ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-65

Beyond tracking and detracking: The dimensions of organizational differentiation in schools

Thurston Domina
University of North Carolina, Chapel Hill

Andrew McEachin
RAND Corporation

Paul Hanselman
University of California, Irvine

Priyanka Agarwal
University of California, Irvine

NaYoung Hwang
University of California, Irvine

Ryan Lewis
University of California, Irvine

Schools utilize an array of strategies to match curricula and instruction to students' heterogeneous skills. While generations of scholars have debated "tracking" and its consequences, the literature fails to account for diversity of school-level sorting practices. In this paper we draw upon the work of Sørenson (1970) to articulate and develop empirical measures of five distinct dimensions of school cross-classroom tracking systems: (1) the degree of course differentiation, (2) the extent to which sorting practices generate skills-homogeneous classrooms, (3) the rate at which students enroll in advanced courses, (4) the extent to which students move between tracks over time, and (5) the relation between track assignments across subject areas. Analyses of longitudinal administrative data following 24,000 8th graders enrolled in 23 middle schools through the 10th grade indicate that these dimensions of tracking are empirically separable and have divergent effects on student achievement and the production of inequality.

VERSION: May 2019

Suggested citation: Domina, T., McEachin, A., Hanselman, P., Agarwal, P., Hwang, N., & Lewis, R. (2019). Beyond tracking and detracking: The dimensions of organizational differentiation in schools (EdWorkingPaper No.19-65). Retrieved from Annenberg Institute at Brown University: http://edworkingpapers.com/ai19-65

Beyond tracking and detracking:

The dimensions of organizational differentiation in schools

Thurston Domina,[1] Andrew McEachin,[2] Paul Hanselman,[3] Priyanka Agarwal,[3] NaYoung Hwang,[3] and Ryan Lewis[3]

Abstract: Schools use an array of strategies to match curricula and instruction to students' heterogeneous skills. While generations of scholars have debated "tracking" and its consequences, the literature fails to account for diversity of school-level sorting practices. In this paper we draw upon the work of Sørenson (1970) and others to articulate and develop empirical measures of five distinct dimensions of within-school cross-classroom tracking systems: (1) the degree of curricular differentiation, (2) the extent to which sorting practices generate skills-homogeneous classrooms, (3) the rate at which students enroll in advanced courses, (4) the extent to which students move between tracks over time, and (5) the relationship between track assignments across subject areas. Analyses of longitudinal administrative data following approximately 20,000 8th graders enrolled in 23 middle schools through the 10th grade indicate that these dimensions of tracking are empirically separable and have divergent relationships with student achievement and the production of inequality.

Acknowledgements: Research reported in this article was supported by the Spencer Foundation (award no. 201400180), by the W. T. Grant Foundation (award no. 183913), and by the Institute for Education Sciences (IES; award R305B120013). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Spencer Foundation, the W. T. Grant Foundation, or IES. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Spencer or W.T. Grant Foundations or IES. The authors thank the following for helpful comments and suggestions: Steven Hemelt, Douglas Lauen, Mark Long, Andrew Penner, Emily Penner, Karolyn Tyson and commenters at the Education and Inequality Seminar at the University of North Carolina, Chapel Hill and the 2016 American Educational Finance and Policy and American Educational Research Association meetings.

[1] Corresponding author. University of North Carolina, Chapel Hill, tdomina@email.unc.edu
[2] RAND Corporation
[3] University of California, Irvine

Schooling may be a "great equalizer" (Mann 1848; Downey et al. 2004; Raudenbush and Eschmann 2015; von Hippel et al. 2018). But at the organizational level, schools are deeply implicated in the production, maintenance, and legitimation of educational inequality. Schools repeatedly sort students, and in the process, they allocate opportunities, resources, and status distinctions unequally across groups (Dreeben and Barr 1988; Kerckhoff 1995; Oakes 1985). One way in which sociologists and other social scientists investigate the mechanisms through which schools generate and perpetuate social inequalities is by studying "tracking," – an umbrella term that refers to a broad array of practices associated with the grouping of students into distinct courses of study.

Several influential studies explore the ideological, political, and technical pressures that lead educators to group students for instruction (Dreeban and Barr 1988; Hallinan 1992; Oakes 1985; Oakes and Guiton 1995; Rosenbaum 1976). Much of this work situates tracking historically. In the American case, this literature suggests that social Darwinist ideas about race and native intelligence enabled the construction of academically differentiated high schools (Cremin 1964; Katz 1975; Kleibard 1995; Oakes 1985) while the Civil Right's movement's ascendance relaxed and reformed school tracking systems (Lucas 1999; Loveless 2011; Powell, Farrar and Cohen 1985). But underneath these broad social trends, school-level analyses reveal that the phenomenon that we refer to as "tracking" actually consists of a dynamic and diverse set of school practices that change over time and space (Dreeban and Barr 1988; Metz 2003; Oakes 1985; Rosenbaum 1999), often in the midst of considerable uncertainty and debate (Lewis and Diamond 2015; Oakes, Wells, Jones and Datnow 1997; Rickles 2011; Watanabe 2006; Wells and Oakes 1996).

The relationship between tracking and the distribution of educational opportunities remains the most persistent question in this literature (Oakes 1985; Argys, Rees, and Brewer 1996; Kerckoff 1986; Carbonaro 2005). Some of this research takes a macro-level approach, comparing outcomes across national educational systems that take different approaches to educational stratification (Buchmann and Park 2009; Hanushek and Woessman 2006; Shavit and Blossfeld 1993). Other recent work takes a micro-level approach, studying individual students' locations in stratified curricular systems and their consequences, (Carbonaro and Gamoran, 2002; Gamoran and Nystrand 1994; Kelly and Carbonaro 2012; Van Houtte 2004). While the sociology of education provides a rich conceptual framework for understanding tracking and its effects, we argue that both macro- and micro-level work on tracking and its consequences has been hindered by a failure to adequately measure the diverse ways in which academic tracking manifests in schools. Some studies compare tracked and untracked schools; others compare students placed in different track locations or courses of study within schools (Betts and Shkolnick 2000; Carbonaro 2005; Duflo, Dupas, and Kremer 2008; Figlio and Page 2002; Gamoran, Nystrand, Berends, and LePore 1995; Kelly and Carbonaro 2012). As a result, the existing empirical literature provides limited information about how specific organizational practices associated with tracking matter for student outcomes and inequality.

In this paper, we draw upon seminal research by Sørenson (1970) and others (Gamoran 1992; Kelly 2007; Lucas 1999; Lucas and Berends 2002) to articulate several dimensions of school-level academic tracking systems. Focusing on middle school mathematics and English courses, we hypothesize that school-level tracking systems differ in at least five important ways: (1) the extent to which schools use distinct courses to differentiate curricula, (2) the degree of within-classroom skills homogeneity school tracking practices create, (3) the proportion of students who enroll in high-track courses, (4) the amount of between-track mobility that occurs as students move from middle to high school, and (5) the extent to which course placements are related across subjects. Our project thus contributes to the research literature on tracking and its consequences by highlighting a handful of distinct ways in which schools sort students for instruction. Just as recent work on between-firm variation in workplace inequality highlights the role that local processes play in the production of inequality (e.g. Stainback, Tomaskovic-Devey, and Skaggs 2014), studying school-level tracking systems highlights specific choices schools make and their implications for the production of educational inequality.

We use a unique set of administrative data from three ethnically- and economically-diverse school districts to measure the dimensions of school tracking systems and study their relation to student academic skills development. These data capture 20,000 8^{th} graders in 23 public middle schools. In contrast to the national probability sample data that are widely used elsewhere in the tracking literature, our data provide detailed longitudinal achievement, demographic, and transcript information for all students enrolled in sample schools. Because these administrative data are drawn from a small number of locations, we cannot make strong claims about the generalizability of our findings beyond this setting. However, these data facilitate the rich description of school tracking systems that emerged from an earlier generation of case studies as well as the inferential rigor of contemporary quantitative analyses. We observe students and all of their peers in middle school math and reading classrooms and collect supplemental qualitative data on the political and administrative contexts in which these classrooms are situated. In doing so, we generate nuanced and time-varying measures of the dimensions of tracking in each sample school and link these measures with subsequent student outcomes.

