Accountability, Rigor, and Detracking: Achievement Effects of Embracing a Challenging Curriculum As a Universal Good for All Students

CAROL CORBETT BURRIS

South Side High School

ED WILEY
KEVIN WELNER

University of Colorado at Boulder

JOHN MURPHY

South Side High School

Background: *This longitudinal study examines the long-term effects on the achievement of students at a diverse suburban high school after all students were given accelerated mathematics in a detracked middle school as well as ninth-grade 'high-track' curriculum in all subjects in heterogeneously grouped classes. Despite considerable research indicating the ineffectiveness and inequities of ability grouping, the practice is still found in most American high schools. Research indicates that high-track classes bring students an academic benefit while low-track classes are associated with lower subsequent achievement. Corresponding research demonstrates that tracks stratify students by race and class, with African American, Latino and students from low-socioeconomic households being dramatically over-represented in low-track classes and under-represented in high-track classes.*

Purpose: *In light of increasing pressure to hold all students to high learning standards, educators and researchers are examining policy decisions, such as tracking, in order to determine their relationship to student achievement.*

Design: *This study used a quasi-experimental cohort design to compare pre- and post-reform success in the earning of the New York State Regents diploma and the diploma of the International Baccalaureate.*

Data Analysis: *Using binary logistic regression analysis, the authors found that there was a statistically significant post-reform increase in the probability of students earning these standards-based diplomas. Being a member of a detracked cohort was associated with an increase of roughly 70% in the odds of IB diploma attainment and a much greater increase in the odds of Regents diploma attainment – ranging from a three-fold increase for White or Asian students, to a five-fold increase for African American or Latino students who were eligible to receive free or reduced-price lunch, to a 26-fold increase for African American or Latino students not eligible for free or reduced-price lunch. Further, even as the enrollment in International Baccalaureate classes increased, average scores remained high.*

Conclusion: *The authors conclude that if a detracking reform includes high expectations for all students, sufficient resources and a commitment to the belief that students can achieve when they have access to enriched curriculum, it can be an effective strategy to help students reach high learning standards.*

INTRODUCTION

High schools are in dramatic need of reform, concluded the 2005 National Education Summit on High Schools (Conklin & Curran, 2005). The Summit participants, who were governors from across the nation, resolved that improvements must include higher academic standards and more challenging curricula—conclusions echoing those of *Standards for Success,* an extensive, multi-year study of the Association of American Universities (AAU) and the Pew Charitable Trusts (Conley, 2005). *Standards for Success* called for dramatic high school curricular upgrades in both skills and content to adequately prepare American students for university learning. Both studies reflect the overwhelming belief among educators, policy makers, and the public that all students should be engaged in highly challenging academic programs.

Students themselves have similarly acknowledged the need for more academic rigor. A high-profile survey of nearly 1,500 public high school graduates found that 76% were neither challenged academically nor adequately prepared by their high school for college or the workplace (Peter D. Hart Research Associates, 2005). Furthermore, those surveyed said that they would have worked harder if their school had demanded more of them.

These calls for more rigorous standards arrive in the wake of a prolonged wave of standards-based accountability reforms, culminating in the No Child Left Behind Act of 2001 (NCLB). Advocates of such policies believe that school accountability legislation acts as a catalyst for

changes in instruction, expectations and curriculum and will thus result in increased academic performance (Berger, 2000; see also Heubert & Hauser, 1999; Natriello & Pallas, 1999). Although specific policies such as NCLB are controversial, most American policymakers appear—on the surface, at least—to share the basic principles of rigor and achievement.

Below this surface agreement, however, lie signs that not all educational policymakers view a rich, challenging curriculum as a universal good or an achievable goal. One clear artifact of these lesser expectations is the continued use of tracking, also known as *ability grouping,* which denies many students access to excellent curriculum and teaching (Oakes, 2005). In fact, one of the tragedies of the standards movement is that schools, desperate to find ways to increase test scores, are relying on practices such as tracking and retention, strategies that are not only counter-productive but also inimical to the very goals of a movement that was intended as a means of closing the achievement gap (Sandholtz, Ogawa, & Schribner, 2004; Thompson, 2001).

Ironically, the most successful response to the pressures of standards-based accountability policies may be the opposite approach: detracking. The longitudinal, quasi-experimental study reported in this article examines the process and outcomes of a diverse suburban district's detracking reform within the context of the standards movement. As this district struggled to help all students achieve high learning standards, it recognized that the standards-based accountability movement shares a core belief with the detracking movement: a challenging curriculum is a universal good and is a benefit to all students (Welner, 2001b). Detracking, when it is done with vigilance and care, can promote both excellence and equity, thus fulfilling the original intent of the accountability movement to depart "radically from the tracking and sorting carried out by the factory-style school of yore" (Thompson, 2001, p. 358).

BACKGROUND

TRACKING AND HIGH LEARNING STANDARDS

In light of increasing pressure to hold all students to high learning standards, policy decisions such as tracking have taken on increased importance (Welner, 2001b). The influence of tracking on student achievement was recently noted when the nation's governors met at the above-mentioned 2005 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 rigor-
> ous college-preparatory program, others into vocational pro-
> grams with less-rigorous curriculum and still others into a gen-
> eral track. Today, all students need to learn the rigorous content
> usually reserved for college-bound students, particularly in math
> and English. (Conklin & Curran, 2005, p. 11)

The governors recognized that, as the job market becomes increasingly
competitive, the sorting and selecting practices of the past, whereby edu-
cators steered students into less rigorous tracks based on perceived abil-
ity and career paths, no longer serve the needs of the nation's youth.

However, although the governors acknowledged the role that tracking
plays in providing less-rigorous educational pathways, they fell short of
calling for its abolishment. Instead, their report refers to appealing to dif-
fering student interests and asserts the need to include a rigorous curric-
ula in all classrooms, implicitly concluding that it is possible to embrace
rigor and low-track 'solutions' at the same time. Research on tracking,
however, suggests that such low-track solutions are destined to disap-
point.

RESEARCH ON TRACKING

Since the turn of the 20th century, schools have placed students in differ-
ent classes based on perceived ability (Goff, 1995; Kliebard, 1995; Oakes,
2005). In the 1980s, and in particular with the 1985 publication of
Jeannie Oakes's *Keeping Track: How Schools Structure Inequality,* researchers
began to seriously question whether this practice was fair or effective.
During the 1980s and 1990s, Oakes and other researchers repeatedly
demonstrated that tracking depresses student achievement and causes
racial and socioeconomic stratification in schools (Braddock & Dawkins,
1993; Gamoran, 1986, 1992; Lipman, 1998).

The asserted purpose of tracking is to tailor the rigor and pacing of
curriculum to meet the specific learning needs of all students (Hallinan,
1994). Therefore, the accuracy of track placement should be a matter of
critical importance to schools. Numerous studies, however, have demon-
strated that the process of assigning students to tracks is influenced by
factors unrelated to student achievement (Garet & DeLany, 1988;
George, 1992; Goyochea, 2000; Lucas, 1999, 1999; Oakes, 2005; Useem,
1992; Wells & Oakes, 1996; Wells & Serna, 1996; Welner, 2001a).

For example, the philosophy of school leaders and the organizational
factors of school, such as the perceived number of vacancies in classes,
often influence the decision as to whether or not a student is placed in

advanced classes (Garet & Delany; Hallinan & Sorenson, 1987; Useem, 1992). In addition, parents with college degrees are more likely to intervene in school experiences, resulting in their child's placement in advanced mathematics classes that lead to the study of calculus in high school (Useem, 1992; see also Wells & Serna, 1996).

There is also ample evidence to show that tracks stratify students by race and class. African American and Latino students are dramatically over-represented in low-track classes and under-represented in high-track classes (Black, 1992; Braddock & Dawkins, 1993; Hallinan, 1992; Oakes, Ormset, Bell, & Camp, 1990; Slavin & Braddock, 1993). Socioeconomic status (SES) has been found to affect track location as well (Lucas, 1999; Lucas & Gamoran, 1993; Vanfossen, Jones, & Spade, 1987). Even after accounting for prior performance, high SES students are over-represented in the academic track, and the effect of SES on track placement extends beyond the effect of SES on student performance (Vanfossen et al.; Welner, 2001a). The theory that instruction in tracked classes is tailored to meet students' academic needs is difficult to support in light of the preponderance of studies that identify the many non-academic factors that influence track placement.

TRACKING AND STUDENT ACHIEVEMENT

Considering the racial and socioeconomic stratification caused by tracking, the practice would seem defensible only if it has a positive effect on student learning. Yet the preponderance of studies indicates that low-track classes are associated with depressed student achievement (Heubert & Hauser, 1999; Oakes, 2005). In addition, these studies indicate that the achievement gap between low- and high-achieving students widens over time in tracked settings (Gamoran & Mare, 1989). This is because low-track classes, "typically characterized by an exclusive focus on basic skills, low expectations, and the least qualified teachers," cause students to fall further and further behind (Heubert & Hauser, p. 282). In contrast, research on accelerated instruction indicates that an enriched curriculum enhances the performance of low achievers and students at risk of failure (Bloom, Ham, Melton, & O'Brient, 2001; Levin, 1997; Mehan, Villanueva, Hubbard, & Lintz, 1996; Peterson, 1989; Singham, 2003).

It is not surprising then, that studies have found higher student achievement in high-track classes than in low-track classes (Epple, Newlon, & Romano, 2002; Heubert & Hauser, 1999; Oakes, Gamoran, & Page, 1992; Welner, 2001a). What is less clear is the effect on high achievers when they study in detracked classrooms. Some studies report that

the learning of higher achievers decreases in detracked, heterogeneous classes (Brewer, Rees, & Argys, 1995; Epstein & MacIver, 1992; Kulik, 1992), while other studies report no significant differences (Burris, Heubert, & Levin, 2006; Figlio & Page, 2002; Mosteller, Light, & Sachs, 1996; Slavin, 1990). In a study of two high schools in England, Boaler (2002) found that traditional, high-track mathematics classes were associated with a disadvantage to high-achieving students—in achievement as well as in enjoyment of mathematics—when compared to a heterogeneous class using reformed curriculum, pedagogy, and assessment.

