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.*

Teachers College Record Volume 110, Number 3, March 2008, pp. 571–607

0161-4681

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-n.Sth lunch, to a 26Latino students who were eli-*

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

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

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 com-

beneficial outcomes: (a) ameliorate the racial and socio-economic stratification 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 combined with high-track curricula on two achievement measures of importance. We describe the results of a longitudinal study of the effects on student 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

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

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.

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

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.

3. E C

		Low SES		
		NO	YES	Total
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.

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 par-.0udeownf predictors isto achie odr-s

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.

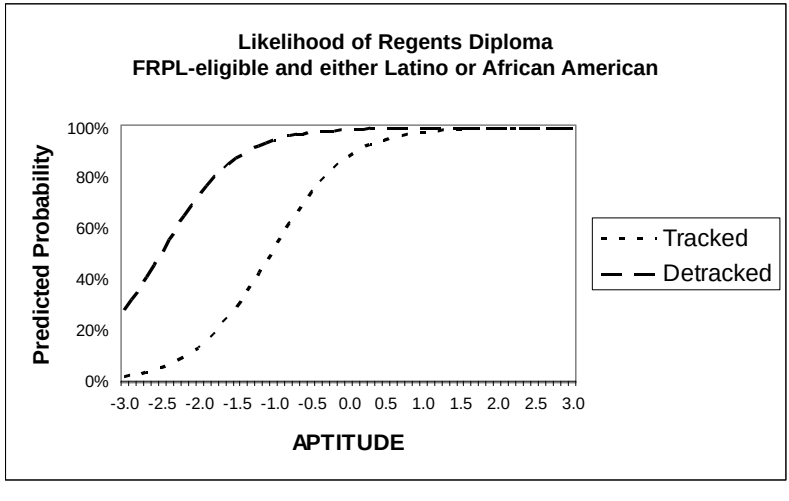
The inference regarding detracking is clear from Model 7 (and is strikingly consistent across all models): being a member of a detracked cohort is associated with substantial increases in the odds of attaining the Regents diploma. For students in Group 2 (non-minority, FRPL eligible) and Group 4 (non-minority, 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).

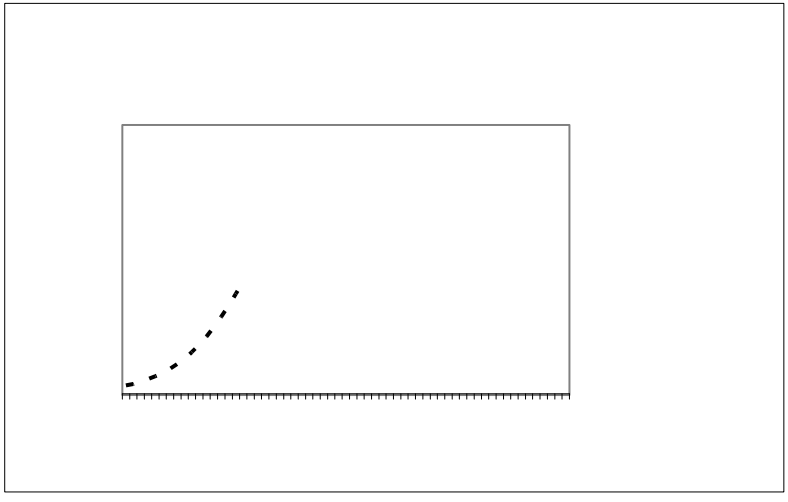
The following illustration, based on coefficient odds ratios, helps to place the magnitude of the effect associated with detracking in context. The odds ratio for PREPOST (3.35) is nearly half as large (47%) as the odds ratio of APTITUDE (7.07). As such, for those students in GROUP2 and GROUP4, being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

F ~ 1.





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

Coefficients	Model 1			Model 2			Model 3			Model 4			Model 5
	B	se(B)	exp(B)	B	se(B)	exp(B)	B	se(B)	exp(B)	B	se(B)	exp(B)	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***
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***
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
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
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
SPED by GROUP1				1.07	431.688	2.90	431.205	431.205	2.15	2.08	########	8.03	1.8
SPED by GROUP2				-1.18	887.032	0.31	878.245	878.245	0.26	0.57	1.5E+03	1.76	

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 PREPOST 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 APTITUDE. 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. 10j.9 -1.2011 TD-.02 e nea

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

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

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]

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-

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 stan-

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

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,*

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.