DOCUMENT RESUME

ED 363 509 SE 053 810

AUTHOR	Hoffer, Thomas B.; Gamoran, Adam
TITLE	Effects of Instructional Differences among Ability Groups on Student Achievment in Middle-School Science and Mathematics.
INSTITUTION	Center on Organization and Restructuring of Schools, Madison, WI.; Northern Illinois Univ., De Kalb.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Apr 93
CONTRACT	MDR-8550085; R117-Q00005
NOTE	63p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Ability Grouping; *Academic Achievement; Educational Research; Inquiry; Junior High Schools; Junior High School Students; Lecture Method; *Mathematics Education; *Middle Schools; Problem Solving; *Science Education
IDENTIFIERS	Middle School Students

ABSTRACT

This paper analyzes the effects of ability grouping on middle school math and science achievement, attempting to account for these effects in terms of measurable classroom experiences of students. It is hypothesized that grouping effects operate through classroom instructional differences. Data collected from a national sample of public school students and teachers reveal large effects of group placement on 8th-to-9th grade achievement growth, and substantial track differences appear for several instructional variables. Adding the instructional variables to the achievement models shows that 75 percent of the high-ability group effect on science achievement, but only 17 to 33 percent of the mathematics track effects are accounted for by the instructional variables. The largest instructional effects on mathematics achievement are associated with emphases on problem solving and understanding principles (as opposed to computational mechanics), the pacing of the course, and the use of lectures to present the material. The largest effects on science achievement are found for measures of "inquiry-based" instruction, the use of lectures and small-group projects, and student assessments of the accessibility of the subject matter. (Author)

trkmech.f93
4/7/93

FINAL DELIVERABLE

Effects of Instructional Differences among Ability Groups on Student Achievement in Middle-School Science and Mathematics

Thomas B. Hoffer

Social Science Research Institute
Northern Illinois University
DeKalb, IL

Adam Gamoran

Center on Organization and Restructuring of Schools
University of Wisconsin
Madison, WI

April, 1993

Address correspondence to Thomas B. Hoffer, Social Science Research Institute, 148 N. 3rd St., Northern Illinois University, DeKalb, IL 60115. Hoffer's work on this paper was supported by the National Science Foundation (Grant No. MDR-8550085). The authors thank Maureen Hallinan, Sally Kilgore, and Linda Pifer for their constructive criticisms and advice on an earlier draft. Cynthia Nelson provided assistance on several phases of the data file construction and analysis. Gamoran's work on the paper was supported by the Center on Organization and Restructuring of Schools, Wisconsin Center for Educational Research, University of Wisconsin-Madison, which is supported by a grant from the Office of Educational Research and Improvement (Grant No. OERI-R117-Q00005). Any opinions, findings, or conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

BEST COPY AVAILABLE

Effects of Instructional Differences among Ability Groups
on Student Achievement in Middle-School Science and Mathematics

ABSTRACT

This paper analyzes the effects of ability grouping on middle school math and science achievement, attempting to account for these effects in terms of measurable classroom experiences of students. The authors hypothesize that grouping effects operate through classroom instructional differences. Data collected from a national sample of public school students and teachers reveal large effects of group placement on 8th-to-9th grade achievement growth, and substantial track differences appear for several instructional variables. Adding the instructional variables to the achievement models shows that 75 percent of the high-ability group effect on science achievement, but only 17 to 33 percent of the mathematics track effects are accounted for by the instructional variables. The largest instructional effects on mathematics achievement are associated with emphases on problem solving and understanding principles (as opposed to computational mechanics), the pacing of the course, and the use of lectures to present the material. The largest effects on science achievement are found for measures of "inquiry-based" instruction, the use of lectures and small-group projects, and student assessments of the accessibility of the subject matter.

Effects of Instructional Differences among Ability Groups

on Student Achievement in Middle-School Science and Mathematics

What distinguishes "low groups" from "high groups" in ability-grouped schools? Most research on grouping has focused on differences in learning outcomes, and usually finds that high-group students learn much more than low-group students. A large part of the difference is due to the more advantaged backgrounds of higher group students, but several studies have concluded that some of the difference is due to group membership itself. As is true of many other types of group differences, survey researchers have not had much success in accounting for the effects of ability grouping. Presumably, the level of a group affects classroom processes, and these classroom differences contribute to inequality in achievement. If the relevant classroom processes cannot be identified, though, then one might well be sceptical of the original claim for grouping effects on learning, since the unexplained effects could amount to nothing more than unmeasured selection factors.

Current policy interest in the consequences of ability grouping give a clear practical importance to these issues. Ability grouping in the middle and upper secondary grades is increasingly viewed as a structure producing inequalities in student educational outcomes. Some writers emphasize the need to find alternatives (Oakes and Lipton, 1992), while others call for improving the way ability grouping is implemented (Gamoran, 1991). Given the current widespread use of ability grouping, it is essential to document how its impact occurs, to help guide future reforms.

This study addresses the issue of the instructional consequences of ability grouping with recently collected data on mathematics and science education in American middle schools. The analysis makes three contributions to current research on the effects of ability grouping. First, studies of the use and consequences of grouping at the middle school level are rare, and attention to that level should contribute to a greater understanding of secondary school tracking, a subject which has received much more attention. A second contribution is that we analyze grouping within specific subjects and across classrooms, a distinctive and arguably more appropriate focus than the somewhat amorphous across-subjects-and-classrooms design of most work on secondary school tracking. Third, this study makes use of information about the conduct of classes collected from students and their teachers to assess the effects of factors hypothesized – but rarely tested – to account for the effects of group placement.

RESEARCH ON THE EFFECTS OF TRACKING ON ACHIEVEMENT

The conceptual model that guides our research is schematically represented in Figure 1. The diagram shows individual student achievement growth (a one year time period is analyzed here) to be most directly affected by learning opportunities provided in courses. We distinguish two general types of opportunity variables. The first, shown on the right-hand side of the diagram, consists of instructional variables: curricular objectives or emphases, time use ("quantity"), and presentational methods ("quality"). The second type consists of what are more properly considered "resource" variables, or what classes are given by the schools which support them. These resources include the background and expectations of the teacher assigned to teach the student's class, as well as factors such as class enrollment size, access to laboratory facilities, and the amount of time scheduled for classes to meet. We hypothesize that both types of opportunities are in turn affected by the ability level or track of the class. Finally, Figure 1 shows that student background variables affect ability group placement. For present purposes, we consider background to include variables determined prior to the start of the achievement growth period (fall semester of the 8th grade in our analysis) which affect growth. Background thus encompasses social-demographic variables such as gender, race-ethnicity, and socioeconomic status; family socialization variables such as parental expectations for school success and further study; and student academic background variables such as initial levels of academic interest, achievement, and grades.

Our review of prior research suggests that the hypothetical model shown in Figure 1 is problematic in three respects. First, there is some question over whether ability group placement really has any effects on student achievement. Second, student experiences and learning opportunities have been shown to vary with ability group rank, but it is not clear how much these differences reflect real effects of group rank as opposed to pre-existing differences among students. Third, the links between the kinds of classroom-based learning opportunities that observers have focused on and student achievement are largely unexamined by research using sample surveys and correlational methods.

Track Effects on Learning

For present purposes, we consider effects of ability grouping or tracking on achievement to consist of outcome differentials between groups that cannot be accounted for by student background differences. While our analytic focus is on ability groups within particular subject domains, specifically eighth-grade mathematics and science, the general model shown in Figure 1 is also applicable to high school tracks that cut across subjects and extend over the entire high school grade span. Most studies of curriculum tracking in secondary schools have found that inequality of achievement across tracks increases over time among students with similar academic and social backgrounds (see Gamoran and Berends, 1987, for a review). These studies mainly relied on global indicators of track positions, such as academic, general, and vocational programs, usually drawn from student reports. A limitation of this type of measure is that in many schools, students are grouped by performance into different classes within such tracks, and furthermore many schools do not have such broad programs at all, though they still sort students by ability for specific subjects (Oakes, 1985).

Studies of ability grouping within subjects have yielded mixed results, with some showing increasing inequality, but others not. In a review, Slavin (1990) concluded that the average effects of ability grouping versus heterogeneous assignments on achievement in secondary schools is zero, and the so-called "differential effects" that appear in some studies -- whereby high-ability students appear to benefit while low-ability students are harmed -- reflect selection bias, that is, uncontrolled differences among students assigned to different groups. Because students are assigned to ability groups on the basis of past and expected future performance in school, what appears to be increasing inequality may simply reflect the impact of pre-existing differences.

Without minimizing the importance of addressing selection bias, another interpretation of the mixed results which Slavin (1990) underscored is also possible. Most of the studies that Slavin reviewed were small-scale and of limited duration. It may be that grouping has different effects under different circumstances; for example, in one school, students at all levels may receive the same instruction, while in another, high-group students may have opportunities not granted to other students. Differing results may thus reflect varied implementation, rather than measurement error. Hoffer (1992a) found that the effects of middle-school ability grouping do in fact vary from school to school, but stopped short of giving an explanation of that variability. Testing an elaborated model of between-school variation in the effects of high school tracking, Gamoran (1992) found that the effects of academic track placement on math and verbal achievement varied between high schools, being larger in schools characterized by lower levels of track mobility and smaller relative academic track enrollments.

The inconsistencies in the numerous small-scale studies Slavin (1990) reviewed thus may reflect variability in the ways grouping is implemented, but it would still be useful to have a summary picture of grouping effects based on a national sample. Unfortunately, national survey data on subject-specific ability grouping have not been available for the United States until quite recently.

Drawing on a panel study of national sample of students which began in 1987, Hoffer (1992b) found consistent evidence for differential effects in middle-school science and mathematics using several different methods of controlling for pre-existing differences. These results compared quite closely with Kerckhoff's (1986) findings from an analysis of a large national sample of British students. Kerckhoff (1986) reported that between ages 11 and 16, students assigned to high-ability classes gained, while students assigned to low-ability and remedial classes fell further and further behind. These findings held despite a variety of controls for pre-existing differences among students assigned to different types of classes.

Track Effect Mechanisms

While evidence of track effects on achievement is found in several studies, the mechanisms generating the effects are less well documented (Oakes, Gamoran, & Page, 1992, p. 599). The most widely-cited study of tracking in recent years (Oakes, 1985) presented a wealth of useful descriptive information along with provocative interpretations of how tracking is implemented, experienced, and understood by students and teachers, but did not pursue quantitative analyses of tracking effects on student outcomes and their mechanisms. On the basis of Oakes' (1985) and other descriptive accounts of track differences in classroom activities, Gamoran (1989) argued that variation across tracks in the quality and quantity of instruction was likely the most powerful mechanism through which tracking produces unequal achievement. In the present study, we examine variation across tracks in three aspects of students' opportunities for learning: instructional objectives, quantity of instruction, and quality of instruction.

Instructional objectives refer to the emphases on specific types of cognitive and affective outcomes reflected in curriculum and instruction. Oakes (1985, ch. 4) showed that lower-track classes emphasized rote learning of formulaic knowledge ("basic skills"), while conceptualization, problem solving, and (in English classes) creative thinking were relatively more important in higher-track classes. Teachers' socialization goals also differed: in low-track classes teachers stressed obedience ("learning to follow rules") whereas they focused more on nurturing self-direction in higher-track classes (Oakes, 1985, pp. 79-90).

Our first hypothesis, then, is that track differences in achievement result in part from variation in teachers' educational goals. Of course, these goals must be somehow communicated to students or translated into action if they are to affect student outcomes. For example, an emphasis on problem-solving implies providing opportunities for students to respond to complex questions instead of simple drills. Because we are frequently unable to observe the activities undertaken to accomplish instructional goals, we specify direct effects from objectives to outcomes in Figure 1, allowing the objectives to serve as proxies for their behavioral manifestations in the analyses.

A second factor that could account for learning differences is the quantity of instruction, which includes "time on task" as well as "content coverage" as discussed in previous studies (e.g., Denham and Lieberman, 1980; Barr and Dreeben, 1983; Kilgore, 1993). In some cases quantity is measured as the amount of time allocated to instruction (e.g., the number to hours a math class meets over the course of a semester or year). Such conditions usually do not vary across classes within schools, so they are unlikely to account for ability-group differences in achievement. More subtle indicators of instructional quantity are applied time variables (e.g., time spent covering new material as opposed to administrative or evaluation tasks, the amount of homework assigned, etc.). These conditions may well vary among different types of classes within a school. Oakes (1985, pp. 98-102) found that high-track teachers report spending a somewhat larger proportion of classroom time on instruction as opposed to discipline, routines, and socializing. High-track teachers also expect their students to spend more time on homework (Oakes, 1985, pp. 102-103). Other writers have noted that high-track classes proceed more rapidly through the curriculum, covering more academic material in the process (see Gamoran & Berends, 1987, for a review).

