Exploring the Relationship Between Content Coverage and Achievement:

Unpacking the Meaning of Tracking in Eighth Grade Mathematics

The Education Policy Center
at Michigan State University

Exploring the Relationship Between Content Coverage and Achievement:

Unpacking the Meaning of Tracking in Eighth Grade Mathematics

William H. Schmidt
Michigan State University

The Education Policy Center
at Michigan State University

TABLE OF CONTENTS

INTRODUCTION

In the wake of the No Child Left Behind Act of 2001, issues of suitable standards for all children and equitable access to adequate learning opportunities have acquired a new urgency in education reform deliberations. States and individual districts are being compelled to make explicit what it means to have high standards for all children and what it means for all children to have equitable opportunities to learn necessary, important, and challenging content (Achieve, 2002b).

These issues of equitable learning opportunities and challenging standards are visible nowhere more keenly than with the case of eighth grade mathematics. The great need in this area is shown, at least in part, by the mathematics performance of U.S. eighth grade students. This performance has been characterized as "lackluster" and "just not good enough" (National Commission on Mathematics and Science Teaching for the 21st Century, 2000; Riley, 1996; Schmidt et al., 1999). Given the lack of focused, coherent, and challenging standards for all eighth grade mathematics students and the somewhat "splintered vision" that appears to inform classroom instruction, this type of student performance is not surprising (Schmidt, McKnight, & Raizen, 1997). This situation may also explain the choice of eighth grade mathematics as the first area of concerted effort and focus by Achieve, Inc., an independent organization to help states raise academic standards (Achieve, 2002a).

One possible explanation for this "lackluster" mathematics performance is the widespread use of tracking in U.S. middle and secondary schools—a process that was found to be relatively rare across the more than 40 countries involved in the Third International Mathematics and Science Study (TIMSS) (Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999).

Tracking in the United States has had an amorphous history, meaning different things to different people at different times (Oakes, 1985). At one point, tracking implied dividing secondary students into rigid curricular programs (e.g., college-preparatory, general, vocational) that spanned all academic subjects (Lucas, 1999; Oakes, 1985). Today, such school-wide curricular programs are rarely overt aspects of school policy. This does not mean, however, that schools do not track students—most do. Rather, instead of overarching curricular programs that keep students in the same track across subjects, both secondary and middle schools now differentiate students within subjects (Lucas, 1999). This implies, for example, that two students in eighth grade taking mathematics may be in two substantively different mathematics classes such as basic arithmetic and algebra. Although the curricular level of one class is often associated with the curricular levels of a student's other classes, tracking can best be understood by examining the specific courses that students take (Friedkin & Thomas, 1997; Heck, Price, & Thomas, 2004; Lucas, 1999; Lucas & Berends, 2002; Stevenson, Schiller, & Schneider, 1994).

Tracking in mathematics, therefore, is considered to be the provision of substantively different mathematics content or curriculum to different students at the same grade level. Tracking is differentiated from ability grouping, where the content is common but the instructional approach, such as the pacing and depth of instruction, may differ. By definition, then, tracking provides different students different opportunities to learn mathematics content.

Tracking in mathematics, as it is typically conceived, implies that some students will eventually have an opportunity to learn advanced mathematics content and some will not. Advocates of tracking argue that this type of curricular differentiation facilitates teaching and learning, as it matches students' ability level to the most suitable curriculum. Tracking theory contends that some students would struggle immensely in high-level curricula while a low-level curriculum would confine others. Tracking, therefore, allows students to be placed into classes where they will—theoretically—make the greatest achievement gains. In turn, this theory posits that tracking, compared to non-tracking, increases overall student achievement and lessens achievement inequality (Gamoran, 1992).

Most research on tracking (focusing only on secondary schooling), however, has found that differentiating the mathematics curriculum tends to adversely affect students in low-level courses compared to their high-tracked peers (Gamoran, 1987; Gamoran & Berends, 1987; Gamoran & Mare, 1989; Gamoran, Porter, Smithson, & White, 1997; Hallinan & Kubistchek, 1999; Hoffer, 1992; Lucas, 1999; Oakes, 1985, 1990; Schneider, Swanson, & Riegle-Crumb, 1998; Stevenson, Schiller, & Schneider, 1994). Students in low-tracked mathematics courses are less likely to expect to go to college, (Alexander, Cook, & McDill, 1978; Alexander & Cook, 1982; Alexander & McDill, 1976; Heyns, 1974; Vanfossen, Jones, & Spade, 1987), less likely to actually attend college, even after controlling for students' post-secondary expectations (Alexander & Eckland, 1975; Rosenbaum, 1980; Vanfossen et al., 1987), and have lower self-images (Alexander & McDill, 1976; Oakes, 1985; Vanfossen et al., 1987). Perhaps most salient, though, is that many studies have found that mathematics tracking tends to exacerbate achievement inequalities between high- and low-tracked students (Gamoran, 1987; Gamoran & Mare, 1989; Hallinan & Kubistchek, 1999; Hoffer, 1992; Ma, 2000; Schneider et al., 1998; Stevenson et al., 1994).

Many studies analyzing the effect of tracking on achievement, however, have had several limitations. To begin, studies using large, nationally representative data sets such as the National Education Longitudinal Study (NELS) or High School and Beyond (HSB)—which provide the data for most of the research done in this area—have for one thing used students' self-reports to indicate track location. This can be problematic, though, as students may be in different curricular track-levels depending on the academic subject. How, then, should students respond to a survey question about their track if they are in a high-tracked mathematics course but a low-tracked English course? Moreover, self-reports assume that students' perceptions of curricular tracks are similar to those of school personnel or researchers. However, Rosenbaum (1980) found that the correlation between students' perceptions of their track and their actual track was only 0.60. Later evidence found that 19.7% of the students in Rosenbaum's study misperceived their placement[1] (Fennessey, Alexander, Riordan, & Salganik, 1981). The second and more serious limitation is that the sampling plan by which students are chosen within the sampled school is not properly defined for studying the issue of tracking. [1]

Research on mathematics tracking and achievement has also mostly focused on high school students. But tracking typically begins—especially in mathematics—during the middle grades (Dauber, Alexander, & Entwisle, 1996; Hallinan, 1992; Useem, 1992). Consequently, studies focusing solely on high school tracking may mask tracking's earlier achievement effects. And although ability grouping in elementary schools may affect a student's subsequent middle school track-location, it is in the middle grades where formal curricular differentiation typically begins. It is thus important to analyze how tracking is related to content coverage and to unpack the relationship of both to the achievement of middle-school students, as these grades are the primary root of U.S. mathematics tracking.

In this paper we address these limitations by examining the effect of tracking on eighth grade mathematics achievement and by defining track location in terms of the actual mathematics content that students covered in that track. We have previously documented the extent of the variation that passes for eighth grade mathematics in terms of course titles, textbooks used, amount of instructional time devoted to specific topics, and the relative difficulty of courses from an international perspective (Cogan, Schmidt & Wiley, 2001). Here we report on a set of analyses designed explicitly to explore the effect of tracking on achievement. Using a unique nationally representative sample of seventh and eighth grade students that allows us to overcome, to some extent, the limitations due to within school sampling (by providing a full characterization of tracking within the sampled schools), we explore the relationship of tracking in eighth grade to what mathematics topics are studied during eighth grade (content exposure) and to what is learned during the year as well as to what is achieved by the end of eighth grade.

---

1 Lucas and Gamoran (2002) have shown that students' self-reports affect mathematics achievement independent of their actual course enrollments. They therefore suggest that it is "unwise to dismiss self-reports as merely perceptual" (p. 175). Instead, they suggest that self-reports may partly measure the social-psychological dimension of tracking. This dimension includes students' attitudes, values, and acceptance of school. Nevertheless, Lucas and Gamoran agree that self-reports are weaker measures of the structural dimension of tracking than actual course enrollments.

BACKGROUND

Mathematics tracking is commonly practiced in the eighth grade (Cogan, Schmidt & Wiley, 2001; Hoffer, 1992). Consequently, eighth grade students often take one of several mathematics-course options, typically algebra, pre-algebra, or general math (Cogan, Schmidt & Wiley, 2001). Which course a student takes, however, is not inconsequential. Each course presents a substantively different curriculum, and in turn affects students' achievement differently. Prior studies have highlighted two important ways that a student's eighth grade course affects his or her subsequent mathematics achievement: positional advantages and differential achievement growth.

Eighth-Grade Tracking and Positional Advantages

A student's eighth grade mathematics course affects which mathematics courses he or she will take in high school (Atanda, 1999; McFarland, 2006; Stevenson et al., 1994). Because mathematics is typically presented as a linear sequence, where entry into one class is contingent on successfully completing the antecedent class (e.g. Algebra I and Algebra II), a student's eighth grade course largely feeds into his or her ninth grade course. Students taking pre-algebra in eighth grade tend to take algebra in ninth grade. Similarly, students taking algebra in eighth grade tend to take geometry in ninth grade (McFarland, 2006). These ninth grade courses then affect students' tenth grade courses, and so on, ultimately affecting "how far" a student gets in the mathematics curriculum (McDonnell, 1995; Riegle-Crumb, 2006; Schneider et al., 1998; Stevenson et al., 1994). Of course, in the absence of a national mathematics curriculum, these course sequences can vary by school (for example, see McFarland, 2006); the hierarchical nature of mathematics, however, tends to produce/result in similar course sequences across the U.S. (Stevenson et al., 1994).

One's eighth grade math course can therefore be viewed as a position in a long sequence of courses (Schneider et al., 1998; Stevenson et al., 1994). Certain positions facilitate students' entry into advanced mathematics courses such as calculus or trigonometry. Other positions may only take students to Algebra II. Some students, then, have a positional advantage, as their position allows them to take higher-level mathematics courses—which in turn significantly increases their chance to attend college (Adelman, 1999, 2003).

Rosenbaum (1978) called the progression through course sequences a tournament, where students in advantaged positions (e.g. taking algebra in the eighth grade) can "win" or "lose." Winning entails maintaining one's position, whereas losing represents a relegation to a lower position. Moreover, once a student is relegated, he or she cannot regain the advantaged position, and thus "the tournament is over" and they "lose forever" (Rosenbaum, 1978, pp. 252). Subsequent research, however, has shown that track positions are not as inflexible as Rosenbaum's tournament (Hallinan, 1996; Lucas & Good, 2001; McFarland, 2006). Hallinan (1996) found that some upward mobility does exist in mathematics tracks, and thus students can regain a positional advantage. Nevertheless, the track a student begins in, generally in the seventh or eighth grade, is a significant predictor of where he or she will be in twelfth grade (Hallinan, 1996; McFarland, 2006; Schneider et al., 1998; Stevenson et al., 1994). Further, students who have access to algebra before high school—typically taking it in eighth grade—have significantly greater access to advanced twelfth grade mathematics courses, even after controlling for tenth grade track location, mathematics achievement, and educational aspirations (Smith, 1996). Tracking in the eighth grade thus sets in motion high school tracking, which in turn affects college attendance; those students who have a positional advantage in eighth grade are more likely to have a similarly advantaged position at the end of high school.