Even in studies finding that high-track classes result in higher achievement, it is not clear why this is so. Researchers have not been able to disentangle the effects of specific factors associated with high-track classes, such as peer effects, better instruction, and more qualified teachers (Kerckhoff, 1986; Oakes 1986; Slavin & Braddock, 1993). Reflecting on the results of his own study, Kerckhoff 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 then suggests that a high-track advantage may be the result of differentiated curriculum, better teachers in high-track classes, or classroom culture. Similarly, Oakes (1982, 1986, 2005) found that students in high-track classes receive higher-quality instruction, and that lessons in high-track classes include higher-level thinking skills rather than drill-and-practice activities. She and other scholars believe that any higher achievement associated with high-track classes results not from grouping practices per se, but from the factors described above (Levin, 1997; Wheelock, 1992). If highly proficient students show lower achievement in heterogeneous classes, it is possible that it is not due to the presence of low- and average-achieving students in the class, but rather to the dilution of high-level instruction as teachers attempt to teach to the perceived middle.

Scholars who support detracking view an accelerated curriculum as a universal good—of benefit to all students. Rather than viewing curriculum adjustment as a *rationale* for tracking, these researchers view it as a *means* by which to successfully detrack schools. Oakes (1990), Slavin and Braddock (1993), Braddock and Dawkins (1993), and Wheelock (1992), for example, propose that detracking occur as a process of "leveling up." These researchers argue that detracking will only work if "the top track" curriculum "becomes accessible to a broader range of students without watering it down" (Slavin & Braddock, p. 15). In addition, other researchers, such as Henry Levin (1997), founder of the Accelerated Schools Movement, contend that accelerating learning, rather than

remediation, is the best method of improving the achievement of struggling, at-risk students.

Administrative progressives and Taylorist educators in the first part of the 20th century held a contrasting view, approaching accelerated instruction with a dual, stratified mindset (Ravitch, 2000). The same curriculum was viewed as a benefit for "smart" students but as a detriment for "slower" students who, according to these proponents of tracking, were likely to feel frustrated. More recently, researchers with a more favorable view of tracking have argued that if students were equitably and accurately assigned to tracks, and if the quality of both curriculum and instruction were improved, then the negative effects of tracking on low-achieving students would likely be eliminated (Hallinan, 1994; Loveless, 1998; see also Gamoran & Weinstein, 1998).

The most common justification for tracking today rests on the belief that high achievers will be hurt by heterogeneous grouping. According to Kulik (1992), providing tracked classrooms for high achievers is part of the American public school tradition of offering "special classes for students with special needs" (p. xiii). Those who favor tracking warn that if there is an influx of low-achieving students in high-track classes, the learning of high achievers might be adversely affected even if the high-track curriculum remains (Gamoran & Hannigan, 2000; White, Gamoran, Porter, & Smithson, 1996).

This difference of opinion outlined above frames the question that is at the heart of the modern tracking debate, and there is now a most compelling reason for this debate to be resolved. With policymakers demanding that all students attain higher learning standards and that achievement gaps close, researchers can assist by determining whether any high-track advantage is due to a more rigorous curriculum, better resources, peer effects, or other, yet unidentified factors.

THE PURPOSE OF THIS STUDY

Our study responds to this question by examining patterns of student achievement when a district gradually detracked its middle and high school, offering all students a rich, accelerated curriculum in heterogeneously grouped classes. If these students prove unsuccessful, a reasonable conclusion might be that attributes of a homogeneous environment account for the positive high-track outcomes seen in some studies. If, on the other hand, students benefit from the enriched curriculum, a reasonable conclusion might be that any potential high-track advantage is primarily due to the curriculum and expectations of high-track classes. If this is so, then detracking with a high-track curriculum would serve two

beneficial outcomes: (a) ameliorate the racial and socio-economic strati-fication associated with tracking, and (b) increase student achievement without denying high achievers access to high expectations and rigorous curricula.

In this study we examine the effects of heterogeneous grouping com-bined with high-track curricula on two achievement measures of impor-tance. We describe the results of a longitudinal study of the effects on stu-dent achievement when low-track classes were gradually eliminated and replaced with heterogeneously grouped classes in a demographically diverse, suburban high school. Specifically, we examine how detracking affected the earning of two diplomas that represent high standards of achievement—the New York State Regents diploma and the diploma of the International Baccalaureate (IB). The first diploma is tied to New York State standards; the IB Diploma is tied to world-class standards. In addition, we provide a detailed description of the context in which this reform occurred. Our intent is that this study will further inform the debate on detracking reforms that include enriched curricula as an instructional strategy and that it will add to the emerging literature on accelerated study as an alternative to remediation.

CONTEXT

This detracking reform took place in a demographically diverse school in a New York State suburban school district. The district's history and demographics, as well as its leaders' philosophy, influenced the reform's origins and implementation. This section describes the district, the ratio-nale for detracking, and the multi-year implementation process.

THE DISTRICT OF STUDY

The school district is located in a suburban community of 28,000 in Nassau County on Long Island. It operates five elementary schools serv-ing grades K–5, a middle school serving grades 6–8, and a high school serving grades 9–12.[1] New York categorizes the district as one of low needs relative to resource capacity. Nevertheless, the high school is cate-gorized as having a greater number of students of poverty than the usual low-needs district. The district student/teacher ratio of 12:1 approxi-mates the county average of 13:1, and spending per student is approxi-mately the same as the mean for the county (slightly more than $15,000 per pupil, reflecting the New York City suburb's high cost of living). All of the district's teachers are certified in their area of teaching, which is

typical of Long Island districts and now mandated by New York State.

Approximately 20% of the high school's nearly 1,200 students are African American or Latino, about 12% of all students qualify for free or reduced-price lunch, and approximately 10% are special-education students. Of those students who receive free or reduced-price lunch, virtually all are minority students—56% of all African American or Latino students participate in the subsidized lunch program.

ELIMINATING THE THIRD TRACK

In 1993, the district's superintendent and the Board of Education established an ambitious goal: *By the year 2000, 75% of all graduates would earn a New York State Regents diploma, in addition to a local diploma.* To earn a Regents diploma, students must pass a minimum of eight rigorous state Regents exams in multiple subject areas in addition to fulfilling all course requirements. This goal reflected the superintendent's strong belief in the assessment of student learning by an objective, external standard, and it also reflected the district's commitment to academic rigor. At that time, the respective the Regents diploma rates for the district and the state were 58% and 38%. The district gradually eliminated low-track courses that did not follow the Regents curriculum, and eased the transition by offering struggling students instructional support classes while carefully monitoring these students' progress. At the same time, the "gates" to study honors courses were opened, and any student who wanted to take a high-track class could do so. Over a period of about four years, the high school replaced a three-level rigid tracking system with one that had two tracks in grades 9–12. The honors classes in the 11th and 12th grades were International Baccalaureate and/or Advanced Placement courses in all subjects.

Although the overall number of Regents diplomas increased after the lowest, third-tier tracks were eliminated during the early 1990s, a disturbing profile emerged of students who were not earning the diploma. These students not reaching this standard were more likely to be African American or Latino, receive free or reduced-price lunch, or have a learning disability. While majority, middle-class, regular-education students made great progress in earning the Regents diploma after the school eliminated the lower track, students of color and poverty, as well as students with learning disabilities, were left behind. If all graduates were to earn the Regents diploma, systemic change would need to occur to close the gaps and ensure that the school met the needs of all students.

ACCELERATED MATHEMATICS IN HETEROGENEOUS CLASSES

School leaders noticed that passing the second mathematics Regents exam appeared to be the most common roadblock for students in earning a Regents diploma. While high-track students met the mathematics requirement by the end of ninth grade and enrolled in the *third* Regents mathematics course in the tenth grade, low-track students did not even begin the *first* Regents mathematics course until grade ten.

In order to provide all students with ample opportunity to pass the needed courses, in 1995 the district decided that all middle-school students would study the accelerated mathematics curriculum formerly reserved for the district's highest achievers. Under the leadership of the assistant principal of the middle school, the school's mathematics teachers revised and condensed the curriculum. The new curriculum was taught to all students, in heterogeneously grouped classes. To assist struggling learners, the school initiated support classes called 'mathematics workshops' and provided after-school help four afternoons a week.

The results were positive. More than 90% of incoming freshmen entered the high school having passed the first Regents mathematics examination. The achievement gap dramatically narrowed. Between the years of 1995 and 1997, only 23% of regular-education African American or Latino students passed this algebra-based Regents exam before entering high school. After universally accelerating all students in heterogeneously grouped classes, the percentage more than tripled—up to 75%. The percentage of White or Asian American regular-education students who passed also greatly increased—from 54% to 98%.

HETEROGENEOUS GROUPING IN THE HIGH SCHOOL

When universal mathematics acceleration began, the district cautiously excluded some special-education students from the first Regents mathematics exam until they completed ninth grade. These students with learning disabilities were placed in a double-period, low-track Sequential I ninth-grade mathematics class, along with low-achieving new entrants. Consistent with the recommendations of researchers who have defended tracking and encouraged its reform (e.g., Hallinan, 1994; Loveless, 1999b), this class was rich in instructional resources—a mathematics teacher, a special-education inclusion teacher, and a teaching assistant. Class size was limited to 15 or fewer students. Yet the low-track culture of the class was an obstacle to learning as teachers spent valuable instructional time addressing behavior-management issues.

District and school leaders decided that this low-track class failed its

purpose, and the high school principal became convinced that tracking was an ineffective strategy, especially for low achievers. The class was eliminated and the district assertively moved forward with several new reforms the following year. For the 1999 ninth-grade year of entry (YOE) cohort, all special-education students, with the exception of those who were developmentally delayed, took the mathematics Regents exam in the eighth grade, with all other regular-education students.[2] The YOE cohort of 1999 also studied science in heterogeneous classes throughout middle school, and it became the first cohort to be heterogeneously grouped in ninth-grade English and social studies classes.

Ninth-grade teachers were pleased with the results. They described their classes as academic, focused, and enriched. Science teachers reported that the heterogeneously grouped middle school science program prepared students well for ninth-grade biology.

Detracking at the high-school level continued, paralleling the introduction of revised New York State standards-based curricula. Students in the YOE 2000 cohort studied the state's new biology curriculum, titled "The Living Environment," in heterogeneously grouped classes. The following September, the state's new Mathematics A curriculum was taught to the cohort of 2001 without tracking; this class was accordingly the first class to be heterogeneously grouped in *all subjects* in the ninth grade.