Again, we do not find in the research literature direct tests of the explanatory power of these factors with regard to learning differences among high- and low-ability group classes. Extending the "quantity of instruction" hypothesis to an analysis of high school track differences, Gamoran (1987) found that academic track students complete more coursework in mathematics and that this could account for about half of the estimated track effect on math achievement. Differential coursetaking did not account for track effects in language arts, however, leading Gamoran to speculate that differences in the quality of instruction are more influential for reading comprehension and writing skills.

A third way in which classes and tracks may differ is the quality of instruction. Quality of instruction refers to the manner in which material is presented. This incorporates what Nystrand and Gamoran (1991) refer to as "instructional discourse," and includes variables such as the coherence of lessons, the teacher's degree of follow up on student questions and answers, and the extent of interactive engagement of ideas (discussion) as opposed to one-way transmission of knowledge and passive reception. Several recent writers have argued that students learn more when they are intellectually challenged and when their ideas are treated as serious and valuable in their own rights (e.g., Nystrand and Gamoran, 1991; Newmann, 1992). Kilgore (1993) conceptualizes instructional quality in similar terms, but includes an additional dimension of "accessibility," or the degree to which the difficulty of the materials is appropriately matched to the students' initial level of mastery.

Oakes (1985, pp. 105-112) also argues that the quality of instruction differs among ability grouped classes. The dimensions of instructional quality she examined included teacher verbal clarity, goal-directedness, and punitiveness; student reports were used to measure the constructs. While the sizes of the differences were not reported, Oakes (1985, pp. 107-109) noted that high-track students gave their teachers significantly higher ratings on these dimensions than their low-group counterparts. Gamoran and Nystrand (1992) reported that high-ability English classes were characterized by more emphasis on discussion and more open-ended questions about literature than average or low-ability classes.

Why might one expect ability groups to differ along the lines of instructional objectives, quantity, and quality? Our principal hypothesis is that these kinds of learning opportunities are likely to be affected by the student's ability group placement, but it is useful to consider the larger issue of why opportunity differences may arise. One explanation of these differences in instruction is that teachers' and students' expectations of what is appropriate differs across tracks. The basic idea here is that placement in a given track carries with it a certain status and set of peer influences which affect students' attitudes and efforts. Meyer (1980, p. 30) suggests, in fact, that the institution of tracking is legitimized to the extent that students in different tracks come to believe that they are essentially "different" and act accordingly. While none of the analyses of national survey data have been able to address this factor directly (see however Gamoran, 1986 for an indirect approach), several of the ethnographic studies have emphasized the social-psychological effects of track placement. Oakes (1985) noted that lower-track students tend to see themselves -- and are seen by higher-track students -- as less intelligent and able to learn. Other observers have argued that lower tracks are also characterized by anti-school norms which pressure students toward low levels of engagement and even disruptive behavior. This pattern of withdrawal and rebellion has generally been interpreted as a defense mechanism against the stigma of being labeled as less intelligent.

Teachers may adjust their teaching style in response to these differences in student orientations and behaviors. Some studies suggest that teachers of lower-track classes place fewer or lower demands on their students, not because they do not want to challenge the students, but because the students simply refuse to do challenging assignments (Powell, Farrar, and Cohen, 1985).

Teachers' low expectations for student achievement in that case are grounded in practical realities. Other studies have argued that teachers could make greater demands but do not, because they believe the students either are incapable of higher levels of performance, or do not need higher levels of skills and knowledge for their putative destinations (Oakes, 1985).

A second explanation is essentially a "scarce resource" argument. The main variant of this is that the quality of instruction is tied to the teachers themselves. This conception of teachers and teaching is a static one, wherein teachers have certain talents and shortcomings, and their teaching behavior flows directly from their characteristics. This view contrasts with the more dynamic view, just discussed, of teachers responding to their classes. Whatever the merits of the hypothesis that teacher background has a significant impact may be, some studies have found these background characteristics vary between tracks. Analyzing a national sample of high school teachers, Talbert (1990) found that about one-third of all teachers teach mostly classes at a single ability level. Another recent national study of secondary-level math and science teachers showed that teachers of high-track classes are more experienced and better qualified in their subjects than low-track teachers in the same schools (Oakes, 1990, pp. 62-67). Rosenbaum's (1976) case study showed that teachers were allocated primarily to one track, and that teachers reputed to be the best were assigned to the higher track classes. Rosenbaum's (1976) and Finley's (1984) case studies noted that the higher track teachers spent more time and effort preparing for classes, and worked harder in class to challenge the students.

While we will not pursue these hypotheses in an exhaustive manner here, we will examine some indicators of teacher expectations and professional backgrounds in a preliminary effort to shed some light on the main findings of our analyses.

DATA BASE AND MEASURES

The data we analyze were collected by the Longitudinal Study of American Youth (LSAY) from fall 1987 to fall 1989. The LSAY is a six-year panel study of middle and high school science and math education which began in fall 1987. The base year samples consisted of 3,116 seventh graders and 2,829 10th graders drawn from 51 pairs of middle or junior high schools and senior high schools. The schools were systematically selected with probability proportional to enrollment size within twelve sampling strata defined by region of the country (four categories) and type of community (rural, suburban, urban). Target samples of sixty students at both grade levels were then randomly selected within the sampled schools. In the fall of each year, students are administered achievement tests in math and science consisting of items drawn from the National Assessment of Educational Progress (NAEP) and a questionnaire asking about background and attitudes. In the spring of each year, the science and math teachers of each student in the sample are mailed questionnaires asking about various aspects of the specific classes in which the sampled students were enrolled. In spring 1988, all of the math and science teachers in the participating schools were also mailed a questionnaire asking about their professional backgrounds and characteristics of the their schools. The LSAY also conducts a telephone interview with an alternating parent of each student in the study each spring, collecting additional information on students' family backgrounds.[1]

Analysis Samples

The data analyzed here are drawn from the middle school student instruments and the spring 1989 teacher questionnaires. The models we estimate for mathematics and science achievement differ in some details, and the analysis sample compositions vary as a result. The math achievement

[1] The LSAY data and codebooks are available for public use at a nominal fee. For more information, contact the project at the following address: LSAY, Social Science Research Institute, Northern Illinois University, DeKalb, IL 60115.

subsample consists of 1,346 younger cohort (originally seventh grade) students who met the following criteria: (1) During the 1988-1989 school year, the student was still enrolled in one of the 46 original LSAY schools (out of the 51 total original schools) using ability grouping for eighth grade mathematics instruction; (2) the student completed the fall 1987 and spring 1989 questionnaires, and the fall 1989 science and mathematics achievement test; (3) the student's eighth grade math teacher completed the spring 1989 class questionnaire; and (4) the eighth grade math teacher completed the spring 1988 teacher background and attitude questionnaire. Students who transferred out of their original school were excluded because the LSAY did not collect ability group or teacher-class data for those respondents, even though they continued to complete tests and questionnaires. Students enrolled in the five original schools which did not use ability grouping were excluded because the the small number of cases precluded any strong conclusions.

The science achievement subsample consists of 1,235 students meeting the following criteria: (1) the student was still enrolled in one of the original LSAY schools during the 1988-1989 school year, regardless of whether the school used ability grouping in eighth grade science; (2) the student completed the fall 1987 and spring 1989 questionnaires, and the fall 1989 science and mathematics achievement test; (3) the student's eighth grade science teacher completed the spring 1989 class questionnaire; (4) the eighth grade math teacher completed the spring 1988 teacher background and attitude questionnaire; and (5) the student was not in a low- or remedial-level science class in eighth grade. We added the last restriction because after the other criteria were met, the sample included only 35 low-group students. Since this was too small a number on which to base any inferences, and since they are much different from average-group students on many measures, we decided to exclude them rather than collapse them into the average group.[2]

[2] Hoffer's (1992b) analysis of ability grouping effects using the LSAY data included the nongrouped students in mathematics and the low-group students in science. His analysis samples were larger than ours, because he did not use the teacher-class or teacher-background data used here.

While the numbers in the math and science analysis samples are close, only about 900 students are included in both subsamples. The reason is that a relatively higher proportion of students were excluded from the science subsample because of science teacher nonresponse to the class questionnaire. Despite following identical administration procedures, the LSAY realized response rates of .71 for the science and .76 for the math teacher-class questionnaires in spring 1989. Thus, while some students were excluded from the math analysis because they were in the small number of schools that did not use tracking in eighth grade math, this number was roughly matched by the loss of cases from the science analysis due to greater teacher nonresponse.

The selection criteria taken together resulted in a sizable loss of students and a potential sample selection bias in the estimation of effects. We address this problem by including an adjustment term in our achievement equations, which in effect controls for unmeasured factors associated with the probability of exclusion from the sample and with achievement (Berk, 1983).[3]

Measures

Academic achievement

Appendix A contains descriptions of the measures used in the analysis, including names of the variables as they appear in the LSAY public-use files. Cognitive achievement in science and mathematics are measured here as composite scores summarizing student performance in the different domains tapped by the NAEP-derived LSAY tests. The scores were estimated by Item Response

[3] In addition to the analyses presented in Tables 3 and 4, we estimated a number of other regression equations to test the robustness of the results. By selecting the analysis subsamples in terms of the presence of different combinations of available data, the approximate case bases can range as high as N=2,200 for some the estimates of ability group effects on achievement ("Model 1" in Tables 3 and 4), N=1,800 for estimates of resource effects ("Model 2"), and N=1,700 for estimates of instruction effects ("Model 3"). The results and conclusions presented in the present paper were not appreciably altered by the alternative specifications. The main difference was that the lower case base used in Tables 3 and 4 results in lower levels of statistical significance and thus more conservative estimates.

Theory (IRT) methods, whereby individual test items are first calibrated and then students are scored according to their response patterns. The scores were scaled so that the average score of the seventh graders was equal to 50 with a standard deviation of 10 (see Miller, et al, 1992 and Suchner, 1988 and 1989 for detailed discussion of the LSAY test construction procedures).

The LSAY public use files include achievement scores with imputed values for cases with missing data. We used these imputed values for cases with missing data on the controls for prior achievement (the 1987 and 1988 science and math tests), but not for the ultimate dependent variables (the 1989 tests). The imputations were made with the "nearest neighbor hot deck" method described by Little and Rubin (1986). Cases with missing data ("recipients") were matched with complete cases ("donors") according to the students' patterns of course work in science and math, and their completed achievement tests. Imputed values for the recipients were then obtained by randomly drawing a single donor from the pool of matched donors. Approximately 4 percent of the cases in our analysis subsamples had one or more imputed scores on the measures of prior achievement.

Ability group placement

The ability levels of the students' mathematics and science classes are measured primarily from master schedules collected each year from the sampled schools. The students provide information on their courses and teachers each semester, and this information is matched with the school schedules to code the courses and teachers. In most cases it was clear from the schedule whether the school used ability grouping and, if so, the levels of each class that semester. If the schedule was not clear, or if teachers provided contradictory information about the level of the class in their questionnaire responses, phone calls to the school were made to clarify the situation.

Student academic and social background

Student background is represented by five variables: prior academic performance, SES, student gender, race-ethnicity, and parental expectations for success in mathematics and science. In addition to the achievement measures discussed above, we used students' self-reported grades as additional controls for prior academic performance. As with the measures of prior achievement, some imputed values were calculated for cases with missing grade data. Here we used a simple OLS regression method, whereby grades were regressed on prior achievement and social background variables, and the regression coefficients then used to calculate predicted grade values for cases with missing information. About 10 percent of the cases have imputed data from this method.

We constructed a composite SES measure by combining information from indices of parental education, occupation, and household possessions. The measures of parental expectations for math and science education were constructed by summing students' baseyear (fall 1987) responses to two sets (one for math and another for science) of three checklist items asking about whether their parents (1) "... have always encouraged me to work hard on math [science]," (2) "... expect me to do well in math [science]," and (3) "...think math [science] is important." Student gender and race-ethnicity were identified from student questionnaire items administered in the base year. Missing data on gender were coded by a check on student names; calls to the schools were made for cases with ambiguous names. Missing data on race-ethnicity were filled in by using school records if available.

