

NBER WORKING PAPER SERIES

SCHOOL CHOICE AND THE DISTRIBUTIONAL EFFECTS OF
ABILITY TRACKING: DOES SEPARATION INCREASE EQUALITY?

David N. Figlio
Marianne E. Page

Working Paper 8055
<http://www.nber.org/papers/w8055>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2000

We appreciate the data and financial support of the U.S. Department of Education, as well as helpful conversations with Laura Argys, Dominic Brewer, David Card, Dennis Epple, Larry Kenny, Dick Murnane, Dan Rees, Steve Rivkin, and Rich Romano, and seminar participants at Berkeley, Columbia, Cornell, Florida, Oregon, South Florida, and SUNY-Albany, and the Econometric Society, APPAM, Population Association of America, and American Educational Finance Association meetings. The financial support of the National Science Foundation through grant SBR-9816629 is gratefully acknowledged. All remaining errors are our own. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by David N. Figlio and Marianne E. Page. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

School Choice and the Distributional Effects of Ability Tracking:
Does Separation Increase Equality?
David N. Figlio and Marianne E. Page
NBER Working Paper No. 8055
December 2000
JEL No. I2

ABSTRACT

Tracking programs have been criticized on the grounds that they harm disadvantaged children. The bulk of empirical research supports this view. These studies are conducted by comparing outcomes for across students placed in different tracks. Track placement, however, is likely to be endogenous with respect to outcomes. We use a new strategy for overcoming the endogeneity of track placement and find no evidence that tracking hurts low-ability children. We also demonstrate that tracking programs help schools attract more affluent students. Previous studies have been based on the assumption that students' enrollment decisions are unrelated to whether or not the school tracks. When we take school choice into account, we find evidence that low-ability children may be helped by tracking programs.

David N. Figlio
Department of Economics
University of Florida
Gainesville, FL 32611-7140
and NBER
figliodn@dale.cba.ufl.edu

Marianne E. Page
Department of Economics
University of California-Davis
Davis, CA 95616-8578

The practice of ability tracking--grouping students together in classrooms according to their ability level--has been criticized on the grounds that it subjects disadvantaged students to a lower quality education than they would receive in heterogeneous classrooms.¹ Indeed, the bulk of empirical research supports the view that tracking benefits high ability students at the expense of those with fewer endowments (examples from the literature include: Alexander and McDill, 1976; Argys, Rees and Brewer, 1996; Betts and Shkolnik, 1997; Gamoran, 1987; Hoffer, 1992; Kerckhoff, 1986; and Vanfossen, Jones and Spade, 1987) and this has led the National Education Association (NEA) to recommend that schools discontinue tracking (Oakes, 1992). A number of prominent politicians, both liberal and conservative, also oppose ability tracking; for example, Florida Governor Jeb Bush points to schools that have detracked as models for school reform. Although there are no published statistics on the degree to which schools are heeding the NEA's recommendation, we estimate that between 1987 and 1993 the number of schools that maintained programs for gifted children declined by seven percent.²

Proponents of ability grouping argue that by narrowing the range of student abilities within a classroom, tracking allows teachers to target instruction at a level more closely aligned with student needs than is possible in more heterogeneous environments. Teachers of high ability students can provide them with more challenging material or

¹ Criticisms include those by Braddock and Slavin (1993), Gamoran (1989), Oakes (1985), Rosenbaum (1980b) and Wheelock (1992).

² These estimates are based on schools included in the Schools and Staffing Survey: in 1987, 74 percent of the high schools in our sample maintained a program for gifted students, and 54 percent maintained both a remedial and a gifted program. In 1993, 69 percent maintained a gifted program and 46 percent had both a gifted and a remedial program. Among elementary schools, the fraction with gifted programs fell from 78 percent to 72 percent, and the fraction with both types of programs fell from 67 percent to 64 percent. While there may not be a one-to-one correspondence between school administrators' interpretations of whether a program for gifted children exists and whether students are grouped by ability level for standard subjects (the usual definition of tracking), we know of no other statistics on changes in "tracking" status over time.

present standard material at a faster pace than would be possible in a classroom where less-able students' needs also have to be met. At the same time, low-ability students are expected to benefit from the slower pace or alternative teaching methods that become feasible when teachers are not simultaneously responsible for engaging the students' high-ability peers. Empirical studies of tracking effects, however, have produced little evidence to support the proponents' claims. While high-ability children appear to benefit from ability specific instruction, students placed in low-ability tracks appear to perform less well than seemingly comparable students placed in heterogeneous classrooms. Critics of tracking programs point to a number of factors that may "undo" the benefits to low-ability children of receiving instruction targeted at their level. The predominant criticism is that when students are segregated by ability, disadvantaged students lose any positive peer effects that might be gleaned from coming into regular contact with more able students. But critics also maintain that schools with tracking programs systematically redistribute resources away from low-ability students towards high-ability students, and that less capable teachers are disproportionately assigned to low-ability tracks. Although there is little evidence to support or refute these claims, we know of no study to date in which the potential benefits to low-ability students of receiving targeted instruction appear to outweigh the potential costs.

There are several reasons that economists should care about the effect of tracking on student achievement. First, it is well known that the wage distribution is spreading (for example, see Bound and Johnson (1992) and Juhn, Murphy and Pierce (1993)), and there is also evidence that student test scores and future earnings are linked (Murnane, Willett and Levy, 1995). If tracking widens the distribution of test scores it may also lead to further increases in wage inequality. Second, economists have long contributed to the literature on education production functions, which focuses on the impact of school

inputs on student outcomes. That literature, which has been largely unsuccessful at identifying factors that affect *between* school variation, also points to the fact that much of the variation in student outcomes is *within* schools. Tracking may partly explain why this within school variation exists. Conversely, if students benefit from being educated among similarly skilled children, then tracking may be a school “input” that can help improve outcomes among America’s most disadvantaged citizens.

In this paper we highlight three empirical problems that have impeded researchers’ ability to assess the effect of tracking programs on students’ test scores. When we address these issues we find no evidence that low ability students are harmed by being grouped together, and conclude that the trend away from tracking has been misguided. In fact, we find that programs targeted to specific parts of the test score distribution have a substantive effect on a school’s ability to attract high-income students, which may benefit low-ability students in a number of ways: by providing funding for additional resources, by attracting better teachers or through positive school-level peer group effects, for example. When we use an instrumental variables procedure to address the possibility that school choice is partly determined by tracking status, we find that tracking programs are associated with test score *gains* for students in the bottom third of the initial test score distribution. We conclude that the move to end tracking may harm the very students that it is intended to help.

The rest of this paper is laid out as follows. The next section provides a review of the tracking literature and discusses the difficulties associated with identifying program effects. An overview of our data is presented in Section III. In Section IV we use our data to re-estimate the impact of tracking on low-ability children, paying special attention to the problems discussed in Section II. Section V concludes.