Table 1. Progression of Detracking Courses: Grades 6-10

YOE Grade 9	Grades 6-8	Detracked Courses Grade 9	Grade 10
1995-1997	English, social studies, foreign language	Foreign language, 3 tracks in mathematics	Foreign language
1998	English, social studies, foreign language, accelerated mathematics for all students	Foreign language, 2 tracks in mathematics	Foreign language
1999	All subjects	English, social studies, foreign language	Foreign language
2000	All subjects	English, social studies, foreign language, science	Foreign language
2001	All subjects	All subjects	Foreign language
2002	All subjects	All subjects	Foreign language, English, social studies
2003	All subjects	All subjects	Foreign language, English, social studies
2004	All subjects	All subjects	Foreign language, English, social studies, math

Reaching Higher: The International Baccalaureate

The International Baccalaureate Diploma Program (IB) was created in 1967 in order to serve the educational needs of students who were geographically mobile, such as the children of military personnel, diplomats, and international executives. These students needed high-quality academic instruction in order to meet the university entrance requirements in their native countries (Duevel, 1999).

In 1983, the Rockville Centre School District introduced the IB program as a highly exclusive program to serve "gifted and talented" students in the high school. Initially, enrollment levels were low. For example, the YOE 1984 cohort had only 9 diploma candidates.

However, in early 1990s the district began to eliminate the lower track and instituted open enrollment in honors classes. Any student could, in theory, take IB courses in their junior and senior years. Effectively, students made this decision very early in their schooling, when they chose whether they would study honors-track English, social studies, mathematics, and science courses in the ninth and tenth grade. The "late bloomer" students, who decided to opt for the IB program in their junior year, were thus at a disadvantage.

As the school progressively detracked the ninth and tenth grades, enrollment in eleventh and twelfth-grade IB courses grew, allowing a majority of students to participate in the program. The structure and philosophy underlying the IB Diploma program was an ideal fit for the detracking initiatives that were underway in the district. Both the school district and the IB program believe 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" (International Baccalaureate Organization, 2005, p. 5). In other words, given the opportunity to study enriched, challenging curriculum that develops higher-order thinking skills, student capacity to learn and think can grow and expand.

Between 1990 and 2005, the total number of IB examinations taken by students in the high school of study increased more than tenfold, from 75 to 993. As indicated in Table 2, even as the program expanded, student scores on IB examinations have remained high. In fact, during the last five years (2001–2005), the percentage of students earning the highest scores (4 and above) exceeds the percentage of students earning high scores in early years (1990–1994) when far fewer students took IB courses. While this table provides descriptive data only, it suggests that the learning of the school's highest achievers was not adversely impacted by an influx of lower achievers into rigorous IB classes. Both in number and proportion, the highest scores of 6 or 7 on these secure and

standardized international examinations increased as the program became more inclusive.

Table 2. Growth in the IB Program 1990-2005: Number of Exams Taken and Grade Distribution

Exam Year	Total # of Exams	7	6	5	4	3	2	1
			Distribution of scores on IB examinations					
2005	933	23	117	268	344	153	28	0
2004	758	21	101	223	298	104	11	0
2003	739	22	107	223	254	114	19	0
2002	673	24	116	237	197	88	11	0
2001	577	16	87	193	200	76	5	0
2000	608	7	68	186	205	129	13	0
1999	497	24	69	141	170	71	22	0
1998	447	17	83	139	115	79	14	0
1997	270	5	32	77	78	45	13	20
1996	278	16	51	85	89	21	10	6
1995	230	10	32	71	78	30	5	4
1994	192	0	14	44	74	44	13	3
1993	221	6	24	49	67	55	17	3
1992	194	7	21	43	75	38	9	1
1991	158	2	20	37	54	39	4	2
1990	75	5	10	18	25	15	2	0
			Percent of Total Examinations					
		7	6	5	4	3	2	1
2001-2005	3680	3%	14%	30%	34%	15%	2%	0%
1990-1994	840	2%	11%	23%	35%	23%	5%	1%

Note: IB scores range from 7 (highest) to 1(lowest)

Over the past dozen years, the district's detracking has progressively expanded. Beginning in the 2005-06 school year, heterogeneous grouping extended into tenth-grade mathematics, and all students now take the second year of Mathematics B (a course in advanced algebra, trigonometry and pre-calculus) in heterogeneously grouped classes.

RESEARCH DESIGN AND METHODOLOGY

COHORTS OF INTEREST

To determine the effects of detracking the district's middle school and high school, we examined demographic and achievement data for six cohorts of students. The first three cohorts entered high school in 1995, 1996, and 1997, the three years prior to universal acceleration in mathematics and also prior to the beginnings of detracking the ninth grade. The final three cohorts were among the first four cohorts in which all students were accelerated and heterogeneously grouped in middle-school mathematics and experienced at least some detracking in grade 9 (see Table 1). Specifically, these cohorts entered high school in 1998, 1999, and 2001. Unfortunately, data from the year-of-entry (YOE) 2000 cohort

could not be included. This cohort took the 10th grade PSAT in October 2001; their examinations were among those lost due to anthrax exposure in a New Jersey post office in the fall of 2001.

During the years of this study, the racial demographics and socio-economic cohort population characteristics remained stable. The percentage of high school students qualifying for free or reduced-price lunch varied between 11% and 14%. In any given cohort, racial/ethnic make-up also stayed within a small band: for African American students, between 7% and 8%; for Latino students, between 10% and 12%; for White students, between 77% and 78%; and for Asian American students, between 3% and 5%. Moreover, the real estate market within the district remained stable—there were no major construction projects and district boundaries did not change.

Selection effects are possible, however, even in stable populations. For example, the inclusion of transfer students who entered the high school after ninth grade could bias study results. One strategy for eliminating such effects is to include only data for cohort members who have the most similar histories – students who obtained their complete high school education through this school's regular program (Cook & Campbell, 1979). Therefore, student data were included only for regular and special-education cohort members who: (a) were continuously enrolled in the school district from the ninth grade to the exit of high school, (b) had a YOE into ninth grade from 1995–2001 (excepting 2000), and (c) were not developmentally delayed.[3]

Applying these criteria, each of the six cohorts ranged in number between 221 and 262 students. There were a total of 1,300 individual student records included in this study.

The study used four measures of student achievement, two of which are based on the Practice Scholastic Assessment Test (PSAT), which students take in the tenth grade, as discussed in the next section of this article. The four measures are as follows: (a) scores on the verbal portion of the PSAT, (b) scores on the mathematics portion of the PSAT, (c) the earning of the New York State Regents diploma, and (d) the earning of the International Baccalaureate diploma. PSAT scores were used to create an independent variable representing student scholastic aptitude (described below); the two indicators of diploma attainment were used as outcome measures of achievement.

THE PRACTICE SCHOLASTIC ASSESSMENT TEST (PSAT)

The PSAT is a standardized test typically given to high school students to provide practice for the Scholastic Assessment Test (SAT) and to qualify

them for National Merit Scholarships. The PSAT measures critical reading skills, mathematics problem-solving skills, and writing skills. Like the SAT, it produces normalized verbal and mathematics scores. The SAT, used primarily for college admission purposes, is a test that is indicative of academic aptitude commonly referred to as *g* (Frey & Detterman, 2004). The school district pays for all of its students to take the PSAT in the 10th grade; therefore, nearly all students have the exam in their records.

In this analysis we use PSAT mathematics and PSAT verbal exam scores to provide a measure of general scholastic aptitude; this will allow us to address the important question of whether the effect of detracking is constant across prior achievement levels. In this sample, overall PSAT mathematics and verbal scores are highly related; the Pearson correlation between the two measures is r=0.730 (p<.001). This strong relationship points to the influence of a general level of aptitude in addition to aptitude specific to each measure's subject area.

PSAT scores could be used in several ways to represent general scholastic aptitude in our statistical analysis; three such strategies are discussed here. First, both the verbal and mathematics measures could be included in a statistical model. However, this strategy is not advisable, since the two predictors are very highly correlated, which would preclude us from interpreting the two individual estimates. A second strategy is to include only one of the two scores in our statistical model. This, too, is not ideal, since any given measure would not only reflect general scholastic aptitude but also scholastic aptitude specific to the designated subject area (and measurement error).

A third strategy—the one ultimately employed in this analysis—is to estimate a general aptitude score from the correlated part of the two individual subject measures. This approach is premised on the assumption that an individual's PSAT mathematics and verbal scores share the influence of that individual's general scholastic aptitude. That is, a given student's mathematics and verbal scores should each be influenced by the same general scholastic aptitude. To estimate this general aptitude we used principal components analysis (PCA) to create an index of component scores from the first principal component taken from the two measures. PCA is a statistical method for transforming correlated variables into new variables that are uncorrelated with each other and best represent the variance shared by original variables (Dillon & Goldstein, 1984). In our PCA analysis of PSAT verbal and mathematics scores, the first principal component represents the correlated part of these two scores—the influence shared by the two assessments rather than specific to one of the designated subject areas. Component scores for the first

principal component—values calculated for each individual based only on the correlated part of the two assessments—were used as a measure of aptitude common across the two assessments (Dillon & Goldstein). This index, measured in standard normal units (mean = 0; standard deviation = 1), is hereafter represented by the independent variable "APTITUDE" and is used to model whether the effects of detracking varied for students at different levels of scholastic aptitude.

NEW YORK STATE REGENTS DIPLOMA

During all but the final year analyzed in this study,[4] in order to qualify for a New York State Regents diploma, students needed to pass a minimum of eight end-of-course Regents examinations including the following: (a) two in mathematics, (b) two in laboratory sciences, (c) two in social studies, (d) one in English Language Arts, and (e) one in a foreign language.[5] All coursework in the above subject areas must be passed as well. The secure Regents examinations are prepared by a committee of New York State teachers in conjunction with education department specialists in the subject matter and in testing (New York State Department of Education, 2001). Examinations are given three times a year, in January, June and August, at a time designated by the New York State Board of Regents. Scores range from 0%–100%. Scores of 65% and above are designated as passing by the New York State Board of Regents, and scores of 85% and above are designated as reflecting mastery. If students pass all of the needed exams, all required coursework, and earn at least the designated number of high-school credits, a Regents diploma is awarded upon graduation. Students who do not meet these standards may receive a local (also called "school") diploma by meeting local and state diploma requirements.

THE INTERNATIONAL BACCALAUREATE DIPLOMA

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. Student learning is measured by criterion-referenced assessments, which are consistent from year to year and applied equally across schools (International Baccalaureate Organization, 2005). International senior examiners grade the student work.[6]

Participating schools must be accredited by the International Baccalaureate Organization (IBO), and must make a substantial commitment to teacher training and development.[7] Colleges around the globe give students credit for IB courses, recognizing the demanding nature of the curriculum and the assessments. Students may elect to become full IB diploma candidates, or they may study individual courses to earn certificates.