Class resource variables

Class resources refer to variables that are allocated to classes either by direct administrative decisions or by more diffuse cultural mechanisms. They do not refer directly to what transpires in classes, but they are likely to affect classroom activities and outcomes. For present purposes, we include two types of variables under this general heading: teacher and class characteristics. Our measures of teacher characteristics are limited, but are consistent with prior research and current policy discussions. One set of measures includes the teacher's professional background, indicated by educational background and years of teaching experience. These variables are measured by teacher self-reports collected in the LSAY 1988 teacher background survey and subsequently merged to the student records on a class-by-class basis. Background indicators include whether the teacher majored in the subject (math or science), the number of college semester-equivalent courses the teachers completed in their subject and in science or mathematics education courses, and the number of years the teachers have taught their subject (science or mathematics).

A second set of teacher characteristics includes indices of teacher expectations. We rely on student reports for these measures. The student questionnaires administered in the spring semester of the eighth grade asked the students to indicate agreement or disagreement with several propositions about their math and science teachers. These questions focused on the teachers' encouragement and expectations. We conducted exploratory factor analyses of the student responses to class-related questions to determine if a smaller set of underlying factors could summarize the larger sets of related indicators. Principal components analyses found that the students' responses to the questions about their mathematics and science classes were patterned very comparably in both subjects. Three factors emerged from the battery: teacher academic push, teacher effectiveness, and teacher's career push (see Appendix A for a list of the specific items used in each scale). Although we used factor analysis to help detect coherent instructional indicators, we constructed the scales as unweighted additive composites.

Class characteristics are measured by three indicators from the spring 1989 science and math teacher class questionnaires. For both subjects, we use the teacher's report of the number of students enrolled the class, and the number of minutes per week the class met. For science, we also use an indicator of whether the course had an additional laboratory class, and if so, how many minutes of laboratory the class had each week.

Instructional variables

We drew our measures of classroom instruction from the spring 1989 surveys of the students' math and science teachers, and from the fall 1988 and spring 1989 student questionnaires. The math and science teacher surveys collected data on the mathematics and science classes of about 70 percent of the sampled students. The questions focused on class objectives, teaching methods, and classroom time use.

Instructional objectives. Our measures of instructional objectives are taken from the teacher questionnaires. The teachers were asked about the relative emphasis they accorded each of several general objectives, rating each from 1=no emphasis to 4=heavy emphasis. Factor analyses of these batteries found that only two multiple-item scale emerged from the teacher reports on instructional goals. From the science teachers' questionnaire, we identified an emphasis on inquiry learning factor measured by five items, and an emphasis on science and society measured by two other questions from the same battery. Both scales were constructed as unweighted averages of the component items. We also used one single-item indicator of objectives from this battery, a question asking how much emphasis "Teaching science facts and principles" received.

Our measures of instructional objectives in mathematics classes consist of single-item indicators, since no interpretable scales emerged from preliminary factor analyses. These items (see the Appendix for specific wordings) generally follow the same lines as the science class inquiry scale, measuring the instructional emphasis on "basic," in this case, computational, skills and "higher order," problem-solving skills.

Instructional quantity. The measures of instructional quantity are also drawn mainly from the teacher questionnaires, but are supplemented with some student reports. Teacher reports of textbook coverage, the percentage of class time devoted to covering new material, and the average amount of homework they assigned each week are the primary measures. We supplement these with student reports of their absenteeism over the school year (from the spring questionnaire) and the amount of homework they had in the class over the year (an average of the student's fall and spring reports). Absenteeism is included here as a control variable, since it is associated with ability group placement, instructional quantity, and learning outcomes.

Instructional quality. Instructional quality is measured with teacher and student reports. The teachers were asked to indicate the hours per week they devote to various kinds of instructional activities, including teacher-directed interactions with the class (lectures and discussions) versus small group work and the individualized activities of seatwork, one-on-one help, and computer tutorials.

We also use two measures of instructional quality from the student questionnaires. The spring 1989 student questionnaires asked the students to grade their math and science classes in terms of how much they like the subject matter, the teacher's clarity, the intellectual challenge, the textbook's clarity, and the overall difficulty. Only one multiple-indicator factor emerged from a principal components analysis of the student course evaluation questions. This factor is reasonably characterized as the intellectual accessibility of the course (the constituent items are listed in the appendix). We also use a single-item indicator of the intellectual challenge of the course, taken from the same battery: "How much does this course challenge you to use your mind?"

STATISTICAL MODELS

Our main hypothesis argues that the effects of group placement are mediated by class resource and instructional variables. Testing this hypothesis presupposes that group placement indeed has an effect on student achievement, and our first task is thus to specify and estimate a model of group effects. As the discussion of analysis samples indicated, our group effects model differs somewhat for mathematics and science. This is because virtually all of the schools in the LSAY sample use ability grouping in eighth grade math but only about half of the schools use ability grouping in eighth grade science. For mathematics, the group effects hypothesis can be formally represented by the following regression model:

$$\text{Math Achievement}_i = \alpha_1 + \beta_1(\text{Background})_i + \beta_2(\text{High Group})_i + \beta_3(\text{Low Group})_i + \epsilon_{1i} \qquad (1)$$

where the dependent variable is student i's level of achievement at the beginning of ninth grade, Background includes multiple measures of prior achievement, grades in school, and social background; and High Group and Low Group are dummy variables indicating the student's ability group placement. The omitted category against which the effects of the group placement dummies is interpreted is the middle ability group.

Group effects on science achievement are estimated for all students (except the small number of students enrolled in low tracks) whether or not enrolled in a grouped school. These estimates are obtained by adding an additional term to represent the students attending schools not using ability grouping in eighth grade science:

$$\text{Science Achievement}_i = \alpha_2 + \beta_1(\text{Background})_i + \beta_2(\text{High Group})_i + \beta_3(\text{Nongrouped})_i + \epsilon_{2i} \qquad (2)$$

The omitted reference category for the dummy variables is thus the average-level students attending schools using ability grouping in eighth grade science. As discussed earlier, we excluded low-track students from the science analyses because of the very small sample size for this group in the LSAY (n=35, after other data availability restrictions were applied).

The theoretical and empirical research on tracking suggests that the effects of the class resource and instruction variables on achievement can be estimated with simple additive elaborations of equations (1) and (2):

$$\text{Math Achievement}_i = \alpha_3 + \beta_1(\text{Background})_i + \beta_2(\text{High Group})_i + \beta_3(\text{Low Group})_i + \beta_4(\text{Class Resources})_i + \beta_5(\text{Instruction})_i + \epsilon_{3i} \quad (3)$$

$$\text{Science Achievement}_i = \alpha_4 + \beta_1(\text{Background})_i + \beta_2(\text{High Group})_i + \beta_3(\text{Nongrouped})_i + \beta_4(\text{Class Resources})_i + \beta_5(\text{Instruction})_i + \epsilon_{4i} \quad (4)$$

Our analyses also address two other important sources of error that much of the prior work on grouping effects has ignored. The first might be called "treatment selection bias," and it concerns the estimates of the group placement effects in equations (1) and (2). Campbell and Erlenbacher (1970) criticized correlational analyzes of the effects of "high" versus "low" program placements on achievement growth for relying too heavily on single pretest controls for initial differences among students The problem is that measurement error on the pretest leads to inflated estimates of (positive) high-group and (negative) low-group effects. This claim could be assessed by including an adjustment for pretest unreliability into the model. Unfortunately, it is usually not possible to determine exactly the reliability of the pretest (and other covariates). The main alternatives for making this adjustment are to use published test reliability estimates or to use multiple indicators of the unobserved "true score." Since the former strategy is not well suited to panel models with repeated measures (Jencks, 1985), we rely on the latter strategy. Thus for both the science and mathematics models, we include seventh and eighth grade math and science test scores plus students' self-reported grades as controls for prior differences, in addition to the standard social background variables. Although no set of controls can eliminate treatment selection effects with certainty, our controls are more extensive than those in most previous non-experimental work on tracking. In other studies with equally extensive sets of controls, Gamoran and Mare (1989) and Hoffer (1992b) showed that by including multiple pretest scores, estimates of bias due to nonrandom selection to tracks could be minimized.

A second potential source of error in this analysis is tied to the problems of nonresponse among students and their teachers. This problem can be thought of as one of "sample selection bias." Analyses not presented here show that nonresponse is not random with respect to the variables in the models, being more likely for lower-achieving and lower-SES students. Precisely how the loss of these cases affects the estimated effects of the variables in the model cannot be determined a priori, and over the past decade analysts have developed special techniques for addressing the problem of potential biases (cf. Berk, 1983 for a discussion). These methods essentially involve estimating two equations, a "selection" model and an "outcome" model. The selection model is specified for all students. It gives estimates of the effects of the independent variables on the likelihood of being excluded from the analysis subsample, and these estimates are then used to assign each respondent a predicted probability of being included in the analysis. The outcome model is estimated for the subset of cases that has complete data. It has the form shown in equations (1) through (4) but also includes the predicted probabilities from the first model. One is thus able to determine whether the likelihood of exclusion is related to the outcome, and to control for the effects of whatever correlation that likelihood has with the other independent variables.

If the selection equation were a linear regression with the same independent variables as the outcome equation, the outcome equation would not be identified. Identification is ordinarily achieved through either or both of two methods: use of a nonlinear logistic or probit estimator for the selection equation, and exclusion from the outcome equation of one or more predictors in the selection equation. We use both methods here, using the logistic estimator for the selection model, and adding indicators of school location (dummy indices of suburban and rural versus urban) to the selection model but not to the outcome model. Urban location proves to be strongly associated with nonresponse in the LSAY data, yet location has virtually no independent effect on achievement and can thus be excluded from the outcome model.[4]

RESULTS

Grouping, Cognitive Growth, and Instruction

The first problem we address is essentially descriptive in nature, and concerns the extent to which ability groups differ on the indicators of classroom functioning available in the LSAY data. To set the context of these comparisons, it is useful to consider first the magnitude of the student achievement and background differences among ability groups. The upper panel of Table 1 (science subsample) and Table 2 (mathematics subsample) shows that average achievement scores differ greatly between ability groups, but change little from seventh to ninth grade. The differences between the high and the low groups in mathematics range from 1.4 to 1.7 pooled standard deviation units at each grade level, and the high-average group differences in both subjects are about half of those magnitudes. Most of the group differences found at the beginning of the ninth grade are thus already

[4] In addition to the school location dummies, the science subsample selection equation included the following measures (see Appendix A for desciptions): Female, SES, Black, Hispanic, 1987 Science Score, 1988 Science Score, 1986 Composite Grades, 1987 Science Grades, and 7th Grade Science Class Ability Group Placement (dummies for high, middle, and low, where nongrouped students are the omitted reference category). The science equation was estimated for 2,310 students. The model fit statistics showed that 60% of the cases had predicted probabilities of exclusion which agreed (when rounded to zero or one) with their actual status.

The mathematics subsample equation included rural and suburban location dummies, Female, SES, Black, Hispanic, 1987 Math Score, 1988 Math Score, 1986 Composite Grades, 1987 Math Grades, and 7th Grade Math Class Ability Group Placement (dummies for high, middle, and low, where nongrouped students are the omitted reference category). The model fit statistics were somewhat better than for science: 65% of the 2,272 cases were predicted correctly.

in place at the beginning of seventh grade, but grouping may still be associated with a growing inequality of what students learn.

The second panels in Tables 1 and 2 show the distribution of student social background variables by ability group. We include these variables in our analyses primarily as controls for differences among students which are established prior to ability group placement and thus classroom instruction. Nonetheless, it is worth bearing in mind that the social compositions of classes are important organizational outcomes in their own right, irrespective of whatever achievement differences may be found after controlling for background. Consistent with findings from studies of school records and many other local and national surveys, these figures show that higher tracks enroll disproportionate numbers of students from higher SES families, and enroll disproportionately lower numbers of African-American and Hispanic youth. The measures of parental expectations for academic success also favor higher-track students. As Gamoran and Mare (1989) observed in an earlier national data set, gender differences among tracks are small and favor girls for higher-group placement. With respect to social class and race-ethnicity, though, it is clear that one consequence of ability grouping is to increase the segregation of students across classes, particularly in the mathematics curriculum.