Our analyses reveal substantial between-school and within-school (over time) variability, with independent variation in the separate dimensions of tracking. We use between school and temporal variation in our five measures of tracking to test their effects on student outcomes, allowing us to account for persistent unobserved differences between schools. Our findings indicate that different dimensions of school tracking systems have independent (and occasionally counter-acting) consequences for student achievement and student achievement inequality. Further, we find suggestive evidence that school-level tracking systems may exacerbate achievement inequalities within schools by providing a boost for high-achievers relative to their lower achieving peers.

Tracking and its implications

The study of tracking and its consequences is central to understanding the role that education plays in the construction of social inequality. Several studies suggest that students in rigidly tracked schools demonstrate no greater academic achievement, on average, than students in untracked schools (Hoffer 1992; Kerckoff 1986; Slavin 1988). However, students in high-track classes enjoy a wide range of educational advantages relative to their peers in low-track classes including access to high-achieving peers, high educator expectations, and rigorous instruction (Carbonaro and Gamoran 2002; Gamoran and Nystrand 1994; Kelly and Carbonaro 2012, Van Houte 2004). These educational advantages translate to higher levels of educational achievement, greater access to post-secondary education, and higher levels of educational attainment (Author 2008; Long, Conger, and Iatorola 2012). Further, poor students, students whose parents have relatively low levels of educational attainment, and students of color are all less likely to enroll in high-track classes. Accordingly, much of the research literature suggests that school tracking practices have negligible average effects on student achievement, but that these practices contribute to achievement inequalities by providing relative educational advantages to the already-advantaged students who enroll in high-track classes. (See Gamoran 2010 for a more thorough review of this literature.)

However, the research literature is by no means unanimous on tracking's impact on achievement and achievement inequality. From a teacher's point of view, differentiation is a technical response to pedagogical challenges associated with schooling large and heterogeneous student populations (Hallinan 1994; Rosenbaum 1999). One might thus expect some forms of tracking to help teachers target instruction to their students' needs, yielding positive effects for a broad range of students, particularly for students with unique needs (e.g. special education and ELL). Consistent with this hypothesis, a handful of studies using experimental and quasi-experimental methods indicate that sorting students into skills-homogeneous classes has positive achievement effects for students across the skills distribution (Betts and Shkolnick 2000; Figlio and Page 2002; Duflo, Dupas, and Kremer 2008). Further, large-scale policy efforts to create more skills-heterogeneous classroom assignments often have unintended negative consequences for high- and low-achieving students alike (Allensworth, Nomi, Montgomery, and Lee 2009; Author 2015).

Conceptualizing "tracking"

We argue that while past research conceptualizes tracking in nuanced ways, the often simplistic measurement and operationalization of tracking in the empirical studies referenced above may contribute to this literature's mixed and ambiguous findings. Scholars utilize a variety of measures to operationalize tracking, including principal reports of school differentiation practices and written school policies related to course assignments (Betts and Shkolnick 2000; Hoffer 1992; Kelly 2007; Kelly and Price 2011), student reports of track assignment (Gamoran and Mare 1989), teacher reports of classroom composition (Argys, Rees, and Brewer 1996), and transcript-verified measures of student course assignments (Lucas 1999). In many cases, these measures impose simplistic categorizations on school tracking systems, classifying schools as "tracked" or "untracked" or dividing students between "vocational" and "academic" tracks. While this literature demonstrates the importance of tracking for educational achievement and inequality, it largely fails to address the ways in which tracking systems likely differ and the consequences of these differences for student outcomes. As an example, Duflo et al. (2008) estimate the effects of an isolated change in one dimension of a school tracking system – the degree to which students are grouped by ability into separate classrooms for instruction – but provides little evidence regarding the relations among this change and other dimensions of school tracking systems. Investigating these relations is essential to understanding the social organization of schooling and designing effective and equitable instructional practices.

A handful of studies attempt to operationalize a more nuanced view of school tracking systems. Hallinan (1992, 1994) documents between-school variation in the number of academic tracks as well as between-school variation in the relationship between placement in high-track courses and student outcomes across schools. Using school course catalogues and assignment policies to measure the several dimensions of tracking systems, Kelly and Price (2011) find that schools with wide variation in student skills are most likely to develop highly differentiated academic tracking systems. Lucas (1999) uses student-level data from the nationally representative High School and Beyond (HSB) to measure the flexibility of secondary school tracking systems, demonstrating that despite the dissolution of an over-arching track system, the curricular experiences of students U.S. high schools remain highly stratified by race and class. Using the same data, Gamoran (1992) provides evidence to suggest that different dimensions of school tracking systems have different consequences for students, demonstrating that the achievement effects of enrolling in high-track courses varies across schools. In particular, Gamoran demonstrates that relatively flexible school tracking systems are associated with high levels of mean student achievement and low levels of cross-track achievement inequality.

These studies point to the potential for a more nuanced view of tracking practices for understanding the role that schools play in the production and reproduction of social inequality. However, each faces substantial data limitations. Lacking access to student-level data, Kelly and Price (2011) are unable to test the relationship between tracking systems and student outcomes. Meanwhile, Lucas (1999), Gamoran (1992), and other scholars who have studied tracking using NCES cohort-based studies (including the HSB, NELS, ELS, and HSLS) are limited by the paucity of available contextual data (Argys et al. 1996; Betts and Sckolnik 2000; Figlio and Page 2002; Lucas 1999; Gamoran 1992). These panel studies generally provide detailed data on 20-50 students sampled from each of approximately 500 secondary schools. While this stratified sampling scheme provides data on a nationally representative sample, it situates the student as the unit of analysis and provides limited direct data on the emergent institutional structures in which students are situated. In particular, these student panel studies provide limited data about the range of courses schools offer and the ways in which schools sort students across those courses, and the demographic and skill composition of students' classroom peers. Moreover, because these studies collect information on only one cohort of students per school, they do not allow consideration of within-school variation in elements of tracking over time. By contrast, our data provide student-level measures as well as school and classroom indicators for every 8th grader enrolled in 23 schools in three districts over three school years. As a result, these data make it possible to observe several highly salient dimensions of school tracking systems that are unobservable in widely-utilized nationally representative panel datasets.

The dimensions of tracking

We understand school tracking systems as the culmination of an array of school-level processes related to the provision of differentiated academic coursework and the allocation of students among the available courses. We thus measure track structures as school-level variables. Building upon Sørenson's theoretical work (1970) as well as prior efforts to measure the dimensions of tracking, we develop a framework for thinking about and measuring school tracking systems. We articulate and measure five conceptually distinct dimensions of school tracking systems; trace the relationship between these dimensions of school tracking systems and student achievement growth; and investigate the extent to which the effects of school tracking systems vary with students' prior achievement.

Central to this undertaking is the supposition that tracking systems vary on multiple dimensions both across schools and over time. In particular, we identify and measure the following five dimensions of school tracking systems. Table 1 provides a summary of these five dimensions, which we describe in detail below:

TABLE 1 ABOUT HERE

1. Degree of curricular differentiation. Sørenson (1970, p. 355) defines organizational differentiation as "the division of a school's student body into subgroups of a permanent character." Some form of organizational differentiation is nearly ubiquitous in contemporary schools. The U.S. public education system sorts children into schools by neighborhood and parental preferences. These schools then sort children by age into grades. However, beyond these basic forms of differentiation, schools vary considerably in the degree to which they differentiate curriculum and instruction. Schools may differentiate curriculum and instruction *horizontally* by providing students with various learning environments in which they can be exposed to different bodies of knowledge, as when a university offers a wide range of graduate seminars focusing on distinct topics. In addition, they may differentiate curriculum and instruction *vertically*, by creating different learning environments that expose students to similar bodies of knowledge but at different paces, levels of rigor, and/or with differing degrees of social status. In our conceptualization, schools that offer students a broad range of classes – whether vertically or horizontally differentiated – display a high degree of differentiation (as measured by the number of course offerings), while schools that offer few classes display a low degree of differentiation (Hallinan 1992). All else equal, one might expect curricular differentiation to have positive consequences for student achievement, since it allows both educators to develop subject-matter and skill-level specializations and students to find classes that match their academic interests and instructional needs. Given that our focus is on tracking in a middle school context, it is not likely that schools will use horizontal differentiation (e.g. students often get access to courses of varying degrees of difficulty).