Eighth-Grade Tracking and Differential Achievement Growth

The fact that mathematics tracking causes students to end up in different curricular locations is not incredibly surprising. After all, a main purpose of tracking is to allocate students into the courses that best match their ability level. Tracking theory posits that these matches will improve students' achievement growth compared to non-tracked courses. But most studies of secondary tracking have shown that tracking tends to have the opposite effect, actually increasing achievement inequalities. This tendency is no less true at the eighth grade, although it is not a consensus.

Research on the effect of tracking on achievement growth generally takes two forms: one compares differences in achievement growth between tracked and non-tracked schools, while the other compares achievement growth between different track-levels. The results of research comparing tracked and non-tracked schools in the middle grades have been mixed. On one hand, Hoffer (1992) found that non-tracked students tend to make greater achievement gains than their tracked peers, although this difference was not statistically significant. On the other hand, in a meta-analysis of 27 studies, Slavin (1993) found the effects of tracking on student achievement in the middle grades to be essentially zero. Slavin did not limit his studies to mathematics achievement, however, nor did he differentiate between ability grouping and tracking.

This does not mean, however, that tracking similarly has a small to negligible effect on all students. Studies comparing tracked schools to non-tracked schools typically aggregate student achievement to the school level and thus mask how individual students within the tracked schools may be differentially affected. Indeed, many studies have found that the effects of tracking in the middle grades depend on the track-level. Using the nationally representative NELS date set, Hallinan and Kubistchek

(1999) found that students in a high-tracked eighth grade mathematics class made significantly more achievement gains by the time they reached tenth grade than their general-tracked peers. Students in the lowest track, however, achieved significantly less than general-track students. Similarly, Hoffer (1992) found that eighth grade students in a high-tracked mathematics course achieved significantly more and low-tracked students achieved significantly less than non-tracked students, even after controlling for seventh grade achievement. In general, most studies of within-school tracking in eighth grade mathematics have found higher rates of achievement growth among high-tracked students than low-tracked students.

Positional advantages and differential achievement growth are not a dichotomy, however, in that achievement is often used to make decisions about future track placement (Useem, 1992). This is especially true when students face curricular turning points, where students can no longer remain in the same track and must choose to move upward or downward (McFarland, 2006). Early tracking decisions, such as those in the eighth grade, can therefore not only disadvantage students in low tracks by relegating them to disadvantaged curricular positions and diminishing their achievement growth, but by also making entry into a high-tracked course less likely.

How Tracking Affects Achievement

Many studies have highlighted the effects of tracking on achievement but few have examined how these effects occur. Three different kinds of mechanisms have been identified that may produce the effects of tracking: social, institutional, and instructional (Gamoran & Berends, 1987; Lucas, 1999). We focus in this paper on instructional effects—specifically on the effect of content coverage.

Perhaps the most salient instructional aspect of tracking is the fact that tracking differentiates students' content exposure. This is of course expected, as the purpose of tracking is to separate students into substantively different courses. Many studies, primarily at the high school level, have found that it is precisely because students take different courses that tracking is associated with increased achievement inequality (Gamoran, 1987; Gamoran et al., 1997; Ma, 2000; Rock & Pollack, 1995; Schneider et al., 1998; Stevenson et al., 1994). These different courses provide students with different opportunities to learn specific mathematics content. Many studies posit that the curricular opportunities available in high-tracked classes facilitate higher achievement rates.

Little is known, however, about the substantive nature of these differences in opportunities to learn or how they mediate the effects of mathematics tracking—especially in the eighth grade. Some studies have analyzed how different mathematics course sequences may differentially affect achievement. But this can be problematic. There is tremendous content diversity in U.S. eighth grade mathematics courses with the same title, diversity which makes any attempt to define what might actually be standard for all students difficult (Cogan, Schmidt & Wiley, 2001). Eighth grade algebra in one school may present different content opportunities than eighth grade algebra in another school.

Hardly any studies, however, have closely examined how specific mathematics content within the classes mediates the effects of tracking; virtually all large-scale, nationally representative data sets lack this information.[ii]

A study of 48 mathematics classes across seven urban high schools (Gamoran et al., 1997) illuminated how differences in course content mediate the achievement effects of tracking. It, however, focused on the match of content between instruction and test and not the level of the content itself and was isolated to urban schools and was consequently non-nationally representative. Further, it focused on students in grades 9-12.

Data from TIMSS allow us a unique opportunity to address these limitations. TIMSS includes a large, nationally representative sample, actual course indicators and achievement results; perhaps most importantly, within course content coverage measures were collected. These data allow us to address the following questions: how does content coverage differ between tracked and non-tracked schools? How does content coverage vary by track position? How does achievement vary by track location and between tracked and non-tracked schools? Lastly, how does a track's content coverage affect achievement? These questions allow us to explore the role of instructional content on the effects of tracking on achievement at a pivotal point in a student's educational career, namely eighth grade mathematics.

DATA SOURCES AND METHODOLOGY

The data collected as part of the U.S. participation in the 1995 TIMSS were used in this study. The TIMSS represents an opportunity to explore these issues with a nationally representative sample of seventh and eighth grade students that employs a common framework for describing the mathematics content covered in each of the sampled classrooms. The TIMSS sample is unique for the purpose of exploring issues related to tracking because of the within school sampling frame which provides detailed within-school tracking information for all classes and students within the school—not just for those classrooms which were drawn to be in the U.S. sample for TIMSS. This provides a detailed characterization of the tracking structure within each of the randomly sampled schools. In that sense it represents a random sample of the structure of tracking, proportional to the size of the eighth grade population.[3] In the U.S. a representative sample of more than 13,000 students in 183 schools was tested and surveyed and their teachers completed lengthy questionnaires about the content of their instruction.[2]

2 After the U.S. as a whole had been stratified, schools were sampled within

Course Differentiation

The TIMSS sample was designed to be representative of the U.S. as a whole but was not explicitly stratified to deal with the widespread tracking policies of many schools. This is a limitation inherent in most other U.S. eighth grade samples as well and is why it is difficult to study the tracking question empirically in a totally suitable way at the national level. The 1995 TIMSS data, although somewhat dated, provide perhaps the only opportunity at this point to adequately explore tracking given our access to the within school class sampling frame, which is not available (at least publicly) for any other nationally representative data sets. Also, combined with the availability of detailed teacher reports of content coverage it is essentially the only data set available for exploring these issues.

As part of the within-school sampling procedure, schools listed all of their seventh and eighth grade mathematics classrooms along with the class titles and the list of the students enrolled in each class. This was used to draw the sample but it also provides complete tracking information for the sampled schools. Using this information, it was possible to specify the within-school course-offering structures (Cogan, Schmidt & Wiley, 2001). It is that within-school course offering structure that defines tracks for purposes of this paper. More than 25 different patterns of school course offerings were identified in the sample, based on six types of classes: remedial, regular, pre-algebra, enriched, algebra, and geometry. Each of the six types of classes defines a track in the sense of providing different content opportunities to learn mathematics. The actual number of tracks is probably much larger.[3]

As very few remedial, enriched, or geometry classes appeared in the actual

---

strata with probabilities proportional to size. TIMSS sampled two eighth grade and one seventh grade mathematics classroom in each selected school and weights were assigned to each student. Sampled student weights sum to the number of U.S. students in the population at that grade (Foy, Rust, & Schleicher, 1996). This procedure yielded a sample of 7087 eighth grade students in 183 schools and 3886 seventh grade students in 179 schools. (Seventh grade students were not tested in four schools.) Additionally 127 seventh grade teachers and 241 eighth grade teachers filled out questionnaires providing data on their content coverage. The response rates for teachers in both grade levels was about 70 percent.

3 By far the most prevalent title for a grade eight mathematics class was simply “math” or “mathematics.” A number of variations on this title were also observed—many of which incorporated the notion of tracking students according to ability. Examples of this approach included “average mathematics”, “basic mathematics”, “advanced mathematics”, “gifted” or “high “ mathematics, “LD mathematics”, “remedial mathematics”, and “resource mathematics” among others. The only other commonly occurring class titles that did not contain either “math” or “mathematics” were “arithmetic”, “pre-algebra” and “algebra.” Courses in some schools carried such unique titles that their content focus and relation to a progressively unfolding mathematics curriculum was unclear (Cogan, Schmidt & Wiley, 2001).

TIMSS sample, a reduced number of course types (regular, pre-algebra and algebra) were employed in these analyses to define tracks and to more fully examine their effects. Based on other data in the study which provides more extensive knowledge of course content, enriched courses were recoded as pre-algebra, remedial as regular, and geometry as algebra (there were only a few such cases).

Students in TIMSS were given 90 minutes to respond to one of eight rotated assessment forms. Approximately 150 mathematics items were distributed across the eight forms providing a broad representation of student knowledge of the mathematics studied around the world at eighth grade. The eight forms were rotated within each classroom so that information on all 150 mathematics items was obtained for each classroom. The test was the same for students in both the seventh and eighth grades. Item Response Theory (IRT)-scaled mathematics scores were created across all countries to have a mean of 500 with a standard deviation of 100. In addition to the TIMSS tests, students completed a background questionnaire related to their home environment from which a composite measure of SES was developed, including education level of the parents, the number of education-related possessions in the home, and the number of books in the home (Schmidt & Cogan, 1996).

Cohort Longitudinal Analysis

For analysis purposes we made the assumption that the cohort of seventh grade students in a school was essentially the same as the cohort of eighth grade students from that same school other than the eighth grade students being simply a year older and having an additional year of mathematics instruction (see Schmidt et al., 2001). This permits the use of the seventh grade score as a pseudo-pre-measure to examine the effect of tracking on student learning at the eighth grade. Through the school mathematics class tracking forms, the appropriate seventh grade class which served as the feeder to each eighth grade class could be identified. This was necessary since some tracking was also done at the seventh grade.