The next panels in Tables 1 and 2 give a picture of how schools allocate instructional time, students, and teachers to the different ability groups. Schools assign teachers with stronger educational backgrounds to the higher ability groups in mathematics but not in science. The LSAY data show substantial differences in college majors for math teachers of different ability groups: 64 percent of high-group math teachers were math majors compared to only 43 and 46 percent in the remedial and average classes, respectively. More of the average-group math teachers have college minors in math, but the low-group teachers are are no more likely than high-group teachers to have minors. Students in remedial math groups thus have teachers which, on average, are well below the levels of formal preparation seen among the teachers of high and middle-group students. In science, however, the inequalities are much smaller. Sixty-eight percent of the high group teachers and 61 percent of the middle group teachers had college majors in science. Virtually all the science teachers had at least a college minor in their subject.

The teachers' reports on their major field are loosely corroborated by their reports of the numbers of undergraduate and graduate courses they have taken in science. High-group science teachers have not taken more science courses than those teaching average groups. Nor is there a clear pattern of inequality across ability groups in the science teachers' reports of their coursework in secondary science education. The ability group differences are sharper in mathematics, where the variability in post-secondary coursework is lower and where there are clear differences among the high, middle, and low groups. High-group math teachers also completed more coursework in mathematics education than the middle and low-group teachers, who are about equal in that respect.

The other dimension of teacher background for which we have measures is teaching experience. There are no meaningful differences in average years of teaching experience among ability groups apparent in Table 1 or 2. We also examined two other measures of experience which are not shown in the tables: (1) total years of teaching experience, regardless of the field now taught, and (2) whether the teacher was in his or her first or second versus third or higher year of teaching overall. Neither of these indicators showed any differences among ability groups.

Average- and high-ability groups differ little in the numbers of students enrolled in classes, but remedial math classes are much smaller than the rest (21 students versus 26). In this case, ability grouping serves to allocate more resources (teacher time per pupil) to the lower group, despite the standardization of regular class instructional time across groups in mathematics. In science, however, class sizes are about equal but high group students are much more likely to have an additional laboratory session.

The other teacher resource variables refer to the more subjective realm of "teacher quality," particularly as evaluated by students. Students rate their mathematics teachers somewhat higher than their science teachers on all three of the measures used here. Both measures of teacher expectations (for academic and career push) show more favorable ratings in the higher ability groups. The other dimension we measured, "teacher effectiveness," also indicates that higher-group students give more positive ratings to their teachers on the items comprising this scale. Contrary to the image of teachers holding low expectations for low-group students, the high levels of teacher academic push (maximum value of 1.0) in all groups mean lower-group students generally report high levels of encouragement even though their levels are comparatively lower.

The lower panels of Tables 1 and 2 show the distributions of teacher and student responses to questions about the relative emphases of their courses, the time allocations in their classes, and instructional methods, broken down by ability group. The emphasis questions show that higher track science teachers place greater emphasis on problem solving and the experimental method, but that the teachers in all three ability groups report quite high average emphases on these variables. A similar pattern holds in mathematics, where we see that problem solving is emphasized more in high track classes, at the same time as it is given high priority by remedial and average classes. The reverse pattern with similar qualifications holds for the math teachers' emphases on computational skills: lower groups place significantly greater emphasis on these skills, but the substantive range on the variable is not great.

The time use variables show only small ability group differences among science classes. Higher ability science classes devote slightly less time to discipline problems, slightly more time to review, and have a half hour or so more homework per week, but the time spent on routines, new material, and testing are roughly equal. The time use patterns show stronger track differences in mathematics. Higher ability mathematics classes spend less time on discipline and review, more time on covering new material, and the students have over an hour more homework to do each week.

The next block of variables refers to instructional methods and how the classes are organized. Higher ability science classes use lectures and small group formats more, but are about equal in the use of discussions, seatwork, and computerized instruction. Teachers of higher science groups also return a greater proportion of the students' homework. Mathematics classes show different patterns than science classes on several of these variables. Higher ability math classes also use lectures more, but use small group and individualized study formats less. Corrected homework is returned less often in higher math classes, perhaps indicating a tendency to use homework more in class and less as a means of social control.

The final variables included in Tables 1 and 2 are more difficult to categorize in terms of the means-end schemes apparent in the other classroom variables. These are the more subjective student evaluations of the accessibility of the material and the intellectual challenge of the courses. Here we find that higher group students in both subjects generally give their teachers and courses more favorable ratings. The largest track differences in both subjects appear to be on the assessments of the intellectual challenge of the course, but all students consider their science and mathematics courses to be very challenging (average ratings of about 4 on a 1-to-5 scale).

In sum, the instructional variables indicate a number of substantial differences across ability groups: Teachers of higher-ability classes (1) are more likely to have majors in their teaching field, (2) are viewed by their students as holding higher performance expectations, (3) place greater emphasis on problem solving and conceptualization, (4) devote more time to covering new material and less on maintaining discipline, and (5) use whole-group instructional methods (lectures and discussions) somewhat more frequently. These descriptive tabulations show some large inequalities among classes at different ability levels. Neither the intended nor the unintended consequences of most of these differences are clear, however, and thus one cannot make a strong case for unfairness or fairness on the basis of Table 1 alone. For example, the distribution of teachers by their educational backgrounds favors the higher tracks, but it is not clear whether teachers with majors are any more effective than those without. Additionally, we do not know whether lower-track teachers could in fact implement a more problem-oriented curriculum in the face of student apathy or recalcitrance, and, if they could, whether their students would actually actually learn more. If so, then the case for unfairness in the distribution of learning goals and the attendant resources is strengthened. A prima facie case for unfairness can be made on some variables, however: Perhaps the clearest examples are the teacher encouragement items. One is hard pressed to think of legitimate reasons why teachers should give less encouragement to lower-track students than they give to higher-track students. Even here, though, it is important to remember that the questions about encouragement were asked of the students, and there may be a tendency to give more negative appraisals of teachers among students who are academically less successful, independently of the teachers' behavior.

Effects of Group Placement on Learning

Before turning to the analysis of the effects of the instructional variables on achievement growth, it is useful to establish the magnitude of the group placement effects which we are trying to explain. The columns labeled "Model 1" in Tables 3 and 4 present the results of regressions of ninth grade math and science achievement on the student background and ability group variables. Since the eighth grade achievement score is included as an independent variable in these models, the effects estimated for the social background and ability group variables are interpreted as effects on eighth-to-ninth grade achievement change, rather than as effects on the level of ninth grade achievement. These results show that ability group placement does, in fact, have a strong effect on learning in both subjects. In the mathematics equation, the reference group against which the effects of high- and low-group placement are interpreted is the middle-ability group. The estimates show that high-group students are 1.8 points above the middle group, while low-group students are 1.9 points below. These differences translate into effect size estimates of about .15 of the total sample standard deviation on the posttest for the one year interval.

The science achievement model applies to grouped and nongrouped students and thus also includes three comparison groups: high track, average track, and nongrouped. The regression equation shown in Table 3 uses the average-group students as the omitted reference category, and the coefficients on the group placement dummies are interpreted as the effects of being in the high group or in a heterogenous instructional setting versus being in the average group in an ability-grouped school. The results show that high-group students gain 1.6 more points than their nongrouped counterparts (ES = 1.6/9.5 = .16, where the 9.5 is the nongrouped sample standard deviation on the posttest). Nongrouped students, in contrast, show no significant difference from average grouped students over the one year period.

Turning briefly to the other variables in the models (the full regression results are not presented here but are available from the authors), the controls for initial achievement clearly dominate the statistical results, as one would expect. The measure of sample selectivity ("probability of exclusion" in Tables 3 and 4) has a strong association with growth in both math and science, indicating that the analytic subsample is in fact different (positively selected with respect to learning) from the full sample. It is worth noting that the effects of social background are uniformly trivial: SES, race-ethnicity, and parental expectations have no significant effects on eighth-to-ninth grade achievement growth in either subject. Any direct effects of social background appear to be confined to science achievement: Girls learn somewhat less than boys, and SES has a weak positive effect on growth.

The main point from Model 1 for present purposes is that ability group placement has effects on how much students learn in their eighth grade math and science courses. While these effects constitute only the point of departure for the main problem we are addressing, these results are noteworthy in light of recent claims about the ineffectiveness of ability grouping (Slavin, 1990). The results for mathematics presented here support the claim that grouping has differential effects, benefiting high-group students while depressing the achievement of low-group students. The results for science also show gains of high group placement relative both to average group students and to students in schools not using ability grouping.

Tracking and Classroom Instruction

The next step in the analysis involves the addition of variables describing the classroom resources and experiences of students. We divide this analysis into two segments, represented by the columns labeled "Model 2" and "Model 3" in Tables 3 and 4. Model 2 includes all of the variables from Model 1 plus measures of what we have referred to as resources: teacher educational background (whether the teacher majored in the field taught), years of teaching experience in the respective field, the number of students in the class, and the student's reports of the teacher's performance expectations and encouragement. In science, we include the presence of a laboratory period as an additional resource. Model 3 builds on Model 2 to include the full set of class variables.

Some of the variables described in Tables 1 and 2 are not included in the regression analysis. Exclusions were made on the basis of preliminary regressions in order to simplify the presentation of results. Teacher background variables excluded were the Science and Math Minor dummies, and the College Course Credits count measures. Neither showed any relation to achievement, and were dropped because they were conceptually redundant with the Science and Math Major variables. We also tested for nonlinear effects (diminishing returns) of Teaching Experience, but found that the model fit was better with the simple linear specification. Several of the Instructional Quantity variables shown in Tables 1 and 2 were also dropped from the analysis. The main factor we are concerned with in this set is "time on task," and this is most clearly represented by the Time on New Materials measure. Since Time on Routines, Discipline, Review, and Testing are in this light simply alternatives to Time on New Material, we relied on the single indicator. Similarly, we considered the teacher-reported Homework Assigned to be conceptually redundant with the student reports of their Homework Time. Since the latter is a better measure of quantity, we used it instead. Exclusions from the set of Quality measures included Seatwork Time, Individualized Time, and Teaching Machine Time. The theoretical concept behind these and the Lecture, Discussion, and Small Group Time variables concerns the direct presence or absence of interactions with the teacher. Work in small groups represents an intermediate context between teacher-directed and individualized class time. We thus excluded the Seatwork, Individualized, and Teaching Machine Time variables since we regard them as the alternative implied by the retained measures.

The results of Model 2 give little support for the hypothesis that inequalities among ability groups in the resource variables leads to learning outcome differences. The only statistically significant effect of Teacher Background in either subject is for whether the science teacher majored in science. As Table 1 showed, high- and average-ability groups differ slightly on this variable, and it thus cannot account for the effect of being in a high group on learning. The small explanatory power found in the resource variables is tied to the other indicators. Science and math class size and, in science, the use of separate laboratory session, show no significant effects. The teacher expectation variables also prove to be only weakly related to student achievement, but two indicators have significant effects: Career Push has a positive effect on science achievement, and the students' rating of their teachers' effectiveness has a positive effect on math achievement.

The results of the Model 3 regression equations show that some instructional variables have effects on achievement growth, while others do not. A key point to note first is that the addition of these variables statistically accounts for most of the estimated effect of ability grouping in science, but relatively little of either the high- or low-group effects in math. In science, the effect of high group placement decreases from +1.6 to +0.4. The insignificant difference between nongrouped students and average-group students seen in Model 1 changes very little with the elaborations of Models 2 and 3. In mathematics, the coefficients for low group placement decrease from -1.9 to -1.6 (17% of the effect accounted for), while the effects of high group placement shrink from +1.8 to +1.2 (33% of the effect is explained).

While the instructional variables do not fully account for the effects of ability group placement, several do have effects on the achievement outcomes in the expected directions. For science, Model 3 in Table 3 shows that greater emphasis on "inquiry" or experiential learning has a positive effect on learning. As Table 1 showed, high track classes are about a full standard deviation higher than the average group and heterogeneous classes on this scale, and this difference is not entirely reducible to a simple correlate of student background differences. For reasons beyond student background differences, high track teachers place greater emphasis on the various components of this scale: problem solving, experimental logic and design, systematic observation and reporting, and mathematical applications in science. The effect on learning, although statistically significant, is not enormous: Students in science classes which are a full standard deviation above the sample average of the scale are predicted to realize an advantage of only 0.05 of a standard deviation unit on the science achievement test. At the same time, that does represent about a third of the total effect of being in a high ability group. It must also be remembered that the time span of these effects is only a year (fall of eighth grade to fall of ninth grade), and one which includes a summer at that.