In order to receive the IB Diploma, students must earn a minimum of twenty-four points on assessments from six IB courses, five of which must come from the five areas of study, referred to as groups 1–5. Three of the courses must be taken at the higher level; in other words, the course must meet for no less than 240 classroom hours. The remaining three courses must meet for a minimum of 150 hours. Students also must successfully complete three central elements: (a) *Community Action Service*, which is a reflective chronicle of their extracurricular/service learning activities; (b) *Theory of Knowledge,* a transdisciplinary epistemology course; and (c) the *Extended Essay*, an extensive independent research project of no more than 4000 words, conducted over the course of two years under the guidance of a faculty mentor.

STUDENT DEMOGRAPHICS

Given the complex interaction between program and social factors in any social science context, analysis in educational research should include consideration of the demographic characteristics of research subjects. Two such variables often included in such analyses are ethnicity and socioeconomic status. Ideally data would support analyses that disentangled the effects associated with each of these variables; such data would support estimation of separate main effects for each variable as well as the interaction between the two. In practice, however, these two variables are often strongly related. For example, African American and Latino students in the United States tend to be more likely to come from families lower in socioeconomic status than do White students (Cabrera & Bernal, 1998). As such it is often impossible to estimate effects as though the two were orthogonal. In this sample, 107 of the 124 students eligible for free or reduced-price lunch (86.3%) are either African American or Latino, whereas only 76 of the 1,176 students not eligible for lunch programs (6.5%) are also African American or Latino (Table 3). Main effects for SES and ethnicity, therefore, cannot be interpreted from this sample, since these two variables are overwhelmingly related.

Table 3. SES and Minority Status Crosstabulation

		Low SES		Total
		NO	YES	
MINORITY	NO	1100	17	1117
	YES	76	107	183
Total		1176	124	1300

To address this issue in our analyses we describe students as represented by one of four independent groups based on the combination of ethnicity and socioeconomic status. Group indicator variables are used in order to estimate main effects of group membership (i.e., group differences in likelihood of diploma attainment) as well as interactions between other variables and group membership (i.e., group differences in the relationship between aptitude or cohort and diploma attainment).

VARIABLES

The variables used to answer the research questions were the following:

> REGDIP—A dependent binary variable of 0 or 1 to indicate whether the student received a Regents diploma.
> IBDIP—A dependent binary variable of 0 or 1 to indicate whether the student received an International Baccalaureate diploma.
> SPED—An independent binary variable of 0 or 1 that represents whether the student received special education services.
> APTITUDE—an independent variable measured in standard normal units that represents estimated general scholastic aptitude.
> GROUP1—Free or Reduced Price Lunch ("FRPL") eligible and either Latino or African American.[8]
> GROUP2—FRPL eligible and either Asian American or White.
> GROUP3—Ineligible for FRPL and either Latino or African American.
> GROUP4—Ineligible for FRPL and either Asian American or White.[9]
> PREPOST—An independent binary variable of 0 or 1 to indicate whether the student was a member of a cohort (1) that entered high school in September 1998 or beyond.

Descriptive statistics for the variables used in this study are presented in Table 4.

Table 4. Descriptive Statistics

			Tracked						Detracked		
			APTITUDE	REGDIP	IBDIP				APTITUDE	REGDIP	IBDIP
Regular	GROUP1 (Minority; LowSES)	Mean	-1.069	36.8%	5.3%	GROUP1 (Minority; LowSES)	Mean		-1.069	85.7%	12.2%
		N	42	57	57		N		36	49	49
		SD	1.048				SD		0.815		
	GROUP2* (Non-Minority; LowSES)	Mean	---	---	---	GROUP2* (Non-Minority; LowSES)	Mean		---	---	---
		N	3	3	3		N		9	11	11
		SD	---				SD		---		
	GROUP3 (Minority; Non-LowSES)	Mean	-0.279	69.7%	21.2%	GROUP3 (Minority; Non-LowSES)	Mean		-0.210	97.8%	20.0%
		N	27	33	33		N		42	45	45
		SD	0.812				SD		0.842		
	GROUP4 (Non-Minority; Non-LowSES)	Mean	0.304	94.2%	29.4%	GROUP4 (Non-Minority; Non-LowSES)	Mean		0.150	98.5%	32.5%
		N	482	521	521		N		549	587	587
		SD	0.923				SD		0.846		
			APTITUDE	REGDIP	IBDIP				APTITUDE	REGDIP	IBDIP
Special Education	GROUP1* (Minority; LowSES)	Mean	---	---	---	GROUP1 (Minority; LowSES)	Mean		-1.488	28.1%	0.0%
		N	9	10	10		N		20	32	32
		SD	---				SD		0.874		
	GROUP2* (Non-Minority; LowSES)	Mean	---	---	---	GROUP2* (Non-Minority; LowSES)	Mean		---	---	---
		N	3	4	4		N		2	2	2
		SD	---				SD		---		
	GROUP3* (Minority; Non-LowSES)	Mean	---	---	---	GROUP3* (Minority; Non-LowSES)	Mean		---	---	---
		N	1	3	3		N		6	8	8
		SD	.				SD		---		
	GROUP4 (Non-Minority; Non-LowSES)	Mean	-0.900	39.3%	0.0%	GROUP4 (Non-Minority; Non-LowSES)	Mean		-0.994	55.3%	0.0%
		N	41	56	56		N		28	38	38
		SD	0.907				SD		0.778		

* Figures not reported due to small sample size

ANALYTIC STRATEGY

Attainment of the two diploma types (REGDIP and IBDIP) was modeled using logistic regression on independent variables representing tracked status (PREPOST), scholastic aptitude (APTITUDE), demographic characteristic grouping (GROUP1, GROUP2, and GROUP3), and special education identification (SPED). Several models were fit for each dependent variable, using a three-stage framework by which we took the following steps: (a) we started with aptitude-related, demographic, and special education variables; (b) we then progressively added main effects as well as interactions with the tracked-status variable; and (c) finally, we progressively removed those variables which do not appear to add explanatory power to the model. For each of our models, we report statistics for individual predictors (e.g., coefficient estimates, significance values, and odds ratios) as well as for the model as a whole (e.g., log likelihoods and pseudo-R^2). We describe results at each step of the process; however, our interpretation of these results is offered only at the end of each diploma discussion, after all the models for that diploma are presented. The final models best balance explanatory power and parsimony, and therefore provide the basis for the conclusions presented in the discussion sections.

RESULTS

REGENTS DIPLOMA ATTAINMENT

Logistic regression models for the attainment of the Regents diploma are provided in Table 5a-b.

The first four models increase in complexity through the progressive addition of independent variables. Model 1 begins with only the main effects for aptitude, special education, and demographic group indicator variables. As expected, APTITUDE is positively associated with Regents diploma attainment, while attainment is less likely for GROUP1, GROUP2, and SPED. Model 2 adds all possible 2-way interactions between these variables. Taken together, these interactions add significant explanatory power to the model $(X^2(7) = 26.09, p<.001)$.[10]

Model 3 is the first model to include the programmatic variable (PREPOST), representing the effect of detracking. Adding PREPOST to the model adds significant explanatory power $(X^2(1) = 35.6, p<.001)$; the likelihood that the effect of detracking showed up due to randomness is less than one chance in one thousand. According to Model 3, students in

Table 5a. Regents Diploma Models (1 of 2)

Coefficients	Model 1 B	se(B)	exp(B)	Model 2 B	se(B)	exp(B)	Model 3 B	se(B)	exp(B)	Model 4 B	se(B)	exp(B)
Constant	4.2***	0.250	67.01	4.8***	0.375	122.02	4.28***	0.392	71.91	4.34***	0.417	76.86
APTITUDE	1.77***	0.189	5.85	2.12***	0.301	8.31	2.3***	0.327	9.99	2.26***	0.344	9.58
GROUP1	-1.08***	0.326	0.34	-2.14**	0.658	0.12	-2.3***	0.682	0.10	-2.43***	0.728	0.09
GROUP2	-2.03**	0.665	0.13	5.22	9.234	184.79	7.57	12.203	1.9E+03	97.29	3.4E+03	1.8E+42
GROUP3	-0.16	0.512	0.85	-1.88***	0.708	0.15	-2.02**	0.732	0.13	-2.19**	0.755	0.11
SPED	-2.26***	0.294	0.10	-3.48***	0.531	0.03	-3.56***	0.544	0.03	-3.43***	0.576	0.03
SPED by GROUP1				1.57*	0.682	4.80	1	0.728	2.73	0.59	0.881	1.80
SPED by GROUP2				-2.51	5.620	0.08	-4.34	7.437	0.01	-71.37	2.5E+03	0.00
SPED by GROUP3				3.54**	1.348	34.64	2.84*	1.324	17.17	2.39	1.499	10.92
APTITUDE by GROUP1				-0.51	0.437	0.60	-0.64	0.464	0.53	-0.4	0.505	0.67
APTITUDE by GROUP2				8.22	9.279	3.7E+03	9.93	12.086	2.1E+04	102.03	3.5E+03	2.0E+44
APTITUDE by GROUP3				-0.8	0.596	0.45	-0.85	0.594	0.43	-0.63	0.646	0.53
APTITUDE by SPED				-0.47	0.421	0.62	-0.48	0.433	0.62	-0.64	0.448	0.53
PREPOST							1.75***	0.319	5.78	1.4*	0.693	4.06
PREPOST by APTITUDE										0.02	0.450	1.02
PREPOST by GROUP1										1.3	0.815	3.68
PREPOST by GROUP2										30.63	1.2E+03	2.0E+13
PREPOST by GROUP3										1.35	1.282	3.85
PREPOST by SPED										-0.48	0.725	0.62
Model Fit												
-2 Log Likelihood	413.59			387.5			351.9			344.99		
Chi-Square Change	396.11***			26.09***			35.6***			6.91		
Degrees of Freedom	5			7			1			5		

Table 5b. Regents Diploma Models (2 of 2)

Coefficients	Model 5			Model 6			Model 7		
	B	se(B)	exp(B)	B	se(B)	exp(B)	B	se(B)	exp(B)
Constant	4.27***	0.404	71.56	4.04***	0.345	57.06	3.91***	0.290	49.94
APTITUDE	2.32***	0.341	10.19	2.14***	0.308	8.46	1.96***	0.211	7.07
GROUP1	-2.38***	0.720	0.09	-2.26**	0.693	0.10	-1.86***	0.457	0.16
GROUP2									
GROUP3	-2.12**	0.746	0.12	-1.88**	0.719	0.15	-1.48*	0.581	0.23
SPED	-3.46***	0.556	0.03	-2.78***	0.325	0.06	-2.8***	0.323	0.06
SPED by GROUP1	0.48	0.875	1.62						
SPED by GROUP2									
SPED by GROUP3	2.32	1.482	10.22						
APTITUDE by GROUP1	-0.48	0.503	0.62	-0.41	0.503	0.66			
APTITUDE by GROUP2									
APTITUDE by GROUP3	-0.7	0.643	0.50	-0.56	0.660	0.57			
APTITUDE by SPED	-0.67	0.446	0.51						
PREPOST	1.46*	0.676	4.30	1.17*	0.552	3.22	1.21***	0.361	3.35
PREPOST by APTITUDE	0.07	0.447	1.07	-0.08	0.436	0.92			
PREPOST by GROUP1	1.31	0.806	3.70	1.48*	0.708	4.38	1.66*	0.674	5.27
PREPOST by GROUP2									
PREPOST by GROUP3	1.34	1.276	3.81	2.78*	1.265	16.13	3.27**	1.150	26.24
PREPOST by SPED	-0.37	0.708	0.69						
Model Fit									
-2 Log Likelihood	359.16			364.51			365.64		
Chi-Square Change	14.17**			5.35			1.13		
Degrees of Freedom	4			4			3		

detracked cohorts have odds of Regents diploma attainment nearly six times greater than their tracked counterparts with corresponding aptitude and demographic characteristics.