II. Previous Research

Tracking studies can be roughly classified into two categories: those that estimate the effect of tracking on *mean* achievement and those that investigate the impact of tracking on the *distribution* of achievement. Those that compare mean outcomes across schools with and without tracking programs generally find that tracking programs have small, and typically insignificant, effects (Slavin, 1990). More recent investigations have considered the effect of tracking on the distribution of test scores across ability groups by using within school variation in individuals' track placement (e.g. Alexander and McDill, 1976; Gamoran, 1987; Vanfossen, Jones and Spade, 1987) or by comparing students in different tracks to those placed in heterogeneous classrooms (e.g. Argys, Rees and Brewer, 1996; Betts and Shkolnik, 2000; Hoffer, 1992; Kerckhoff, 1986). These studies estimate the following type of regression model

$$(1) \quad \Delta A_i = \alpha + \beta_1 H_i + \beta_2 M_i + \beta_3 L_i + \gamma X_i + \varepsilon_i$$

where ΔA_i represents the change in student i 's test score from time $t-1$ to time t , H_i , M_i and L_i are dummy variables representing placement in a high ability, medium ability or low ability classroom, and β_1, β_2 and β_3 are interpreted as the degree to which being placed in a high, medium or low ability classroom affects test score growth relative to being placed in an untracked classroom. X_i represents a vector of other student-level covariates, usually including the student's base-year test score. The results produced by these studies almost universally support the hypothesis that tracking programs benefit the upper tail of the ability distribution at the expense of the lower tail. Consequently, a number of policy analysts have called for an end to tracking practices (Oakes, 1992).

As has been noted by other researchers, however, at least two factors complicate the identification of differential tracking effects across student types. First, there is a

great deal of ambiguity about what it means to say that a school “tracks,” particularly at the high school level. If a secondary school offers its seniors courses in both algebra and calculus, and if high-ability students tend to sort into calculus and low-ability students sort into algebra, are the calculus students in a high-ability track and the algebra students in a low-ability track? Or does tracking mean that college bound students take an entirely different *set* of courses from those who are not college bound? While researchers and policy makers agree that tracking involves ability grouping, rarely do policy discussions or studies of tracking effects clarify specifically which types of school programs “count” as tracking programs and which do not. Furthermore, school principals, parents and students also have ambiguous views about what constitutes tracking (Gamoran, 1987; Rosenbaum, 1980a; Rees, Argys and Brewer, 2000),³ and estimates of tracking effects based on individual survey responses, therefore, vary, depending on who is being surveyed and depending on whether and how the survey questions define tracking programs (and track placement). As a result, it is often difficult to interpret estimated tracking coefficients. Although our study is also subject to this potential problem, we add to the literature by investigating the sensitivity of our results to a number of different definitions of tracking.⁴

The second problem faced by researchers is that estimates of $\beta_1 - \beta_3$ are likely to suffer from omitted variables bias, even when a prior test score is included in the regression equation, because other factors unobservable to the researcher will affect track placement and some of these factors may be correlated with test score growth. Oakes

³ Indeed, using the National Educational Longitudinal Study of 1988, we find a great deal of within school variation in students’ reported 8th grade tracking status, which suggests that using student reports would lead to noisy measures of true tracking status.

(1986) provides evidence that teachers' evaluations of student abilities, track assignments in previous years, and student motivation all affect track placement. Several innovative studies have recently emerged that use sophisticated econometric techniques to try to address this problem, but none of the approaches has been entirely satisfactory. Betts and Shkolnik (2000), for example, compare students in tracked classrooms to students in untracked classrooms in which average test scores are similar. While their approach has the advantage of comparing outcomes across students who are similar, it is unclear why we should expect students in classrooms that are differentiated by their label (tracked vs. untracked) but not differentiated by the composition of their classmates to produce different outcomes. Hoffer (1992) uses a propensity score method to divide students into quintiles based on similar backgrounds, and then compares outcomes for students in high, medium and low tracks to the average heterogeneously grouped student within each quintile. The estimates produced by this method will still be subject to omitted variables bias, however, unless the quintiles are chosen based on all relevant characteristics. Argys, Rees and Brewer (1996) account for selection into particular tracks using regional indicators, urbanicity indicators and student body characteristics as instruments in a two-stage least squares procedure, but the authors do not provide evidence that these variables are exogenous with respect to the dependent variable. When we estimate an expanded version of equation (1), we find that some of these instruments are correlated with the change in individuals' test scores. Although all of these researchers have made valiant attempts to address the problem, it remains unclear whether their finding that children placed in low-ability tracks have lower test scores is a causal effect.

⁴ Rees, Argys and Brewer (2000), in a comment on Betts and Shkolnik (2000), criticize papers that use principal-reported measures of tracking as unreliable, given the heterogeneity in the ways in which school tracking policies could be classified. Their preferred method involves relying on teacher reports of whether a specific *class* is tracked. In this paper, we employ both teacher-reported and principal-reported measures of ability tracking, thereby reducing the validity of this criticism as it pertains to our own work.

We propose a simpler and cleaner strategy for circumventing the selection problem, which is to eliminate track placement from the model altogether and, instead, estimate the effect of attending a tracked school separately for students in the top, middle and bottom thirds of an initial test score distribution. This allows us to identify the effect of tracking solely from variation across schools, rather than from variation *both* across school types and across student types. Provided that school choice is unrelated to the school's tracking status, this strategy will yield unbiased estimates of the effect of tracking for each group.⁵ Even if school choice is associated with school tracking status, our approach will eliminate the bias associated with within-school track placement.

Previous empirical studies have been based on the assumption that students' enrollment decisions are unrelated to a school's tracking status, but if this assumption is incorrect, then tracking may appear to be unequalizing even if it is not. For example, if schools in disadvantaged neighborhoods use tracking programs to attract high-income students, then the estimated tracking coefficient may be biased upward for the high ability group and biased downward for the low ability group.⁶ Understanding the degree to which tracking programs affect school choice decisions is also important to the tracking debate because the presence of high socioeconomic status students may generate positive externalities for their less advantaged peers. Recent work by economic theorists (Epple, Newlon and Romano, 1997) suggests that if household utility is increasing in student achievement, and student achievement is partly influenced by the average ability of one's peers, then public schools will raise the socioeconomic composition and average

⁵ An alternative approach would be to use the full sample to estimate a model of test score growth as a function of the school's tracking status and an interaction term between the school's tracking status and the individual's initial ability. We find, however, that the coefficient estimates on many of the control variables vary substantially across groups. Furthermore, we found that the estimated effect of tracking on student achievement is not linear with respect to the student's initial test score.

⁶ These biases would occur if high income students that move to tracked schools are more motivated than non-movers and/or if low ability students in disadvantaged neighborhoods have lower test scores (or test score growth) than similar students living in less disadvantaged neighborhoods.

ability level of their student body by maintaining a tracking program. This finding suggests that low-ability children may benefit from tracking even if they do not benefit from experiencing more specialized instruction. Furthermore, if students from high-socioeconomic status families are attracted to schools with tracking programs, they may bring both pecuniary and non-pecuniary resources (such as parental involvement) with them, and these resources may benefit all students.

Our third contribution to the literature is that we document a (positive) relationship between changes in a school's tracking status and changes in the socioeconomic composition of its student body. Our investigation suggests that the relationship is a causal one. Since this implies that estimates of the effect on achievement gains of attending a tracked school are subject to selection bias, we then apply a two-stage least squares strategy to our test score analysis, and find suggestive evidence that tracking programs may actually benefit low-ability students.

III. Data

A. National Educational Longitudinal Study of 1988

We use two datasets in our analyses. Our analyses of the relationship between tracking and achievement gains are based on data from the National Education Longitudinal Study (NELS). The NELS began in 1988 when a nationally representative sample of schools was surveyed. In that year, a random sample of 8th grade students was interviewed within each school, and these students were then re-interviewed in subsequent years, resulting in a dataset that contains test scores and other individual, family background and school level variables for the same nationally representative set of students in 8th, 10th and 12th grade. A comparable dataset covering elementary grades does not exist.