2. Classroom skills homogeneity (ability grouping). By sorting students across learning environments according to their measured skills, many tracking strategies attempt to simplify the task of instruction. While teachers in skills-heterogeneous (or ungrouped) classrooms may struggle to deliver instruction that is at the appropriate level for a wide range of students (Rosenbaum 1999), skills-homogenous grouped classrooms may allow teachers to provide instruction that is more appropriately tailored to their students (Eccles and Roeser, 2011). Schools vary in the extent to which their assignment processes generate skills-homogeneous classrooms. Some schools attempt to assign students to courses exclusively on the basis of their prior test scores (Dougherty et al. 2015; Kelly 2009). However, scheduling constraints and limited resources often restrict educators' discretion over students' classroom assignments. Further, many schools allow teacher recommendations as well as parent and student preferences to influence classroom assignments (Oakes and Guiton 1995; Rickles 2011). As a result, even in otherwise "tracked" schools, students with very different skills levels may sit in the same academic classrooms (Mickelson 2001). Conversely, even in explicitly "untracked" schools, informal pathways may develop that lead students to be grouped based on skills levels across classrooms (Horvath 2015; Watanabe 2006). Building upon Sørenson's notion of "selectivity," we conceptualize the degree to which schools assign students to skills-homogeneous classrooms as a distinct dimension of tracking systems.[4]

This dimension of tracking systems likely has mixed consequences for students. While skills-homogeneous classroom assignments may allow teachers to target their instruction to student skills; such grouping strategies may broaden skills gaps by exposing high-achieving students to positive peer effects and low-achieving students to negative peer effects (Becker 1987; Epple, Newland, and Romano, 2002; but c.f. Marsh 1987; Zimmer 2003). Further, skills-homogeneous classroom assignments may create status hierarchies in schools, creating inequalities in learning opportunities and academic expectations across high- and low-achieving classrooms (Authors 2017; Carbonaro 2005; Kelly and Carbonaro 2012; Dreeben and Barr 1988; Metz 1978; Nystrand and Gamoran 1997; Oakes 1985; Page 1991).

3. Track exclusiveness. Over the last several decades, policy-makers and educators have undertaken a concerted effort to intensify academic curricula in American schools. Nonetheless,

[4] Indeed our conception of skills-homogeneity is nearly identical to Sorenson's notion of selectivity, which he defines (1970, p. 363) as "the amount of homogeneity that educational authorities intend to produce by the assignment, in terms of the index of learning used, shall be denoted the selectivity of the assignment."

schools vary in the extent to which they expose students to high-level academic content (Author 2012, 2015; Stein et al. 2011). Some schools enroll all students in courses previously reserved for relatively high-achieving students; others allocate relatively advanced or academically rigorous instruction to some students, and less advanced and rigorous instruction to others (Author 2016). We label this dimension of school differentiation systems "track exclusiveness." Our conceptualization of exclusiveness focuses on the relative size of the lower track.

If enrolling a student in a more advanced course exposes her to a more rigorous curriculum, one might expect track exclusiveness to limit student achievement. However, track exclusiveness has the potential to positive affect student outcomes. If, for example, inclusive systems expose some students to material beyond their preparation, exclusivity could protect overmatched students' learning. Alternatively, if teachers adjust instruction in high-track classes to accommodate underprepared students, decreased exclusivity could undermine high-achieving students' learning (e.g. Cronbach and Snow 1977; Snow 1989). Further, decreases in track exclusivity might depress achievement for students left in low-track classes by creating new stigmas associated with these classes (Gamoran 1992).

4. Track stability. School tracking systems likely also vary in the extent to which they create opportunities for students to move between tracks over time. We describe this dimension of school tracking systems as "track stability," and seek to distinguish between schools in which track placements are fairly permanent and students have few opportunities to move up or down in a track system from schools in which track placements are relatively fluid over time. Rosenbaum's classic portrayal of tracking at "Grayton High" (1976) provides an example of a "tournament-style" track system, in which few students move from low-track courses to high-track courses and upward track mobility is thus exceedingly rare. Less rare, however, is downward mobility, or the phenomenon of students moving from high-track courses to low-track courses. Subsequent analyses suggest that this description may not always hold, indicating that some schools provide opportunities for both upward and downward track mobility (Hallinan 1996; Lucas 1999; Lucas and Good 2002; McFarland 2006).

Systems that allow for high degrees of track mobility may be particularly effective at matching students with instruction. If so, exposure to a low-stability track system may boost student achievement. However, these positive effects may be less common in "tournament" track systems, where upward mobility is rare and downward mobility is common. It is possible that tournament mobility systems may also boost achievement by facilitating an appropriate match between students and instructional offerings and motivating students. Alternatively, one might expect a high degree of tournament mobility to depress student achievement and broaden inequalities by stigmatizing track mobility and associating it with failure.

5. Track scope. The tracking system that was common in American secondary schools throughout the first half of the 20th Century sorted students to vocational, general, college preparatory tracks, which typically defined students' secondary school curricula. One distinguishing characteristic of this system, as well the between-school tracking systems that are common in secondary education in much of Europe and Asia, is that it places students into overarching tracks such that students who are exposed to high-level instruction in one subject tend to be subject to high-level instruction in all areas (Hanushek and Woessman, 2006; Lucas, 1999). As such, this system can be said to have a high degree of "scope." As Lucas (1999) documents, American schools dismantled this overarching track system during the 1960s and 1970s, creating a system that theoretically allows students to take high-track classes in some subjects and low-track classes in others. Although Lucas's analyses suggest that track scope remained high in American high schools through the 1980s, he shows that track scope varies considerably across schools. Further, Lucas demonstrates that track placements in socioeconomically diverse schools tend to be higher scope than in homogeneous schools. We consider “scope” as a fifth dimension of contemporary tracking systems.

One might expect scope to relate negatively with student achievement, if schools with high degrees of track scope find it difficult to match students with instruction appropriate for their course-specific skills (Sørenson 1970; Hallinan 1994). High-scope tracking systems may also intensify a tendency toward social closure – or cliquishness – in student peer networks, since it limits the extent to which students have the chance to socialize in class with peers outside of their academic track (Hallinan and Sørenson 1985). The resulting social processes may increase the extent to which students identify with their academic track position, exacerbating the association between track assignments and achievement inequality.

Relations among the dimensions of tracking

These five dimensions of tracking are conceptually linked, in the sense that each can be understood as a contributing factor to school tracking systems. Consider for example, a school that offers a highly differentiated curriculum, in which students are grouped into skills-homogeneous classrooms, all but the highest-achieving students are excluded from high-track classes, track placements are highly stable over time, and track scope is high. One would clearly describe such a school as highly tracked. In practice, however, the dimensions of tracking need not vary together across schools or over time. As Oakes observed in her landmark study of tracking in 25 high schools (1985, p. 43) “tracking in schools is *not* an orderly phenomenon in which practices, even within a single school, are consistent or even reflective of clearly stated school or district policies.” Since tracking is a collection of practices that schools accumulate, debate, reject, and reform over time, the same school may rank high on one dimension of tracking and low on another. For example, a school can offer a highly differentiated curriculum composed of a wide array of distinct courses even as it places students into highly skills-heterogeneous classrooms. Another school can maximize track scope by having students spend the entire school day with the same set of peers even as it minimizes track stability by changing students' location in the track system year after year.

In this paper, therefore, we view the relations among the dimensions of tracking as well as the extent to which these dimensions are stable within schools over time as an empirical question. We measure these dimensions, explore their correlation across schools and over time, and investigate their independent links to student achievement and achievement inequality.