Other variables showing statistically significant effects on science growth in Model 3 include student absenteeism, the use of small groups, and the student's rating of the accessibility of the course materials. From Table 1, high group students have lower absenteeism and their teachers use small groups more often, and thus these factors also account for some of the reduction in the estimated high-group effect from Model 1 to Model 3. The ratings of course accessibility, in contrast, show no effect of track, and the contribution of this factor does not mediate the ability grouping effects.

The strongest effects on mathematics achievement are associated with emphasis on problem-solving, the rate of coverage, and the accessibility of the material. Of these, high-ability classes have a much greater emphasis on problem solving than average and low group classes, which tend to be about the same. The positive effect of problem solving on achievement thus accounts for some of the high-ability group effect, but none of the low-ability group deficit. In contrast, high- and average-ability groups are about the same in terms of coverage, but are higher than the remedial group. The positive effect of more coverage thus accounts for some of the remedial deficit, but none of the high group's advantage relative to the average group. The teacher's emphasis on "facts and principles," and the teacher's greater use of lecture formats, also contribute positively to achievement, whereas reported emphasis on computation tends to depress achievement. Of these, the greatest contribution to explaining the ability group differences is tied up with the emphasis on computation. In this case, Table 2 showed that the three ability groups are ordered in a clear hierarchy of decreasing emphasis with increasing ability. This factor thus accounts for both part of the remedial deficit and part of the high group advantage.

Tests for Nonlinear and Multiplicative Effects

While the models we have estimated are consistent with the conceptual model of Figure 1, one might question whether the effects of class resources and instruction fit this assumed linear, additive formulation. One alternative hypothesis is that the effects of resources and instruction on learning will depend on students' levels of engagement in their coursework: The greater the student's engagement, the greater the benefit frc a additional resources and more and/or better instruction. As Newmann (1992) has discussed, engagement is difficult to measure with survey questions. Nonetheless, student engagement is likely to be associated with ability group placement, with the result that the effects of additional resources and greater instructional quantity and quality should be higher in higher ability groups.

A second hypothesis is that our measures of class resources and instruction have different meanings in the different ability groups. For example, an emphasis on "problem-solving" could refer to low-level formulaic objectives in lower-ability groups, but more complex cognitive processes in higher-ability groups. If this is the case, then we would again expect to find different effects in the different ability groups.

A third hypothesis is that the effects of resources and instruction may show nonlinear effects due to the common phenomenon of diminishing returns to productive inputs. In contrast to those typically modeled in the production function which economists apply to the firm, diminished returns in the present case are more likely tied to students' bounded capacities or willingness to assimilate new material. In contrast to the other hypotheses, this suggests the effects of resource and instruction variables could be lower in higher-ability groups, where the levels of those variables are more likely to approach their productive limits.

These hypotheses suggest a simple alternative formulation to equations (3) and (4), one where the effects of resources and instruction are allowed to vary across ability groups:

$$\text{Math Achievement}_i = \alpha_5 + \beta_1(\text{Background})_i + \beta_2(\text{High Group})_i +$$
$$\beta_3(\text{Low Group})_i + \beta_4(\text{Class Resources})_i + \beta_5(\text{Instruction})_i +$$
$$\beta_6(\text{High Group x Class Resources})_i +$$

β_7(Low Group x Instruction)$_i$ + ϵ_{5i} (5)

Science Achievement$_i$ = α_6 + β_1(Background)$_i$ + β_2(High Group)$_i$ +

β_3(Nongrouped)$_i$ + β_4(Class Resources)$_i$ + β_5(Instruction)$_i$ +

β_6(High Group x Class Resources)$_i$ +

β_7(Low Group x Instruction)$_i$ + ϵ_{6i} (6)

We first re-estimated Models 2 and 3 with the full set of ability group-by-class variable interaction terms included. Most of these proved statistically insignificant, but a few showed effects which introduce some qualifications to the linear, additive formulation. To simplify the results, we re-estimated the equations excluding all interaction terms with significance levels of $p > .10$. For present purposes, we provide only a verbal summary of this supplementary analyses; the full regression results are available from the authors.

The results show some consistency between science and math, and some disjunctures. The estimates for Science Model 2 show that Teaching Experience has a positive effect for high-group students, but is zero or slightly negative for the others. In Math Model 2, three variables had significant interactions. First, only high-group students whose teachers had math majors learned more than students whose teachers had non-math majors. Second, larger class size has a significant negative effect for high-group math students, but no effect for anyone else. Third, teacher career push has a negative effect for low-group students, but no effect for anyone else.

The results for Model 3 show three significant interactions in both subjects. In science, Emphasis on Inquiry has a large positive effect for both high- and middle-group students (b=3.15 vs. 1.02 in Table 3), but no net effect for non-grouped students (b=-3.82; net effect= 3.15 - 3.82= .67 ns). Homework Time has a positive effect for high-group students, but no effect for anyone else.

Finally, Lecture Time has positive results for high- and low-group students, but no net effect for the nongrouped students.

Model 3 for math shows interactions for problem-solving, absenteeism, and homework. Additional emphasis on problem-solving is effective for low- and middle- but not high-ability group classes. Absenteeism hurts only high-group students, and additional homework helps only high-group students.

The coefficients in Tables 3 and 4 for the variables which did not have significant ability group interactions show no important changes with the addition of the interaction terms. Our summary assessment of the interaction results is that there is no clear pattern in the interactions, and the hypotheses we outlined have little support. The results introduce a few important qualifications, but raise further questions that we could only speculate on at this point: Why do nongrouped students receive no benefits from greater emphasis on inquiry and more use of lectures in science? Why do high math group students show no benefits from greater emphasis on problem solving, while low- and average-group students are helped?

In any case, there are some difficult interpretive issues even for the the effects which were consistent with our hypotheses. One key problem is the ambiguity around whether these interactions represent ability-by-treatment effects versus effects of differences among groups in what a nominally-identical treatment actually involves. The former effect would mean, for example, that if a mixed group of students did the same homework assignment, the slower students would get nothing out of it while the faster students learned something. The latter effect could mean, for example, that higher-group students benefit from more homework because their teachers assign them challenging problems; lower-group students, in contrast, learn nothing from their homework because the homework assigned is little more than busywork.

If the ability-by-treatment hypothesis is correct, then we might find significant ability-by-treatment interactions within ability groups. To test this possibility we added pretest (eighth grade science or math score)-by-instruction multiplicative terms to the models which already included the group-by-instruction interactions. The regression results showed that none of these new multiplicative terms had significant effects, and in no case did they explain away the significant group-by-instruction interactions. Thus we conclude that the later interactions are most likely reflections of different implementations of practices, rather than masking differences in students' capacities to benefit from those practices.

SUMMARY AND CONCLUSIONS

A large body of research has documented the importance of instructional grouping within schools for student academic success, but has stopped short of giving a systematic account of the differences among groups responsible for the different outcomes. The analysis presented here addressed this problem with recently collected survey data. These data are noteworthy in several respects: a national sample of middle schools and students, objective measures of ability group structure and student placements, reasonably reliable outcome measures, and relatively detailed data from teachers about their course objectives and routines. These data were then analyzed in terms of a straightforward ordinary least squares regression framework, which included an adjustment for sample selection bias.

The results provide limited support for the theoretical lines identified in previous research. At a descriptive level, Tables 1 and 2 showed several ways in which higher-track classes differ from lower-track classes, generally in directions which appear to favor students in the high group. Inclusion of these variables in the regression equations explained about 75% of the effect of being in a high-ability science class on eighth-grade achievement growth. Most of the statistical explanation is tied to the greater encouragement, emphasis on inquiry, and use of team projects in high-ability science classes.

The explanatory power of the class resource and instruction variables in mathematics was much less, accounting for 33% of the high-ability group effect, and only 18% of the low-group effect on growth. Most of the explained part of the high-group effect is tied to the greater emphasis on problem-solving and the lower emphasis on computation. Most of the explained portion of the low-group (relative to the average group) effect is tied to the slower pace of coverage in low-group classes.

Why does the model not account for all of the effects of ability grouping on achievement? At least three reasons seem possible. First, our controls for pre-existing conditions may not have addressed some aspects of selection into the different tracks that are associated with achievement. In that case, apparent track effects that remain could reflect differences among students instead of variation in what occurs in the different tracks. We think this is the least likely explanation, first because our controls are extensive, including four measures of prior achievement, two measures of prior grades, and the standard battery of social background controls. Furthermore, the estimated reliabilities of our academic and social background control variables are similar to indicators used in other studies.

Assuming the remaining track effects are real, a second reason our model does not explain them fully may be that our measures do not capture all aspects of instruction that differ across tracks. For example, the LSAY did not ask teachers or students which specific topics they covered. Moreover, aspects of instruction which are addressed in the data are measured imperfectly. We rely on teacher and student reports rather than on classroom observations, and we allow questions about instructional objectives to proxy for a range of important activities that may occur in class. More extensive information about instruction might allow us to explain a greater portion of the effects of tracking on achievement.

Third, there are undoubtedly other mechanisms at work, even if differential instruction is the most important as we have argued here. Assignment to different tracks may affect the way students view themselves and their schoolwork, thus influencing the amount of effort they put in, the way they behave in class, and the extent of their achievement (see Gamoran and Berends, 1987). In our model, the only indicator of effort was time spent on homework, and we included no measures of academic self-concepts or attitudes toward schoolwork. We indirectly tested for this specification error by assessing whether the effects of instructional variables were greater in the higher ability groups, but found only limited and inconsistent support for that prediction.

Aside from their value in accounting for track effects, the effects of instructional conditions are interesting in their own rights. We observed several similarities in the patterns of effects in the two subjects. In math, significant instructional effects on achievement were those of emphasizing problem-solving, textbook coverage, lecture time, and students' perceptions of the class as accessible to them. Taken at face value, these results suggest that a more conventional approach in which the teacher presents material to students is more effective than interactive or "student-centered" methods. Much more work is needed, particularly work that includes classroom observation, to illuminate these speculations.

In science, teachers who devoted more time to small-group work produced higher achievement, whereas teachers who lectured more did not consistently have greater success. Textbook coverage also had no effects on achievement. Students further benefitted when they found class materials accessible and when their teachers emphasized inquiry methods, although an emphasis on science in everyday life did not appear productive. Both the use of small groups and greater accessibility go against the grain of some traditional models of science pedagogy. These results indicate that students learn more when they work together on tasks as opposed to the traditional emphasis on individual work; the greater success associated with higher accessibility stands against the increasingly-criticized view that it is the solely the student's responsibility to make sense of the traditional canon. Again, these findings need to be supplemented by observational studies of classroom activities.

Consistent with Oakes' (1990) report, we find evidence of considerable variation within schools in resources for and implementation of secondary math and science instruction. Some of this variation appears consequential for student achievement. Our models suggest that if instruction in average and (in math) low tracks were brought to the level of that in high tracks, less inequality within schools would emerge. Our supplementary analyses of interaction effects place some qualifications on these inferences, for we did find that the benefits of a few variables do not hold in all ability groups. Generally we can say, however, that instruction which is effective on average--for example, inquiry methods in science, and more textbook coverage in math--is equally important for students in the different ability groups and at varying levels of prior performance within groups. If further work with observational data or more incisive survey questions confirms this conclusion, we could use high-track classes, which are currently more effective in both subjects, as models for redesigning instruction for all students. Lacking such information, we cannot yet be certain that this is the best course to follow.

Similarly, our models for eighth grade science achievement also hold some interesting suggestions about whether heterogeneous instruction would be more effective than current tracking systems. The science data showed that students in heterogeneous classes achieved at levels similar (actually lower, but not significantly) to those in average classes and less than those in high-track classes, controlling for prior performance levels. The descriptive data showed that these classes tend, as one might expect, to show levels of class resource and instructional variables which are intermediate between high and average classes within the ability grouped schools. Our supplemental analyses of interaction effects found that, however, that inquiry methods and lecture formats seem to work less well in the nongrouped contexts. This suggests that teachers of heterogeneous classes face special problems which require some different approaches. Hoffer's (1992b) comparison of ability-grouped and nongrouped mathematics classes in the LSAY data showed a similar pattern to science in terms of achievement outcomes, but LSAY does not have teacher data on a large enough number of heterogeneous classes to carry out the analyses pursued in this study. Hence, we need more information about the limits and possibilities of heterogeneous classes before we can say that mixed-ability teaching is less effective for some students than grouping by ability in middle school math and science.

REFERENCES

Barr, Rebecca and Robert Dreeben. 1983. How Schools Work. Chicago: University of Chicago Press.

Berk, Richard A. 1983. "An Introduction to Sample Selection Bias." American Sociological Review, 48: 386-397.