Our dependent variable is the change from 8th to 10th grade in the student's item response theory (IRT) math score.⁷ We estimate the effects of tracking on mathematics because there is a well-documented link between math test scores and labor market outcomes (e.g. Bishop (1994); and Murnane, Willett and Levy (1995)), and because math test scores are the outcome most frequently studied in the tracking literature. The NELS provides standardized test scores in addition to IRT scores, but we chose the IRT score because it concretely characterizes students' progress from one year to the next, whereas the standardized test score is a further transformed measure, and merely reflects individuals' *relative* positions in the test score distribution. In our sample, the mean IRT score in eighth grade is 35.8 points, and its standard deviation is 11.8 points. The tenth grade IRT score for our sample is 42.4 points with a standard deviation of 13.9 points.

As discussed in Section II, our ability to distinguish “tracked” from “untracked” schools is complicated by the fact that tracking means different things to different people. At the high school level, tracking may consist of a series of classes intended for pre-college vs. vocational students, or it may refer to the grouping of particular classes by ability level. Furthermore, a school may not *formally* group its students by ability level, but may offer a set of subject specific classes that are aimed at different ability groups (e.g. calculus, pre-calculus, algebra) and into which students of different ability levels naturally sort themselves. Like other datasets, survey questions that address tracking in the NELS are vague about what exactly tracking is. For example, school principals are asked “Does your school track in mathematics?” but they are not told what types of programs constitute tracking in mathematics.

⁷ IRT scores are essentially raw test scores in which responses to more difficult questions are weighted more heavily. The NELS uses item response theory to construct its test scores in order to reduce the potential problems associated with ceiling effects. Raw test scores are not provided.

We are interested in assessing the effect of being schooled in a classroom with similarly skilled students relative to the effect of being schooled in a classroom that has a larger variance in student abilities. Of the tracking-related questions asked in the NELS, principals' answers to the above question seem to us to provide the most direct information on whether or not mathematics classes are grouped by ability level, but because the definition of tracking is not explicit, we may misclassify some schools in which ability-grouping takes place. In order to explore the sensitivity of our results to the way in which tracking is defined, we have created several alternative definitions of tracking, which are based on teacher reports of classroom homogeneity, and on teacher reports of the official "track" of the class that he or she teaches. The more consistent results are across varying definitions of ability tracking, the more confident one can be that an estimated tracking effect is genuine.

The alternative definitions of tracking are based on the following information: for each student surveyed in the NELS, two of her teachers were also surveyed. The teachers provided instruction in either math, science, English or history and were asked to answer questions about the math, science, English or history class in which the student was enrolled. Two of these questions are particularly relevant to our study. First, each teacher was asked whether he or she "best describes the achievement of the [eighth] tenth graders in this class compared with the average [eighth] tenth grade student in the school" as "higher achievement levels," "average achievement levels," "lower achievement levels," or "widely varying achievement levels." Second, teachers also reported on whether the track that best described the class was "advanced," "academic," "general" or "vocational." We used the responses to these questions to create additional tracking variables, which we discuss in the next section. The answers to these questions have the

advantage of providing information on the composition of a particular classroom, but are potentially more subjective or noisier than the principals' responses.⁸

Our sample includes all public school students with 10th grade school identifiers who have both 8th and 10th grade test scores, information on tracking, and information on the following covariates: indicators for whether the student is white, black, or Hispanic, dummies for the student's family income (under \$15,000; between \$15,000 and \$35,000; between \$35,000 and \$50,000; between \$50,000 and \$100,000; over \$100,000; and a missing income indicator), indicators for whether the highest educational attainment of the student's most highly educated parent is a high school graduate or college graduate, region-of-country indicators (four Census regions), and indicators of central city or suburban residence.⁹ In order to reduce the possibility that our tracking estimates are picking up something else about the school environment, we also include a set of school-level characteristics in our regressions. These include: the fraction of students who are white, the fraction of students who are free lunch-eligible, the number of days in the school year, the student-teacher ratio, the size of the tenth grade cohort, and the highest teacher salary.¹⁰ Our sample excludes students who are missing these variables. Finally, we eliminate students who are missing the instrumental variables that we require for the two stage least squares analysis described in Section IV.¹¹ This leaves us with a sample

⁸ Argys, Rees and Brewer (1996) provide evidence that tracking measures created from these two sets of questions may be quite distinct. In other words, a teacher will not necessarily characterize her class as being both "high achieving" and "advanced," or as both "low-achieving" and "general."

⁹ We have experimented with including many other covariates, including measures of personal possessions, measures of family involvement in school, etc., and have found that these variables do not affect our estimates.

¹⁰ Note that our empirical results are very similar whether or not we include these school level variables in our regressions. Some of these results are described later in the text; those not described are available upon request from the authors.

¹¹ We lost 354 observations due to missing instrumental variables; this is because our instruments are measured at the county level, and we are missing the ability to identify county for some of the schools. There does not appear to be any systematic difference between the included and excluded students. For example, there are no statistically significant differences in their family background variables. The estimates presented in Tables 2 and 3 are not substantively affected by the inclusion of these additional observations.

of 7,676 students with principal provided information on whether their school tracks in mathematics.¹² When we use other definitions of tracking based on teacher reports our sample contains only 5,948 students because mathematics teachers were not surveyed in every school.

Table 1 provides summary statistics for the variables included in our regressions, by students' tracking status, using the principal-reported measure of tracking. The two groups of students differ in a number of ways: for example, the proportion of students coming from families with incomes greater than \$50,000 (in 1988 dollars) is nearly fifty percent higher in the set of tracked schools than in the set of untracked schools, and the fraction of tracked students with college-educated parents is one-third higher than the corresponding fraction of untracked students. Likewise, students attending tracked schools tend to have higher eighth grade test scores and exhibit higher test score growth than their counterparts: average eighth grade test scores are 1.6 points higher in tracked schools, and average test score gains are 0.2 points higher. The difference in eighth grade test scores between the two groups of students is not surprising given that the students attending tracked schools tend to come from higher SES families. With the exception of the difference in test score growth between the two groups, these differences are all statistically significant at conventional levels and suggest that tracking programs may influence the selection of different types of students into different types of schools.¹³

¹² We arrive at this sample size as follows: there are 15,217 public school observations in the NELS with 10th grade school identification numbers. Of these, 12,817 observations have data on 8th and 10th grade mathematics tests. Of these, 12,008 have principal-reported measures of tracking. Of these, 11,642 have all of the individual level covariates. 354 observations are lost due to missing instrumental variables, and another 3612 observations are lost due to missing school level characteristics. We have also run all of our NELS regressions without the school level variables, which allows us to boost our sample sizes to 11,288 (principal reported tracking measure) and 9,516 (teacher reported tracking measures). The estimates produced by those regressions were virtually identical to those presented in Tables 2,3 and 5. We prefer to report estimates based on the smaller samples to help reduce possible concerns that our estimates reflect the effect of unobserved school characteristics.

¹³ These patterns are also observed when the other definitions of tracking are used.

Alternatively, these differences may indicate that student body composition influences a school's decision to track.