Model 4 includes all possible main effects and 2-way interactions. The main effect for PREPOST is again positive, and large odds ratios for the PREPOST interactions with GROUP1 and GROUP3 provide some indication of differential effects of detracking for these groups, though they are not statistically significant when the full set of predictors is included in this model.

Models 5–7 represent a paring down of predictors to achieve parsimony in addition to explanatory power. The main effects and interactions with GROUP2 are removed for Model 5. Only 17 individuals are either Asian American or White yet are FRPL eligible; this small sample size cannot support separate estimates for this group, as evidenced by the lack of statistical significance and extreme odds ratio estimates. The same

problem is addressed in Model 6 for SPED, although the main effect is left in because it is strongly significant (reflecting the fact that fewer special education students received Regents diplomas).

A potentially important 2-way interaction—the interaction between PREPOST and APTITUDE—is also removed, along with other APTITUDE interactions, in Model 7, with virtually no change in explanatory power ($X^2(1) = 1.13$, p>.05). The lack of significance of this interaction is particularly important because it suggests a counter to one of the main arguments against detracking – that detracking helps low-aptitude students at the expense of students at the upper end of the aptitude spectrum. Had this been true in the current study, the PREPOST x APTITUDE interaction would have had a significant negative effect on Regents diploma attainment. This is clearly not the case according to Model 7. Conditional on the other variables in the model, the positive effect of detracking encompasses (is not statistically different for) both low- and high-aptitude students in the earning of a Regents diploma.

Accordingly, Model 7 represents what we believe to be the best balance between explanatory power and parsimony. The remaining terms are both statistically and practically significant. Each of the remaining coefficients has a p-value less than 0.05; each corresponding odds ratio ($\exp(B)$) is far from 1.0.

The inference regarding detracking is clear from Model 7 (and is strikingly consistent across all models): being a member of a detracked cohort is associated with substantial increases in the odds of attaining the Regents diploma. For students in Group 2 (non-minority, FRPL eligible) and Group 4 (non-minority, non-FRPL-eligible), the benefit is a 3-fold increase. The impact of detracking appears to be even greater for those students in Group 1 (minority, FRPL-eligible) and Group 3 (minority, non-FRPL-eligible). For these students, detracking appeared to improve the odds of diploma attainment by factors of greater than 5 and greater than 26, respectively—nearly compensating for the negative main effect of GROUP1 status and more than compensating for the negative effect of GROUP3 status. In sum, detracking is associated with positive results for all students, with even greater results shown for those who, in the State of New York, are far less likely to earn a Regents diploma (Mills, 2004).

The following illustration, based on coefficient odds ratios, helps to place the magnitude of the effect associated with detracking in context. The odds ratio for PREPOST (3.35) is nearly half as large (47%) as the odds ratio of APTITUDE (7.07). As such, for those students in GROUP2 and GROUP4, being a member of a detracked cohort gives an improvement of the odds of Regents diploma attainment similar in magnitude to

an increase of .47 standard deviations of APTITUDE. This means that detracked students at the 25th percentile of APTITUDE would share the same Regents odds ratio as their tracked counterparts at the 42nd percentile of APTITUDE. Similarly, detracked students at the 45th percentile of APTITUDE would share the same Regents odds ratio as their tracked counterparts at the 64th percentile of APTITUDE. These analyses suggest the powerful role that detracking played in helping this school district substantially increase the proportion of its students earning the New York State Regents diploma. For members of GROUP 1, detracked students at the 25th percentile of APTITUDE would share the same Regents odds ratio as their tracked counterparts at the 80th percentile of APTITUDE. And for members of GROUP 3, detracked students at the 25th percentile of APTITUDE would share the same Regents odds ratio as their tracked counterparts at the 95th percentile of APTITUDE.

Figures 1–3 demonstrate the positive effect of being a member of a detracked cohort for non-special education students in each of the groups in this sample, based on Model 7. For each demographic group, the likelihood of Regents diploma attainment is plotted against APTITUDE for both tracked and detracked cohorts. In each case the detracked cohort has a substantially greater likelihood of receiving the Regents diploma at virtually every level of APTITUDE.
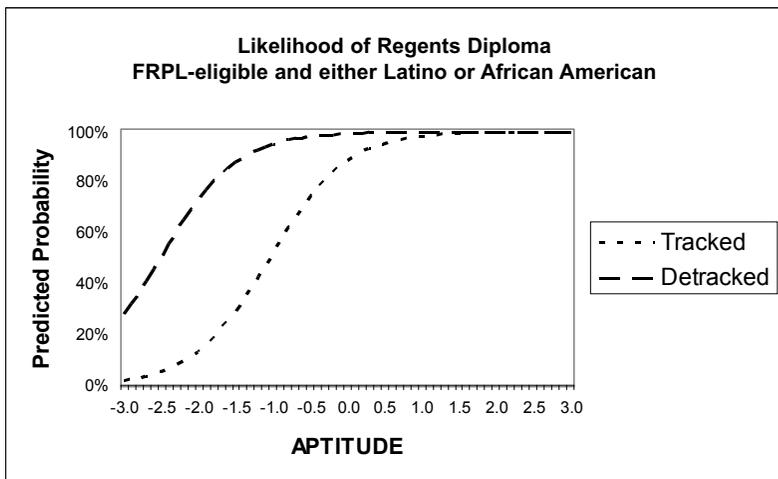
Figure 1.

Likelihood of Regents Diploma
FRPL-eligible and either Latino or African American

Figure 2.

Likelihood of Regents Diploma
Ineligible for FRPL and either Latino or African American

Figure 3.

Likelihood of Regents Diploma
Asian American or White, combined FRPL-status
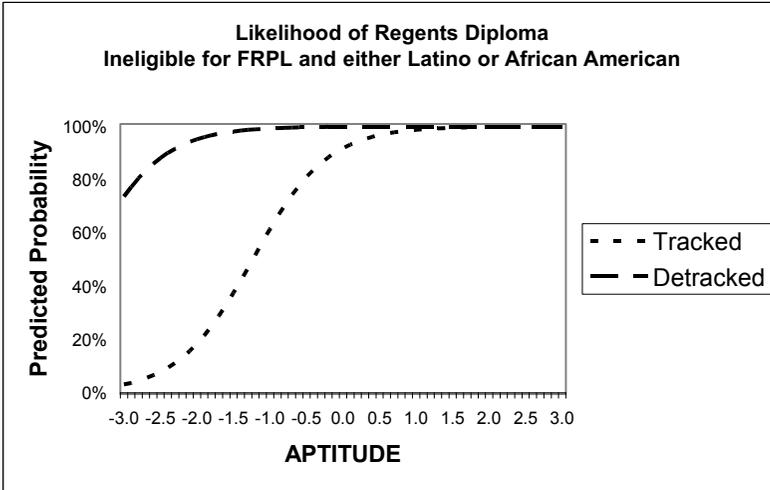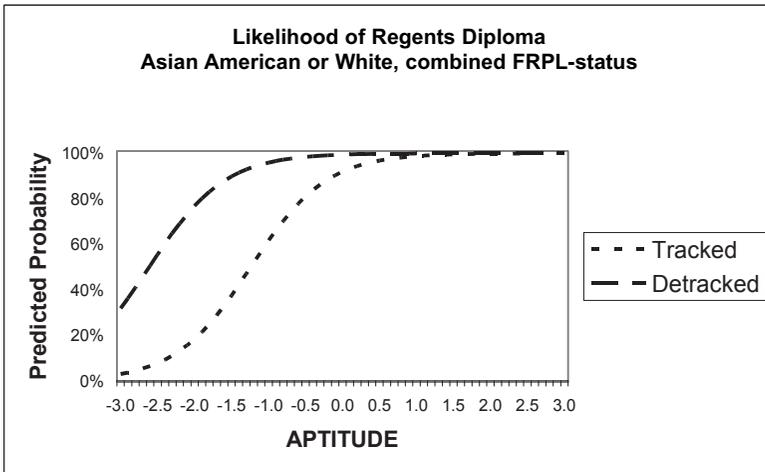
INTERNATIONAL BACCALAUREATE DIPLOMA ATTAINMENT

Logistic regression models for the attainment of the International Baccalaureate Diploma (IBDIP) are provided in Table 6a-b.