Data

We operationalize the above five dimensions of tracking using administrative panel data gathered from all students enrolled in three Southern California public school districts. These data follow more than 20,000 students enrolled as 8th graders during the 2010-11, 2011-12, and 2012-13 school years in 23 Southern California middle schools from their 7th grade through 10th grade years. We strategically selected these three districts for inclusion in this study because they enroll a diverse group of students, have distinct histories related to mathematics and ELA course placement practices, and were willing to participate in qualitative and quantitative primary data collection. These data include: student demographics (gender, race/ethnicity, language status, free/reduced lunch eligibility); student scores on California Standards Tests (CST) in mathematics and English Language Arts (ELA) administered in the spring of 7th, 8th, 9th, and 10th grades; transcript data on student middle and high school math and ELA course assignment and performance; course title, teacher ID, and course period data for these middle and high school courses; and California High School Exit Exam (CAHSEE) scores, which provide a standardized measure of student math and ELA achievement in the spring of 10th grade. We supplement these data with qualitative data gathered in interviews with administrators from each district and approximately 25 teachers who teach 8th grade mathematics courses in the three districts.[5]

Table 2 provides a summary of the longitudinal student-level administrative data that we collected from our sample schools during the 2009-10, 2010-11, and 2011-12 school years. Our sample is by no means nationally representative, and in particular our sample schools enroll a disproportionately large number of Latino and Asian-American students and a correspondingly small number of white and African-American students. However, the sample is racially, ethnically, and economically diverse. Districts A and B, both of which are among the 10 largest public school districts in California, are situated in inner-ring suburban communities that include both middle class and relatively poor neighborhoods. District C spans an affluent beach community as well as a considerably poorer inland city. The share of students in our sample eligible for the federal Free and Reduced Lunch Program, based on their family incomes, roughly matches the state average (55% in 2010-11).[6]

TABLE 2 ABOUT HERE

Methods

These data provide a unique opportunity to develop nuanced measures of school tracking systems. Since our three partner districts provide a census of transcript, achievement, and

[5] We conducted fifteen formal/informal interviews and two focus groups with district administrators, math coaches, and teachers during 2014-2015 school year in the three districts that are part of this study. During the summer of 2014, the team observed six professional development sessions for the three districts and a pilot testing effort at one of the school district with the director of assessments. Interviews consisted of open-ended questions on the district policies and challenges pertaining to student testing, student course placement, curriculum changes and the implementation of new state standards. Observations were done naturalistically and recorded using field-notes. Some focus groups and interviews were audio recorded and transcribed, while others were recorded manually.

[6] http://www.cde.ca.gov/

demographic data for three cohorts of students enrolled as 8th graders in 23 schools, including teacher and period identifiers, we can identify the classrooms in which students took core academic courses and each of their peers in these classrooms. In addition, we draw upon school and district course listings and academic policy documents as well as interviews with educators at the school and district levels to contextualize these transcript and administrative data. In the analyses that follow, we draw upon these quantitative and qualitative data to measure the degree of curricular differentiation, classroom skills homogeneity, track inclusiveness, track mobility, and track scope for 8th graders in 23 schools across three school years. We describe these measures and their properties at the beginning of the results section.

We first analyze these measures at the school/year level (N=69). To explore the extent to which "tracking" as implemented in contemporary schools is a single practice or a collection of independent practices, we estimate a correlation matrix for our measures of the dimensions of tracking. If tracking is best conceptualized as a single institutional practice, one might expect the dimensions of tracking to correlate highly across schools and over time. Alternatively, weak correlations among the dimensions of tracking suggest that tracking may be better conceptualized as a diverse set of structural elements and practices that are realized in different ways across schools and over time.

In this multidimensional conception of tracking, the school-level practices that define the social organization of instruction likely result from time-variant contextually-specific technical, political, and cultural factors. As such, it seems likely that different school-level factors predict different dimensions of tracking. To test this notion, we estimate a series of mixed models of the following form:

(1) $Y_{sdt} = \beta_0 + \beta_1 X_{sdt} + \alpha_t + \alpha_d + u_s + e_{sdt}$

where Y_{sdt} measures the dimensions of organizational differentiation in 8th grade math and ELA for school s in district d at year t; X_{sdt} is a set of time-varying school-level covariates describing observable characteristics of s at time t including: school enrollment, an index of socioeconomic disadvantage (calculated as the mean of the standardized proportion of black and Hispanic students in the school, the standardized proportion of students who qualify for free and reduced lunch, and the standardized proportion of students who are English-Language learners,) students' mean prior achievement levels,[7] and dispersion in students' prior achievement; α_t is a vector of year fixed effects; α_d is a district-level fixed effect; u_s represents school-level random effects; and e_{sdt} is the time-varying school-level error term. While the results of these analyses are limited in external validity, since we can only explore the correlates of tracking systems in the 23 California schools for which we have data, they complement prior research from Kelly and Price (2011) and Oakes (1985) about the school-level correlates of tracking systems.

A multidimensional conception of tracking suggests a more nuanced set of answers to historically contentious questions regarding the effects of tracking for student achievement and inequality. If tracking is actually a collection of conceptually and empirically separable practices, it may be possible to develop school structures that realize the potential benefits associated with instructional differentiation while avoiding the costs that are commonly associated with tracking. To address these questions, we use student-level data to investigate the relationship between exposure to the dimensions of tracking in 8th grade and students' 10th grade achievement scores. These models take the following general form:

[7] Standardized percent black or Hispanic correlates with standardized percent free or reduced lunch at 0.86 and standardized percent English Language Learner at 0.74. Standardized percent free or reduced lunch correlates with standardized percent English Language Learner at 0.86. School mean test score measures correlate at 0.96. Since school-level standard deviations in math and ELA test scores correlate less closely (0.61), we enter these variables separately into the models.

$$(2)\ Y_{icsdt} = \gamma_0 + \gamma_1(Tracking_{sdt}) + \gamma_2(X_{icsdt}) + \alpha_t + \alpha_d + u_c + u_s + r_{icsdt}$$

In these analyses Y_{icsdt} is the 10th grade math and ELA test scores as measured on the California High School Exit Exam (CAHSEE) for student *i*, in classroom *c*, school *s*, district *d*, and year *t*. This exam is administered to all students in the spring of their 10th grade year. At the time of its administration to the students in our sample, the CAHSEE was a requirement for high school graduation.[8] As such, it is a high stakes test for all students in our sample that was administered in a consistent form throughout the study period to virtually all students regardless of their skill level, postsecondary plans, and course enrollments. Since this test primarily measures skills aligned to 6th-8th grade level standards, it may not fully capture the effects of tracking for high achieving students, potentially introducing a negative bias in our findings. In supplementary analyses, we test this bias by estimating parallel models predicting 8th and 9th grade ELA CST scores.[9] However, this bias is likely small since approximately five percent of students in the sample scored at ceiling on any one achievement test. X_i is a set of student-level characteristics including: demographics and prior achievement as measured by students' 7th grade test scores and grade; α_t is a vector of year fixed effects; α_d are district fixed effects; u_c is a class-level random effect; $u_s$is a school-level random effect; and r_{icsdt} is the residual time-varying student-level error term.[10]

The coefficients of interest in this model, γ_1, represent the relationship between school-by-year measures of the dimensions of tracking and students' achievement, independent of the

[8] California has since reversed course on the requirement that students pass the CAHSEE.
[9] We are unable to estimate similar models using Mathematics CST scores, because beginning with Algebra, California middle and high school students take different mathematics tests based on their math course enrollment.
[10] In both equations (1) and (2), the fixed effects terms represented by α_t and α_d absorb all time-invariant district-level variation in the outcomes. By contrast, the school-level random effects term u_s (as well as the course-level random effects term u_c in equation 2) simply account for the non-independence of repeated observations of schools across time.

other relevant measures of the dimensions of tracking as well as district and year fixed effects and student-level controls. If the dimensions of school tracking systems are unrelated to student characteristics and other characteristics of schools, conditional on the demographic and lagged achievement measures we control for, these models generate unbiased estimates of the independent effects of these dimensions of school tracking systems. However, it seems likely that students vary across schools on a wide range of unmeasured characteristics that could potentially confound the observed relation between the dimensions of tracking and student achievement.