Table 1 also indicates that students attending schools with tracking programs are more likely to live in urban environments than are students who attend schools without tracking programs. Regional variation in the prevalence of tracking programs also appears to exist, and suggests that something other than a desire to attract high-ability students may affect a school's decision to track. Identifying the full range of such factors deserves further research, but it is worth noting here that their omission from the model will only affect our OLS estimates if they are also correlated with average test score growth across school types. The omission of such factors from the two-stage least squares analysis will have no effect on the estimated tracking coefficients since IV eliminates omitted variables bias.

B. Schools and Staffing Survey

We investigate the degree to which tracking programs influence school composition using the Schools and Staffing Survey (SASS), which consists of three nationally representative surveys of schools and school districts that were conducted in 1987-88, 1990-91 and 1993-1994. For each year, the SASS provides information on the socioeconomic composition of the student body and whether the school maintained programs aimed at specific ability groups. We measure the socioeconomic composition of the student body using the fraction of students that qualify for free lunches.¹⁴ The

¹⁴ We also conducted our analysis using the fraction of students who are white as a measure of student body composition. Although the sign and magnitude of the estimated coefficients produced by this analysis were generally consistent with our hypothesis that tracking programs influence the school choice decisions of high SES families, they were not usually statistically distinguishable from zero. The weaker findings produced by these analyses may reflect individuals' relatively stronger aversion to racial integration compared to economic integration.

mean value of this variable is 0.26 in 1987 (the standard deviation is 0.22) and 0.30 in 1993 (the standard deviation is 0.24).

Each wave of the SASS provides an approximately ten percent sample of the population of American schools. As a result, even though the SASS is not designed as a panel, many schools are surveyed in multiple years. By linking schools across survey years we can form a panel with which we estimate the relationship between changes in a school's tracking status and changes in its composition. The advantage of using this "difference-in-differences" estimation approach is that it allows us to implicitly control for school characteristics that are fixed during the six year period we study, and that might be correlated with the propensity to maintain a tracking program.

Unfortunately, the SASS is missing explicit information on tracking. We, therefore, proxy for the presence of a tracking program by using information on whether the school maintains gifted and/or remedial programs to create three alternative definitions of "tracking." First, we label a school as "tracked" if it maintains a program for gifted students (in 1987, 76 percent of schools had a gifted program, while 71 percent had a gifted program in 1993.) Second, we label a school as "tracked" if it maintains either a gifted or a remedial program (in 1987 (1993), 90 (88) percent had one or the other.) Our third definition of tracking measures the "degree" of tracking by counting the number of programs provided by the school. This variable takes on a value of 2 if the school has both a gifted and a remedial program, a value of 1 if only one program is available, and a value of 0 if no programs are available (the mean value of this measure was 1.38 (1.34) in 1987 (1993).) We also estimate the relationship between a school's composition and the presence of a remedial program, only (62 percent had a remedial program in both 1987 and 1993.) While we would expect high socioeconomic status students to be more responsive to the presence of a gifted program than to the presence of

a remedial program, we can use the estimated response to remedial programs to test for whether our measures of tracking are endogenous. We will discuss this in more detail in the next section.

Using the presence of gifted and remedial programs to identify tracking is admittedly a second best alternative. We would prefer to base our definition on information regarding ability grouping by subject, but there are no panel data sets that provide both a good definition of tracking and information on student body composition. We are encouraged by the fact that among public *middle schools* in the NELS the existence of gifted and remedial programs is positively correlated with teacher reports on the within classroom distribution of student ability.¹⁵ Ninety one percent of schools in which all surveyed teachers report that all of their students are homogeneously grouped (a third of all schools) have either a remedial or gifted program, for example, and 82 percent have both. In schools where at least some teachers report that their classes are heterogeneous by ability, 71 percent have *neither* a gifted nor a remedial program. This makes us hopeful that the presence of gifted and remedial programs will closely proxy the presence of tracking programs. Furthermore, parents whose school choice decisions are influenced by the presence of these programs are also likely to be influenced by the existence of ability grouped classes.

Our SASS sample consists of 545 public schools, of which 313 contain secondary grades (grades 8 through 12) and 274 contain elementary grades (grades 1-4). In addition to conducting our analysis for the full sample of schools, we also examine whether there is a differential response to tracking across elementary and secondary schools by breaking our sample down according to whether the school has any grades between 8 and

¹⁵ We cannot examine this relationship among high schools in the NELS because the NELS does not report gifted program status for high schools.

12 (secondary) or any grades between 1 and 4 (elementary).¹⁶ We do this for two reasons. First, what constitutes a tracking program surely varies across school types. In secondary schools tracking is likely to be done by subject, whereas in elementary schools tracking often means that the high ability students will be together for the entire school day and that the low ability students will be in the same classroom for the entire day. Second, parental interest in tracking may vary according to the age of the child, or the socioeconomic composition of the school; high schools typically contain more diverse student populations than elementary schools, and as a result, tracking programs may be relatively more important mechanisms for retaining upper middle class students.

IV. Do Tracking Programs Harm Low-Ability Students?

A. Addressing the Endogeneity of Track Placement

To begin with, we replicate earlier studies by estimating equation (1). Our dependent variable is the 8th to 10th grade change in the student's IRT math score. If the student's math teacher characterizes the student's class as being of "above average" ability then the student is classified in the high track, if the teacher characterizes her class as "average" then the student is classified as being in the middle track and if the teacher characterizes her class as "below average" then the student is classified as being in the low track. Individuals whose teachers describe the ability level of their students as "widely varying," are classified as untracked. This is the same classification that Argys, Rees and Brewer (1996) use. The results produced by this exercise are shown in the first column of Table 2. Like the existing literature, the estimates in column 1 suggest that students placed in low-ability tracks experience smaller achievement gains and students placed in high-ability tracks experience larger achievement gains than students with

¹⁶ We have found that our results differ very little when we use different definitions of secondary and

similar test scores and family background characteristics who are placed in heterogeneous classrooms. The point estimate of -6.0 on low-track placement suggests that students in low-ability tracks experience achievement gains that are three-quarters of a standard deviation lower than similar students in heterogeneous classrooms, and the point estimate of 2.2 on high-track placement suggests that students placed in high ability tracks gain about one-third of a standard deviation above similar students in heterogeneous classrooms.¹⁷ Of course, the question is whether these observed differences are causal--they are also consistent with inherent, unobserved differences in student characteristics that affect their selection into particular tracks.

In the next three columns of Table 2, we attempt to address this selection problem by dividing the sample into the top, middle and bottom thirds of the 8th grade test score distribution, and then estimating for each subsample the effect of attending a tracked school. In other words we estimate

$$(2) \quad \Delta A_i = \phi + \alpha T_i + \gamma X_i + \varepsilon_i ,$$

where T_i indicates whether individual i 's principal reports that individual i 's school tracks in mathematics and X_i is a vector of control variables. Relative to existing studies, the advantage of this approach is that $\hat{\alpha}$ is identified only from variation in tracking status across schools. Estimates based on equation (1) are identified both from variation in tracking status across schools and from variation across student types. As discussed in Section II, some of the test score variation across student types probably results from

elementary schools.

¹⁷ Using the same dataset and controlling for eighth grade test score, Argys, Rees and Brewer (1996) find that students placed in high ability tracks have 10th grade math achievement scores that are about 25 percent of a standard deviation higher than similar students placed in heterogeneous classrooms, and students placed in low ability tracks have 10th grade test scores that are about 25 percent of a standard deviation

factors other than the track into which they were placed. As shown in the next three columns of Table 2, the estimates produced by this model lend themselves towards a different conclusion from those in column 1.¹⁸ The estimated coefficient on tracking is negative but trivial in magnitude for high, middle, and low ability students (-0.19, -0.06 and -0.40) and none of the estimates are significantly different from zero. These results suggest that the lower test score gains observed among students in low-ability tracks stem not from their track placement, but rather from unobserved factors correlated with track placement.