For example, a seventh grade pre-algebra class would be an appropriate feeder to an eighth grade algebra class but not to an eighth grade regular class. Employing the appropriate seventh grade class score as a pre-test for eighth grade achievement at the class level has the impact of allowing the study to explore the effect of tracking on what students learned during eighth grade and not merely on the status of the eighth grade achievement score.[4] This could only be done at the class level and not at the individual student level and only after a careful matching of seventh and eighth grade classrooms based on the school tracking form. Given that there was only one seventh grade class sampled per school, out of necessity we imputed seventh grade average achievement scores for additional classrooms necessary to serve as surrogate pre-test measures.[5] Previous uses of these data support the validity of this approach (Schmidt et al., 2001).

CONTENT COVERAGE

TIMSS also surveyed the mathematics teachers of the sampled classes. [6] They were asked to indicate the number of periods over the year in which they taught each of 21 mathematics topics. For each content area, teachers checked a box indicating whether they had taught a topic for "1-5", "6-10", "11-15", or "> 15" periods or "not taught" the topic at all during the year.[7]

Although the focus of this paper is solely on the U.S. practice of tracking, the TIMSS curriculum data from approximately 50 countries provided an empirical non-ideological basis on which to develop an index of topic difficulty. Such a quantitative index is essential for statistical modeling. This index is referred to as the "international grade placement" index or IGP. There is an IGP value for each specific content topic in the taxonomy. The index gives a value between 1 and 12 indicating the grade, averaged across over 40 countries, at which the specific topic received its greatest instructional focus, taking into account the grade at which it was first introduced. This scale has been found to have strong face validity as well as construct validity (Achieve, 2004; Cogan, Schmidt & Wiley, 2001).

It seems a reasonable assumption that topics receiving their instructional focus in later grades are more difficult than those receiving their focus in earlier grades, given the hierarchical nature of school mathematics and the fact that this value is estimated over a large number of countries. Thus, the IGP provides an estimate of rigor for each topic, at least in terms of school mathematics.

This index was used as a weight to estimate the difficulty of the delivered curriculum as described by teachers. This was done using the data from the teacher questionnaire in which they indicated the number of periods of coverage associated with a set of topics, which was in turn used to determine the content coverage profile over 21 topics by estimating the percent of the school year associated with the topic. These estimated teacher content profiles were then weighted by the corresponding IGP values and summed across all topics. This produced a single value that was an estimate of the level of demand associated with the implemented curriculum in mathematics for each teacher as illustrated in the following equation.

$$IGP_{class} = \sum_{i}^{21} (\textit{Instruction Time for Topic}_i \times \textit{IGP for Topic}_i)$$

Thus the weighted content coverage index is a multi-faceted measure that is based on three distinct aspects of Opportunity to Learn (OTL): 1) the mathematics content itself (topic coverage—yes/no), 2) instructional time for each topic, and 3) rigor or content difficulty (as estimated from international curriculum data). Therefore the IGP measure of the mathematics taught in the classroom is a measure of content-specific OTL defined at the classroom level. The metric of the index is defined in terms of grade levels and as a result is directly interpretable.

RESULTS

The Prevalence of Tracking in U.S. Schools

An analysis of the school tracking forms revealed two types of schools. The first type offered a single type of mathematics course to all eighth grade students. The second type of school offered multiple courses or tracks into which different students were assigned.

Non-Tracked Schools. Approximately 27 percent of U.S. eighth grade students attended a school in which there was only one course available to them in mathematics. Although these schools might group students into different sections based on ability, they do not formally track students using the definition employed in this paper. The content at least by policy is the same for all students attending eighth grade in that school. This included those attending schools that offered only algebra (4.3 percent); only pre-algebra (.8 percent—including .4 percent of the students who attended school with only an enriched course which we recoded as pre-algebra); and only regular mathematics (21.7 percent—including .3 percent who attended school with both regular and remedial courses—the latter recoded as regular) (see Cogan, Schmidt & Wiley, 2001).

For these non-tracked schools, content coverage should be the same for all eighth grade students in the same school, at least in terms of official school or district policy, even if there are multiple sections of the same course offered. One might also expect content coverage to be the same across schools that offer the same type of course. However, that is a different issue from tracking, having to do with the absence of national standards and policies of local control. As teachers define the content of their actual instruction, content coverage can and does vary across sections of the same course even within the same school. The same variation can also occur across schools even though they define eighth grade mathematics as the same course. (We explore this issue in a later section).

Tracked Schools. The other type of school attended by the vast majority of eighth grade students (73 percent) offered two or more different types of mathematics courses or tracks covering different aspects of mathematics for different eighth grade students. The combinations of tracks offered within a school based on the three course types (which itself is a simplification) are many. For example the popular impression that most tracked schools offer the three basic types of courses including regular mathematics, pre-algebra and algebra was true for only 30 percent of U.S. eighth graders who attended tracked schools. Some schools did offer those three tracks (attended by around one-fourth of all eighth graders) but other schools offered different paired combinations of the three types with the most common being regular mathematics and algebra. This type of school was attended by one quarter (25.2 percent) of the eighth graders.

TRACK DIFFERENCES IN CONTENT COVERAGE

Previously specific-topic differences were described among the various tracks (Cogan, Schmidt & Wiley, 2001). Here, however, we employ the IGP index as an indicator of overall content demand for the entire year. Recall that the metric of the index is grade level.

Statistically significant differences were evident in the IGP index across the three types of courses (whether offered within a tracked or non-tracked school): regular mathematics, pre-algebra, and algebra ($p < .0001$).[viii] The estimated contrast of the algebra course with the combination of the pre-algebra and regular mathematics courses was statistically significant (see Table 1 for the summary statistics). Using a 95% confidence interval, the estimated value indicated an almost one year difference (.88) between the algebra and the other two types of courses. Perhaps what may be surprising, but is consistent with earlier analyses, is that the estimated confidence interval for the orthogonal contrast between the regular and the pre-algebra classes was not statistically significantly different from zero ($p < .06$). Thus, in spite of the presumed difference implied by the course titles, this result suggests that, although there may be ability differences defining who is taking which type of course, the difficulty of the content coverage is essentially the same—at least from the international perspective as reflected in the IGP index. The data then imply, relative to the issue posed at the start of this section, that there are real differences in the educational opportunities afforded students in the different types of courses—at least between algebra and the other two course types.

Table 1.
Means, standard deviations*, and sample sizes for schools and classes by type of school and class track*

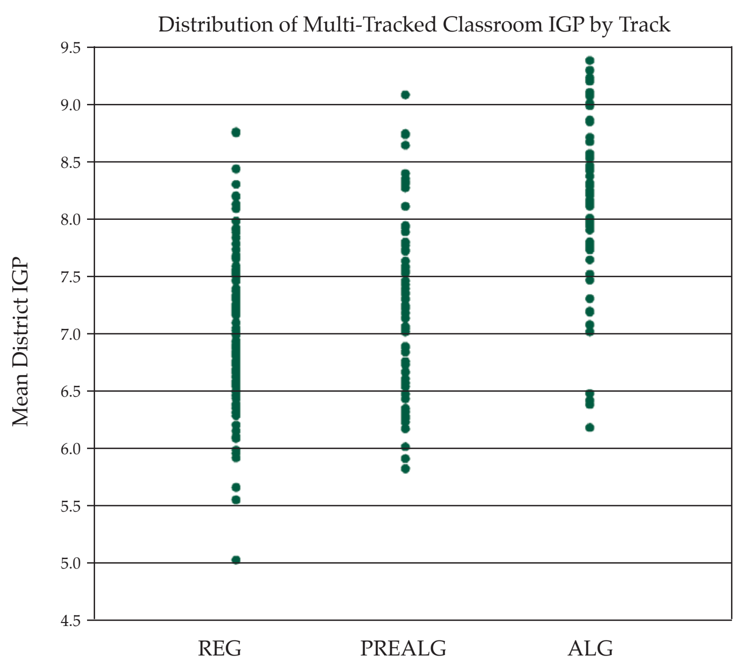
	Schools	IGP		Classes	International Scaled Score		Students
	N	*Mean*	*Std Dev*	*N*	*Mean*	*Std Dev*	*N*
Non-tracked schools	47	7.28	0.57	85	484	93	1822
Regular		7.27	0.43	70	481	90	1509
Pre-Algebra		6.70	0.65	3	522	96	81
Algebra		7.51	1.02	12	495	110	232
Tracked Schools	134	7.45	0.84	258	505	90	5124
Regular		7.11	0.63	132	469	79	2506
Pre-Algebra		7.28	0.76	67	517	85	1356
Algebra		8.20	0.75	59	565	79	1262
Totals	181**			343			6946

*Weighted
** Two schools were dropped due to lack of information

This issue may also be addressed by taking into account the school structure in which the course occurs. To do this, separate analyses of variance were done on the same IGP index for non-tracked and tracked schools. Overall there were no differences in IGP between tracked and non-tracked schools ($p < .09$). The means were almost identical—7.45 vs. 7.28—which is equivalent to about a two month difference in content difficulty.

In tracked schools, the algebra track was statistically significantly different from the other two tracks in content rigor ($p < .0001$). The difference between the pre-algebra and regular mathematics tracks was also significant ($p < .02$). The estimated contrasts indicate that the algebra track classrooms were covering content slightly over one grade level higher (1.09) than the regular track and almost one grade level (.92) more advanced than the pre-algebra track. The estimated contrast indicates about a two months (.17) difference for the content difficulty between the pre-algebra and regular tracks. These results are generally consistent with the analysis cited over all schools. Figure 1 reveals that the variation in the IGP index is very large even within each of the three tracks and that the three IGP distributions have extensive overlap. This large variation and overlap is indicative of the point made earlier that course labels can be misleading in terms of what is actually covered.

Figure 1.
Distribution of IGP by Track



For the non-tracked schools the same pattern emerges with respect to algebra. The content difficulty of the coverage for classrooms in schools (n=12) that offer only algebra is not significantly different ($p < .14$) from the coverage for classrooms (n=70) that are in schools that offer only regular mathematics. That difference is only about a fourth of a year. However, the average values of the IGP index for the algebra classes offered in non-tracked schools is about three-fourths of a year ($p < .03$) less rigorous than for the algebra classes offered in tracked schools. On the other hand, the content of the regular mathematics classes in tracked schools is less rigorous by about two months than that of the regular classes in non-tracked schools.