Campbell, D.T. & Erlenbacher, A.E. 1970. "How Regression Artifacts in Quasi-experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In J. Hellmuth (Ed.) Compensatory education: A national debate (Vol. 3). The disadvantaged child. New York: Brunner/Mazel.

Denham, C., and A. Lieberman, editors. 1980. Time to Learn. Washington, DC: National Institute of Education.

Dreeben, Robert and Rebecca Barr. 1988. "Classroom Composition and the Design of Instruction." Sociology of Education 61: 129-142.

Finley, M.K. 1984. "Teachers and Tracking in a Comprehensive High School." Sociology of Education, 57: 233-243.

Gamoran, Adam. 1986. "Instructional and Institutional Effects of Ability Grouping." Sociology of Education 59: 185-198.

Gamoran, Adam. 1987. "The Stratification of High School Learning Opportunities." Sociology of Education 60: 135-155.

Gamoran, Adam. 1989. "Measuring Curriculum Differentiation." American Journal of Education 97: 129-143.

Gamoran, Adam. 1991. Alternative uses of ability grouping in secondary schools. Madison, WI: Center on Organization and Restructuring of Schools.

Gamoran, Adam. 1992. "The Variable Effects of High School Tracking." American Sociological Review 57: 812-828.

Gamoran, Adam and Mark Berends. 1987. "The Effects of Stratification in Secondary Schools: Synthesis of Survey and Ethnographic Research." Review of Educational Research 57: 415-436.

Gamoran, Adam and Robert D. Mare. 1989. "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" American Journal of Sociology 94: 1146-1183.

Gamoran, Adam and Martin Nystrand. 1992. "Taking Students Seriously." Pp. 40-61 in F. Newmann (ed.) Student Engagement and Achievement in American Secondary Schools. New York: Teachers College Press.

Hoffer, Thomas B. 1992a. "A Multilevel Model of the Effects of Ability Grouping on Student Achievement." DeKalb, IL: Social Science Research Institute.

Hoffer, Thomas B. 1992b. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." Educational Evaluation and Policy Analysis 14: 205-228.

Jencks, Christopher. 1985. "How Much Do High School Students Learn?" Sociology of Education 58: 128-135.

Kerckhoff, Alan C. 1986. "Effects of Ability Grouping in British Secondary Schools." American Sociological Review 51: 842-858.

Kilgore, Sally B. 1993. "The Organizational Context of Learning: Framework for Understanding the Acquisition of Knowledge." Sociology of Education 66: 63-87.

Little, R. J. A. and Rubin, D. B. 1987. Statistical Analysis with Missing Data. New York: Wiley.

Meyer, John W. 1980. "Levels of the Educational System and Schooling Effects." Pp. 15-63 in C. E. Bidwell and D. M. Windham (Eds.), The Analysis of Educational Productivity. Volume 2: Issues in Macroanalysis. Cambridge, MA: Ballinger.

Miller, Jon D., Suchner, Robert W., Hoffer, Thomas B., Brown, Karen G., & Nelson, Cynthia. 1992. LSAY codebook: Student, parent, and teacher data for longitudinal years one through four (1987-1991). Public Opinion Laboratory, Northern Illinois University, DeKalb, Illinois.

Newmann, Fred M. 1992. Student Achievement and Engagement in American Secondary Schools. New York: Teachers College Press.

Nystrand, Martin, and Adam Gamoran. 1991. "Student Engagement, Instructional Discourse, and Literature Achievement." Research in the Teaching of English 25: 261-290.

Oakes, Jeannie. 1985. Keeping Track: How Schools Structure Inequality. New Haven: Yale University Press.

Oakes, Jeannie. 1990. Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science. Santa Monica, CA: RAND Corporation.

Oakes, Jeannie, Adam Gamoran, and Reba N. Page. 1992. "Curriculum Differentiation: Opportunities, Outcomes, and Meanings." Pp. 570-608 in Handbook of Research on Curriculum, edited by Philip W. Jackson. New York: Macmillan.

Oakes, Jeannie, and Martin Lipton. 1992. "Detracking Schools: Early Lessons from the Field." Phi Delta Kappan 74: 448-454.

Powell, Arthur G., Eleanor Farrar, and David K. Cohen. 1985. The Shopping Mall High School. Boston: Houghton Miflin Company.

Rosenbaum, Ja .s E. 1976. Making Inequality: The Hidden Curriculum of High School Tracking. New Y .k: John Wiley & Sons.

Slavin, Robert E. 1990. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." Review of Educational Research 60: 471-499.

Suchner, Robert W. 1988. "Modeling NAEP Items by Cognitive Process." Paper presented to the Annual Meetings of the American Educational Research Association, New Orleans, LA.

Suchner, Robert W. and Bianchi, Leonard J. 1989. "Two-Stage Testing in the Longitudinal Study of American Youth." A paper presented to the annual meetings of the American Educational Research Association, San Francisco, CA.

Talbert, Joan E. 1990. "Teacher Tracking: Exacerbating Inequalities in the High School." Center for Research on the Context of Secondary Teaching, Stanford University, Stanford, CA.

Appendix A. Description of Variables Used in the Analysis

Variable Name	Description (LSAY codebook variable names in parentheses)
	Student Achievement Outcomes
1989 Math Score	IRT score for composite test consisting of NAEP items (EMTHIRT).
1989 Science Score	IRT score for composite test consisting of NAEP items (ESCIIRT).
	Student Social Background
Female	Student gender (1=female, 0=male) (AA29):
Black, Hispanic	Student race (0-1) and ethnicity (0-1)
SES	Family composite socioeconomic status; constructed from parental education (maximum of MOTHED and FATHED), occupation (maximum of Duncan SEI variables BH281S, BH287S, DH267AS, and DH267BS), and household possession index (sum of BA15A, BA15B, BA15C, BA15F, and BA15I: daily newspaper, a specific place to do homework, a typewriter, a room of your own, and a weekly news magazine). The three components were separately standardized, and the standard scores averaged to form an equally-weighted composite for each student (SES3).
Parent Push	Parent push for success in science (math) classes: Student report from year 1. Sum of student's agreements with AA19F /AA19E ("My parents have always encouraged me to work hard on science /math"), AA19O /AA19N ("My parents expect me to do well in science /math"), and AA19R /AA19Q ("My parents think science /math is important") (PSCPH1, PMHPH1). The alpha reliabilities are .71 for the math scale and .73 for the science scale.
	Student Academic Background
1987 Math Score	IRT composite score (missing values imputed) (AMTHIRT).
1988 Math Score	IRT composite score (missing values imputed) (CMTHIRT).
1987 Science Score	IRT composite score (missing values imputed) (ASCIIRT).
1988 Science Score	IRT composite score (missing values imputed) (CSCIIRT).
1986 Grades	Student-reported composite GPA (missing values imputed) (AA33).
1987 Science Grades	Student-reported science grades (missing values imputed) (CA27C).
1987 Math Grades	Student-reported math grades (missing values imputed) (CA27B).
	Teacher Background
Science (math) major	Teacher majored in field (0-1) (BE87B1, BE87B2, BE88B1, BE88B2).
Science (math) minor	Teacher minored in field (0-1) (BE87C1, BE87C2, BE88C1, BE88C2).
Science (math) credits	Total # of graduate & undergrad. courses in field (TOTSCI, TOTMTH).
Sci. (math) educ. credits	Total # of grad. & undergrad. secondary education courses in field (TOTSSE, TOTSME).
Teaching Experience	Years teaching experience in field (science or mathematics) (BF2, BG2).
	Class Characteristics
Class Size	Total # of students in class (DI3C, DJ3C).
Weekly Class Time	Minutes of class per week (DI4, DJ4).

Science Lab — Additional lab period for science course (0-1) (DI5A).
Science Lab Time — Minutes of additional science lab per week (DI5B).

Teacher Expectations & Encouragement: Student Reports

Academic Push — Science (math) teacher pushes student to work hard: student reports. Constructed by summing individual student responses to the following spring 1989 true-false-don't know (coded 1, 0, -1, respectively) questions about "My science (math) teacher this semester..." (alpha =.73):

- o ... expects me to do my best all the time (DA2B, DA3B).
- o ... encourages me to do extra work when I don't understand something (DA2C, DA3C).
- o ... expects me to work hard on science (math) (DA2D, DA3D).
- o ... expects me to complete my homework every night (DA2E, DA3E)
- o thinks it is very important that I do well in science (math) (DA2P, DA3P).

Career Push — Teacher pushes student toward scientific or mathematical career: student reports. Constructed by summing individual student responses to the following spring 1989 true-false-don't know (coded 1, 0, -1, respectively) questions about "My science (math) teacher this semester..." (alpha =

- o ... has talked to me about the kind of job I might want to do (DA2G, DA3G).
- o ... expects me to go to college (DA2H, DA3H).
- o ... has encouraged me to take all the science [math] I can get in school (DA2M, DA3M).
- o ... has encouraged me to think about a career in math or science (DA2N, DA3N).

Effectiveness — Teacher is conscientious & effective: student reports. Constructed by averaging student responses to four true-false-not sure items (α = .71 for science, .73 for math): "My science (math) teacher this semester..."

- o ... really enjoys teaching science [math].
- o ... is a very good teacher.
- o ... gives me extra help when I don't understand something.
- o ... really seems to like me

Instructional Objectives: Teacher Reports

Emphasis on Inquiry — Science course emphasizes problem solving & inquiry; average of science teachers' responses to the following four-point Likert scale items asking about the relative emphasis given to different objectives in the particular class: Teaching of experimental logic and design (DI10C), Developing problem solving/inquiry skills (DI10E), Developing systematic observation skills (DI10H), and Teaching applications of mathematics in science (DI10I). The alpha coefficient for the scale estimated at the student level is .75.

Science in Everyday Life — Science course emphasizes the importance of science for everyday life, especially environmental issues. Constructed by averaging teachers' responses to items asking about the relative emphasis given to: Increase awareness of the importance of science in daily life (DI10G),

	and Learning about applications of science to environmental issues (DI10L).
Emphasis on Facts	Science (math) course emphasis: Teaching science (math) facts and principles (DI10B, DJ10B).
Emph. on Problem-Solving	Math course emphasis: Teaching math problem solving skills (DJ10D).
Emphasis on Computation	Math course emphasis: Developing computational skills (DJ10E).

<u>Instructional Quantity: Teacher & Student Reports</u>

Text Coverage	% of text covered this year: Teacher report (DI7, DJ7).
Days Absent	Days absent this school year: Student report (DA17).
Time on Routines	% class time: daily routines (DI11A, DJ11A).
Time on Discipline	% class time: getting students to behave (DI11B, DJ11B).
Time on New Material	% class time: presenting new material (DI11C, DJ11C).
Time on Review	% class time: review or student practice of skills (DI11D, DJ11D).
Time on Testing	% class time: testing and evaluation (DI11E, DJ11E).
Homework Assigned	Hrs of homework/week assigned: Teacher report (DI22, DJ22).
Homework Time	Hrs of homework/week: Student report (CASCIJ, DASCIJ; CAMTHJ, DAMTHJ).

<u>Instructional Quality: Teacher Reports</u>

Lecture Time	Class hours/week: Lecturing to the class (DI9A, DJ9A).
Discussion Time	Class hours/week: Leading discussions (DI9B, DJ9B).
Small Group Time	Class hours/week: Student work in small groups or laboratory (DI9C, DJ9C).
Seatwork Time	Class hours/week: Having students do seatwork on homework, workbook, or text assignments (DI9D, DJ9D).
Individualized Time	Class hours/week: Providing individualized instruction (DI9E, DJ9E).
Teaching Machine Time	Class hours/week: Having student use teaching machines or computer-assisted instruction (DI9F, DJ9F).
Homework Returned	% homework assignments corrected and returned to students (DI24, DJ24).

<u>Instructional Quality: Student Reports</u>

Class Accessibility — Class materials & ideas are accessible: student reports. Constructed by averaging student responses to five questions about the student's current science and math courses (all items are coded to range from 1="F" to 5="A."):

- o How much do you like the subject matter of this course? A means you really like the subject; F means you hate it.
- o How clear is the teacher in explaining the material? A means very clear; F means not clear at all.
- o How useful do you think this course will be to you in your career? A means that it will be very useful; F means that it will be of no use.
- o How clear is the textbook for this course? A means very clear; F means hard to understand.

o How difficult or easy is this course for you? A means that it is very easy; F means that it is very difficult.