Rees, Argys and Brewer (2000) argue that principal-reported measures of tracking will be unreliable because of ambiguities about what “tracking” means. In the spirit of their argument, we, therefore, re-estimate equation (2) using six alternative definitions of tracking that are constructed from teacher reports. The first three measures are based on teachers’ descriptions of their classes as being of “high,” “average,” “low” or “widely-varying” ability, while the others are developed from teachers’ descriptions of their classes as being “advanced,” “academic,” or “general.”¹⁹ Specifically, as a first alternative, we define a school as a “tracked” school if at least one teacher reports that his class is “high-achieving” and at least one teacher reports that his class is “low-achieving.” Our second alternative is based on the number of different types of responses (“high-achieving,” “average-achieving,” and “low-achieving”) observed at a given school and can take on values ranging from zero to three. Our third definition counts a school as “tracked” if at least one teacher reports that his class has “high-achieving” students. This measure is somewhat analogous to the “gifted” classification we use in our SASS

below similar students placed in heterogeneous classrooms. This result is based on specifications that account for selection into tracks, as described in section III.

¹⁸ Because our “treatment” variable does not vary below the school level, we adjust our standard error estimates to account for error correlation within-schools, using the Huber standard error correction.

analysis. The last three measures are similar to the first three except that we replace “high-achieving” with “advanced,” “average-achieving” with “academic,” and “low-achieving” with “general.” In all six cases, we add as an additional covariate the number of unique classes observed in the school because the likelihood of observing a particular type of class is higher as we observe more classes.

The estimated treatment effects are reported in Table 3. Across the specifications, there is no evidence that tracking harms low-ability children. Compared to the estimates in the first column of Table 2, the coefficient estimates for this group are all small in magnitude, ranging from -0.40 to $+0.06$, and none of them is statistically different from zero. Some of the alternative definitions do produce more economically (and nearly statistically) significant estimates for the high and medium ability groups, but the estimates are still substantially smaller than the estimated coefficients on the track placement variables. Taken together, the estimates in Table 3 reiterate the conclusions that we drew in Table 2: much of the effect associated with track placement can be ascribed to unobservable factors correlated with track placement, and once the endogeneity of track placement is addressed there is no evidence that low-ability children are hurt by tracking.²⁰

B. Do Tracking Programs Affect Student Composition?

The results presented in Section IV.A. suggest that tracking programs are probably not harmful to low-ability students, but, like other studies, these estimates are based on a model in which the school’s tracking status is assumed to be exogenous with

¹⁹ Teachers are also given the choice of describing their classes as “vocational.” Only 2 percent of teachers classified their classes in this way, however, suggesting that the existence of vocational classes is not indicative that the school has a tracking policy.

²⁰ None of these conclusions are driven by our inclusion of school-level covariates in the regression. Were we to exclude school-level regressors, for instance, the estimated treatment effects of tracking for low-

respect to student test scores. If the factors affecting school choice cannot be thoroughly controlled then estimates of the return to tracking based on equation (2) will still be biased. In this section, we estimate the impact of a school’s tracking program on the socioeconomic composition of its student body. This exercise is important not only because it affects our interpretation of the tracking coefficient estimates produced by other researchers (and our own estimates in Table 2), but also because tracking programs may be a useful device for retaining high-income students, which may in turn affect the school environment in a number of ways.

We use the SASS to estimate the following model:

$$(3) \quad C_{st} = \gamma T_{st} + \alpha_s + \beta_s t + \delta_t + \varepsilon_{st}$$

where C_{st} represents the socioeconomic composition of school s (percent of students eligible for free lunches) in year t , T_{st} is a dummy variable indicating whether or not the school tracks its students, α_s is a school specific fixed effect, β_s is a school specific time trend and δ_t is a year dummy. We estimate equation (3) using first differences, which allows us to control for static, unobserved school characteristics (α_s) that might be correlated with the school’s tracking status. Our model also includes a set of school specific trends, $\beta_s t$, to help control for the possibility that schools add or eliminate tracking programs in response to changes in school demographics. The coefficients β_s pick up anything about a school that is changing at a constant rate over the six year period, including changes in demographic characteristics, which usually occur gradually. In contrast, changes in a school’s tracking status will be abrupt.

The top panel of Table 4 summarizes our estimates of the relationship between a school’s tracking policy and its socioeconomic composition. Across all specifications,

achieving students would range from -0.15 to $+0.08$, and remain far from statistical significance at any

the presence of a tracking program is estimated to reduce the fraction of students who are free lunch eligible by between two and three percentage points. Almost all of the estimates are statistically different from zero,²¹ and are of similar magnitude across elementary and secondary schools. Since about thirty percent of a school's student body is free lunch eligible, on average, our results suggest that when a school eliminates its tracking program, the proportion of its students who come from low income families will increase by about 8 percent. In fact, we may be underestimating the magnitude of the response because we observe it within only a few years of the change in tracking status.

Note that the estimated effect of gifted and remedial programs is approximately the same. This provides some evidence that our estimates are not driven by reverse causality. If these programs were added/eliminated in response to reductions/increases in the fraction of less-affluent students, then one would expect the presence of a remedial program to be positively associated with our dependent variable.²²

We investigate our estimates' robustness by looking at whether the effect of tracking varies across schools according to the degree of competition they face. If the estimates in the top panel of Table 4 are truly picking up a school choice response, then we would expect the results to be stronger for schools located in counties where there are more choices. We split our sample into the set of schools with a) greater than and b) fewer than, the median number of schools per county²³, and test the hypothesis that the response to tracking is larger in counties where there is more competition. Because our

commonly-used threshold.

²¹ We use the White correction to ensure that our standard estimates account for a potentially heteroskedastic dependent variable.

²² Of course, it is possible that more-affluent students are attracted to schools with remedial programs because, like the presence of gifted programs, remedial programs provide these students with a means of segregating from low-SES students. It is unlikely, however, that schools add remedial programs in direct response to increases in the fraction of students who are more affluent unless it is because the presence of these students increases funding for such programs, since more affluent students are less likely to participate in remedial programs than their less affluent peers.

²³ The number of schools per county comes from the Common Core of Data.

basic results are so similar across high schools and elementary schools, we base our specification checks on the full sample of schools, which increases the precision of our estimates, but does not substantively alter their magnitude. As shown in the middle panel of Table 4, we find that the response to tracking is stronger in counties where more public schools are available. Across the three specifications the average difference is about two percentage points. This bolsters our hope that our estimates of γ are picking up a tracking effect.

