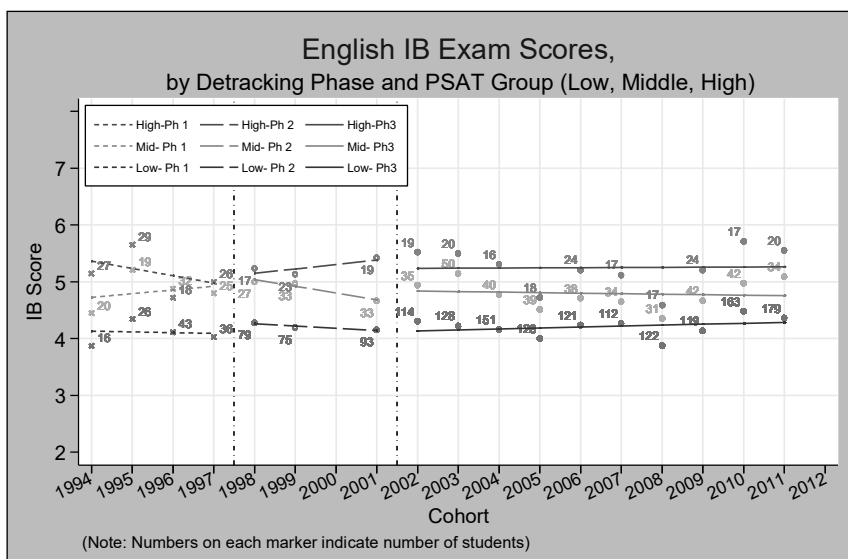

Figure C3. English: IB scores as a function of cohort and phase, separately by prior verbal achievement groups (interrupted time series analysis)
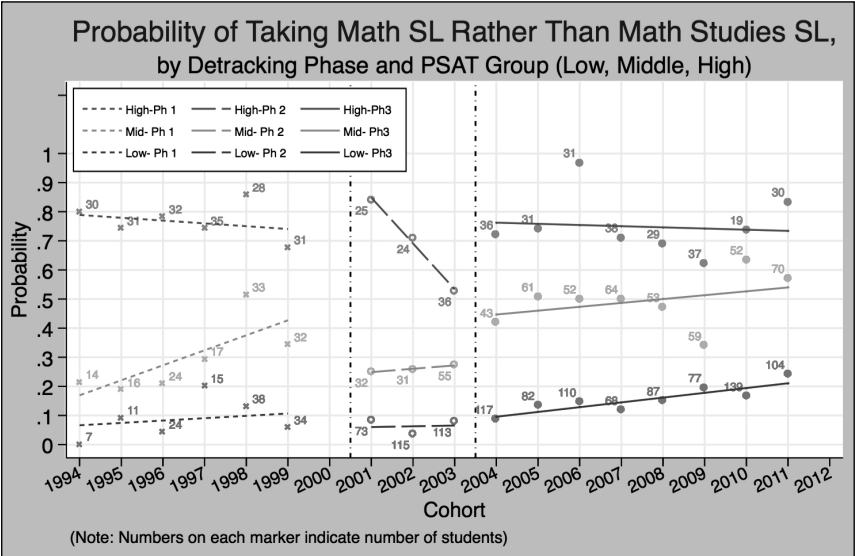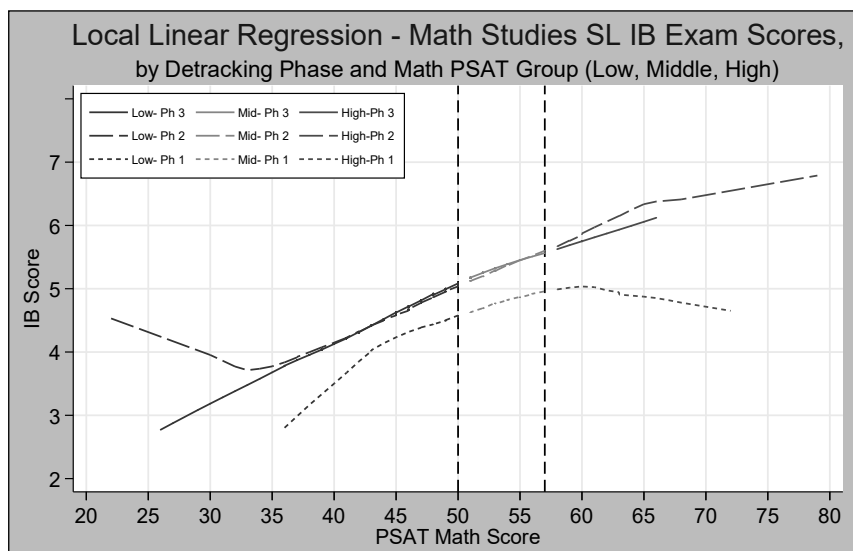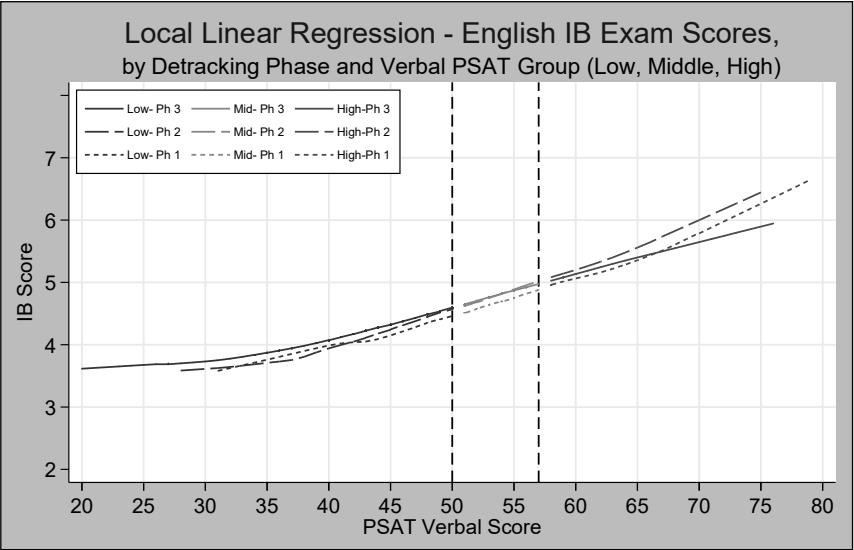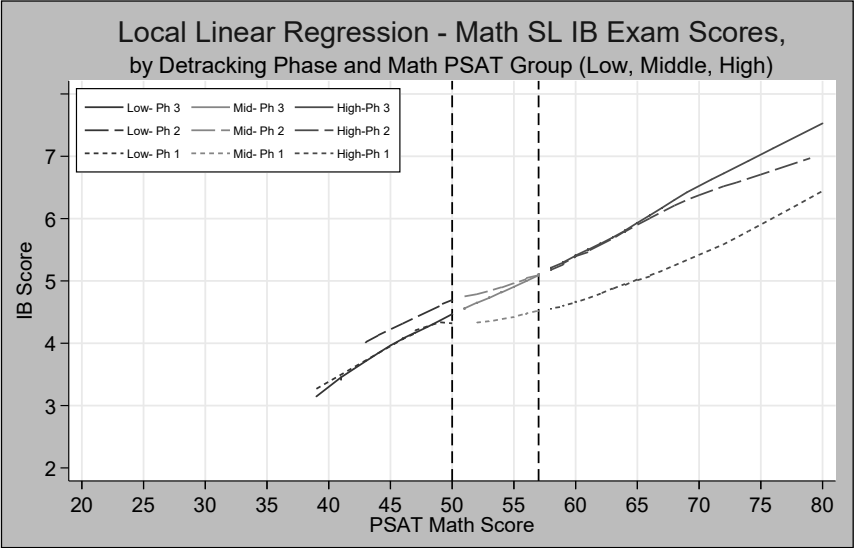
Figure C4. Probability of taking Math SL rather than Math Studies, separately by prior math achievement groups (interrupted time series analysis)

APPENDIX D

Figures 4, 5, and 6 NONPARAMETRICALLY

We address the possibility that our chosen cut scores for prior achievement groups may be driving the observed trends with the following local linear regression results. Although we do plot the results using the color and line patterns used throughout the article to indicate ability group and detracking phase, we performed the underlying local linear regression (nonparametric function) as a single analysis across all ability groups. That is, we did not force the function to break at our designated ability group boundaries, but rather permitted it to run uninterrupted across the range of PSAT scores. As illustrated in the plots, we observe very similar patterns to those from the analysis in the main article. For example, in the case of the Math Studies scores, we see that students in the lower range (left side) of the lowest group of PSAT scores appeared to do better in Phase II than in Phase III, that students in the highest group of PSAT scores appeared to do similarly well in Phase II and Phase III, and that across all PSAT scores, students tended to do better in Phases II and III than in Phase I. Because these results tell a similar, though nonlinear, story to those of our linear functions, we find some evidence that our choice of cut scores for ability groups did not drive our results.