To address this potential bias, we fit additional models in which we center each of the tracking measures on their school-level mean. These models, which are equivalent to estimating school fixed effects models, thus estimate the effect of tracking exclusively off of the within-school variation in tracking systems. These models generate unbiased estimates of the effects of these school tracking systems on students' achievement, if the following two assumptions hold: 1. That students do not select into school on the basis of cross-year variation in their 8th grade math and ELA tracking systems, 2. that changes in school tracking systems are not associated with confounding changes in school organization. While we cannot directly test either of these assumptions, our qualitative data suggest that both are plausible. The vast majority of students in our sample attend their middle schools due to neighborhood zoning, rather than school choice. While students likely sort into schools and their feeder neighborhoods based on broad reputations, we believe it is unlikely that they sort into schools based on contemporaneous and difficult-to-observe changes in school tracking regimes. Our interviews with school and district leaders suggest that most schools draw upon teacher, student, and parent observations and input to continuously modify school tracking systems, rather than as a part of broader systematic school reform efforts. We do not observe, for example, systematic links between leadership change or curriculum adoption in a school and changes in school tracking systems.

Finally, to understand the extent to which tracking practices work to exacerbate achievement inequalities within schools, we add an interaction between students' 7th grade test scores and the school-mean centered version of the school dimension of tracking. Positive values on these interaction terms suggest that tracking practices magnify the association between 7th grade test scores and 10th grade test scores, as one would expect if tracking increases achievement inequality. For the purposes of simplicity, we interpret results in terms of the predicted associations between 7th and 10th grade achievement under different tracking regimes and report these interactions graphically.

Results

Measuring the dimensions of tracking

Based on a review of school course catalogues as well as conversations with educators at sample schools and districts, we categorize 8th grade math and ELA courses into three levels: advanced, college prep, and remedial. We refer to the middle track as "college prep" since it is designed to prepare students to complete the high school course sequence required for admission to the four-year colleges in the University of California and California State University systems. As Figure 1 indicates, schools tend to place relatively high-achieving students in advanced and honors courses, students at the middle of the test score distribution in college prep courses, and low-achieving students in remedial courses. However, we also find evidence of considerable skills-heterogeneity among students in each of these tracks.[11]

[11] We compared observed track placements to hypothetical course assignments based solely on prior content-relevant achievement, holding course enrollments constant. Observed placements correlated with these hypothetical assignments at approximately .7 in 8th grade mathematics and ELA.

FIGURE 1 HERE

In this paper, we move beyond the broad representation of tracking systems represented in Figure 1 and empirically measure each of the five dimensions of school tracking systems. Since we have access to testing and transcript data in mathematics and ELA for every student in our sample schools, we can identify the title and level of all courses that sample schools offer to 8th graders in these key academic areas. In addition, by identifying students who take the same class with the same teacher during the same school period, we can identify every peer in 8th graders' math and ELA classrooms. These data allow description of schools' tracking systems and students' places in these systems. Table 3 provides a descriptive overview of our measures of the five dimensions of tracking.

TABLE 3 HERE

We measure the **degree of curricular differentiation** as the number of different course titles schools make available to 8th graders in any given year. As Table 3 reveals, the schools in our sample offer an average of four mathematics classes during the study period. However, schools vary appreciably on this measure. We observe schools that offer as few as two distinct 8th grade mathematics courses (Algebra and Pre-Algebra) and schools that offer as many as seven (including a remedial General Mathematics Skills course, Pre-Algebra courses in English and Spanish, Algebra courses in English and Spanish, an Honors Algebra course, and a doubly-advanced Honors Geometry course.) While our sample schools offer slightly fewer ELA courses to 8th graders during the study period, we observe no less cross-school variation in 8th grade ELA course offerings.

We measure the degree of **classroom skills homogeneity** in schools' 8th grade math and English classes by using students' 8th grade classroom assignments to predict their 7th grade standardized test scores within each school and year for which we have data. The intraclass correlation (ICC) from this multi-level model captures the amount of between class variation that exists within a given school-by-year based on students' prior achievement. We interpret this ICC as the degree of skills homogeneity in 8th grade mathematics and English classrooms in a school in a given year on a zero-to-one scale. This measure has a mean of 0.52 in our sample schools and a standard deviation of 0.17 for mathematics and a mean of 0.50 and a standard deviation of 0.18 in ELA.

Conversations with school and district leaders reveal substantial variation in course assignment policies, both across schools and within schools over time. Educators articulate a wide range of opinions about tracking. One teacher tells us "I like it that the advanced kids are separated, because then those kids are not in a class where it is constantly a behavior thing happening… I think they've earned that." Another teacher, in the same school, disagrees, saying "I don't believe in the honors system. I believe that all kids should be deserving of a high quality curriculum." The profound variation that we observe across schools and within schools over time reflects this debate and the very different ways that schools and districts approached tracking over the study period. During the three years our data cover, District B encouraged schools to enroll students in 8th grade math and ELA exclusively on the basis of the prior test scores. While teachers report that they occasionally overruled the district's placement formulae, our analyses indicate that classroom assignments are relatively skills-homogeneous in District B over time. However, the test score and grade thresholds that District B used as benchmarks to guide course placements changed over the study period. In the study's first year, the district used fairly inclusive standards in an attempt to boost enrollments in accelerated and Honors classes. When many students had to repeat the courses they were placed into based on these inclusive standards, District B revised its placement guidelines. By contrast, Districts A and C gave schools relatively little guidance regarding course placements. In District A, schools typically used an informal approach to course assignments, allowing teachers, parents, and teachers to place students independently of their prior test scores. In this district, placement practices varied considerably across schools and over time, as school personnel changed positions in the school hierarchy and engaged in constant negotiations with parents and students over assignments. Finally, schools in District C experimented with an array of course assignment practices over time, ranging from explicitly skills-heterogeneous course assignments to rigid test-score based assignments.

Figure 2 provides an illustration of our measure of homogeneous classroom assignments, plotting the distribution of 7th grade mathematics test scores by 8th grade mathematics classroom for 8th graders in one District C school in 2010 and 2012. During this period, this school moved from an informal course placement system to a system that explicitly attempts to create skills-heterogeneous classrooms in middle-track mathematics. In the process, the schools' skills-homogeneity measure decreased from 0.51 to 0.24, a change equivalent to approximately 1.5 standard deviations in the sample-wide distribution. There is considerable overlap across classrooms in the distribution of student achievement in both years. However, in 2010, the bulk of students scored within half of a population-wide standard deviation of their classroom mean. The distribution of scores within classrooms is considerably broader in 2012, especially in the 9 middle-track mathematics classrooms where a large proportion of students score more than a full standard deviation higher or lower than their classroom mean.

FIGURE 2 HERE

Track exclusiveness refers to the extent to which schools assign students to high-track courses. We measure exclusiveness as the proportion of 8th graders enrolled in remedial or college prepcourses in our sample schools.[12] As Table 3 indicates, we observe a lower degree of track exclusiveness in mathematics in our sample schools than in ELA. This is likely largely due to California's use of its accountability policy to incentivize schools to boost 8th grade Algebra enrollments. While the state began to move away from this effort as it transitioned to the Common Core State Standards in both math and ELA, California schools continued to enroll students in 8th grade Algebra – a course we consider accelerated since it puts students on a track to complete Calculus by the end of 12th grade – at a considerably higher rate than their peers across the U.S. (Author, 2015). The state's Algebra-for-all effort limits the degree of variation in math track exclusiveness in our sample schools. However, we observe a large degree of both between-school variation as well as within-school temporal variation in ELA track exclusiveness, where the mean is 0.37 and the standard deviation is 0.27.