Because our model includes school specific trends, which should capture all but abrupt changes in school composition, we think it is unlikely that our estimates reflect reverse causality. In addition, our finding that the presence of remedial programs reduces the fraction of students who are eligible for free lunches is the opposite of what one would expect if the program change was driven by changes in the school's demographic composition. Nevertheless, we attempt to further address the potential endogeneity of a school's tracking status by conducting a two stage least squares analysis which uses as instruments two and three way interactions between three variables: the number of academic courses required for state graduation, the number of schools in the county (measured in 1987), and the fraction of voters in the county who voted for President Reagan in the 1984 election.^{24 25} The interactions between these variables should be uncorrelated with changes in a school's socioeconomic composition (and when we include them as additional regressors in equation (3) none of their estimated coefficients are statistically distinguishable from zero, nor are they jointly significant at conventional levels), but they do explain some of the variation in the tracking variable. Depending on

²⁴ These variables come from the Digest of Education Statistics, the City and County Data book and the Common Core of Data.

the definition of tracking, the first stage partial R^2 runs from 0.011 to 0.047. In addition, the p-values of the marginal explanatory power of the instrumental variables in the first stage is never bigger than 0.002. The partial R^2 are small enough that our IV strategy may yield imprecise estimates, but they are large enough to make us hopeful that the estimates will not display a gross inconsistency of the type discussed in Bound, Jaeger and Baker (1995).

Why do the interactions between these variables make sensible instruments? When state academic requirements are minimal, low ability and high ability students are likely to select into different classes, but as the potential for mixing increases with the number of academic classes required for high school graduation, we hypothesize that parents of high ability children will be more likely to pursue schooling options that separate their children from low-ability peers. As a result, increased academic requirements may provide schools with more of an incentive to track, and this incentive should be larger when parents have more schools from which to choose. Presumably, this differential incentive will be further influenced by parental tastes: we would expect that schools in communities where the residents have less taste for mixing should be more likely to track as competition increases, than would schools in communities where the residents have more taste for mixing. These interaction terms vary over time because there is a nontrivial amount of variation in course requirements over time: for example, between 1987 and 1993, fourteen states altered their course requirements. The average change among these states was 2.5 additional courses. We include the three-way interaction because it increases the amount of variance in the endogenous variable that can be explained.

²⁵ Note that the interaction between the number of schools in the county in 1987 and the fraction of voters in the county who voted for President Reagan in 1984 is a variable that does not vary over time. Thus, it is subsumed in the school-specific fixed effect.

We would, therefore, expect that tracking will increase with the degree of competition as state requirements increase, and that this increase will be stronger in communities with less taste for mixing. That is, we expect that the value of the second-order interaction between state requirements and competition should be positive (over the range of relevant values), and that the value of the third-order interaction between state requirements, competition, and percent voting for Reagan in 1984 (our proxy for parental tastes) will be positive if more conservative voters tend to have less taste for mixing. Indeed, we find that the third-order interaction is positive and statistically significant in the first stage, regardless of the measure of tracking employed. For instance, when the presence of a gifted program is used as our measure of tracking, the coefficient estimate on the third-order interaction is 0.00012 and is significant at the five percent level. The second-order interaction itself is negative (point estimate of -0.0048) and insignificant; however, it is difficult to directly interpret this coefficient because the effect it is intended to capture is partly incorporated in the third-order interaction. When we estimate the model without including the three-way interaction, the two-way interaction between state requirements and the number of schools in the county is positive (0.0013) and modestly significant (at the 11 percent level).

Interpreting these two coefficients together, we find that in a 25th percentile county in the voting distribution (55 percent voting for Reagan in 1984), a one-standard-deviation increase in the number of high schools leads to a three percentage point increase in the probability that the school will have a gifted program, *for each one-course increase in state-imposed graduation requirements*, and in a 75th percentile county (69 percent voting for Reagan in 1984), a one-standard-deviation increase in the number of high schools leads to a seven percentage point increase in the probability that the school will have a gifted program, *for each one-course increase in state-imposed graduation*

requirements. These two figures are each statistically distinct from zero, as well as from one another.

In addition to the interaction terms discussed above, we also include the two-way interaction between the state requirements and the fraction of 1984 county voters who voted for Reagan. We have no *a priori* expectations about whether Republican voters are more or less responsive to changes in state requirements, but we do expect that tastes should influence the degree to which changes in state requirements will affect the tendency to track. The first stage coefficient estimate on this variable is 0.0016, but is not statistically significant when the three-way interaction is also included in the model. However, when the three-way interaction is excluded from the model, the coefficient estimate is 0.004 and is modestly significant at the 11 percent level.

Although we include the non-interacted versions of our instrumental variables in the analysis, they are not included as separate instruments. Sensible arguments can be made that these variables are not exogenous with respect to student outcomes. For example, some authors have argued that school competition increases student test scores. Likewise, parents' tastes may be correlated with student test scores. Since two of the three variables do not change over the period of time that we consider, they are subsumed into a school fixed effect and are, therefore, implicitly included in both stages of the regression. The third variable (state graduation course requirements) is directly included in both stages. Since the two-way interaction between the number of schools in the county and the fraction of county voters who voted for President Reagan in 1984 does not change over time, this variable is also subsumed in the school fixed effect.

The bottom panel of Table 4 provides our two stage least squares estimates, which are generally significantly different from zero at conventional significance levels. These results suggest that the OLS estimates presented in the top panel are not driven by reverse

causality. Two stage least squares estimation strategies frequently produce standard error estimates that are much larger than those produced by OLS, however, and ours are no exception: the point estimates range from 7 to 17 percentage points, but the confidence intervals around these estimates also include much smaller and much larger values.

Although our IV estimates provide evidence that tracking programs affect the socioeconomic composition of a school's student body, they are not precise enough to clarify the magnitude of the effect.

Taken together, the results produced by our analyses indicate that schools' can attract relatively higher income students by adopting gifted and remedial programs. Of course, school choice decisions may be more affected by the presence of gifted/remedial programs than by the presence of subject specific ability grouping, but if student mobility is influenced by gifted and remedial programs it is likely to be influenced by tracking policies as well.

C. Estimating Achievement Gains Taking Student Mobility into Account

Our finding that tracking programs affect the socioeconomic composition of the student body suggests that the estimates in Tables 2 and 3 may be biased. We investigate this possibility by applying two stage least squares, using the same instruments that we used in our SASS analysis: two and three way interactions between state graduation requirements (number of academic classes required for high school graduation), the number of schools in the county in 1987, and the fraction of voters in the county who voted for President Reagan in the 1984 election. As discussed in the previous section, these instruments are correlated with a school's tracking status, but are unlikely to be correlated with student outcomes.

Table 5 reports our second stage coefficient estimates, together with the p-value of the instruments from the first stage regressions, the first stage partial R^2 , and the p-value of the Hausman exogeneity test. The first stage results indicate that our instruments are correlated with our tracking measures, with p-values lower than 0.001 across all definitions of tracking. Because the partial R^2 's are 0.012 or above, we are encouraged that our IV estimates are unlikely to display the type of inconsistency discussed by Bound, Jaeger and Baker (1995).

We find that the effect of tracking is, if anything, *positive* for members of the low ability group, regardless of whether we employ the principal-reported measure of tracking or one of our teacher-reported measures. The estimates for the middle group are not distinguishable from zero, and the estimates for the high ability group are actually negative (though usually insignificant) when we apply IV. Using the principal-reported measure of tracking, the coefficient estimate for the low-ability group is 6.8 points, an effect equivalent to moving a student from the 20th percentile of the eighth grade ability distribution to about the 40th percentile of the tenth grade ability distribution. Although this estimate is not precise, the standard error estimate is small enough to reject (at the two percent level) the null hypothesis that tracking has no effect, and the confidence interval around the estimated effect for low-achieving students does not include the estimates reported in Tables 2 and 3. In other words, our IV estimate is both statistically different from zero, and larger in magnitude than the estimates based on regressions that ignore the impact of tracking policies on school choice. Using constructions of tracking measures based on teacher reports yields similar point estimates, but they are less precise: one can reject the null that the effect of tracking is zero at the nine to fifteen percent level, depending on specification. The wider confidence intervals around the estimates derived from teacher reported tracking measures are probably related to the noisiness of

those measures. In sum, we find evidence that ability tracking may lead to *improved* outcomes for low-achieving students, rather than the diminished outcomes generally believed to exist.