Classroom Variation in Content Coverage

One way in which to examine how the track structure influences the variation in content coverage across classrooms is by estimating the variance components associated with each level of the school structure. Using the IGP index as a reflection of the complexity of the content coverage and hence in one sense content coverage itself, standard statistical algorithms were used to estimate the variance components associated with schools, tracks within schools and classrooms within tracks for the tracked schools (see Table 2 for a listing of the particular pairings of class types that enabled the estimation of the different variance components).[9] In the case of the non-tracked schools, variance components were estimated for course type across schools, schools and classrooms within schools. Tables 3 and 4 present the results of this analysis for non-tracked schools and tracked schools respectively.[x]

Table 2.
Types of 8th grade classes sampled by school 8th grade course offering pattern (Number of Schools=181)

	8th Grade Sampled Class Type(s)—Number of Schools					
8th Grade School Course Offering Pattern	*Regular*	*Pre-Algebra*	*Algebra*	*Regular/ Pre-Algebra*	*Regular/ Algebra*	*Pre-Algebra/ Algebra*
Regular only	38					
Pre-Algebra only		2				
Algebra only			7			
Regular/Pre-Algebra	3	1		15		
Regular/Algebra	23		3		21	
Pre-Algebra/Algebra		5	3			10
Regular/Pre-Algebra/Algebra	10	5	3	16	10	6
Totals	74	13	16	31	31	16

Table 3.
Variation in the IGP in non-tracked schools

Source	*Estimated Variance*	*Estimated Percentage of Variance*
Course Type	.019	5
Schools within course type	.066	17
Classrooms within schools	.302	78
Total across all classrooms	.387	

Table 4.
Variation in the IGP in tracked schools

Source	*Estimated Variance*	*Estimated Percentage of variance*
Schools	.214	25
Tracks within schools	.336	40
Classrooms within tracks	.300	35
Total across all classrooms	.851	

Non-tracked schools do not all offer the same course type. Some offer only regular mathematics, some only algebra and very few only pre-algebra. Therefore, some of the variation among non-tracked schools in content coverage reflects the fixed differences that stem from the different courses being offered by different schools. The majority of the variance in IGP at the school level reflects variation across schools in which the same course type is offered. This likely results from different interpretations across schools as to what constitutes a course, for example, in regular mathematics (variation which is likely associated with the U.S.'s lack of a national curriculum [see Cogan, Schmidt & Wiley, 2001; Schmidt et al., 1997]. An estimate of the variance among schools in the rigor of the content coverage of eighth grade mathematics suggests that one-fourth of the total variation in the international grade placement of the topics taught is of the cross-school variety. Only about 20 percent of the total cross-school variance is related to the fact that different non-tracked schools offer different courses. The remaining three-fourths of the total variation exists within schools across classrooms, likely reflecting teacher differences in interpretation of what constitutes a particular course, from differences related to the textbook used or some form of ability grouping, or a combination of these differences.

The results for the tracked schools as found in Table 4 clearly reflect the impact that tracking has on content coverage. In schools with tracks, 40 percent of the total cross-classroom variation in content coverage, as indicated by the IGP index, is attributable to track differences. Around 25 percent of the variation is due to school differences in terms of course offerings. Around 35 percent of the variation is attributable to cross-classroom within track within school variation, mostly reflective of teacher differences in interpretation of the content or other adjustments made relative to textbooks or students characteristics.

The estimated total across classrooms variation for non-tracked schools (.302) is less than half the size of the same component for tracked schools (.336 + .300). This implies that tracking actually increases the variation in content coverage across classrooms. The increased variation occurs primarily because of the track level, as the actual value of the class component within tracks is essentially the same between the tracked and non-tracked schools, thus increasing the overall cross-classroom variation.

Track Differences in Achievement

Since tracking starts in the seventh grade and continues through eighth grade, it is desirable to also look at differences in the gain in achievement over the two year period for different track patterns. Given this goal, the best type of data would be longitudinal data on the same students so that the tracking effect at eighth grade could be separated from prior learning. TIMSS did not provide such longitudinal data, but did provide cohort-longitudinal data. As described previously, we assume no major cohort differences within the same school for students following the same track other than the additional year of schooling (Schmidt et al., 1999, see pp.29-30).

Using this assumption we paired eighth grade classrooms in each tracked school with an appropriate seventh grade feeder classroom from that same school as defined by the school course offering structure as noted in the tracking form. Here the imputed seventh grade data were used to provide the appropriate seventh grade classroom when none was sampled at that particular school. Four types of paired track patterns were formed. The first pattern was a seventh grade regular mathematics class leading into an eighth grade regular mathematics class, while a second track pattern was where the same seventh grade course led into pre-algebra at eighth grade. The other two patterns both end up with eighth grade algebra. One starts with regular mathematics at seventh grade while the other starts with seventh grade pre-algebra. These were the dominant patterns available in the data.

Table 5 shows the average mathematics scores for these four track patterns. The unit for these analyses was the paired classrooms on which two measures were available—the mean seventh grade achievement score and the mean eighth grade score, both averaged over the students in the pair of classrooms and then averaged across all schools with that pattern. For each pattern, the data in the table give three values: the seventh grade mean of the feeder classroom, the eighth grade mean and the gain defined as the difference between the two means.

Table 5.
Classroom means for four patterns of mathematics tracks, including their appropriate seventh grade feeder class, and eighth grade gain.

Reg 7- Reg 8 Track			*Reg 7—PA8 track*			*Reg 7—A8 track*			*PA7-A8 track*		
7th Grade Regular Mean	*8th Grade Mean*	*Gain*	*7th Grade Regular Mean*	*8th Grade Mean*	*Gain*	*7th Grade Regular Mean*	*8th Grade Mean*	*Gain*	*7th Grade Pre-Algebra Mean*	*8th Grade Mean*	*Gain*
462	469	7	468	517	48	462	550	88	524	569	45

Obvious differences are apparent in the mean achievement across the two seventh grade tracks and across the three eighth grade tracks. Seventh grade classrooms teaching "regular" mathematics scored around 60 points lower than the pre-algebra classrooms. Correspondingly, there is about an 80 to 100-point difference between eighth grade regular track classrooms and algebra track classrooms with pre-algebra classrooms falling in between (517). This nearly 100-point difference between the algebra and the regular mathematics track at eighth grade represents about a one standard deviation difference in the test score. The difference between the pre-algebra and regular mathematics track was estimated as roughly one-half of a standard deviation.

One important observation from Table 5 is that the difference in mean scores across regular seventh grade classrooms that serve as feeders for the three different eighth grade tracks is trivial. In other words, if a regular mathematics course is all that is available in seventh grade then the sorting process of who takes which kind of eighth grade mathematics does not seem particularly related to seventh grade achievement. Other research on tracking and the assignment of students to a particular mathematics track suggests that these decisions are often based on some estimation of students' mathematical ability (Oakes et al., 1992). Certainly these data do not support the notion that eighth grade tracking assignments have been made of the basis of differences in prior achievement. This raises the question of what the basis for such assignments was in many of these schools.

The estimated classroom level gains point out additional interesting patterns. Very little gain appears to occur in regular eighth grade mathematics classes that have regular seventh grade mathematics as their feeder class. However, large gains of around one-half to almost one full (.9) standard deviation are noted for those students in either an eighth grade algebra or pre-algebra course when coming from a regular seventh grade mathematics course. This is especially true for the algebra course. This result, when combined with essentially no gain from a regular seventh grade class to a regular eighth grade class, calls into question the wisdom of having any student take "regular" mathematics (which is mostly arithmetic) at eighth grade. The differences among the eighth grade track gains were also statistically significant ($p < .0001$).

The above analyses were done only on the tracked schools. Previously we indicated that IGP differed only slightly between the tracked and non-tracked schools on average. Similarly, the difference in eighth grade mean achievement is also small (505 vs. 484)—about two-tenths of a standard deviation difference—but is statistically significant ($p < .0001$). See Table 1 for the summary statistics related to eighth grade achievement. A more careful examination of the pattern indicates an interaction effect. For algebra classes the 70-point difference in mean achievement between those in tracked schools versus non-tracked schools is significant ($p < .003$) but the differences in mean achievement for the other two types of courses are not significant. Finally, across the non-tracked schools there were no significant differences in eighth grade achievement for the three different types of courses ($p < .38$).

Classroom Variation in Achievement

The variance components for the achievement scores are of interest here as a comparison to those reported for content coverage. This is true since the issue of tracking is not just about equality in learning opportunities but also about equality in attained learning. Tables 6 and 7 present the results of this analysis. This analysis is again estimated separately for non-tracked and tracked schools. In the case of achievement a fourth source of variation is estimable—student variation within classrooms.

In the tracked schools, about 40 percent of the variation in achievement across students is related to track differences. This estimated track component is very similar in magnitude to the 40 percent estimate for topic coverage as defined by the IGP index. For achievement variation, the fourth variance component reflects within-classroom or individual student variability. This component was estimated to be around half of the total variability for both tracked and non-tracked schools. In other countries in TIMSS such as Japan, Korea, Norway and Sweden the estimated variance component for students within classrooms was around 90 percent (Schmidt, McKnight, Cogan, Jakwerth & Houang, 1999, p. 174).

Taking into account track differences reduces the classroom component to under 10 percent from the 30 percent earlier reported in (Schmidt, McKnight, Cogan, Jakwerth & Houang, 1999), where the track component was not estimated. In non-tracked schools, the class component was around one-fourth of the total variation.

Table 6.
Variation in the mathematics scaled score in schools having no tracks.

Source	Non-Tracked Schools	
	Score Variance	*Score Variance (%)*
Course type	0[11]	0
Schools	1328	16
Classrooms within schools	2179	25
Students within classrooms	5074	59
Total across all students	8531	

Table 7.
Variation in the mathematics scaled score in schools having tracks.

Source	Tracked Schools	
	Score Variance	*Score Variance (%)*
Schools	0[xi]	0
Tracks within schools	3431	40
Classrooms within tracks	652	8
Students within classrooms	4414	52
Total across all students	8497	

Relationship of Tracking to Achievement

The pattern of achievement differences across classrooms indicates that on average the achievement level of a class is related to the track of the class. This is certainly consistent with many other studies cited in a previous section and is not particularly surprising. The analyses presented previously in Table 5 did not control for the selection bias introduced by the fact that students were not randomly distributed across the different tracks within schools. Differences in achievement levels across tracks could be attributed to the selection bias associated with who is counseled into the various tracks or who self-select into them. How to disentangle those effects from other potential effects such as a curriculum effect related to content coverage is difficult. We explore this relationship in several ways using different statistical adjustments in an attempt to understand the nature of the relationship between track as a curriculum issue and student achievement in mathematics.