Track stability refers to the extent to which students' remain in the same track levels as they progress through school. Because differentiated courses often begin in 8th grade, we operationalize track stability over the ensuing transition to high school. This transition is partly influenced by high school policies, but it is largely a function of the organization of middle school courses. For example, the inclusivity of 8th grade groupings is closely related to track stability between 8th and 9th grade (see below). While our sample schools enrolled a large proportion of students in advanced courses in both math and ELA during their 8th grade year, these middle school placements do not ensure that students will remain on an advanced track through high school. Consistent with Rosenbaum's observations in "Grayton High" (1976), we

[12] Our data include text variables describing the titles of each of the mathematics and ELA courses students enrolled in during their 8th grade year. We sort these courses into three levels: Remedial, college prep, and advanced. In most cases, these categories appeared to be self-evident based on course titles. Courses labelled as "remedial" or "developmental" clearly belonged in the lowest level, while courses labelled as "advanced" or "honors" clearly belonged in the highest level. However, in all cases, we triangulated our course labels in conversation with district-level mathematics and ELA curricular specialists and as well as school-level department chairs.

find that virtually no students in our sample schools move from 8th grade remedial classes to 9th grade college prep classes or 8th grade college prep classes to 9th grade advanced classes. However, 41 percent of the students in our sample schools experienced downward mobility in mathematics between 8th and 9th grade and 34 percent experienced downward mobility in ELA.[13]

We use the proportion of a school's 8th graders who remained in the same track during their 9th grade year to measure track stability.[14] In interviews, teachers and district leaders report that they prefer to place students in relatively high-level middle school courses, so as not to foreclose students' opportunities to take advanced courses later in their educational careers. There is some evidence to suggest that state policy around 8th grade Algebra reinforced this tendency (see Author 2015 for more detail), leading schools to place students into accelerated 8th grade Algebra classes even as many students retook it as 9th graders. As a result, we observe a somewhat higher level of track stability in ELA than in mathematics. However, track stability varies appreciably within and between schools in both mathematics and ELA.

Finally, **Track Scope** refers to the relation between students' classroom assignments during one part of the school day and their assignments during the rest of the day. Following Lucas (1999), we measure scope as the correlation between a ranked measure of 8th grade mathematics course placements and a similar ranked measure of 8th grade ELA courses placements. (In these ranked measures, remedial or low-track courses are coded as 1, mid-track courses are coded as 2, and high-track, honors, or accelerated courses are coded as 3.) In schools that approach 1 on this measure, students who are assigned to high-track mathematics courses

[13] Many of educators we interviewed expressed frustration at the lack of upward track mobility for there students. Curricular planners in Districts A and C have dedicated particular attention to attempting to facilitate upward mobility by creating multiple "course acceleration" opportunities, including double-dose and summer courses. However, these efforts to create upward mobility paths were not in place in sample schools during the study period.
[14] Most downward mobility in mathematics occurred when students took Algebra in the 8th grade and retook it in the 9th grade. In ELA, a the most common example of downward mobility is from 8th grade Honors course to a 9th grade College Prep course.

are typically also assigned to high-track ELA courses. In schools that approach 0 on this measure, mathematics and ELA courses placements are largely unrelated. On average, this measure of scope is fairly high in our sample schools, and students' math course assignments correlate with their ELA course assignments at 0.67. This correlation corresponds closely with Lucas's (1999) findings regarding track scope in a nationally representative sample of U.S. high schools. Underlying this measure, however, we find considerable variation in track scope between schools as well as temporally within schools. In some schools, students' math track placements rarely diverge from their ELA course placements while in others it is not uncommon for students to enroll in advanced math and college preparatory ELA courses (or vice-versa.)

Testing a multi-dimensional conception of tracking

If "tracking" is a unified practice or set of practices in schools, one might expect schools that score high on one of the five dimensions of tracking to score high on the remaining four dimensions. However, as we note above, these dimensions are conceptually distinct. Further, prior researchers have noted that the school practices that underlie tracking systems are often developed and debated within schools in an ad hoc fashion (e.g. Oakes 1985). The correlation matrix reported in Table 4 investigates the extent to which the theoretically separable dimensions of school tracking systems are separable in practice among our 69 school/year observations. We observe close associations between our measures of track exclusiveness and track stability. Schools that enroll small proportions of students in advanced courses in 8th grade tend to have more students who remain in the same track location as they move into 9th grade. This correlation is particularly pronounced in ELA, at 0.95.

More generally, however, Table 4 indicates that the correlations among the dimensions of tracking are fairly modest. For example, while schools that sort students into relatively skills-homogeneous classes tend to have lower levels of enrollment in advanced math classes and higher levels of track stability, these associations are fairly small at -0.18 and 0.16 respectively. We observe positive associations between the degree of curricular differentiation in schools and the degree of within-classroom skills homogeneity, consistent with the idea that curricular differentiation facilitates the sorting of students into skills-homogeneous classrooms. In both mathematics and ELA, we find that as the number of courses schools offer increases so too does skills homogeneous classroom assignments. However, these associations are modest, at 0.36 and 0.44 respectively.

TABLE 4 HERE

The multilevel models reported in Table 5 indicate that associations between school characteristics and school tracking practices vary across the dimensions of tracking. In these models both the dependent variables and the independent variables are standardized, so that the coefficients can be interpreted as the expected increase in the dimensions of tracking (expressed in standard deviation terms) associated with a one standard-deviation increase in each of the independent variables, conditional on all other controls.

While we find evidence to suggest that the degree of mathematics curricular differentiation varies significantly across districts and over time, none of our measured school characteristics significantly predict the number of different mathematics courses offered by schools in our sample. Similarly, we find no significant association between school characteristics and skills homogeneity in mathematics classrooms. Indeed, the only relatively consistently significant school-level predictor of school mathematics tracking systems is schools' total enrollment. In particular, these analyses indicate that relatively large schools tend to have less exclusive enrollments in advanced 8th grade math courses. Further, students in large schools tend to experience relatively low rates of track stability between 8th and 9th grade. Since the relatively small school-level sample size limits the power in these analyses, several nonsignificant conditional associations between the concentration of socioeconomically disadvantaged students and mathematics tracking dimensions are worth noting. These non-significant associations indicate that schools that educate relatively large proportions of poor, minority, and EL students may tend to offer fewer mathematics courses and place students in relatively heterogeneous mathematics classes.

TABLE 5 HERE

The pattern of school-level predictors of the dimensions of tracking in ELA is somewhat different. We find that schools with relatively disadvantaged student populations tend to offer significantly *more* 8th grade ELA courses than more advantaged schools, net of controls. The concentration of socioeconomically disadvantaged students is also significant predictor of track exclusiveness and track stability in ELA. Consistent with mathematics, we find that school mean prior achievement relates negatively with curricular differentiation and classroom skills homogeneity, while correlating positively with track stability in ELA. Finally higher school enrollment is negatively associated with track stability in ELA.

The dimensions of tracking and student achievement

In light of the above evidence suggesting that the dimensions of tracking are empirically separable, the remaining analyses examine the links between these dimensions and student achievement. Table 6 reports the results of a series of multilevel models regressing students' 10th grade math achievement on the dimensions of mathematics tracking systems in students' 8th grade middle schools; Table 7 reports the results of parallel analyses in ELA. All dependent and independent variables are standardized in each of the models reported in both tables, such that each has a mean of zero and a standard deviation of one in the student population under analysis.

The first model in Table 6 provides a fully unconditional look at these relationships, while the second model investigates the relationship between the dimensions and math achievement controlling for the other dimensions. The third model adds student-level demographic and prior achievement controls as well as indicator variables that account for commonalities among students enrolled in the same school district (district fixed effects) and students in the same grade cohorts (cohort fixed effects). Finally, in the fourth model, we mean-center the time-varying school-level measures of the dimensions of tracking around schools' 3-year mean scores on these measures. Doing so controls for time-invariant school characteristics that may confound the link between the dimensions of school tracking regimes and student achievement.[15]

TABLE 6 AROUND HERE

The first model of Table 6 indicates that the unconditional relationship between the dimensions of tracking and student math achievement varies across dimensions. Mean 10th grade math achievement is slightly higher for students from schools that offer many math 8th grade courses than students from schools that offer few. Similarly we find that math achievement is slightly higher for students from schools that maintain relatively exclusive access to high track mathematics courses. Neither of these associations are significant after controlling for the other dimensions of tracking. Further, they remain nonsignificant with the addition of demographic controls and school mean centering. The models reported in Table 6 further suggest that there is

[15] This is a correlated random-effects model. All models include school and classroom level random effects terms to adjust standard error estimates for the clustering of students in schools and classrooms.

no significant association between 8^{th} grade mathematics track stability and 8^{th} grade track scope and student 10^{th} grade mathematics achievement.