V. Conclusions

The current trend away from ability tracking results largely from the perception that tracking is harmful to low-ability students. Previous empirical research has concluded that high-ability students gain from tracking at the expense of their less-able schoolmates, but those studies have not adequately addressed the possibility that track placement and tracking programs may be endogenous with respect to student outcomes. Research on tracking effects has also suffered because of ambiguity over the definition of tracking.

We estimate the effects of tracking on students of different ability levels by comparing achievement gains across similar students attending tracked vs. untracked schools. By dividing the sample into groups based on their 8th grade test scores instead of using their track assignment we are able to identify the effect of tracking on different ability groups using only variation across types of schools (those with tracking vs. those that do not track) rather than using variation *both* across *school* and *student* types. This empirical strategy produces estimates that are strikingly different from those produced by comparing individuals schooled in different tracks—in particular, our estimates provide no evidence that tracking harms low ability students. This finding is robust to a number of different tracking definitions.

We also demonstrate that gifted and remedial programs help schools attract students who are from medium-high income families (as opposed to those who are free lunch eligible). This result suggests that even if students do not benefit from attending

homogeneous classes per se, schools may wish to employ tracking policies as a means of maintaining an economically diverse student body. Low-ability students may indirectly benefit from tracking policies if the retention of higher-income students increases school expenditures, improves teacher quality or creates other positive externalities. These should be kept in mind by school administrators when considering the potential costs and benefits of detracking.

Our finding that tracking affects school choice suggests that our comparison of achievement gains across tracked vs. untracked students may produce biased estimates, and so we supplement our results with two stage least squares estimates. The results from this exercise suggest that low-ability students may actually experience larger test score gains when they are schooled in tracked settings. Although the point estimates produced by the IV procedure are not very precise and should be regarded cautiously, when they are taken together with their standard error estimates we can find no evidence that detracking America's schools, as is currently in vogue, will improve outcomes among disadvantaged students. This trend may instead harm the very students that detracking is intended to help.

References

- Alexander, Karl L. and Edward L. McDill (1976), "Selection and Allocation within Schools: Some Causes and Consequences of Curriculum Placement," American Sociological Review, 41, 963-980.
- Argys, Laura M., Daniel I. Rees, and Dominic J. Brewer (1996), "Detracking America's Schools: Equity at Zero Cost?" Journal of Policy Analysis and Management, 15(4), 623-645.
- Betts, Julian R. and Jamie L. Shkolnik (2000), "The Effects of Ability Grouping on Student Math Achievement and Resource Allocation in Secondary Schools," Economics of Education Review, 19 (1), 1-15.
- Bishop, John (1994), "Is the Test Score Decline Responsible for the Productivity Growth Decline?" American Economic Review, 79 (1), 178-197.
- Bound, John and George Johnson (1992), "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," American Economic Review, 82 (3), 371-392.
- Bound, John, David A. Jaeger and Regina M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," Journal of the American Statistical Association, 90 (430), 443-450.
- Braddock, Jomills H. and Robert E. Slavin (1993), "Why Ability Grouping Must End: Achieving Excellence and Equity in American Education," The Journal of Intergroup Relations, 20(1), pp. 51-64.
- Epple, Dennis, Elizabeth Newlon and Richard Romano (1997), "Ability Tracking, School Competition, and the Distribution of Educational Benefits," unpublished mimeo.
- 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.
- Hoffer, Thomas B. (1992), "Middle School Ability Grouping and Student Achievement in Science and Mathematics," Educational Evaluation and Policy Analysis 14(3), pp. 205-227.
- Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce (1993), "Wage Inequality and the Rise in Returns to Skill," Journal of Political Economy, 101 (3), 410-442.

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

Murnane, Richard J., John B. Willett and Frank Levy (1995), "The Growing Importance of Cognitive Skills in Wage Determination," Review of Economics and Statistics, 77, 251-266.

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

Oakes, Jeannie (1992), "Can Tracking Research Inform Practice? Technical, Normative and Political Considerations," Educational Researcher 21(4), pp. 12-21.

Rees, Daniel I., Laura M. Argys, and Dominic J. Brewer (2000), "How Should We Measure the Effect of Ability Grouping on Student Performance?" Economics of Education Review, 19 (1): 16-20.

Rosenbaum, J.E. (1980a), "Track Misperceptions and Frustrated College Plans: An Analysis of the Effects of Tracks and Track Perceptions in the National Longitudinal Survey," Sociology of Education 53(2), 74-88.

Rosenbaum, J.E. (1980b), "Social Implications of Educational Grouping," Review of Research in Education, 8, 361-401.

Slavin, Robert E. (1990), "Achievement Effects of Ability Grouping in Secondary Schools: A Best Evidence Synthesis," Review of Educational Research 60(3), 471-499.

Vanfossen, Beth E., James D. Jones and Joan Z. Spade (1987), "Curriculum Tracking and Status Maintenance," Sociology of Education, 60, 104-122.

Wheelock, Anne (1992), Crossing the Tracks: How "Untracking" Can Save America's Schools (New York: The New Press).

Table 1: Sample means of variables included in achievement equations

<u>Variable</u>	Mean in tracked schools (standard deviation)	Mean in untracked schools (standard deviation)	p-value of difference
Change in math test score from 8 th to 10 th grade	7.290 (6.667)	7.117 (6.520)	0.364
8 th grade math test score	36.640 (11.880)	35.059 (11.438)	0.000
White	0.729	0.768	0.002
Black	0.096	0.093	0.596
Hispanic	0.108	0.087	0.016
Income under \$15,000	0.154	0.208	0.000
Income \$15-35,000	0.341	0.384	0.002
Income \$35-50,000	0.213	0.186	0.026
Income \$50-100,000	0.147	0.110	0.000
Income over \$100,000	0.066	0.035	0.000
Parent high school graduate	0.915	0.894	0.011
Parent college graduate	0.293	0.217	0.000
Central city	0.205	0.138	0.000
Suburb	0.453	0.367	0.000
Northeast	0.223	0.125	0.000
South	0.318	0.403	0.000
West	0.187	0.147	0.001
Days in school year	179.664 (3.333)	178.962 (3.018)	0.000
Percent white in school	74.318 (29.309)	79.276 (27.178)	0.000
Percent free lunch eligible in school	18.871 (19.087)	23.664 (20.441)	0.000
Highest teacher salary in school (1000s)	39.803 (8.111)	36.044 (7.707)	0.000
Cohort enrollment (1000s)	0.347 (0.216)	0.267 (0.189)	0.000
Student-teacher ratio	15.881 (3.675)	15.425 (3.583)	0.000

Note: The variables for which standard deviations are not provided are dichotomous variables.

