Abstract Title Page
Not included in page count.

Title:

Relationship of Tracking to Content Coverage and Achievement: A Study of Eighth Grade Mathematics

Author(s):

William H. Schmidt, Leland S. Cogan, Richard T. Houang, Michigan State University;

Background/Research Question/Focus of Study

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 8th grade mathematics. The great need in this area is shown, at least in part, by the mathematics performance of U.S. 8th 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 8th 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). 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, et. al., 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 students into rigid curricular programs (e.g., college-preparatory, general, vocational) that spanned all academic subjects (Lucas, 1999; Oakes, 1985). Today, 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, schools now differentiate students *within* subjects (Lucas, 1999). This implies that two students in the same grade taking mathematics may be in two substantively different mathematics classes such as pre-algebra 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 is therefore 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 mathematical content. Tracking in mathematics, as it is typically conceived, implies that some students will eventually have an opportunity to learn advanced mathematics content such as calculus 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). Many studies analyzing the effect of tracking on achievement have had several limitations. To begin, many studies using large, nationally representative data sets such as the National Education Longitudinal Study (NELS) or High School and Beyond (HSB) have 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. 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, et. al, 1996; Hallinan, 1992; Useem, 1992a). Consequently, studies focusing solely on high school tracking may mask tracking's earlier achievement effects. Prior research has shown that 8th grade tracking affects students in two important ways. The 8th grade is often one of the first years of formal tracking and thus one of the first years for students to obtain a position in a course sequence. Some of these positions will facilitate students' eventual transition into advanced mathematics courses such as calculus or trigonometry, which will in turn assist their entrance into college. Other 8th grade positions will lead the majority of its students to less advanced mathematics courses. Eighth grade tracking also differentially affects students' achievement growth. Students in high-tracked classes tend to learn more than their low-tracked peers.

Setting/Population/Participants

Data from TIMSS allow us to address this issue. TIMSS includes a large, nationally representative sample of more than 13,000 7th and 8th grade students, actual course indicators, achievement results, and perhaps most importantly, within course topic coverage measures. These data allow us to ask 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 the instructional content on the effects of tracking at a pivotal point in a student's educational career, namely 8th grade mathematics.

Course Differentiation

Like most samples at this grade level, the TIMSS sample was designed to be representative of the U.S. as a whole but was not explicitly designed to deal with the widespread stratification created by U.S. tracking policies.

The class tracking form, a byproduct of the U.S. TIMSS sampling design, provides the best data yet to address this deficit. As part of the within-school sampling, schools listed all of their 7th and 8th grade mathematics classrooms along with the class titles and the number of students enrolled in each. This was used to draw the sample but it also provides complete tracking information on selected schools which themselves comprised a nationally representative sample of 7th and 8th grade students. Using this information, it was possible to specify the within-school course-offering structures (Cogan, et.al., 2001). This within-school course offering structure defines tracks for purposes of this paper.

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. The test was the same for students in both the 7th and 8th grades. IRT-scaled mathematics scores were created across all countries to have a mean of 500 with a standard deviation of 100. For analysis purposes we made the assumption that the cohort of 7th grade students in a school was essentially the same as the cohort of 8th grade students from that same school other than the 8th 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 7th grade score as a pseudo-pre-measure to examine the effect of tracking on student learning at the 8th grade. Through the school mathematics class tracking forms, the appropriate 7th grade class which served as the feeder to each 8th grade class could be identified.

Content Coverage

TIMSS also surveyed the mathematics teachers of the sampled classes They were asked to indicate the number of periods over the year in which they taught each of 21 mathematics topics.[i] 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.[ii]

While the focus of this paper is on the US 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 mathematical modeling. This index was 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 one and 12 indicating the grade, averaged across almost 50 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 and construct validity (Achieve, 2004, Cogan et al, 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 some 50 countries. Thus, the IGP estimates 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 22 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 rigor of the implemented curriculum in mathematics for each teacher. This is what was used to characterize content coverage in this study.

Results

An analysis of the classroom tracking forms revealed two types of schools. This includes those that offered a single type of mathematics course to all 8th grade students. The second type of school offered multiple courses or tracks into which different students were assigned.

Non-Tracked Schools. Approximately 27% of US 8th grade students attended a school in which there was only one course available to them in mathematics (Cogan et al, 2001). Although they 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 8th grade in that school.

Tracked Schools. The other type of school attended by the vast majority of 8th grade students (73 percent) offered two or more different types of mathematics courses or tracks covering different aspects of mathematics for different 8th 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 US 8th graders who attended tracked schools. Some schools did offer those three tracks (attended by around one-fourth of all 8th graders) but other schools offered different paired combinations of the three types with the most common being regular mathematics and algebra which was attended by one quarter (25.2 percent) of the 8th graders.

Track Differences in Content Coverage

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). The estimated contrast of the algebra course with the combination of the pre-algebra and regular mathematics courses was statistically significant. Using a 95% confidence interval, the estimated value indicated an almost one year difference (.85) between 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 < .07). Thus, in spite of the presumed difference implied by the 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.