By contrast, we find that students in schools that have relatively skills-homogeneous 8^{th} grade math classroom assignment practices score significantly less well on 10^{th} grade courses than their peers in schools where math courses are less rigidly grouped by student achievement. That association holds after controlling for other dimensions of tracking in Model 2 and remains statistically significant in the most stringent model with school-mean centering(Model 4). Accordingly, we interpret the analyses presented in Table 6 as suggesting that that homogeneous math course assignment practices may have small negative average effects on students' mathematics achievement.

Model 1 in Table 6 suggests that there is a weak positive association between 8^{th} grade school math track exclusiveness and students' 10^{th} grade mathematics achievement. This relationship is nonsignificant when we control for the other dimensions of tracking and demographic controls Models 2 and 3. After controlling for time-invariant school characteristics in Model 4, we find evidence to suggest that attending middle schools with highly exclusive 8^{th} grade math tracking systems significantly *increases* student achievement by the 10^{th} grade. This estimate suggests that, all else equal, a student who moves from a school that enrolls 27 percent of students in advanced math classes to a school that enrolls 15 percent in advanced math classes will experience a test score increase of approximately $1/10^{th}$ of a standard deviation. Supplementary analyses, available by request, confirm that these results hold when investigating changes in track exclusiveness without controlling for other dimensions of school tracking systems. While somewhat counter-intuitive, these findings are consistent with evidence elsewhere in the research literature suggesting that efforts to intensify middle school mathematics curricula may have unintended negative consequences for students' achievement (Clotfelter, Ladd and Vigdor 2014; Author 2015).

Table 7 reports parallel models exploring the relation between the dimensions of school ELA tracking systems and students' 10th grade ELA achievement. In general the results reported here indicate that ELA test scores are less sensitive to the dimensions of tracking than math scores. We find that 8th grade ELA exclusiveness and stability are both associated with 10th grade ELA scores. Model 3 indicates that students in schools with more exclusive ELA track placement systems demonstrate significantly lower 10th grade ELA scores than demographically similar students in schools that allow more access to Honors and other high-track English classes. At the same time, we find that students in schools with a high degree of track stability between 8th and 9th grade score significantly higher than students in contexts in which students face greater prospects of downward track mobility. However, Model 4 indicates that neither of these associations are significant after controlling for student characteristics and time-invariant school characteristics, suggesting that these relationships are largely a function of unmeasured school effects.

TABLE 7 AROUND HERE

Taken together, the results reported in Tables 6 and 7 suggest that the constellation of practices researchers often refer to as "tracking" have mixed and modest average effects on student achievement. We find that placing students into skills-homogeneous 8th grade mathematics classrooms has a small negative effect on students' 10th grade mathematics achievement. However, our findings regarding the average effects of track exclusiveness suggest that efforts to detrack mathematics instruction by enrolling all students in accelerated courses may have unintended negative consequences. Meanwhile, we find no evidence to suggest that any of the dimensions of 8th grade ELA tracking systems influence student achievement on the 10th grade ELA California high school exit exam.

Sensitivity analyses

One possible explanation for the weak link between school tracking practices and mean student achievement is that the dependent variable in these analyses – 10th grade exit exam scores – is measured two years after students experience their 8th grade learning environments.[16] While the structure of state testing in California schools precludes intermediary analyses in mathematics, we can analyze the relation between the dimensions of school tracking systems in 8th grade and achievement on standardized tests in ELA administered in the spring of students' 8th grade year. The results of these analyses, reported in Appendix Table 1, provides limited evidence of a short-term effect for the dimensions of tracking and in particular, our school-mean centered analysis reported in Model 4 suggests that attending a school in which students are assigned to skills-homogeneous 8th grade ELA classrooms may slightly but significantly decrease 8th grade achievement. However, consistent with the 10th grade ELA findings reported in Table 6, we find no robust evidence of a link between the other dimensions of tracking and students' 8th grade ELA scores.[17]

A related empirical concern is that students' disparate high school experiences could attenuate the observed relation between middle school tracking systems and 10th grade achievement. Each of the middle schools for which we have data enroll 7th and 8th graders

---

[16] However, as noted above, the CAHSEE, especially in math, is a function of 6 to 8th grade standards. Therefore the relationship between 8th grade tracking practices and 10th grade achievement is especially relevant since these practices occurred at the same time students who largely exposed to the material tested on the CAHSEE.

[17] Our operationalization of stability (organizational patterns in the transition to high school) requires special attention for the interpretation of this coefficient in models with 8th grade achievement as an outcome. From the perspective of 8th grade experiences, this stability reflects the structural prospects afforded by curricular differentiation practices in middle school. Our results in initial models suggest that arrangements with greater prospects of (downward) mobility are negative, but that this association is does not hold when accounting for time-invariant school characteristics.

exclusively. All but one of these schools explicitly serves as a feeder for a single high school. (The exception to this rule is a middle school that feeds into two neighboring high schools.) In most cases our focal middle schools occupy a separate building on a shared or a closely neighboring space with their linked high school. As a result, more than 80 percent of the students enroll in the same high school as the bulk of their middle school classmates. Appendix Table 2 reports a series of supplementary analyses in which we estimate the link between the dimensions of 8th grade tracking and 10th grade achievement separately for students who follow their middle school's modal high school and for students who move out of that modal middle-to-high school path. The results of these analyses are strikingly similar to one another and to the results we report in Tables 6 and 7.

Do high- and low-achieving students experience tracking differently?

Since the analyses reported in the first four models in Tables 6 and 7 focus on the mean effects of school-level tracking systems, they neglect crucial questions regarding to the relationship between of tracking systems and achievement inequality. Model 5 in Tables 6 and 7 address the equity effects of tracking by taking a closer look at one key dimension of school tracking systems – the degree to which schools group students into classrooms based on their prior test scores. In these models, we investigate the extent to which the effects of school-level ability grouping vary with students' 7th grade test scores. To ease interpretation, we also report these interactions in Figures 3 and 4.

FIGURES 3 AND 4 AROUND HERE

The y-axis in Figure 3 represents students' z-scored predicted 10th grade mathematics achievement scores, while the x-axis represents students' z-scored 7th grade mathematics scores. The dashed line represents the predicted relation between 7th grade achievement and 10th grade achievement in mathematics for students in schools that have implemented ability grouping to an above-average degree in 8th grade mathematics classrooms. The solid line, meanwhile, represents that same relation in schools that have implemented a below-average degree of ability grouping in 8th grade mathematics classrooms. The shaded areas around both lines represent 95% confidence intervals. Consistent with the results indicating a negative average effect of classroom skills homogeneity in 8th grade mathematics reported in Table 6, the dashed line is lower than the solid line across the 7th grade math test score distribution in Figure 3. Notably, however, the disadvantage associated with attending a school in which students attend largely skills-homogeneous 8th grade mathematics courses is particularly pronounced for students at the bottom of the 7th grade mathematics test score distribution. Put differently, this figure suggests that low-achieving students disproportionately bear the achievement costs associated with ability grouping in middle school mathematics. Model 5 in Table 6 indicates that this interaction term is statistically significant, if small in magnitude. All else equal, this model suggests that enrolling in a school with a high degree of ability grouping will increase the gap between students who come into the 8th grade 1 standard deviation above and below the math test score average by approximately 0.08 standard deviations. The models reported in Figure 4 and model 5 in Table 7 indicate that these inequitable effects are more pronounced in ELA.[18]

Discussion

This study measures multiple dimensions of tracking and identifies their effects on student achievement. Building on the work of Sørenson (1970) and others (Gamoran 1992;

[18] Since these analyses are only available for students who have remained in school through the 10th grade, they understate the degree to which students' prior achievement moderates the effects of school tracking regimes. Approximately 20 percent of the students for whom we have tracking measures in the 8th grade leave our sample before taking the 10th grade exit exam. While our data do not allow us to differentiate between students who move to other districts or private schools and those who drop out of school, we suspect that high school dropout is an important source of attrition. Consistent with that explanation, students in the bottom quartile on 7th grade achievement are approximately twice as likely to leave the sample than their peers in the top quartile.