The first four models increase in complexity in exactly the same manner as those for REGDIP, as presented in the last section. Model 1 again begins with only the main effects for aptitude-related and demographic

Table 6a. IB Diploma Models (1 of 2)

Coefficients	Model 1 B	se(B)	exp(B)	Model 2 B	se(B)	exp(B)	Model 3 B	se(B)	exp(B)	Model 4 B	se(B)	exp(B)	Model 5 B	se(B)	exp(B)
Constant	-1.45***	0.098	0.23	-1.44***	0.100	0.24	0.143	0.143	0.17	-1.87***	0.175	0.15	-1.87***	0.174	0.15
APTITUDE	1.78***	0.116	5.93	1.77***	0.122	5.87	0.126	0.126	6.28	1.96***	0.186	7.07	1.96***	0.185	7.10
GROUP1	0.42	0.442	1.53	0.67	0.541	1.96	0.543	0.543	2.01	-1.66	1.712	0.19	-1.66	1.714	0.19
GROUP2	-0.83	1.100	0.44	-0.84	1.113	0.43	1.116	1.116	0.37	-13.37	1.2E+03	0.00			
GROUP3	0.22	0.338	1.24	0.28	0.334	1.32	0.336	0.336	1.29	0.92	0.522	2.52	0.93	0.522	2.53
SPED	-13.77	277.03	1.0E-06	-14.09	340.61	7.6E-07	341.85	341.85	8.3E-07	-14.78	638.23	3.8E-07	-14.86	641.47	3.5E-07
SPED by GROUP1				1.07	431.688	2.90	431.205	431.205	2.15	2.08	#######	8.03	1.8	1.0E+03	6.02
SPED by GROUP2				-1.18	887.032	0.31	878.245	878.245	0.26	0.57	1.5E+03	1.76			
SPED by GROUP3				-0.97	788.143	0.38	784.137	784.137	0.27	-0.22	1.5E+03	0.80	-0.5	1.4E+03	0.61
APTITUDE by GROUP1				1.38	1.004	3.96	0.981	0.981	3.69	2.31	1.311	10.09	2.32	1.311	10.14
APTITUDE by GROUP2				-1.39	1.567	0.25	1.597	1.597	0.19	-1.96	1.697	0.14			
APTITUDE by GROUP3				-0.39	0.486	0.68	0.494	0.494	0.64	-0.43	0.489	0.65	-0.43	0.489	0.65
APTITUDE by SPED				-2.1	206.259	0.12	209.118	209.118	0.11	-2.67	332.531	0.07	-2.66	332.201	0.07
PREPOST							0.159	0.159	1.75	0.68**	0.212	1.98	0.68**	0.211	1.97
PREPOST by APTITUDE										-0.22	0.245	0.80	-0.23	0.244	0.79
PREPOST by GROUP1										3.35	2.064	28.47	3.36	2.066	28.80
PREPOST by GROUP2										12.43	1.2E+03	2.5E+05			
PREPOST by GROUP3										-1.05	0.668	0.35	-1.05	0.668	0.35
PREPOST by SPED										-2.15	9.5E+02	0.12	-1.76	842.730	0.17
Model Fit															
-2 Log Likelihood	1062.52			1058.22			1045.51			1037.15			1039.45		
Chi-Square Change	468.33***			4.3			12.71***			8.36			2.3		
Degrees of Freedom	5			7			1			5			4		

Table 6b. IB Diploma Models

Coefficients	Model 6 B	se(B)	exp(B)	Model 7 B	se(B)	exp(B)	Model 8 B	se(B)	exp(B)	Model 9 B	se(B)	exp(B)
Constant	-1.87***	0.174	0.15	-1.79***	0.146	0.17	-1.72***	0.139	0.18	-1.75***	0.137	0.17
APTITUDE	1.96***	0.185	7.10	1.84***	0.126	6.27	1.8***	0.120	6.02	1.82***	0.118	6.18
GROUP1	-1.66	1.714	0.19	-1.63	1.663	0.20	-1.69	1.662	0.18			
GROUP2												
GROUP3	0.93	0.522	2.53	0.86	0.507	2.35						
SPED	-14.83	244.83	3.6E-07	-14.88	244.64	3.4E-07	-14.88	245.19	3.4E-07	-13.62	275.83	1.2E-06
SPED by GROUP1												
SPED by GROUP2												
SPED by GROUP3												
APTITUDE by GROUP1	2.32	1.311	10.14	2.29	1.302	9.86	2.33	1.302	10.26			
APTITUDE by GROUP2												
APTITUDE by GROUP3	-0.43	0.489	0.65	-0.44	0.488	0.64						
APTITUDE by SPED												
PREPOST	0.68**	0.211	1.97	0.55***	0.165	1.74	0.49**	0.159	1.64	0.54***	0.158	1.72
PREPOST by APTITUDE	-0.23	0.244	0.79									
PREPOST by GROUP1	3.36	2.066	28.80	3.39	2.017	29.68	3.45	2.017	31.52			
PREPOST by GROUP2												
PREPOST by GROUP3	-1.05	0.668	0.35	-0.95	0.658	0.39						
PREPOST by SPED												
Model Fit												
-2 Log Likelihood	1039.45			1040.35			1043.44			1052.48		
Chi-Square Change	0			0.9			3.09			9.05*		
Degrees of Freedom	4			1			3			3		

variables. APTITUDE provides most of the predictive power in Model 1. Although it is not statistically significant, SPED is also associated with a strong negative effect – according to this model, the odds of IB diploma attainment by special education students were nearly zero.
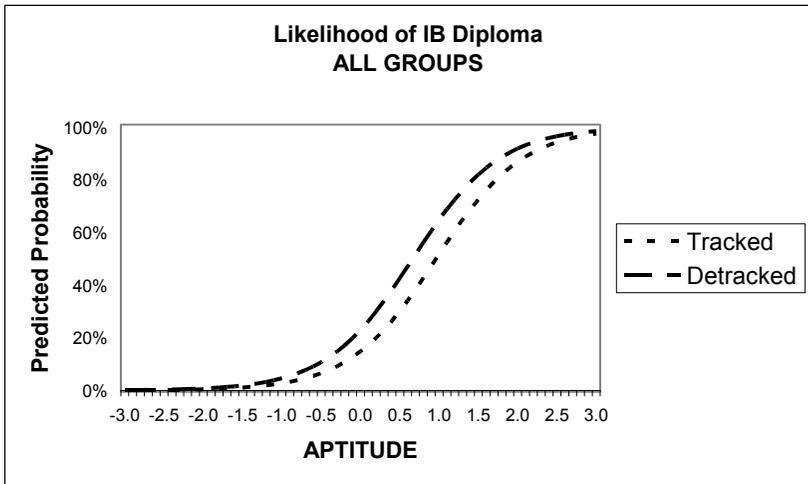
Model 2 introduces interactions between the variables in Model 1; these appear to add little to the model ($X^2(7) = 4.3$, p>.05). Adding PRE-POST to Model 3 significantly increases explanatory power ($X^2(1) = 12.71$, p<.001) and presents an odds ratio of 1.75, suggesting that students in detracked cohorts have 75% greater odds of attaining an IB diploma than their tracked counterparts. Similar to the case for Regents diploma attainment above, interactions of PREPOST with aptitude-related and demographic variables add little explanatory power in Model 4. Models 5-9 exclude predictors in the name of parsimony. The removal of GROUP2 main effects results in no significant loss of explanatory power, nor does the removal of GROUP2 its interactions (Model 5). Removing all interactions with SPED (Model 6), while keeping the main effect in the model, also results in no significant loss of explanatory power. Model 7 removes the interaction between PREPOST and APTI-TUDE. Consistent with Model 6 of the Regents diploma analysis above, Model 7 demonstrates that little explanatory power is gained from this interaction—effects of detracking on IB diploma attainment appear to be uniform and positive across aptitude levels.

Models 8 and 9 continue the progression toward parsimony by removing main effects and interactions for GROUP3 and GROUP1, respectively. Although the magnitude of GROUP1 effects and corresponding odds ratios appear large (and similar in magnitude to those from the Regents analysis), they are not statistically significant in this model.

As with Model 7 for the Regents diploma analysis, here Model 9 represents our best candidate to balance explanatory power and parsimony. Once again, the inference regarding detracking is the same: being a member of a detracked cohort is associated with an increase of roughly 70% in the odds of IB diploma attainment (based on an odds ratio (exp(B)) of 1.72). This PREPOST odds ratio is 27% as large as the odds ratio of APTITUDE (6.18), meaning that detracking is associated with an improvement of the odds of IB diploma attainment similar in magnitude to an increase of .27 standard deviations of APTITUDE. Detracked students at the 25th percentile of APTITUDE would share the same IB odds ratio as their tracked counterparts at the 35th percentile of APTITUDE, while detracked students at the 45th percentile of APTITUDE would share the same IB odds ratio as their tracked counterparts at the 56th percentile of APTITUDE.

Figure 4 demonstrates the positive effect of detracking for all demo-

Figure 4.

graphic groups in this sample, based on Model 9. The plot in Figure 4 represents the likelihood of IB diploma attainment at various levels of APTITUDE for both tracked and detracked cohorts. The detracked cohort has a greater likelihood of receiving the IB diploma at virtually every level of APTITUDE.

EXPLORATION OF OTHER POTENTIAL INFLUENCES ON THE OUTCOME

As the school detracked, students were expected to pass more rigorous courses and examinations. Could it be that the increase in the proportion of students graduating with Regents and/or IB diplomas was the result of fewer students graduating overall? In other words, was the increased rigor of the detracked curricula associated with student discouragement and an increase in "dropping out" of high school? In order to answer this question, we examined the district's dropout rates from 1996 to 2005. We found that progressive detracking was not associated with more students leaving school prior to graduation. The reverse was true – detracking was associated *with a decrease* in the rate of students dropping out of school although it should be noted that this school had a very low dropout rate over the entire period.

To put this rate in a statewide perspective, only 67% of all public high

Table 7. Dropout rates 1999-2005

Graduation year	1999	2000	2001	2002	2003	2004	2005
Dropout rate (number of students/ total high school population)	0.3%	0.3%	0.6%	0.1%	0.2%	0.1%	0.1%

school students in New York State who entered ninth grade in 2000 graduated in 2004. In the school of study, 98% of all students who entered high school in 2000 graduated in 2004. Of the remaining five students, two were developmentally delayed students who will remain until the age of 21, one student dropped out, and the remaining two students graduated one year late, in 2005. It would not appear, therefore, that the increase in the rigor of coursework led to students being left behind or being pushed out of school.

We also considered the possibility that the school's rise in diploma rates reflects a broader trend of increases in Regents diploma rates, or that the district of study may have begun with an unusually low rate and then dramatically increased as Regents examinations became high school exit exams for New York State students. To test this hypothesis, we compared the Regents diploma attainment rates of the district's students with all students in New York State and with students in similar schools. Between the years of 2000–2002, there was a sharp increase (48% to 56%) in the attainment of Regents diplomas by graduates of New York State Public Schools, as the state phased in earning a score of 55 on selected Regents examinations as a graduation requirement for non-special education students. The district of study's increase during those early years was smaller (84%–88%). During the years (2002–2004) that the progressively detracked cohorts began to graduate, however, increases in the Regents diploma rate statewide were substantially smaller (56%–57%), but the rate for the district of study in the same time period accelerated (88%–94%).