Class Challenge

Class is intellectually challenging: student report (DASCIE, DAMTHE). Responses range from "'A', it challenges you a lot," to "'F', it never challenges you," coded to range from 1="F" to 5="A."

Appendix B. Effects of Ability Group Placements on Classroom Variables.

Tables 1 and 2 showed several differences between ability groups in the levels of the instructional variables, but did not answer the question of whether ability group placement has an effect on these variables. To answer that question, it is necessary to determine whether students with comparable academic backgrounds, but who are in different ability groups, differ in their classroom experiences. As we have noted, schools and teachers are likely to adapt instruction to the preparation and orientation of their students. It may be the case, then, that while group placement is associated with instructional differences (as shown in Tables 1 and 2), group placement has no independent effect on students' instructional experiences, once the effects of individual student background differences are taken into account.

We address this issue by regressing the instructional variables on students' ability group placements, prior achievement and grades, and social background. The coefficients on the ability group variables in these regressions thus give the estimated effects of group placements on the students' instructional experiences, controlling for academic and social background. The problem we address in this analysis should not be confused with the related, but still distinct, organizational issue of why the instruction variables differ between classes. To answer that issue properly, one would estimate class-level equations, wherein variability in instruction is some function of ability group level and the distribution of student background characteristics in the class. Estimating such an equation would indicate whether class characteristics (instruction) are affected by the ability group of the class, net of the distribution of the background characteristics of the students enrolled in the class. While that would be a useful analysis, limitations of the LSAY sample do not permit it, for there are not enough sampled students in most classes to obtain reliable estimates of the background distributions in each class. Another way of thinking about the difference between our analysis and a class-level analysis is that we are trying to answer questions about the allocation of opportunities to students, while the later addresses an issue of the production of opportunities.

The results of the regressions of the class variables on the group placement dummies and background controls are shown in Tables B-1 and B-2.

Appendix B-1. Effects of Eighth-Grade Students' Science Ability Group Placement on Science Class Resource and Instructional Variables: Metric Coefficients

Dependent Variable	Nongrouped minus Average	High Group minus Average	Adjusted R^2
Resources			
Teacher Background			
Science Major	0.04	-0.09*	.05
Science Minor	-0.02	0.06	.05
Science Credits	-2.48***	-2.66**	.05
Science Educ. Credits	-3.15***	-3.15***	.19
Teaching Experience	-4.91***	-1.46*	.12
Class Characteristics			
Class Size	0.83**	-0.19	.06
Class Time	0.23	7.82**	.13
Science Lab	0.09**	0.20***	.10
Lab Time	5.91**	8.41**	.05
Teacher Expectations			
Teacher Academic Push	-0.05*	-0.02	.03
Teacher Career Push	0.03	0.30***	.07
Teacher Effectiveness	-0.08**	0.06	.04
Instruction			
Objectives			
Emphasize Inquiry	0.10**	0.53***	.15
Emphasis on Facts	0.25***	0.23***	.05
Emphasize Sci in Life	-0.02	0.46***	.12
Quantity			
Text Coverage	-3.55***	2.92*	.07
Days Absent	-0.05	-0.28**	.11
Time on Routine	-0.33	-1.29**	.07
Time on Discipline	-0.72	-2.42***	.03
Time on New Materials	-3.91***	-2.98**	.05
Time on Review	3.05***	4.43***	.06
Time for Testing	1.72***	1.72**	.09
Homework Assigned	-0.46***	0.72***	.13
Homework Time	0.02	0.39**	.02
Quality			
Lecture Time	-0.12**	0.12	.06
Discussion Time	-0.55***	0.54***	.26
Small Group Time	0.27***	0.23**	.06
Seatwork Time	0.46***	-0.04	.15
Individualized Time	0.29***	0.34***	.13
Teaching Machine Time	0.08***	-0.04*	.04
Homework Returned	-0.77	0.47	.09
Class Accessibility	0.04	-0.04	.10
Class Challenge	0.06	0.27**	.03

* $p < .10$ ** $p < .05$ *** $p < .001$

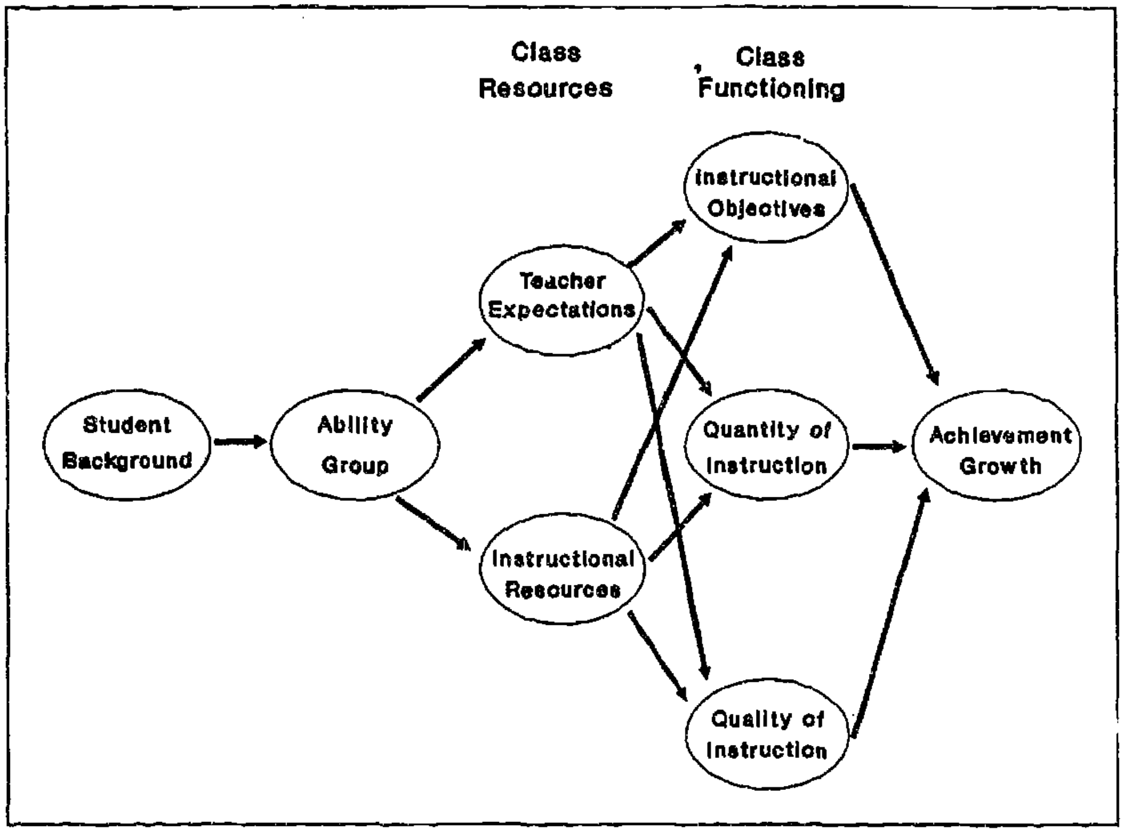
Note: Effects are estimated from student-level regressions of the science class resource and instruction variables on students' ability group placements (dummy variables for nongrouped and high group) and controls for seventh- and eighth-grade science and math achievement, sixth-grade composite grades, seventh-grade grades in science, gender, race-ethnicity, parental encouragement and expectations, and the predicted probability of being excluded from the analysis sample.

Appendix B-2. Effects of Eighth-Grade Students' Math Ability Group Placements on Math Class Resource and Instructional Variables: Metric Coefficients

Dependent Variable	Low Group minus Average	High Group minus Average	Adjusted R^2
Resources			
Teacher Background			
Math Major	0.02	0.23***	.05
Math Minor	-0.13**	-0.13	.05
Math Credits	-1.88***	2.26	.07
Math Educ. Credits	0.41	2.80***	.06
Teaching Experience	0.79	-0.63	.13
Class Characteristics			
Class Size	-4.31***	0.90*	.13
Class Time	-8.09**	0.34	.06
Teacher Expectations			
Teacher Academic Push	-0.05*	0.04	.05
Teacher Career Push	-0.02	0.19***	.06
Teacher Effectiveness	-0.05	0.12**	.04
Instruction			
Objectives			
Emphasis on Problem Solving	-0.00	0.31***	.12
Emphasis on Facts	-0.10**	0.02	.07
Emphasis on Computation	0.23***	-0.30***	.15
Quantity			
Text Coverage	-3.02**	1.29	.14
Days Absent	0.01	-0.17*	.12
Time on Routine	-0.43	-1.13**	.07
Time on Discipline	1.79**	-4.52***	.12
Time on New Materials	-1.65	7.37***	.13
Time on Review	1.65	-0.54	.04
Time for Testing	-1.11*	-1.01*	.02
Homework Assigned	-0.05	1.19***	.14
Homework Time	0.09	1.06***	.12
Quality			
Lecture Time	-0.10*	0.05	.03
Discussion Time	-0.25***	-0.03	.05
Small Group Time	0.03	-0.06	.01
Seatwork Time	-0.10*	-0.00	.09
Individualized Time	0.13**	-0.16**	.06
Teaching Machine Time	0.16**	-0.03	.12
Homework Returned	15.5***	2.5	.04
Class Accessibility	0.18**	-0.09	.07
Class Challenge	0.05	0.44***	.04

* $p < .10$ ** $p < .05$ *** $p < .001$

Note: Effects are estimated from student-level regressions of the math class resource and instruction variables on seventh- and eighth-grade science and math achievement, sixth-grade composite grades, seventh-grade grades in math, gender, race-ethnicity, parental encouragement and expectations, and the predicted probability of being excluded from the analysis sample.



NOTE: Direct paths from Student Background, Ability Group placement, and Class Resources to Achievement are estimated in the data analysis, but are not shown here for clarity of presentation.

FIGURE 1. Schematic Model of Within-School Influences on Achievement.

Table 1. Means and Standard Deviations (in Parentheses) of Academic Achievement, Background, and Instructional Variables: Science Subsample.

Variable	Total (N = 1,235)	Nongrouped (N = 686)	Average (N = 377)	High (N = 172)
Dependent Variable				
1989 Science Score	56.9 (10.3)	56.9 (9.5)	54.7 (10.2)	62.1 (11.8)
Background Variables				
Female	0.48 (0.47)	0.46 (0.46)	0.51 (0.49)	0.51 (0.48)
SES	0.02 (0.70)	-0.04 (0.69)	-0.04 (0.66)	0.36 (0.71)
Black	0.11 (0.30)	0.09 (0.27)	0.14 (0.34)	0.12 (0.31)
Hispanic	0.09 (0.26)	0.09 (0.27)	0.09 (0.28)	0.05 (0.22)
Parent Push	1.66 (1.14)	1.63 (1.09)	1.56 (1.21)	1.97 (1.12)
1987 Math Score	51.5 (8.86)	50.8 (8.53)	49.8 (8.23)	57.5 (9.25)
1988 Math Score	52.9 (9.94)	53.6 (8.82)	49.6 (9.88)	58.0 (11.8)
1987 Science Score	51.6 (9.19)	51.2 (8.96)	49.6 (8.62)	57.4 (9.22)
1988 Science Score	53.5 (10.6)	53.7 (9.80)	50.8 (10.8)	58.9 (11.1)
1986 Grades	5.30 (1.42)	5.29 (1.38)	4.96 (1.47)	6.07 (1.12)
1987 Science Grades	5.12 (1.68)	5.03 (1.69)	4.81 (1.70)	6.14 (1.18)
Resources				
Teacher Background				
Science Major	0.68 (0.44)	0.69 (0.42)	0.66 (0.47)	0.68 (0.45)
Science Minor	0.31 (0.44)	0.31 (0.42)	0.33 (0.46)	0.28 (0.43)
Science Credits	16.8 (8.15)	15.6 (9.05)	18.6 (6.64)	17.0 (6.33)
Science Educ. Credits	2.68 (3.70)	1.4 (1.89)	4.61 (5.13)	2.28 (3.14)
Teaching Experience	13.0 (7.85)	10.8 (7.75)	15.7 (7.61)	14.5 (6.52)
Class Characteristics				
Class Size	25.6 (5.07)	25.6 (5.06)	25.2 (4.58)	26.4 (5.97)
Class Time	233 (26.7)	236 (26.5)	231 (28.1)	224 (22.2)
Science Lab	0.28 (0.42)	0.27 (0.41)	0.20 (0.40)	0.50 (0.48)
Lab Time	18.6 (34.7)	17.8 (31.8)	14.1 (36.7)	31.6 (38.4)
Teacher Expectations				
Teacher Academic Push	0.73 (0.37)	0.71 (0.40)	0.74 (0.35)	0.78 (0.30)
Teacher Career Push	-0.36 (0.50)	-0.40 (0.47)	-0.42 (0.51)	-0.10 (0.52)
Teacher Effectiveness	0.39 (0.52)	0.35 (0.53)	0.41 (0.52)	0.54 (0.47)