From a policy perspective, the disentanglement of selection bias from other interpretations of track differences is critical. Analyses in a previous section clearly demonstrated content coverage differences across tracks, as reflected in the IGP index. The estimated effects suggest a grade level difference in terms of the profile of topic complexity for the algebra track compared to the other two tracks and previous studies have amplified that difference in terms of specific-topic coverage (Cogan, Schmidt & Wiley, 2001; Schmidt et al., 1999).

The critical question remains, are the observed mean level differences in achievement across tracks related to these differences in content coverage? Or do such achievement differences simply reflect an underlying difference in students' mathematics ability? If these achievement differences are primarily the consequence of the latter, then policies requiring all students to study the profile of topics found in the algebra track might not be particularly effective, at least towards the goal of improving overall mathematics achievement in the U.S. On the other hand, if these curricular differences do play a prominent role in creating the observed achievement patterns, then such a policy would not only seem desirable but prudent.

The fact that the profile of topics covered in the algebra track classrooms is more like the profiles found in the classrooms among the top-achieving countries suggests the curriculum-based argument is the more plausible of the two proposed explanations (Schmidt et al., 1999; Schmidt et al., 2001).

To explore this issue more fully, a three level hierarchical linear model was fitted separately for the tracked schools. The conceptual model defining the analysis model and the particular choice of variables follows the framework defined in TIMSS (Schmidt, Jorde, Cogan, Barrier, Gonzalez et al., 1996). The three levels included schools, classrooms nested within schools and students nested within classrooms. The track designation was included as a dummy variable at the classroom level. The model also included several covariates at each of the levels in the design. The student-level model included racial/ethnic identity (with the contrasts centered on the other category) and the composite SES measure. The class-level model included the appropriate seventh grade pre-measure, mean SES, and track. The school-level model included the school-level mean SES, and three variables derived from the school questionnaire including the percent minority enrollment at the school, the location of the school (rural, suburban, or urban) and the size of the school as measured by the number of eighth grade students. The model was specified as follows:

Student Level Model

$$Y = P_0 + P_1\,(\text{White}) + P_2\,(\text{Black}) + P_3\,(\text{Hispanic}) + P_4\,(\text{Asian}) + P_5\,(\text{SES}) + \epsilon$$

Classroom Level Model

$$P_0 = B_0 + B_1\,(7^{th}\text{ grade achievement for feeder class}) + B_2\,(\text{SES}) + B_3\,(\text{pre-algebra}) + B_4\,(\text{algebra}) + R_0$$

School Level Model

$$B_0 = G_0 + G_1\,(\%\text{ minority}) + G_2\,(8^{th}\text{ grade enrollment}) + G_3\,(\text{Rural}) + G_4\,(\text{Urban}) + G_5\,(\text{SES}) + U_0$$

Results are presented in Table 8. At the individual student level the racial/ethnicity identity of the student as well as the SES of the family were significantly related to their performance on the TIMSS test ($p < .001$). The racial/ethnic estimated coefficients indicated a large negative relationship to achievement associated with being African American or Hispanic American.

After adjusting for the student level relationships, the estimated class level model indicated a statistically significant relationship for track even when controlling for the aggregate SES of the class and the mean level performance of the seventh grade feeder classroom ($p < .0001$). Although not an entirely perfect solution, adjusting for the prior achievement at the class level and the SES both at the class and individual level should remove a substantial portion of the likely student selection bias. This makes the estimated track effects less reflective of selection bias and, therefore, more likely to reflect differences in instruction such as content coverage and/or some other social or institutional effects related to track membership—all of which are related to schooling.

The significant track effect is present both in terms of differences between the pre-algebra track and the regular track ($p < .0001$) as well as in achievement differences between the algebra track and the regular track ($p <.0001$). The estimated effect for the pre-algebra track controlling for the other variables in the model is approximately 30 (6.97 se). The typical student in the algebra track reflects an achievement level over 60 points higher than that of the typical student in the regular track and correspondingly some 30 points higher than what is average in the pre-algebra track.

Recall that the standard deviation for the mathematics score used in these analyses is 100. Thus the estimates of effects reported here indicate about one-third of a standard deviation difference in achievement between each of the three tracks (again, controlling for the other variables in the model). This implies two-thirds of a standard deviation difference in performance between the typical regular track student and his or her counterpart in the algebra track. In the international context, this two-thirds of a standard deviation is not inconsequential as it is the difference between the mediocre, i.e., at the international mean, U.S. eighth grade performance and the performance of two of the top achieving countries—the Czech Republic and Flemish-speaking Belgium.

At the school level after controlling for SES and racial/ethnic differences at the student and classroom levels, there are no further differences in achievement related to the (school) aggregate SES measure ($p < .10$) or to the percent of minority students attending the school ($p < .563$). The only statistically significant effect on achievement to remain at the school level is related to the location of the school ($p < .007$). School size is not significant ($p < .781$). The estimated location effect indicates that students attending schools in rural areas outperform their counterparts attending mid-size city and suburban schools. This might seem counter-intuitive but it should be noted that this estimated effect is conditional on schools, which have

Table 8.
Student, classroom, and school variables employed in the three-level HLM analysis of eighth grade mathematics achievement in tracked schools.

Students-level variables	*Estimated Coefficient*	*SE*
Race: White (D)	2.41	(7.10)
Race: Black (D)	-26.53***	(7.70)
Race: Hispanic (D)	-21.01**	(8.02)
Race: Asian (D)	-2.50	(8.92)
Socioeconomic Status (SES) (C)	-2.85***	(0.44)
Classroom-level variables		
7th grade achievement	1.00*	(0.49)
Mean classroom SES	20.18***	(2.69)
Class type: Algebra (D)	62.19***	(6.29)
Class type: Pre-Algebra (D)	31.05***	(6.97)
School-level variables		
Mean school SES	-4.97	(3.00)
8th grade enrollment	0.01	(0.02)
Minority enrollment (%)	-0.06	(0.11)
Location: Urban (D)	4.40	(6.06)
Location: Rural (D)	19.73**	(6.92)

Note: D denotes dichotomous indicator variables. C denotes centered variable. SE denotes standard errors. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Appropriate TIMSS sampling weights were used in the analysis.

been equated relative to the selection bias factors and with respect to differences in tracks. Rural schools, in general, do not as frequently provide an algebra track for students as do schools in the other settings. Less than 60 percent of students attending schools in rural areas attend a school that offers algebra compared to over 80 percent of students attending schools in suburban or mid-sized cities (Cogan, Schmidt & Wiley, 2001).

Does Content Coverage Mediate the Effects of Tracking?

The analyses presented in the previous section demonstrate the effect of tracking on residual achievement gain from seventh to eighth grade after adjusting for likely sources of selection bias. The estimated effect sizes are large and important. The question is, how do these effects occur?

Researchers have suggested three different mechanisms, one of which is instructional, which we further define for purposes of this study as the level of demand or rigor of the content coverage.

One thing is clear from the analyses reported on in a previous section of this paper. From a content point of view, on average there is a major difference between the level of demand associated with the content covered in the algebra track compared to the other two. That difference was estimated to be about one year. This seems reasonable on the face of it since the algebra track is considered to be the equivalent of high school algebra I, which most U.S. students take at ninth grade - one year later than eighth grade regular mathematics. The difference between pre-algebra and regular mathematics is smaller, and for the most part not statistically significant.

To examine this further a discriminant function analysis was done to see if we could further differentiate the content coverage across the three tracks. The variables we used to build the discriminant function were the 21 content areas on which we had teacher reported coverage. These were the same 21 used with the IGP weights to define the level of content demand (IGP index) reported throughout this paper. In this analysis we did not use the weights but the teachers' reported coverage of each area over the school year. The topics covered the full range of eighth-grade mathematics found in the 40 plus countries studied in TIMSS. This included topics such as fractions, proportions, linear functions, statistics, congruence, and similarity, among others.

The results further affirm how inextricably tied content coverage is to the definitions of the tracks. The canonical correlation was .69 (s.e. = .04) accounting for almost 50% of the variation of placing classrooms into the tracks based on the 21 content categories. This led to 77 percent of the algebra classrooms and 68 percent of the regular classrooms being correctly classified. In fact almost all the misclassification is associated with the pre-algebra track. If regular and pre-algebra classes were combined the correct classification for the combined track would be over 90 percent. In fact the classification of pre-algebra classes based on content was wrong about three-fourths of the time. From a content point of view pre-algebra and regular mathematics are essentially the same. Given the centrality of content coverage in defining the opportunities available to learn mathematics, it essentially becomes the core meaning of the tracks with the possible exception of pre-algebra. This of course is reflected in the labels used to define the tracks which, themselves are content laden.

The first latent root was statistically significant ($F= 3.79$, $p <.0001$). The corresponding latent vector (or canonical weights) suggests a pattern consistent with this interpretation and more importantly points out how closely intertwined the two concepts are (see Table 9). The discriminant function is mainly a contrast between the amount of time over the full school year allocated to fractions, decimals and percentages—the substance of a typical regular eighth grade mathematics class in the U.S.—as contrasted with the amount of time allocated to slope, functions and equations—the substance of a typical algebra I class. From the analysis of the misclassification probabilities, apparently pre-algebra is most like regular mathematics at least from a content point of view.