But what about suburban schools with resources similar to the school of study? New York State categorizes districts and schools by a ratio of needs to resources, thus creating similar groups of districts and schools. The district we studied belongs to Group 6, which is described by New York State as districts that serve students with low student needs in relation to district resource capacity. Its high school belongs to Group 54 – secondary schools in Group 6 that have relatively high student needs (New York State Board of Regents, 2003). Mirroring statewide trends, in Group 54 there was a sharp increase in the average Regents diploma rate between 2000–2002 (66%–77%) followed by an increase of only one

percentage point from 2002 to 2004 (77%–78%), the years in which the Regents diploma rate at the school of study rate increased by 6 percentage points.[11]

More telling, perhaps, is the comparison in growth of the earning of Regents diplomas by African American and Latino students during those three years. Such students who were members of progressively detracked cohorts experienced dramatic increases in earning Regents diplomas, from 52% in 2002 to 83% in 2004 (Burris & Welner, 2005). Detracking proved to be an effective strategy for "closing the gap" in Regents diploma attainment. To put this rate in perspective, according to a 2004 New York State report, of those students who graduated in 2003, 23% of Black students and 26% of Hispanic students earned Regents diplomas (Mills, 2004). And these dismal numbers were calculated using as a denominator only those students who received a high school diploma. The official New York State dropout rates for that cohort were 14% (Black students) and 17% (Hispanic students), and an additional 6% and 5%, respectively, transferred to GED programs (Mills).

LIMITATIONS OF THE STUDY

Two potential limitations of this study are its quasi-experimental design (Cook & Campbell, 1979) and its generalizability. As explained recently in a National Research Council publication, the interrupted time-series quasi-experimental approach used here, "represents a relatively strong research design that is often able to provide internally valid evidence about the causal effects of an intervention" (Commission on Behavioral and Social Sciences and Education, 1998, p. 15). This strength is in part due to the fact that "intervention studies generally fulfill the temporal ordering criteria (i.e., the cause precedes the effect)" (Commission on Behavioral and Social Sciences and Education, 1998, p. 15). However, because this was not a true experiment, we cannot categorically attribute all of the increases in the earning of the two diplomas to detracking alone. While we attempted to account for other factors in the previous section, some of the increases in the probability of earning Regents and IB diplomas could conceivably reflect other, unaccounted-for influences not considered by the authors.

Another potential limitation of this case study concerns the generalizability of the findings reported here. The district that implemented this reform has historically allocated generous resources to students who struggle and has the resources to attract highly qualified teachers. Yet many of the key resources existed both before and after the detracking reform. Both before and after detracking there were support classes,

extra help periods, and a highly qualified faculty. Prior to detracking such support was not enough. However, the combination of detracking and support was likely an important factor in the success of this reform. A reasonable deduction is that replication of its success should include both elements.

Another key to the implementation of such a reform concerns values and commitment. A successful equity-minded reform, such as the one described in this article, depends on school leaders' willingness to challenge longstanding practices and assumptions (Sirotnik & Oakes, 1986). Within the district there were shifts in beliefs, curriculum, pedagogy and school culture, changes that accompanied the mechanics of detracking and that educators at the school have seen as essential to the growth in both Regents and IB diplomas. While an explanation of the role of all of these factors in a detracking reform is beyond the scope of this study, it would be incorrect to assume that achievement gains will be realized simply by eliminating tracks. Educators need to sincerely hold and communicate a belief—supported by this research—that many more students can achieve the highest levels if they have the proper curriculum, teaching, and support.

The district's commitment to reform also manifested itself in the reform's breadth. The detracking reform was part of a long-term district strategy. There is only one middle school in the district, and that middle school is now also detracked. One might imagine that implementation of this reform in a large district with several feeder middle schools would be more difficult and would require additional strategies for success.

Taken together, this district's experience will be most generalizable to districts that share basic values, and that are willing to challenge traditional perspectives and attitudes regarding so-called "ability"[12] and learning. Also needed are the resources that must be dedicated in order to provide support to faculty, students, and even parents.

Would detracking be as effective in a district with fewer resources to support struggling students, fewer qualified teachers, or in a district in which more students struggle academically? Certainly such conditions would make the reform more difficult to implement. However, in our opinion, such challenges can be at least partially overcome. Implementation will differ in each new context. Gains may even be reduced. But there is little reason to believe that districts with greater numbers of poor students would not gain achievement benefits from comparable detracking initiatives. In fact, a recently completed longitudinal study in an urban American high school with a far greater proportion of students from low-income households shows results that are remarkably similar to those of this study—when detracking was com-

bined with a common rigorous curriculum in mathematics, stude achievement increased (Boaler & Staples, this issue). Similar results were found in a longitudinal study of two high schools in England (Boaler, 2002).

CONCLUSION

The title of a 1999 article by a prominent advocate of tracking asks, *Will Tracking Reform Promote Social Equity?* (Loveless, 1999a). The Loveless article challenges researchers to demonstrate two benefits of detracking— that reformed schools close the gap between education's "haves" and "have nots" and, importantly, that they do so without adversely affecting the learning of the "haves."

Responding to the first element of that challenge, this article offers a case study showing a dramatically closed achievement gap in the earning of the Regents diploma. Detracking by itself cannot ameliorate social inequities such as poverty, inadequate health, and under-funded urban schools – inequities known to have a deleterious effect on student achievement (Berliner, 2005; Rothstein, 2004). Yet schools also can do a great deal to provide all students with fair access to the best curricula, teachers and instruction that they have to offer. Numerous studies, both past and present, tell us that the best resources are usually dedicated to schools' high track classes (Haycock, 2000; Oakes, 2005; Oakes, et al., 1990). Within a particular school, then, detracking reform can address inequities in educational opportunity.

Moreover, responding to the second element of Loveless' challenge, this longitudinal case study demonstrates that a well-executed detracking reform can help increasing numbers of students reach state and world-class standards without adversely affecting high-achieving students. As shown in Table 2, the percentage of students who scored the highest scores on IB exams (5, 6 and 7) *increased* even as the enrollment in IB classes expanded from an exclusive gifted program to a program for the majority of students.

The findings from this study should help to alleviate the concerns of those who fear that high achievers will learn less if they are placed in classes with low-achieving students and that lower achievers will be frustrated when given high-track curriculum (Brewer, Rees, & Argys, 1995; Kulik, 1992; Loveless, 1998, 1999a). Our findings with regard to those students who had been lower achievers are consistent with what researchers know about the potential of accelerated curriculum and the damage done by rote curriculum (Adelman, 1999; AERA, 2004; Levin, 1997; Singham, 2003). These findings are noteworthy nevertheless, because of

the impressive, documented improvement in their academic outcomes. We are not naïve enough, however, to fail to recognize that, as a political and policy matter, the more important finding of this study is the continued success of the students who had been high-achievers. As evidenced by their performance on IB examinations, and in the earning of the IB diploma, high achievers continue to successfully meet international standards. Case studies such as this, documenting success and grounded in carefully collected and analyzed data, are now emerging and should give confidence to future reformers (Boaler & Staples, in this issue).

Whether measures of accountability are established by states or by federal legislation, such as NCLB, educators are currently presented with the challenge of helping all students, including low-achieving students, meet high learning standards. Increased student achievement is possible, but will not happen without improvements in classroom instruction. Darling-Hammond (2003) concludes that, in order to meet challenging accountability goals, American students must have access to high-quality curriculum, teaching, and resources (see also Wells & Oakes, 1996). Overall adequacy of resources, however, is no more important than the distribution of those resources. The case study presented here shows that combining high-track curriculum with detracked classes can have a positive impact on helping students achieve on measures that matter.

Notes

1   The authors of this article include the high school's principal and the assistant principal and facilitator of the school's IB program.
2   This is the cohort of students who would normally graduate in June 2003. We use ninth-grade YOE to identify the students because a small number of students take four or five years to graduate. Because our concern is the particular phase of the detracking reform that a student experienced, the YOE approach is the most accurate.
3   The high school has a program for developmentally delayed students who receive an IEP certificate and exit high school at the age of 21. These students have a specially designed program that combines academics with job skill training. They were not included in this study. The number of such students who are developmentally delayed is small—typically fewer than four students per cohort.
4   In order to include data from the final year, we used the old, more rigorous standard (as described in the main text) for those cohort members.
5   A five-year sequence in the arts or business may be substituted for the courses and examination in foreign language.
6   For an extensive discussion of the grading practices of the IB program see Diploma Programme assessments: Principles and practice available online at: http://web3.ibo.org /ibis/documents/dp/d_x_dpyyy_ass_0409_1_e.pdf
7   The IBO, which was established in 1968, is a non-profit 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).

8   African American and Latino students are nationally under-represented in high-track classes (Oakes, Gamoran & Page, 1992).

9   The GROUP4 variable actually serves as the null case and is not entered into any statistical model.

10   That is, the chi-square change, with 7 degrees of freedom, is 26.09, which is significant at the 0.001 level. See the final two rows of Table 5, under the Model 2 column.

11   In 2004, the high school Regents diploma rate of 94% was *the highest Regents diploma rate of the 97 high schools in Group 54*. Comparative data for schools in New York State can be found by accessing databases and reports available at: http://www.emsc.nysed.gov/irts/reportcard

12   In their 1999 essay entitled, Access to knowledge: Challenging the techniques, norms, and politics of schooling, Oakes and Lipton argue that terms such as *ability* are merely human constructs and do not represent fixed, objective measures of human potential for learning.

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns and bachelor's degree attainment.* Washington, DC: U.S. Department of Education, Office of Educational Research. Retrieved from http://www.ed. gov/pubs/Toolbox.

AERA. (2004, Fall). Closing the gap: High achievement for students of color. *Research Points, 2*(3). Available online at http://www.aera.net/uploadedFiles/Journals and Publications/Research_Points/RPFall-04.pdf

Berger, J. (2000). Does top-down, standards-based reform work? A review of the status of statewide, standards-based reform. *NASSP Bulletin, 84,* 57-65.

Berliner, D.C. (2005). Our impoverished view of educational reform. *Teachers College Record,* Date Published: August 02, 2005. http://www.tcrecord.org; ID Number: 12106; Date Accessed: 8/24/2005.

Black, S. (1992). On the wrong track. *Executive Educator, 14*(12), 46–49.

Bloom, H.S., Ham, S., Melton, L., & O'Brient, J. (2001). *Evaluating the accelerated schools approach: A look at early implementation and impacts on student achievement in eight elementary schools.* New York: Manpower Demonstration Research Corporation.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Braddock, J.H., II, & Dawkins, M.P. (1993). Ability grouping, aspirations, and attainment: Evidence from the National Educational Longitudinal Study of 1988. *Journal of Negro Education, 62,* 324–336.