Table 1. Continued

Variable	Total	Nongrouped	Average	High
Instruction				
Objectives				
Emphasize Inquiry	2.83 (0.49)	2.79 (0.49)	2.68 (0.41)	3.28 (0.42)
Emphasize Sci in Life	3.13 (0.66)	3.13 (0.67)	3.08 (0.64)	3.27 (0.64)
Emphasis on Facts	3.55 (0.53)	3.64 (0.52)	3.40 (0.50)	3.59 (0.55)
Quantity				
Text Coverage	75.8 (14.0)	74.7 (15.1)	77.0 (12.06)	76.8 (13.0)
Days Absent	2.79 (1.17)	2.75 (1.12)	2.95 (1.28)	2.57 (1.04)
Time on Routine	8.55 (3.69)	8.16 (2.96)	9.21 (3.93)	8.44 (5.27)
Time on Discipline	8.15 (6.29)	7.90 (5.47)	9.16 (7.42)	6.62 (6.37)
Time on New Materials	42.9 (12.3)	41.7 (11.4)	45.0 (11.96)	42.1 (15.4)
Time on Review	25.5 (9.55)	26.3 (9.76)	23.0 (7.63)	28.4 (11.2)
Time for Testing	14.7 (6.20)	15.6 (6.56)	13.6 (5.11)	14.0 (6.46)
Homework Assigned	2.25 (1.02)	1.98 (0.65)	2.34 (1.22)	3.01 (1.31)
Homework Time	1.66 (1.23)	1.59 (1.06)	1.62 (1.30)	2.01 (1.60)
Quality				
Lecture Time	1.42 (0.71)	1.32 (0.70)	1.50 (0.73)	1.61 (0.63)
Discussion Time	1.13 (0.66)	0.87 (0.47)	1.33 (0.73)	1.61 (0.74)
Small Group Time	1.05 (0.67)	1.15 (0.70)	0.86 (0.60)	1.13 (0.59)
Seatwork Time	1.00 (0.58)	1.20 (0.59)	0.78 (0.44)	0.73 (0.46)
Individualized Time	0.74 (0.65)	0.87 (0.79)	0.54 (0.26)	0.72 (0.54)
Teaching Machine Time	0.07 (0.20)	0.11 (0.24)	0.03 (0.14)	0.02 (0.09)
Homework Returned	92.1 (12.5)	93.0 (8.10)	92.4 (14.67)	87.5 (19.0)
Class Accessibility	3.80 (0.80)	3.82 (0.78)	3.72 (0.84)	3.94 (0.79)
Class Challenge	3.96 (1.11)	3.96 (1.05)	3.89 (1.22)	4.14 (1.05)

Table 2. Means and Standard Deviations of Academic Achievement, Background, and Instructional Variables: Math Subsample.

Variable	Total (N = 1,346)	Remedial (N = 227)	Average (N = 744)	High (N = 375)
Dependent Variable				
1989 Math Score	57.4 (11.4)	48.1 (10.0)	55.9 (9.49)	67.1 (9.33)
Background Variables				
Female	0.49 (0.48)	0.42 (0.50)	0.51 (0.48)	0.49 (0.47)
SES	0.10 (0.72)	-0.13 (0.78)	0.04 (0.70)	0.41 (0.62)
Black	0.11 (0.30)	0.20 (0.41)	0.09 (0.28)	0.08 (0.26)
Hispanic	0.09 (0.27)	0.16 (0.38)	0.08 (0.26)	0.04 (0.19)
Parent Push	2.14 (1.04)	1.85 (1.25)	2.11 (1.03)	2.42 (0.86)
1987 Math Score	52.0 (9.51)	44.5 (7.75)	50.7 (8.25)	60.3 (7.37)
1988 Math Score	53.5 (10.3)	45.6 (8.67)	51.7 (8.52)	63.0 (8.24)
1987 Science Score	52.0 (9.8)	45.9 (8.69)	50.5 (8.57)	59.5 (8.76)
1988 Science Score	53.8 (10.8)	45.8 (9.99)	52.9 (9.89)	61.7 (8.30)
1986 Grades	5.34 (1.46)	4.45 (1.62)	5.14 (1.38)	6.39 (0.88)
1987 Math Grade	5.43 (1.60)	4.67 (1.87)	5.28 (1.62)	6.29 (0.88)
Resources				
Teacher Background				
Math Major	0.50 (0.48)	0.43 (0.51)	0.46 (0.48)	0.64 (0.45)
Math Minor	0.34 (0.46)	0.27 (0.45)	0.39 (0.47)	0.29 (0.43)
Math Credits	9.59 (5.29)	7.60 (4.39)	9.54 (5.73)	11.1 (4.34)
Math Educ. Credits	3.64 (4.92)	3.70 (5.61)	2.96 (3.38)	5.00 (6.54)
Teaching Experience	13.81 (7.43)	13.7 (8.55)	13.68 (7.30)	14.13 (6.96)
Teacher Expectations				
Teacher Academic Push	0.84 (0.32)	0.75 (0.41)	0.83 (0.33)	0.93 (0.18)
Teacher Career Push	-0.22 (0.54)	-0.27 (0.59)	-0.28 (0.54)	-0.08 (0.49)
Teacher Effectiveness	0.59 (0.49)	0.50 (0.55)	0.56 (0.51)	0.72 (0.36)
Class Characteristics				
Class Size	24.9 (5.74)	21.2 (6.87)	25.4 (4.47)	26.3 (6.27)
Class Time	236 (29.4)	234 (30.1)	237 (28.3)	235 (30.9)

Table 2. Continued

Variable	Total	Remedial	Average	High
<u>Instruction</u>				
<u>Objectives</u>				
Emphasis on Facts	3.70 (0.56)	3.66 (0.54)	3.71 (0.58)	3.69 (0.53)
Emphasis on Problem Solving	3.42 (0.63)	3.28 (0.73)	3.32 (0.63)	3.72 (0.44)
Emphasis cn Computation	3.15 (0.76)	3.52 (0.65)	3.20 (0.71)	2.79 (0.79)
<u>Quantity</u>				
Text Coverage	84.0 (11.1)	78.8 (12.4)	83.8 (11.1)	88.3 (8.21)
Days Absent	2.73 (1.15)	2.91 (1.23)	2.82 (1.19)	2.44 (0.93)
Time on Routine	7.41 (4.71)	7.93 (5.94)	7.97 (4.68)	5.85 (3.50)
Time on Discipline	6.27 (6.81)	9.07 (7.87)	7.17 (7.19)	2.36 (2.68)
Time on New Material	38.4 (14.7)	32.9 (14.3)	36.3 (13.0)	46.8 (15.1)
Time on Review	32.3 (12.4)	35.3 (13.9)	32.3 (11.6)	30.2 (12.4)
Time for Testing	15.5 (6.64)	14.9 (6.28)	16.1 (6.81)	14.8 (6.42)
Homework Assigned	[illegible].08 (1.70)	2.62 (1.64)	2.75 (1.58)	4.11 (1.60)
Homework Time	2.38 (1.69)	2.01 (1.64)	2.06 (1.42)	3.35 (1.87)
<u>Quality</u>				
Lecture Time	1.21 (0.59)	1.07 (0.61)	1.20 (0.59)	1.32 (0.57)
Discussion Time	0.78 (0.52)	0.61 (0.47)	0.84 (0.56)	0.77 (0.42)
Small Group Time	0.63 (0.59)	0.71 (0.74)	0.63 (0.58)	0.55 (0.49)
Seatwork Time	1.23 (0.71)	1.29 (0.74)	1.28 (0.75)	1.09 (0.59)
Individualized Time	0.86 (0.62)	1.06 (0.85)	0.88 (0.58)	0.69 (0.51)
Teaching Machine Time	0.11 (0.28)	0.27 (0.43)	0.08 (0.21)	0.06 (0.24)
Homework Returned	69.2 (37.2)	83.2 (31.0)	65.5 (38.6)	67.5 (35.8)
Class Accessibility	4.07 (0.74)	4.14 (0.75)	4.03 (0.79)	4.12 (0.63)
Class Challenge	4.22 (1.07)	4.21 (1.21)	4.11 (1.12)	4.46 (0.84)

Table 3. Estimated Effects of Group Placement and Class Variables on Eighth Grade Learning: Science (Standard errors in parentheses)

	Model 1	Model 2	Model 3
GROUP PLACEMENT			
Nongrouped (vs. average)	-0.52 (0.49)	-0.57 (0.51)	-0.77 (0.57)
High (vs. average)	1.56 (0.77)**	1.25 (0.78)	0.39 (0.87)
RESOURCES			
Teacher Background			
Science Major		1.26 (0.47)**	0.63 (0.55)
Teaching Experience		0.01 (0.03)	0.03 (0.03)
Class Characteristics			
Class Size		-0.03 (0.04)	-0.04 (0.04)
Science Lab		0.75 (0.51)	0.39 (0.56)
Teacher Expectations			
Academic Push		-0.84 (0.60)	-0.90 (0.61)
Career Push		0.79 (0.44)*	0.81 (0.44)*
Teacher Effectiveness		0.47 (0.44)	0.06 (0.50)
INSTRUCTION			
Objectives			
Emphasis on Inquiry			1.02 (0.57)*
Science in Everyday Life			-0.40 (0.42)
Emphasis on Facts			0.58 (0.42)
Quantity			
Text Coverage			0.006(0.02)
Days Absent			-0.38 (0.18)**
Time on New Material			0.01 (0.02)
Homework Time			-0.04 (0.17)
Quality			
Lecture Time			0.39 (0.41)
Small-Group Time			0.99 (0.45)**
Discussion Time			0.30 (0.39)
Class Accessibility			0.71 (0.31)**
Class Challenge			-0.14 (0.19)
Probability of Exclusion	-18.51 (4.47)**	-19.6 (4.58)**	-18.54 (4.94)**
ADJUSTED R^2	0.55	0.55	0.56

* $p < .10$, ** $p < .05$.

Note: All three equations include controls for seventh- and eighth-grade science and math achievement, sixth-grade composite grades, seventh-grade grades in science, gender, race-ethnicity, and parental encouragement and expectations.

Table 4. Estimated Effects of Group Placement and Class Variables on Eighth Grade Learning: Mathematics (Standard Errors in Parentheses)

	Model 1	Model 2	Model 3
GROUP PLACEMENT			
Low (vs. Average)	-1.88 (0.57)**	-2.07 (0.59)**	-1.55 (0.60)**
High (vs. Average)	1.84 (0.57)**	1.75 (0.58)**	1.21 (0.61)
RESOURCES			
Teacher Background			
Math Major		0.33 (0.41)	0.12 (0.43)
Teaching Experience		0.04 (0.03)	0.04 (0.03)
Class Characteristics			
Class Size		-0.05 (0.04)	-0.05 (0.04)
Teacher Expectations			
Academic Push		0.16 (0.71)	0.01 (0.70)
Career Push		-0.26 (0.39)	-0.26 (0.39)
Teacher Effectiveness		1.05 (0.48)**	0.90 (0.50)*
INSTRUCTION			
Objectives			
Emphasis on Problem-Solving			0.71 (0.35)**
Emphasis on Facts & Principles			0.75 (0.39)*
Emphasis on Computation			-0.55 (0.31)*
Quantity			
Text Coverage			0.07 (0.02)**
Days Absent			-0.24 (0.18)
Time on New Material			0.01 (0.01)
Homework Time			0.01 (0.12)
Quality			
Lecture Time			0.66 (0.35)*
Small-Group Time			0.30 (0.38)
Discussion Time			0.62 (0.41)
Class Accessibility			0.58 (0.29)**
Class Challenge			0.03 (0.19)
Probability of Exclusion	-5.95 (1.75)**	-4.87 (1.86)**	-3.26 (1.91)*
ADJUSTED R^2	0.62	0.62	0.63

* $p < .10$ ** $p < .05$.

Note: All three equations include controls for seventh- and eighth-grade science and math achievement, sixth-grade composite grades, seventh-grade grades in science, gender, race-ethnicity, and parental encouragement and expectations.