Table 9.
Relationship between the 21 teacher time variables and the canonical discriminant function

Variable	Function 1[a]
Whole numbers	0.999
Common and Decimal Fractions	0.996
Percentages	0.986
Number Sets	-0.929
Number Theory	0.582
Estimation and Number Sense	0.995
Measurement Units	0.989
Perimeter, Area, Volume	0.970
Estimation and Measurement Error	0.917
Geometry: Basics	-0.874
Transformations & Symmetry	0.878
Congruence and Similarity	0.964
3D Geometry	0.918
Ratio and Proportion	0.652
Slope & Trigonometry	-0.976
Functions, Relations, Patterns	-0.998
Equations & Formulas	-0.997
Statistics and Data	0.998
Probability and Uncertainty	0.965
Sets and Logic	-0.901
Other Math Topics	-0.916

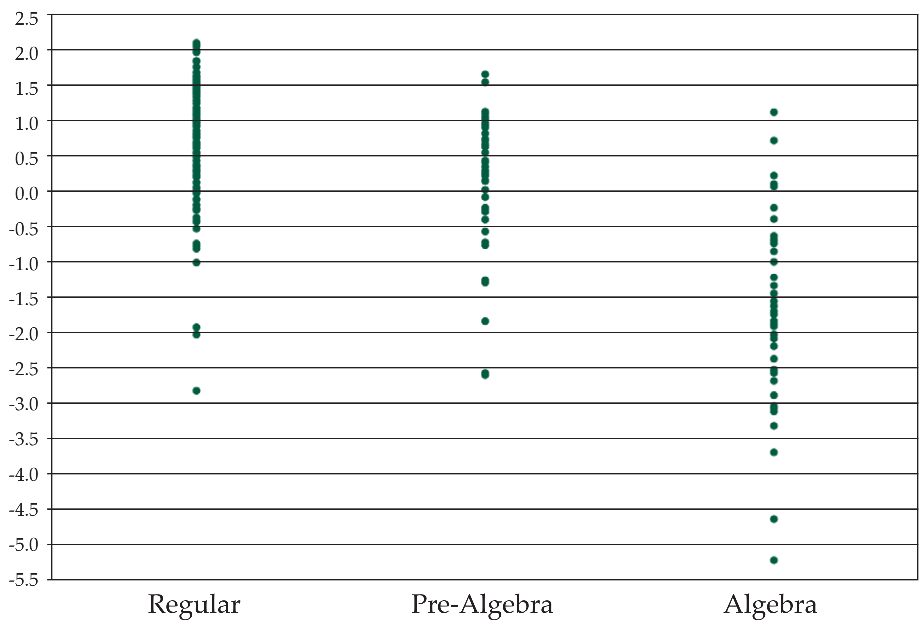
[a] Eigenvalue = 0.932, $F = 3.79$, $p < 0.001$

The above analyses describe what is true in the aggregate—on average the tracks differ with respect to the content covered—measured either in terms of the IGP or looking at the content coverage in each of the 21 areas separately. However, Figure 1, which shows the variation in IGP within each of the tracks, suggests enormous within-track variation in the level of content demand. Even among algebra classes there are some with an IGP suggesting content coverage which internationally is at the sixth grade level while for others the content coverage is at the ninth grade level. In fact, the variation is so large the three distributions have a surprising level of overlap.

This is illustrative of a general problem: without national standards, course labels across schools do not always have the same meaning. This likely implies that many of the classrooms defined as algebra by the school are not covering content that would lead mathematicians to label them an algebra course. This implies that at least from a content point of view the number of tracks in reality may be some larger number. All of this variation in content coverage (which, as we have argued, is the core of the definition of the tracks) makes the study of how content coverage mediates the effects of tracking on achievement gain indeed a very complex problem. The self-reported coverage of the content areas by the teachers does not allow for an estimation of the level of depth associated with content coverage. As a result, what might appear to be equal content coverage can only be considered to be equal in that the topic is covered during classroom instruction, but not how deeply the topic may have been covered. This only further complicates the problem raising the question, what constitutes an algebra class?

This issue is further represented in Figure 2, which plots the distribution of the classrooms defined by the discriminant function around each of the three centroids.

Figure 2.
Distribution on first canonical discriminant function values by track.



With the above difficulties in mind, we did a hierarchical linear model analysis in which we placed both track and IGP in the same classroom level model. All of the other variables as described in Table 8 were incorporated in the model. Those results are summarized in Table 10. Both were statistically significantly related to residual gain.

Central to the point of this analysis is whether the estimated track effect is significantly reduced by the inclusion of a measure of content coverage in the model. Without control for the background of the students or content coverage, the estimated track effect for the algebra track compared to the regular mathematics track was a full standard deviation difference in achievement. After controlling for prior achievement and SES, the estimated effect for the algebra track compared to the regular mathematics track was about two-thirds of a standard deviation (.60—see Table 8).

Controlling for instructional content—one of the proposed mechanisms by which tracking has its impact—the estimated track effect for algebra is further reduced by about 15 percent to slightly more than one-half of a standard deviation (.51). Instructional content is a mediating factor in how tracking influences academic achievement—this is clear. The fact that the track effect is not reduced further by the inclusion of a content measure seems surprising given how inextricably intertwined they are—almost by definition.

One hypothesis for this surprising result centers on the data presented in Figures 1 and 2. Because of the heterogeneity of the IGP index within each of the tracks, and the resulting high degree of overlap across the three distributions, the relationship of the IGP to the residual gain in achievement might well differ across the three tracks. In effect an IGP value of 8 in the algebra track might have a different meaning than an eight in the regular track, In short an "8" is not necessarily an "8."

Table 10.
Student, classroom—including IGP—and school variables employed in the three-level HLM analysis of eighth grade mathematics achievement in tracked schools.

Students-level variables	*Estimated Coefficient*	*SE*
Race: White (D)	2.48	(7.15)
Race: Black (D)	-26.00***	(7.68)
Race: Hispanic (D)	-21.58**	(8.05)
Race: Asian (D)	-2.51	(8.95)
Socioeconomic Status (SES) (C)	-2.84***	(0.44)
Classroom-level variables		
7th grade achievement	1.00*	(0.49)
Mean classroom SES	20.40***	(2.69)
Class IGP	9.53*	(4.38)
Class type: Algebra (D)	51.02***	(7.11)
Class type: Pre-Algebra (D)	30.56***	(6.98)
School-level variables		
Mean school SES	-5.99	(2.91)
8th grade enrollment	-0.01	(0.02)
Minority enrollment (%)	-0.05	(0.12)
Location: Urban (D)	6.45	(6.07)
Location: Rural (D)	19.88**	(6.71)

Note: D denotes dichotomous indicator variables. C denotes centered variable. SE denotes standard errors.
* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Appropriate TIMSS sampling weights were used in the analysis.

One way in which this can occur is due to the limitations of the IGP index. Some of the topics such as linear equations especially at the higher end of the scale include a wide range of subtopics representing increasing levels of content demand. For example solving a simple linear equation such as 3x = 12 is coded as the same level of the IGP as solving two linear equations in two unknowns.

To address this we divided each of the three tracks into 3 subtracks based on the IGP. To accommodate the "8 is not an 8" problem, we divided the classrooms in the tracked schools within each track into three groups defined by high, medium and low content demand. The definitions for each group were different for each track. This creates a nested design. In effect we created nine tracks.

Seventh grade achievement, SES, track, IGP nested within track and IGP were included in a general linear model at the classroom level in order to easily accommodate the nested factors. [12] The results of that analysis are presented in Table 11. The estimated model was significant (F=46.39, p<.0001) with an R2 of .68. The parameter estimates for the nested effects were significant (or marginally significant, p<.10) for those defined for the algebra and regular tracks but not for the pre-algebra track. Their significance supports our "8 is not an 8" hypothesis as related to residual gain. Controlling for those differences in the model adjusts for the within track variation in IGP which consequently results in a substantially larger estimate for the IGP parameter (25.2, s.e.=7.6, p<.001) than was the case for the model presented in Table 10. In fact, it is almost three times the value, implying an effect size of around one-fourth of a standard deviation. This, however, is not the point of the analysis for we have already established the significance of IGP using a more sophisticated model.

More importantly, the estimated effect size for the track is now reduced by half from that estimated in Table 8 (27.9, s.e. = 13.9, p<.046). Given the clustering effect this is probably no longer statistically significant. Even if it were, the size of the remaining effect is greatly reduced, which likely includes other instructional factors such as teacher content knowledge as well as social and institutional factors suggested by Gamoran and Berends (1987).

Table 11.
Analysis of variance for predicting classroom mean mathematics score as a function of SES, 7th grade score, IGP, Track, and IGP blocks within Track.

Predictor	Estimated Coefficient	SE
SES	-17.9***	1.9
7th Grade Achievement	1.6**	0.4
IGP	25.2**	7.6
Track: Regular	-27.9*	13.9
Track: Pre-Algebra	1.7	12.8
Regular Track, Low IGP	24.2	14.6
Regular Track, Mid IGP	24.3*	10.3
Pre-Algebra Track, Low IGP	26.4	16.9
Pre-Algebra Track, Mid IGP	14.5	13.2
Algebra Track, Low IGP	36.5	20.3
Algebra Track, Mid IGP	27.7*	13.0
R^2 .68		

SE denotes standard errors. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Appropriate TIMSS sampling weights were used in the analysis.

RETURNING TO THE VARIANCE COMPONENTS

A variance decomposition analysis of the achievement scores like the one presented in the previous section but now statistically adjusted for factors such as SES and seventh grade achievement as well as a measure of the rigor of the instructional content (IGP) allows us to look at the issue of mediation in another way.[13] The analyses in the previous section focused primarily on the effect of content coverage on achievement at the mean level. Here we examine its relationship to variability in achievement.

Given the limitations of the TIMSS data in which the seventh grade achievement is only available at the class level, the within-classrooms variance component cannot be adjusted for prior achievement. This would necessitate truly longitudinal data. Estimating the adjusted variance components and contrasting them with the unadjusted components provides another way of attempting to disentangle the track effect from other factors.

Tables 12 and 13 report the results of a series of variance component analyses on achievement for both the non-tracked and tracked schools, respectively. Consider first the non-tracked schools. After adjusting for both SES and seventh grade achievement, the school variance component decreased by 86 percent and as a result now accounts for only three percent of the variation. The student variance component also decreases but only slightly, resulting in a percentage increase to now account for three-fourths of the total variation in achievement.

Accounting only for the measure of content difficulty—IGP—reduces the overall variation in achievement by 5 percent. Since the IGP is a classroom-defined measure and, as a result, does not affect the student variation, this has the result of increasing the student percentage of variation accounted for to 60 percent. When both content coverage and background factors are controlled for, the overwhelming majority of the achievement variance resides at the student level—77.3 percent. This is more consistent with results obtained for countries such as Japan, Korea, Spain, Czech Republic, and France among others.

Table 12.
Variation in the mathematics score explained by prior achievement, SES, and IGP in non-tracked schools.

	Non-tracked schools			
	Score	*% Reduction in Score Variance*		
Source	*Variance*	*Prior Achievement, SES*	*IGP*	*Prior Achievement, SES, & IGP*
Course type	0[xi]	---	---	---
Schools	1310	86	14	80
Classrooms within schools	2333	40	26	52
Students within classrooms	4936	4	0	4
Total	8551	26	5	28

Table 13.
Variation in the mathematics score explained by prior achievement, SES, and IGP in tracked schools.