Table 2: OLS estimates of the effects of ability tracking on mathematics test score gains

Variable	Specification including track placement	Bottom third of 8 th grade test distribution	Middle third of 8 th grade test distribution	Top third of 8 th grade test distribution
Placement in bottom track	-6.000 (1.112)			
Placement in middle track	-0.633 (0.557)			
Placement in top track	2.226 (0.570)			
School ability tracks in mathematics (principal report)		-0.404 (0.338)	-0.063 (0.395)	-0.185 (0.322)
8 th grade math test score	-0.323 (0.027)	0.201 (0.041)	-0.048 (0.040)	-0.249 (0.018)
White	-0.423 (0.545)	1.180 (0.617)	-1.753 (0.634)	-1.489 (0.378)
Black	-2.178 (0.958)	0.077 (0.616)	-1.963 (0.836)	-2.488 (0.794)
Hispanic	-1.238 (0.967)	0.523 (0.576)	-1.774 (0.693)	-2.244 (0.640)
Income under \$15,000	0.183 (0.740)	-0.951 (0.469)	0.679 (0.609)	0.369 (0.610)
Income \$15-35,000	-0.242 (0.553)	-0.213 (0.451)	1.456 (0.593)	0.256 (0.483)
Income \$35-50,000	0.060 (0.553)	0.457 (0.538)	2.129 (0.587)	0.402 (0.482)
Income \$50-100,000	-0.185 (0.622)	-0.360 (0.672)	1.420 (0.662)	0.180 (0.499)
Income over \$100,000	-0.005 (0.736)	0.746 (0.932)	1.745 (0.907)	-0.010 (0.544)
Parent high school graduate	0.500 (1.299)	0.988 (0.313)	0.836 (0.530)	1.439 (0.836)
Parent college graduate	1.154 (0.371)	2.126 (0.476)	0.650 (0.349)	0.996 (0.274)
Days in school year	0.000 (0.063)	-0.075 (0.034)	-0.056 (0.053)	-0.035 (0.037)
Percent white in school	0.005 (0.010)	0.001 (0.006)	-0.006 (0.007)	0.000 (0.007)
Percent free lunch in school	-0.020 (0.014)	0.016 (0.008)	-0.025 (0.012)	-0.016 (0.010)
High teacher salary (1000s)	-0.008 (0.030)	-0.024 (0.022)	-0.012 (0.023)	0.025 (0.020)
Tenth grade enrollment (1000s)	1.598 (0.935)	1.191 (0.922)	2.221 (1.013)	0.296 (0.668)
Student-teacher ratio	-0.001 (0.001)	0.003 (0.004)	0.003 (0.001)	-0.000 (0.001)

Note: Standard errors (in parentheses) are corrected to account for within-school clustering of errors. All models also include a constant term, region dummies, and central city/suburb dummies, the coefficients of which are omitted due to space constraints.

Table 3: Estimated treatment effects of ability tracking, using alternative measures of school tracking status

<u>School is ability tracked if:</u>	Bottom third of 8 th grade test distribution	Middle third of 8 th grade test distribution	Top third of 8 th grade test distribution
Principal reports school tracked in mathematics	-0.404 (0.338)	-0.063 (0.395)	-0.185 (0.322)
Teachers report both “high-achieving” and “low-achieving” classes	0.062 (0.489)	0.520 (0.555)	0.751 (0.531)
Number of tracks observed (among “high,” “average,” and “low-achieving” classes)	-0.089 (0.169)	0.007 (0.211)	0.031 (0.169)
Any teacher reports a “high-achieving” class	-0.091 (0.323)	0.676 (0.399)	0.483 (0.384)
Teachers report both “advanced” and “general” classes	-0.271 (0.316)	0.280 (0.400)	-0.314 (0.301)
Number of tracks observed (among “advanced,” “academic,” and “general” classes)	-0.274 (0.241)	0.267 (0.318)	-0.314 (0.230)
Any teacher reports a “advanced” class	-0.099 (0.318)	0.275 (0.413)	-0.398 (0.338)

Note: Standard errors (in parentheses) are corrected to account for within-school clustering of errors. All models also include all covariates from Table 2, an indicator of the number of unique classes observed in the school (for teacher reported measures of tracking), and a constant term. The **total** sample sizes using the principal-reported measure of tracking is 11,288, while the **total** sample size using the teacher-reported measures of tracking is 9,516.

Table 4: Ability tracking and school composition changes: panel evidence from the SASS

Sample	<i>Tracking measure</i>			
	School has a gifted program	School has a remedial math program	Number of programs (gifted + remedial)	School has either a gifted or remedial program
Full sample	-0.026 (0.010)	-0.019 (0.007)	-0.020 (0.006)	-0.027 (0.012)
number of schools	545	545	545	545
Schools with secondary grades	-0.023 (0.011)	-0.026 (0.009)	-0.022 (0.007)	-0.022 (0.014)
number of schools	411	411	411	411
Schools with elementary grades	-0.023 (0.013)	-0.008 (0.010)	-0.014 (0.008)	-0.030 (0.017)
number of schools	274	274	274	274

SCHOOL COMPETITION AND THE COMPOSITIONAL EFFECTS OF ABILITY TRACKING

Greater than median schools in county	-0.049 (0.016)	-0.027 (0.011)	-0.033 (0.009)	-0.068 (0.020)
Fewer than median schools in county	-0.012 (0.013)	-0.015 (0.010)	-0.013 (0.007)	-0.008 (0.014)
Difference	-0.038 (0.020)	-0.012 (0.015)	-0.021 (0.012)	-0.059 (0.024)

INSTRUMENTAL VARIABLES ESTIMATES OF THE EFFECTS OF ABILITY TRACKING

Instrumental variables coefficient	-0.138 (0.067)	-0.103 (0.086)	-0.071 (0.043)	-0.167 (0.087)
p-value of first stage instrument explanatory power	0.002	0.000	0.000	0.000
Partial-R ² in first stage	0.011	0.014	0.016	0.047
p-value of Hausman overidentification test of instrument exogeneity	0.674	0.546	0.636	0.909

Note: White standard errors are in parentheses.

Table 5: Instrumental variables estimates of the effects of ability tracking on math performance

Tracking measure:	Principal-reported measure of tracking	Number (advanced, academic, general) observed in school	Advanced class observed in school	Both advanced and general classes observed in school
BOTTOM THIRD OF INITIAL SCORE DISTRIBUTION				
Treatment effect	6.844 (2.710)	3.362 (2.155)	4.582 (2.750)	3.781 (2.624)
p-value of Hausman exogeneity test	0.468	0.188	0.112	0.121
Partial R ² in first stage	0.015	0.016	0.012	0.013
first stage p-value	0.000	0.000	0.000	0.000
MIDDLE THIRD OF INITIAL SCORE DISTRIBUTION				
Treatment effect	-3.439 (2.147)	1.456 (1.693)	2.369 (2.351)	2.302 (2.291)
p-value of Hausman exogeneity test	0.809	0.381	0.414	0.420
Partial R ² in first stage	0.019	0.019	0.013	0.014
First stage p-value	0.000	0.000	0.000	0.000
TOP THIRD OF INITIAL SCORE DISTRIBUTION				
Treatment effect	-3.819 (3.075)	-1.425 (0.956)	-1.835 (1.883)	-1.636 (1.180)
p-value of Hausman exogeneity test	0.575	0.750	0.488	0.661
Partial R ² in first stage	0.019	0.056	0.026	0.050
First stage p-value	0.000	0.000	0.000	0.000

Note: Standard errors (in parentheses) are corrected to account for within-school clustering of errors. All models also include all covariates from Table 2, an indicator of the number of unique classes observed in the school (for teacher-report-based measures of tracking.)