Local Linear Regression - Math Studies SL IB Exam Scores, by Detracking Phase and Math PSAT Group (Low, Middle, High)

Local Linear Regression - Math SL IB Exam Scores, by Detracking Phase and Math PSAT Group (Low, Middle, High)

Local Linear Regression - English IB Exam Scores, by Detracking Phase and Verbal PSAT Group (Low, Middle, High)

ORDERED PROBIT ANALYSIS OF QUESTION 3

Table D1. Math Studies: IB Scores as a Function of PSAT Math Score and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Phase 2	1.129	0.687 ***	1.665 **	0.356	-0.441	0.958
	(0.733)	(0.162)	(0.615)	(0.804)	(0.441)	(0.962)
Phase 3	0.629	0.753 ***	1.964 **	-0.116	-0.584	1.545
	(0.699)	(0.134)	(0.739)	(0.768)	(0.449)	(1.071)
PSAT Score	0.114 ***	0.068	-0.034	0.129 ***	0.039	-0.094
	(0.031)	(0.056)	(0.051)	(0.032)	(0.058)	(0.060)
PSAT x Phase 2	-0.033	0.047	0.109	-0.038	0.089	0.221 **
	(0.034)	(0.079)	(0.071)	(0.035)	(0.082)	(0.080)
PSAT x Phase 3	-0.007	0.006	0.160	-0.023	0.048	0.361 ***
	(0.033)	(0.066)	(0.083)	(0.033)	(0.068)	(0.104)
N	1063	411	136	1063	411	136
Covariates?	No	No	No	Yes	Yes	Yes

Table D2. Math SL: IB Scores as a Function of PSAT Math Score and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Phase 2	2.766	0.558 *	1.100 ***	-4.059	-1.031	1.238 *
	(3.381)	(0.262)	(0.268)	(4.271)	(0.664)	(0.520)
Phase 3	2.461	0.532 **	1.231 ***	-0.105	-1.254	1.142 *
	(2.781)	(0.197)	(0.223)	(3.402)	(0.642)	(0.546)
PSAT Score	0.222	0.072	0.079 ***	0.133	0.103	0.107 ***
	(0.117)	(0.093)	(0.019)	(0.140)	(0.100)	(0.020)
PSAT x Phase 2	-0.106	-0.038	0.046	0.178	0.018	0.038
	(0.149)	(0.130)	(0.037)	(0.182)	(0.145)	(0.039)
PSAT x Phase 3	-0.104	0.022	0.062 *	-0.018	0.013	0.056
	(0.121)	(0.098)	(0.030)	(0.144)	(0.106)	(0.032)
N	153	303	389	153	303	388
Covariates?	No	No	No	Yes	Yes	Yes

Table D3. English: IB Scores as a Function of PSAT Verbal Score and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low V PSAT	Mid V PSAT	High V PSAT	Low V PSAT	Mid V PSAT	High V PSAT
Phase 2	-0.708	0.075	0.232	-0.351	-0.244	0.224
	(0.623)	(0.160)	(0.358)	(0.661)	(0.358)	(0.541)
Phase 3	-0.202	-0.043	-0.413	0.420	-0.646	0.074
	(0.534)	(0.126)	(0.248)	(0.612)	(0.524)	(0.756)
PSAT Score	0.049	0.071	0.127 ***	0.049	0.083	0.130 ***
	(0.026)	(0.056)	(0.024)	(0.026)	(0.056)	(0.025)
PSAT x Phase 2	0.045	0.073	0.020	0.046	0.058	0.018
	(0.031)	(0.081)	(0.044)	(0.031)	(0.081)	(0.045)
PSAT x Phase 3	0.025	-0.005	-0.063 *	0.028	0.001	-0.061
	(0.026)	(0.062)	(0.032)	(0.027)	(0.064)	(0.033)
N	1705	574	351	1704	573	351
Covariates?	No	No	No	Yes	Yes	Yes