	Tracked Schools			
	Score	*% Reduction in Score Variance*		
Source	*Variance*	*Prior Achievement, SES*	*IGP*	*Prior Achievement, SES, & IGP*
Schools	0[xi]	---	---	---
Tracks within schools	3559	40	29	53
Classrooms within tracks	532	16	7	21
Students within classrooms	4463	2	0	2
Total	8497	19	12	24

In the tracked schools (Table 13), the added dimension of tracking looms relatively large in accounting for achievement variation. Without adjustments for content and background factors, as indicated earlier, tracking accounts for around 40 percent of the variation in achievement. Adjusting the variance component estimates for SES and prior achievement results in a forty percent drop in the value, while the overall variance estimate drops 18.6 percent. As a consequence, the relative amount of achievement variation accounted for by tracking drops to 30.8 percent. Adding the IGP measure of content difficulty reduces the estimated variance component for tracking by 52.5 percent which also reduces the total variance by 24.3 percent. As a result the relative contribution of tracking reduces to 33.6 percent.

The variance component analyses yield implications consistent with the previous analyses. Content coverage does mediate some of the tracking effect but does not eliminate it, implying that other factors associated with tracking are likely involved, such as the motivational impact of putting the best students together with other instructional factors such as teacher subject matter knowledge and the interaction of this factor with content coverage.

DISCUSSION

In this article we set out to examine the effect of content coverage on mathematics achievement and to study its relationship to eighth grade tracking. Previous analyses have shown tracking to be quite prevalent at eighth grade unlike much of the rest of the world. The U.S. practices both between- and within-school tracking, leading to very different content coverage for different students. Without judging the merits of the instructional or other theories which led to this practice, the results presented in this paper explore the consequences of such a policy. Using the IGP index we found statistically significant differences in content coverage across the different types of courses, especially between algebra and the other two course types.

The results presented in this paper challenge the wisdom of this practice. Why should different students study different content, either as a result of being sorted within a school into different content tracks or by the de facto result of different non-tracked schools deciding what the one type of course (e.g., algebra versus regular mathematics) they will offer for all their students? The latter case also effectively tracks students into different content coverage but for students attending different schools. The effect is the same—different content coverage for different students.

Tracked versus Untracked Schools

Twenty-seven percent of U.S. eighth grade students attend schools with no within-school tracking. How does content coverage in these schools compare to that found in the tracked-schools? For the typical student there is no difference in the rigor of the content coverage provided by the tracked-schools compared to the non-tracked schools. However, the variation across students in content coverage is much larger for students in tracked schools than it is for students in non-tracked schools. In fact, the variation is over twice as large and the difference is the direct result of tracking as the classroom and school variation is comparable across the two types of schools. The consequence is simple—within tracked schools, the chances of two students having the same content coverage is about half that of what it is for two students in non-tracked schools, even though the chances that two students, one from each type of school, having the same content exposure is very high.

A closer examination of the data indicates that the content rigor of algebra in tracked schools is about one-half a year more rigorous than what it is in the non-tracked schools that offer algebra. On the other hand, those students in the regular mathematics track find the rigor of their content coverage about two months behind that of students taking regular mathematics in the non-tracked schools. In other words, for the typical student in either type of school there are no differences in the level of demand of the content coverage, but for the typical student in the different tracks there are differences. In tracked schools the elite students gain—the regular students lose. All of these differences in content coverage occur in the context of how representative the typical student is for all students in the two types of schools. Clearly what is typical is fairly common for most students in the non-tracked schools but hardly the case in the tracked schools. To the benefit of the elite students, the rest are exposed to greater inequalities in content coverage and for many lower level content, not just compared to those in the algebra track within their own school but to those students taking the same course in a non-tracked school. This suggests that in terms of content coverage or simply opportunities to learn important mathematics, average students would be better off in non-tracked schools. In tracked-schools only the elite students benefit.

The SES composition and the average seventh grade achievement are essentially identical across the two types of schools, indicating the students essentially come from the same population—U.S. eighth graders. Yet their curricular experiences will be quite different depending on which type of school they were required to attend given the nature of local school attendance boundaries. In the end the difference between the typical students' mathematics achievement in tracked schools is small compared to that of the typical student in non-tracked schools—504 versus 492. In summary, the types of students, the content covered, and what they know about mathematics when they leave are all the same for the typical student attending either type of school. It all seems harmless until one looks beyond that typical student.

For tracked schools, students who take algebra perform substantially better on the achievement measure than those in the other two tracks. In fact, the difference between the algebra track students and the regular mathematics track students is more than one complete standard deviation difference in performance. Across all students in the tracked schools, 40 percent of the resulting variation in achievement is attributable to tracking. By contrast, for students attending different schools which are not tracked but which offer different eighth grade mathematics, such differences in content exposure account for only 18 percent of the variance. Clearly tracking works if the goal is to isolate the elite students and to give them the strongest curriculum (and most likely, together with the most knowledgeable teachers) in order to have them perform better than all the rest of the students. The policy question, however, is at what cost—what is the consequence for the rest (the majority) of the students?

What if all students were given the same curricular experiences? In fact, this is one policy option being advocated by many that call for "algebra for all in eighth grade." This is consistent with what is done around the world, at least in virtually all of those 50 countries studied in TIMSS. Two results from the present study contribute to this debate. First of all, looking at the achievement results for students in the four track patterns found in the data, the largest classroom level gain in achievement was made by those students who in seventh grade took regular mathematics and in eighth grade took algebra. Not only was the gain the largest but mean achievement at eighth grade was only 14 points lower than the mean achievement of those students in the situation where they would have taken eighth grade algebra but who had taken pre-algebra in seventh grade.

Furthermore, the average seventh grade achievement of three of the tracks leading to three different eighth grade courses was essentially the same but what they gained in achievement at eighth grade depended on what course they took at eighth grade. Those who took regular mathematics, again on average, gained nothing while those who took pre-algebra also had a significant gain like those who took algebra. This raises the serious question of why anyone should take the regular mathematics track at eighth grade after having had regular mathematics at seventh grade, yet half of the American eighth graders who attended tracked schools took such a sequence and virtually learned nothing, at least on average.

By striking contrast, 83 percent of the students in non-tracked schools attended a school in which only regular mathematics was offered at both the seventh and eighth grade. They, however, gained on average 21 points on the achievement measure and ended up with a mean of 491 as compared with 465 for those students who took the same sequence of courses but in a tracked school. Obviously, regular mathematics is not regular mathematics, at least between the two types of schools.

A central issue of this paper focuses on the relationship of the curriculum differences created by tracking to the corresponding differences in mathematics achievement. Given that students are not placed in the tracks in a random fashion but rather are selected into them, the inherent selection bias must be adjusted for in an attempt to disentangle it from other effects such as content coverage. This is a complex problem with the complexity arising from the fact that in the U.S. the selection or background factors are inextricably confounded with the content differences associated with the tracks. Content differences are further confounded with teacher differences, especially in terms of their mathematics and mathematics pedagogy backgrounds.

The non-random assignment of students to track and its consequential effect on students' achievement is a particularly important issue to sort out as much as possible since sensible public policy depends on it. If achievement differences among tracks are primarily an issue of selection bias, then the track structures essentially serve as a sorting mechanism that would not necessarily increase the achievement variation but only redistribute it across the tracks. Aside from the issues of its effect on motivation and the social desirability of such a segregation by ability, the tracking in this case does not serve to develop gaps but only to make them visible if not to reify them. On the other hand, if there is a large and significant difference in content coverage, tracking not only segregates but also increases the variability, resulting in a curriculum gap between the tracks. This results in more deleterious effects.

In fact both are true. Selection bias in terms of who is placed or chooses to be in the algebra track is present but only accounts for about 40 percent of the estimated effect size of algebra compared to regular mathematics. An additional one-third of the effect size is attributable to differences in the level of demand of the content coverage. After accounting for student selection the estimated effect size for track was reduced by more than one- half due to differences in content coverage, implying that curricular content is a major contributor to differences in achievement across the tracks. Some variation does remain likely due to other instructional factors, such as teachers as well as various social and institutional mechanisms.

All of this points in the direction of challenging tracking as a reasonable policy, especially given the vision of No Child Left Behind (NCLB). While the nation struggles with work preparation issues so as to compete economically in a highly competitive global environment and to ensure the American dream of people finding gainful employment, as well as to have a quantitatively literate citizenry to make reasonable decisions facing the society in a democratic fashion, can we afford not to provide all of our citizens with focused, coherent and rigorous content coverage in mathematics? In light of these results, "algebra for all" in eighth grade seems more reasonable than continuing the current tracking policies which by design only exacerbate differences and consequently leave many behind.

REFERENCES

Achieve. (2002a). *Foundations for success: Mathematics expectations for the middle grades.* Washington, DC: Achieve, Inc.

Achieve. (2002b). *No child left behind: Meeting challenges, seizing opportunities, improving achievement* (Achieve Policy Brief No. 5). Washington, D.C.: Achieve, Inc.

Achieve. (2004). Do graduation tests measure up? a closer look at state high school exit exams: A report from the American diploma project. Washington, D.C. Achieve, Inc.

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 and Improvement.

Adelman, C. (2003). *Postsecondary attainment, attendance, curriculum, and performance: Selected results from the NELS:88/2000 Postsecondary Education Transcript Study (PETS), 2000* (NCES 2003-39). U.S. Department of Education, Institute of Education Sciences, Washington, DC: Government Printing Office.

Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, 47(5), 626-640.

Alexander, K. L., & Eckland, B. K. (1975). School experience and status attainment. In S. D. Dragastin & G. H. Elder, Jr. (Eds.), *Adolescence in the life cycle: Psychological change and social context* (pp. 171-210). Washington, D.C.: Hemisphere Pub. Corp.

Alexander, K.L., & McDill, E.L. (1976). Selection and allocation within schools: Some causes and consequences of curriculum placement. *American Sociological Review, 41*(6), 963–980.

Alexander, K.L., Cook, M., & McDill, E.L. (1978). Curriculum tracking and educational stratification: Some further evidence. *American Sociological Review, 43*(1), 47-66.

Atanda, R. (1999). *Do gatekeeper courses expand education options?* (NCES Report 99-303). Washington, DC: National Center for Education Statistics.

Cogan, L. S., Schmidt, W. H., & Wiley, D. E. (2001). Who takes what math and in which Track? Using TIMSS to characterize U.S. students' eighth-grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis*, 23 (4), 323-341.

Dauber, S. L., Alexander, K. L., & Entwisle, D. R. (1996). Tracking and transitions through the middle grades: Channeling educational trajectories. *Sociology of Education*, 69(4): 290-307.