Burris, C.C., & Welner, K. G. (2005). Closing the achievement gap by detracking. *Phi Delta Kappan, 86*(8), 594–598.

Brewer, D.J., Rees, D.I., & Argys, L.M. (1995). Detracking America's schools: The reform without cost? *Phi Delta Kappan, 77,* 210–212, 214–215.

Brewer, D.J., Rees, D.I., & Argys, L.M. (1996). The reform without costs? A reply to our critics. *Phi Delta Kappan, 77,* 442–443.

Burris, C.C., Heubert, J., & Levin, H. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal,43*(1), 103–134.

Cabrera, A.F., & Bernal, E.M. (1998). *Association between SES and ethnicity across three national databases.* University Park, PA: Center for the Study of Higher Education.

Commission on Behavioral and Social Sciences and Education. (1998). *Work-related musculoskeletal disorders: A review of the evidence.* Washington, DC: National Academy Press.

Conkin, K.D., & Curran, B.A. (2005). *National education summit on high schools: An action agenda for improving America's high schools.* Washington, DC: Achieve Inc.

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: Jossey-Bass.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record.* Retrieved March 5, 2003 from http://www.tcrecord.org (ID Number: 11109).

Dillon, W.R., & Goldstein, M. (1984). *Multivariate analysis, methods and Applications.* New York: Wiley.

Duevel, L. M. (2000). *The International Baccalaureate experience: University perseverance, attainment, and perspectives on the process.* Dissertation Abstracts International -A 60/11, p. 3852.

Epple, D., Newlon, E., & Romano, R. (2002). Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics, 83*(1), 1–48.

Epstein, J. L., & MacIver, D. J. (1992). *Opportunities to learn: Effects on eighth graders of curriculum offerings and instructional approaches* (Report No. 34). Washington, DC: Office of Educational Research and Improvement.

Figlio, D.N., & Page, M. E. (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.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59,* 185–198.

Gamoran, A. (1992). Synthesis of research: Is ability grouping equitable? *Educational Leadership, 50*(2), 11–17.

Gamoran, A., & Hannigan, E.C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis, 94*(3), 241–254.

Gamoran, A., & Mare, R.D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement or neutrality? *American Journal of Sociology, 94,* 1146–1183.

Gamoran, A., & Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education, 106,* (3), 385–415.

Garet, M.S., & Delany, B. (1988). Students, courses, and stratification. *Sociology of Education, 61,* 61–77.

George, P. (1992). *How to untrack your school.* Alexandria, VA: Association for Supervision and Curriculum Development.

Goff, G.N. (1995). Assessing the impact of tracking on individual growth in mathematics achievement using random coefficient modeling. *Dissertation Abstracts International, 56*(03), 855. (University Microfilms No. 9523572).

Goycochea, B.B. (2000). College prep mathematics in secondary schools: Access denied. *Dissertations Abstracts International, 61*(4), p. 1333.

Hallinan, M.T. (1992). The organization of students for instruction in the middle school. *Sociology of Education, 65*(2), 114–127.

Hallinan, M.T. (1994). Tracking from theory to practice. Exchange. *Sociology of Education, 57*(2), 79–84.

Hallinan, M.T., & Sorensen, A.B. (1987). Ability grouping and sex differences in mathematics achievement. *Sociology of Education, 60*(2), 63–72.

Haycock, K. (2000). Honor in the boxcar: Equalizing teacher quality. In P. Barth (Ed.), *Teaching K-12 4*(1). Washington, DC: Education Trust.

Heubert, J.P., & Hauser, R.M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Research Council.

International Baccalaureate Organization. (2005). *Education for life.* Retrieved March 26, 2005 from: http://www.ibo.org/ibo/index.cfm.

Kerckhoff, A.C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*(6), 842–58.

Kliebard, H.M. (1995). *The struggle for the American curriculum: 1893-1958* (2nd ed.). New York: Routledge.

Kulik, J.A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives.* Storrs, CT: National Center of the Gifted and Talented.

Levin, H.M. (1997). Raising school productivity: An x-efficiency approach. *Economics of Education Review, 16*(3), pp. 303–12.

Lipman, P. (1998). *Race, class and power in school restructuring.* New York: SUNY Press.

Loveless, T. (1998). The tracking and ability grouping debate. *Thomas B. Fordham Foundation, 2*(8). Retrieved July, 1999 from http://www.edexcellence.net/library/track.html.

Loveless, T. (1999a). Will tracking reform promote social equity? *Phi Delta Kappan, 56*(7), 26–32.

Loveless, T. (1999b). *The Tracking Wars: State Reform Meets School Policy.* Washington DC: Brookings Institution Press.

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

Lucas, S.R., & Gamoran, A. (1993). *Race and track assignment: A reconsideration with course-based indicators of track location.* Washington, DC: Office of Educational Research and Improvement.

Mehan, H., Villanueva, I., Hubbard, L., & Lintz, A. (1996). *Constructing school success: The consequences of untracking low-achieving students.* New York: Cambridge University Press.

Mills, R.P. (2004). *New York: The state of learning: A report to the governor and the legislature on the educational status of the state's schools.* Albany: New York State Education Department. Retrieved December 2005 from http://www.emsc.nysed.gov/irts/655report/2004/Volume1/combined_report.pdf

Mosteller, F., Light, R.J., & Sachs, J.A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review, 66,* 797–843.

Natriello, G., & Pallas, A.M. (1999). *The development and impact of high stakes testing.* (ERIC Document Reproduction Service No. ED 443 871).

New York State Board of Regents. (2003). *What is a similar school? Albany: New York State Education Department.* Retrieved December 2005 from http://emsc33.nysed.gov/repcrd2003/information/similar-schools/guide.html

Oakes, J. (1982). The reproduction of inequity: The content of secondary school tracking. *Urban Review, 14*(2), 107–120.

Oakes, J. (1986). Keeping track, Part 1: The policy and practice of curriculum inequality. *Phi Delta Kappan, 68,* 12–18.

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

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J., Gamoran, A., & Page, R. (1992). Curriculum differentiation, opportunities, outcomes, and meanings. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570–607). New York: Maxwell Macmillan International.

Oakes, J., Ormseth, T., Bell, R., & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.* Santa Monica, CA: Rand.

Peter D. Hart Research Associates. (2005). *Rising to the challenge: Are high school students prepared to work?* Washington, DC: Author/Public Opinion Strategies. Available online at http://www.achieve.org/achieve.nsf/StandardForm3?openform&parentunid=ABC3A652CD3B736785256F9E00783B86.

Peterson, J.M. (1989). Remediation is no remedy. *Educational Leadership, 46*(6), 2425.

Ravitch, D. (2000). *Left back: A century of failed school reforms.* New York: Simon & Schuster.

Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap.* Washington, DC: Economic Policy Institute.

Singham, M. (2003). The achievement gap: Myths and reality. *Phi Delta Kappan, 84*(8), 586–591.

Slavin, R.E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60,* 471–499.

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

Sandholtz, J.H., Ogawa, R.T., & Scribner, S.P. (2004). Standards gaps: Unintended consequences of local standards. *Teachers College Record, 106*(6), 1177–1202.

Sirotnik, K., & Oakes, J. (1986). Critical inquiry for school renewal: Liberating theory and practice. In K. A. Sirotnik & J. Oakes (Eds.), *Critical perspectives on the organization and improvement of schooling* (pp. 3–94). Hingham, MA: Kluwer-Nijhoff Publishing.

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82*(5), 358–62.

Useem, E.L. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. *American Journal of Education, 100,* 325–353.

Vanfossen, B.E., Jones, J. D., & Spade, J.Z. (1987). Curriculum tracking and status maintenance. *Sociology of Education, 60,* 104–122.

Wells, A.S., & Oakes, J. (1996). Potential pitfalls of systemic reform: Early lessons from research on detracking. *Sociology of Education, 69,* 135–143.

Wells, A.S., & Serna, I. (1996). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review, 66*(1), 93–118.

Welner, K.G. (2001a). *Legal rights, local wrongs: When community control collides with educational equity.* Albany: SUNY Press.

Welner, K.G. (2001b). Tracking in an era of standards: Low-expectation classes meet high-expectation laws. *Hastings Constitutional Law Quarterly, 28*(3), 699–738.

Wheelock, A. (1992). *Crossing the tracks: How "untracking" can save America's schools.* New York: New Press.

White, P., Gamoran, A., Porter, A. C., & Smithson, J. (1996). Upgrading the high school math curriculum: Math course-taking patterns in seven high schools in California and New York. *Educational Evaluation and Policy Analysis, 18,* 285–307.

CAROL CORBETT BURRIS is the principal of South Side High School, a diverse suburban high school located on Long Island, approximately 25 miles from New York City. Dr. Burris' research interests are detracking and educational equity. Recent publications include: Accelerating Mathematics Achievement Using Heterogeneous Grouping which she co-authored with Jay P. Heubert and Henry M. Levin (*American Educational Research Journal* Spring, 2006, volume 43, no. 1) and Closing the Achievement Gap by Detracking, which she co-authored with Kevin G. Welner (*Phi Delta Kappan*, April, 2005).

ED WILEY is Chair of the Research and Evaluation Methodology Program at the University of Colorado at Boulder, where he also serves as assistant professor of quantitative methods in educational policy. His policy interests include school accountability and teacher quality and compensation; his methodological research focuses on nonparametric and computational statistics. He has recently completed a practitioner's guide to value-added assessment and is currently leading a study of Denver Public School's "ProComp" (Professional Compensation System for Teachers) reform.

KEVIN WELNER is associate professor at the University of Colorado, Boulder School of Education, specializing in educational policy, law, and program evaluation. He is director of the CU-Boulder Education and the Public Interest Center (EPIC). A former attorney, Dr. Welner's research examines the intersection between education rights litigation and educational opportunity scholarship. His current research includes studies focusing on small-school reform, detracking, school choice, and tuition tax credits. His recent work includes two books, *Under the Voucher Radar: The Emergence of Tuition Tax Credits for Private Schooling* (Rowman & Littlefield, forthcoming), and *Small Doses of Arsenic: A Bohemian Woman's Story of Survival* (with Sylvia Welner, Hamilton Books, 2005).

JOHN MURPHY is the assistant principal and IB coordinator at South Side High School in Rockville Centre, NY. He currently teaches Theory of Knowledge, and has taught English for grades 7-12, including IB English. His research interests include differentiated instruction, Multiple Intelligence Theory, and authentic assessment.