Table D4. Math Studies: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Phase 2	-0.242	-0.096	-0.437	20.215	23.345	39.567
	(0.268)	(0.391)	(0.796)	(18.936)	(19.274)	(29.767)
Phase 3	-0.676 **	-0.555	-0.389	-12.117	-14.541	-23.184
	(0.237)	(0.341)	(0.600)	(10.768)	(10.906)	(16.819)
Cohort	0.120	0.155 *	0.085	-1.496	-1.462	-2.789
	(0.063)	(0.065)	(0.094)	(1.248)	(1.283)	(1.997)
Cohort x Phase 2	-0.046	-0.050	0.305	-8.515	-9.805	-16.556
	(0.102)	(0.149)	(0.296)	(7.618)	(7.805)	(12.106)
Cohort x Phase 3	-0.043	-0.053	0.006	2.798	3.141	5.296
	(0.065)	(0.072)	(0.113)	(2.549)	(2.585)	(3.982)
N	1063	411	136	1063	411	136
Covariates?	No	No	No	Yes	Yes	Yes

Table D5. Math SL: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low M PSAT	Mid M PSAT	High M PSAT	Low M PSAT	Mid M PSAT	High M PSAT
Phase 2	-1.182	-1.943 **	-0.564	-4.431	-14.838	34.284 *
	(0.973)	(0.610)	(0.410)	(61.269)	(36.583)	(15.980)
Phase 3	-0.059	-0.329	-0.088	0.534	5.453	-18.405 *
	(0.812)	(0.400)	(0.330)	(2.445)	(20.729)	(9.056)
Cohort	0.204	0.281 **	0.189 ***	0.742	1.128	-2.101
	(0.268)	(0.107)	(0.053)	(12.527)	(2.413)	(1.073)
Cohort x Phase 2	0.241	0.507 *	0.136	1.512	5.209	-13.642 *
	(0.383)	(0.246)	(0.175)	(23.341)	(14.733)	(6.520)
Cohort x Phase 3	-0.259	-0.271 *	-0.182 **	-0.782	-1.964	4.393 *
	(0.271)	(0.111)	(0.063)	(9.965)	(4.921)	(2.138)
N	153	303	389	153	303	388
Covariates?	No	No	No	Yes	Yes	Yes

Table D6. English: IB Scores as a Function of Cohort and Phase, Separately by Prior Achievement Groups (Interrupted Time Series Analysis)

	Models Without Covariates			Models With Covariates		
	Low V PSAT	Mid V PSAT	High V PSAT	Low V PSAT	Mid V PSAT	High V PSAT
Phase 2	0.242	0.332	-0.001	-68.194 **	-113.651 ***	-130.285 ***
	(0.216)	(0.304)	(0.354)	(24.125)	(26.796)	(24.588)
Phase 3	-0.136	-0.063	0.199	-28.041 **	-48.288 ***	-54.097 ***
	(0.186)	(0.259)	(0.323)	(10.307)	(11.442)	(10.496)
Cohort	0.019	0.100	-0.162	-5.292 **	-8.968 ***	-10.719 ***
	(0.099)	(0.103)	(0.093)	(1.925)	(2.162)	(2.006)
Cohort x Phase 2	-0.081	-0.270 *	0.263	35.161 **	58.438 ***	67.917 ***
	(0.113)	(0.137)	(0.148)	(12.510)	(13.916)	(12.793)
Cohort x Phase 3	0.006	-0.113	0.165	8.501 **	14.252 ***	16.822 ***
	(0.100)	(0.104)	(0.097)	(3.078)	(3.439)	(3.172)
N	1705	574	351	1704	573	351
Covariates?	No	No	No	Yes	Yes	Yes

ALLISON ATTEBERRY is an assistant professor of research and evaluation methods at the CU-Boulder School of Education. Dr. Atteberry conducts research on teacher- and school-level interventions designed to improve the quality of instruction experienced by historically underserved students.

SARAH E. LACOUR is a Ph.D. candidate in Educational Foundations, Policy, & Practice at the University of Colorado Boulder. Her research primarily uses quasiexperimental designs to explore the impacts of law and larger educational policy on students and communities.

CAROL BURRIS is presently the Executive Director of the Network for Public Education. She was the principal of South Side High School during the years of this study.

KEVIN WELNER is a professor at the University of Colorado at Boulder School of Education, specializing in educational policy and law.

JOHN MURPHY is the Principal at South Side High School in Rockville Centre, New York.