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 < .10). The means were almost identical—7.35 vs. 7.22. 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 (1.08) than the regular track and almost one year (.81) more advanced than the pre-algebra track. The estimated effect size indicates about a quarter of a year (.27) difference for the content difficulty between the pre-algebra and regular tracks. These results are generally consistent with the analysis cited over all schools at the beginning of this section except that the significant difference between pre-algebra and regular classes was only marginally significant over both types of schools.

For the non-tracked schools the same pattern emerges with respect to algebra. The content difficulty of the coverage for classrooms in schools (n=10) that offer only algebra is significantly different from the coverage for classrooms (n=69) that are in schools that offer only regular mathematics (p < .05). That difference is only about a third of a year. However, the average values of the IGP index for the algebra classes offered in non-tracked schools is about one-half a year (p < .003) 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.

Relationship of Tracking to Achievement

A fairly clear pattern of achievement differences across classrooms emerges indicating 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 and is not particularly surprising. This analysis did not control for the selection bias introduced by the fact that students were not randomly distributed across the different tracks within schools.

To explore this issue more fully a three level hierarchical linear model was fitted separately for the tracked schools. The three levels included schools, classrooms nested within schools and students nested within classrooms. The track designation was included as a classroom level variable. The model also included several covariates at each of the levels in the design. The student-level model included racial/ethnic identity and the composite SES measure. The class-level model included the appropriate 7th grade pre-measure, mean SES, and track. The school-level model included the school-level mean SES, the percent minority enrollment at the school, the location of the school (rural, suburban, or urban) and the size of the school (the number of 8th grade students).

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 7th 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.

Recall that the standard deviation for the mathematics score used in these analyses is 100. Thus the estimated statistical significant effects indicate about one-third of a standard deviation difference in achievement between each of the three tracks. This implies two-thirds of a standard deviation difference in performance between the typical regular track student and his/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. 8th grade performance and the performance of two of the top achieving countries – the Czech Republic and Flemish-speaking Belgium.

The analyses presented in the previous section demonstrate the effect of tracking on achievement gain from 7th to 8th grade even 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 rigor of the content coverage. A parallel HLM analysis was done to address the issue of how these effects occur, which included the IGP index of content rigor at the classroom level together with the tracking variables. The IGP index was positively and statistically significantly related to the residual gain in achievement, controlling for all the variables including SES both at the individual and at the composite level ($p<.05$).

Central to the point of this analysis is the fact that 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. Controlling for instructional content – one of the proposed mechanisms by which tracking has its effect – the estimated track effect for algebra is further reduced by about 25 percent to slightly more than .4 of a standard deviation. This is a clear indication that instructional content is one of the mediating factors in how tracking influences academic achievement.

Discussion/Conclusion

In this article we set out to examine the effect of tracking in mathematics. Previous analyses have shown it to be quite prevalent at 8th grade unlike much of the rest of the world. The US 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 present data as to 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. In light of these results, "algebra for all" in 8th grade seems more reasonable than continuing the current tracking policies which by design only exacerbate differences and as a result leave many behind.

Appendixes
Not included in page count.

Appendix A. 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.

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.

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: Vol. Volume I: Design and Development* (pp. 4-1-4-17). Chestnut Hill, MA: Boston College.

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

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis, 22* (3), 241-254.

Lee, V. E., & Bryk, A. S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education, 61* (2), 78-94.

Lindquist, M., Dossey, J. A., & Mullis, I. V. S. (1995). *Reaching Standards: A Progress Report on Mathematics* (Policy information perspective No.). Princeton, N.J.: Policy Information Center, Educational Testing Service.

National Center for Education Statistics (NCES). (1994). *Curricular differentiation in public high schools* (No. NCES 95-360). Washington, D.C.: U.S. Department of Education. National Center for Education Statistics.

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., 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 Pub. Co.

Oakes, J., & Lipton, M. (1990). Tracking and ability grouping: A structural barrier to access and achievement. In J. I. Goodlad & P. Keating (Eds.), *Access to knowledge: An agenda for our nation's schools:* (pp. 187-204). New York: College Entrance Examination Board.

Oakes, J., with Ormseth, T., Bell, R., & Camp, P. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: RAND.

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.

Riley, R. W. (1997). *Mathematics equals opportunity* (White Paper No.). Washington, D.C.: U.S. Department of Education.

Schmidt, W. H. (2002, June, 2002). *Too little too late: American high schools in an international context.* Paper presented at the meeting of, Brookings Institute.

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: Vol. Volume I: Design and Development* (pp. 5-1-5-22). Chestnut Hill, MA: Boston College.

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.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*: 323-355.

Spade, J. Z., Columba, L., & Vanfossen, B. E. (1997). Tracking in Mathematics and Science: Courses and Course-Selection Procedures. *Sociology of Education, 70* (2), 108-127.

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

Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States* (No. NCES 99-074). Washington, DC: U.S. Department of Education. National Center for Education Statistics.

End Notes

[i] 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.

[ii] 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 almost 50 countries. An 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.