Fennessey, J., Alexander, K.L., Riordan, C., and Salganik, L.H. (1981, Oct.). Tracking and frustration reconsidered: Appearance or reality. *Sociology of Education*, 54(4), 302-309.

Foy, P., Rust, K., & Schleicher, A. (1996). Sample Design. In M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report: Volume I: Design and Development* (pp. 4-1-4-17). Chestnut Hill, MA: Boston College.

Friedkin, N.E., & Thomas, S.L. (1997). Social positions in schooling. *Sociology of Education*, *70*(4), pp. 239-255.

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review, 57*(6), 812–828.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education, 60* (3), 135-155.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research, 57*(4), 415-435

Gamoran, A., and Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology*, 94(5), 1146-1183.

Gamoran, A., Porter, A.C., Smithson, J., & White, P.A. (1997). Upgrading high school mathematics instruction: improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325–338.

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

Hallinan, M.T. (1996). Track mobility in secondary school. *Social Forces, 74*(3), 983–1002.

Hallinan, M. T., & Kubistchek, W. N. (1999). Curriculum differentiation and high school achievement. *Social Psychology of Education,* 3(1), 41-62.

Heck, R.H., Price, C.L., & Thomas, S.L. (2004, August). Tracks as emergent structures: A network analysis of student differentiation in a high school. *American Journal of Education,* 110(4), 321-353.

Helscher, M. L., Levine, D., Moore, D., Rizzo, L., Roey, S., Smith, C., et al. (2000). *The 1995 Third International Mathematics and Science Study and United States Technical Report* (NCES Report No. 2000-18). Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Heyns, B. (1974). Social selection and stratification within schools. *American Journal of Sociology, 79*(6), 1434-1451.

Hoffer, (1992). Middle school ability grouping and student-achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14*(3), 205–227.

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

Lucas, S. R., & Berends, M. (2002). Sociodemographic diversity, correlated achievement, and de facto tracking. *Sociology of Education*, *75*(4), 328–348.

Lucas, S.R., & Good, A.D. (2001). Race, class, and tournament track mobility. *Sociology of Education*, *74*(2), 139–156.

Ma, X. (2000). A longitudinal assessment of antecedent course work in mathematics and subsequent mathematical attainment. *Journal of Educational Research,* 94(1), 16-28.

McDonnell, L.M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*(3), 305-322.

McFarland, D.A. (2006). Curricular flows: Trajectories, turning points, and assignment criteria in high school math careers. *Sociology of Education, 79* (3),177-205.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late* (No.). Washington, DC: U.S. Department of Education.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and ability grouping on access to science and mathematics education*. Santa Monica, CA: RAND.

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

Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York: Macmillan Publishing.

Riegle-Crumb, C. (2006). The path through math: Course sequences and academic performance at the intersection of race-ethnicity and gender. *American Journal of Education, 113*(1), 101-122.

Riley, R. W. (1996, November 20). *Press conference on eighth-grade TIMSS results remarks.* Paper presented at the meeting of Press Conference on Eighth-Grade TIMSS Results, Washington, DC.

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., & Nicol, C. (1993). *Curriculum frameworks for mathematics and science*. Vancouver: Pacific Educational Press.

Rock, D. A., & Pollack, J. M. (1995). *Mathematics course-taking and gains in mathematics achievement*. (NCES Report 99-714).Washington, D.C.: U.S. Dept. of Education, National Center for Education Statistics.

Rosenbaum, J. E. (1978). The structure of opportunity in school. *Social Forces, 57*(1), 236-256.

Rosenbaum, J. E. (1980). Track misperceptions and frustrated college plans: An analysis of the effects of tracks and track perceptions in the national longitudinal survey. *Sociology of Education, 53*(2), 74-88.

Schmidt, W. H., & Cogan, L. S. (1996). Development of the TIMSS Context Questionnaires. In M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report: Volume I: Design and Development* (pp. 5-1-5-22). Chestnut Hill, MA: Boston College.Schmidt, W. H., Jorde, D., Cogan, L. S., Barrier, E., Gonzalo, I., Moser, U., & et al. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. New York: Springer.

Schmidt, W.H., McKnight, C., Cogan, L.S., Jakwerth, P.M., & Houang, R.T. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Dordrecht/Boston/London: Kluwer.

Schmidt, W.H., McKnight, C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht/Boston/London: Kluwer.

Schmidt, W.H., McKnight, C., Valverde, G.A., Houang, R.T., & Wiley, D.E. (1997). Many visions, many aims, Volume I: A cross-national investigation of curricular intentions in school mathematics. Dordrecht/Boston/London: Kluwer.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Schneider, B., Swanson, C.B., and Riegle-Crumb, C. (1998). Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education*, *2*(1), 25-53.

Slavin, R. E. (1993). Ability grouping in middle grades: Achievement effects and alternatives. *Elementary School Journal, 93*(5), 535-552.

Smith, J. (1996). Does an extra year make any difference? The impact of early access to algebra on long-term gains in mathematics attainment. *Educational Evaluation and Policy Analysis, 18*(2), 141-53.

Stevenson, D. L., Schiller, K. S., & Schneider, B. (1994). Sequences of opportunities of learning. *Sociology of Education*, *67*, 184-198.

Useem, E. L. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. *American Journal of Education*, *100*(3), 325-353.

Vanfossen, B.E., Jones, J.D., & Spade, J.Z. (1987, April). Curriculum tracking and status maintenance. *Sociology of Education, 60*(2), 104-122.

ENDNOTES

1 To properly study tracking and its effects, the tracks within a school should be used as strata and a random sample of classrooms should be drawn from each track or stratum.

2 Only recently have such large scale data sets with such detail become available. TIMSS, ADHealth, and LSAY are three such examples. New efforts are also underway to incorporate such measures of content in future longitudinal data collections.

3 This allows analyses that are closer to the ideal which would include within-school stratification by course type with a corresponding random sample of two classrooms from each stratum. Such data have never been created because sampling for such nationally representative large scale studies of education focus on obtaining an appropriate estimate of the national mean and not a study of variation. For those purposes the sample is totally appropriate. The presence of the within school sampling frame for all sampled schools allows us to at least obtain reasonable estimates of the variance components under general assumptions.

4 We imputed the seventh grade scores for the missing classes needed to account for the tracking structure at the eighth grade. We compared the imputed classroom means with the actual observed means for those classes where both were available. This produced a high degree of overlap thus supporting the validity of the approach.

5 The most difficult methodological issue was that one must align the appropriate seventh grade class with each eighth grade class given tracking in U.S. schools. For example, the achievement of a seventh grade pre-algebra class could serve as the "pre" measure for an eighth grade algebra class but not for an eighth grade regular mathematics classroom. This was not a trivial problem since only one seventh grade classroom was sampled at each school. We were forced to deal with this problem statistically. We conducted a conditional multiple imputation to impute class means at the seventh grade. The TIMSS sample contained 183 schools with 367 classes. The classes were cross-classified by school, grade, and class type combined with stream, yielding a total of 2928 cells (183 schools by two grades by eight class types/streams [three class types by three streams yields nine categories reduced to eight]). The grades were seventh and eighth. The class types were regular, pre-algebra, and algebra. Streams were defined using a school-determined ability ordering of specific courses. Many cells were empty since some schools had limited offerings. In particular, seventh grade algebra classes were seldom offered. A "main class" ANOVA model was fitted to the class achievement means. The fitted model yielded an estimated value and a residual for every cell. Imputed values for missing classes were then calculated by using the fitted values for the cell of the missing class (school-grade-class type combination) plus a randomly selected value from the distribution of residuals.

6 TIMSS weights for teacher and mathematics classroom data are sums of the weights assigned to the students linked to a particular teacher/class. These student level weights were then adjusted for teacher non-response. Thus weighted teacher data are not representative of teachers but representative of the students associated with a teacher.

7 The 21 topics in the teacher content questionnaires and the tests themselves were based on the TIMSS mathematics content framework (Robitaille et al, 1993) which spelled out in detail the specific contents covered in school mathematics across approximately 50 countries. A hierarchical array of specific mathematical topics within ten broad topic areas was developed to cover K-12 mathematics. In addition to the topic aspect of content, there was also a taxonomy of expected performances on the part of K-12 students. These taxonomies were developed and used for analyzing policy and standards statements, analyzing and comparing mathematics textbooks, specifying the content teachers taught and categorizing achievement test items. They were developed in a cross-national context and received a broad cross-national consensus and, in that sense, they constitute a well-vetted tool for achieving specificity in mathematics content.

8 This and all analyses throughout the paper include school in the design so that any clustering effects are appropriately accounted for.

9 The sampling plan for the 1995 U.S. TIMSS called for drawing one 7th grade (Population 1) class and two 8th grade (Population 2) mathematics classes from each of the sampled schools. The 7th grade classroom was selected randomly with equal probability. For 8th grade, a systematic sample was drawn from a listing of the mathematics classes. If the school organized classes by track, the procedure was to "number the classes, 1, 2, etc. with the most advanced course having the highest number." Thus depending on the number of classes within a particular track one or both of the sampled classes could be in the same track. This was because the classes were not stratified by track. (U.S. TIMSS Technical Report, April 1999, pages F-4 and F-14.) This procedure was designed to efficiently estimate the population mean because it minimized the between class variance.

10 Variance components were computed using the maximum likelihood estimation procedure available from the SAS procedure PROC MIXED. Where appropriate, student level sampling weights were also used. In the case of course type, for convenience, it was treated as a component of variance.

11 The unconstrained estimated variance component was negative which was then set to zero. The other estimated variance components are conditional on that value. Additional analyses show the variation across this factor to disappear when adjusting for the other levels in the design.

12 The most straightforward and simplest way to fit this more complex model with a nested design factor, main effect (track) and several covariates, SES, seventh grade achievement and IGP was using generalized least squares procedures with the 249 classroom means. The estimates of the parameters are likely to vary very little from a hierarchical linear model, as fitting the model in Table 9 in this fashion produced an estimated algebra track effect of 0.47 compared to the estimated 0.51 found in Table 9. The problem of course is with the estimated standard errors due to the clustering. The main finding here with respect to the reduction in the estimated coefficient for the track variable is clear. However, it does call into question the statistical significance of the remaining track effect which was significant only at the .046 level.

13 These are conditional variance components conditioned on each of the particular variables noted in tables 7 and 8.

The Education Policy Center
at Michigan State University

The Education Policy Center at Michigan State University

College of Education, 201 Erickson Hall, East Lansing, Michigan 48824-1034

Phone: (517) 355-4494 • Fax: (517) 432-6202

www.epc.msu.edu