Lucas 1999; Kelly 2007; Becker 1987), we identify five theoretically distinct dimensions of school math and ELA tracking systems: (1) the degree of curricular differentiation, (2) the degree of skills homogeneity in classrooms, (3) track exclusiveness, (4) track stability, and (5) track scope. We take advantage of a unique set of student-level administrative data gathered from three public school districts, as well as qualitative data gathered from administrators and educators in these three districts, to measure the dimensions of tracking systems in 23 schools and the ways in which these tracking systems changed over the course of three years.

In these 23 schools at least, "tracking" is a multidimensional phenomenon. We observe considerable variation on each dimension both between our sample of 23 middle schools and within these schools over time. In addition, we find that the dimensions of school tracking systems do not correlate highly with one another. Further, consistent with a multidimensional conception of school tracking practices, our analyses indicate that the predictors of school tracking systems vary across the dimensions of tracking. These findings resonates with a long tradition of case study research on school tracking systems (Dreeban and Barr 1988; Metz 2003; Oakes 1985; Rickles 2011; Rosenbaum 1999; Watanabe 2006; Wells and Oakes 1996), while they also suggest that many recent quantitative studies considering individual track locations likely obscure important organizational variation in track practices and their consequences.

Our investigation of the dimensions of tracking reveals new insights into the ways in which school tracking systems influence student achievement, at least in the 23 schools for which we have data. In the area of ELA, our findings are largely consistent with earlier sociological research on the effects of school tracking. While we find little evidence to suggest that the five dimensions of school tracking systems have an effect on student achievement, this null effect conceals statistically significant if substantively modest inequality-producing consequences of school tracking systems. In particular, we find that when schools group students into ELA classes based on their prior achievement, high-achieving students tend to experience more rapid test score growth in ELA than low-achieving students (e.g. Argys, Rees, and Brewer 1996; Gamoran and Mare 1989).

Furthermore, the dimensions of tracking have potentially cross-cutting effects on students' mathematics achievement growth. We find evidence to suggest that students experience lower levels of mathematical achievement growth in schools that place students into skills-homogeneous mathematics classrooms, and that this negative effect is particularly pronounced for low-achieving students. This finding suggests that students may benefit from placement in relatively skills-heterogeneous classrooms for secondary mathematics instruction. Based on this finding, one might be tempted to endorse recent policy efforts that attempt to expose all students to high-quality instruction and high-achieving peers by universalizing accelerated course placements. However, our analyses also reveal positive effects of math track exclusiveness and stability on student mathematics achievement. Consistent with several recent policy analyses (Allensworth et al. 2009; Clotfelter, Ladd, and Vigdor 2015; Author 2015; Stein et al. 2011), these findings suggest that efforts to detrack instruction by enrolling more students in accelerated courses can have negative effects. From a practitioner's perspective, therefore, our findings point to a tension between the benefits of skills-heterogeneous learning environments and the shortcomings of instruction that is insensitive to student skills. Curricular reform efforts that simultaneously provide disadvantaged students with access to higher achieving peers *and* sufficient skill-building opportunities provide one promising strategy for resolving this tension (Nomi and Allensworth 2012; Nomi and Raudenbush 2016).

In interpreting these findings, it is important to note that our data can only speak directly to the experiences of approximately 20,000 students enrolled in 23 middle schools in three Southern California public school districts during the 2010-11 to 2012-13 school years. While we have no reason to suspect that these middle schools are idiosyncratic, we make no strong claims regarding our findings' generalizability. Future research should investigate the dimensions of tracking in other settings. Doing so is data intensive, since many of our measures of the dimensions of school tracking systems require at least two years of longitudinal data on all students enrolled in a school as well as indicators for students' classroom placements and the title or content of the classes in which students enroll. However, the growing availability of student-level administrative data in American public K-12 schooling, combined with new approaches to machine learning and text-based data analysis, provide new opportunities for investigating the dimensions of tracking on a larger scale. In particular, we believe that future research should explore and refine track stability in secondary school. While our analyses indicate that track stability matters, we provide an admittedly limited view of stability since we follow students from middle school to high school, but do not observe students after the 10th grade. In addition to attempting to replicate this study's findings about the dimensions of tracking and their links to student achievement, larger scale research should investigate the link between school resources and organizational characteristics and the development of school tracking systems.

Furthermore, future researchers should work to articulate and study other dimensions of school tracking systems. We note that our measures capture just a few of the important ways in which tracking systems may vary. For example, our analyses do not consider the extent to which student ascriptive characteristics such as race, class, and gender align with course placements and the consequences of such alignment for identity formation and ascriptive inequalities in the school (Frank, Muller, Schiller, Riegle-Crumb, Mueller, Crosnoe, and Pearson 2008; Legewie and DiPrete 2012; Lewis and Diamond 2015; Tyson 2011). Our data also provide limited information on the extent to which schools allow students to choose their own track locations; a dimension of tracking that is sometimes referred to as track electivity (Lucas 1999; Powell et al. 1985; Sørensen 1970). Relatedly, due to the age of our sample, we are not able to separate horizontal from vertical differentiation as the former does not occur widely in middle schools. One may be able to test these in other k-12 contexts like high school science. Similarly, our analyses provide no information about the sorts of between school tracking that occur in contexts in which school admissions are selective or school enrollments are otherwise unequal (c.f. Hanushek and Woessman 2006; Van de Werforst and Mijs 2010).

Despite these limitations, this paper suggests a set of strategies for incorporating a key insight that emerges from the long history of qualitative and case-study research on school tracking systems into the quantitative study of educational inequality. Much quantitative research on tracking – and indeed, much research in the sociology of education and inequality – takes an individualistic approach, focusing on the consequences of students' track locations. By contrast, our analyses challenge sociologists and educators to consider a broad array of school organizational practices and the ways in which these practices cohere into a wide variety of educational tracking systems. Careful study of this organizational variation can help to elucidate and evaluate the complex and interacting mechanisms through which schools produce, reproduce, and even ameliorate social inequality. Ultimately, such an approach may point to promising strategies for building more effective and equitable organizations.

References

Allensworth, Elaine, Takako Nomi, Nicholas Montgomery, and Valerie E. Lee. 2009. "College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago" *Educational Evaluation and Policy Analysis 31*(4): 367-391.

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.

Becker, Henry Jay. 1987. "Addressing the Needs of Different Groups of Early Adolescents: Effects of Varying School and Classroom Organizational Practices on Students from Different Social Backgrounds and Abilities. Report No. 16." Baltimore: Johns Hopkins University Center for Research on Elementary and Middle Schools.

Betts, Julian R., and Jamie L. Shkolnik. 2000. "The effects of ability grouping on student achievement and resource allocation in secondary schools." *Economics of Education Review 19*(1): 1-15.

Buchmann, Claudia, and Hyunjoon Park. 2009. "Stratification and the formation of expectations in highly differentiated educational systems." *Research in Social Stratification and Mobility* 27(4): 245-267.

Carbonaro, William. 2005. "Tracking, students' effort, and academic achievement." *Sociology of Education* 78(1): 27-49.

Carbonaro, William J., and Adam Gamoran. 2002. "The production of achievement inequality in high school English." *American Educational Research Journal* 39(4): 801-827.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2015. "The Aftermath of Accelerating Algebra Evidence from District Policy Initiatives." *Journal of Human Resources 50*(1): 159-188.

Cremin, Lawrence A. 1965. *The genius of American education*. Pittsburgh: University of Pittsburgh Press.

Cronbach, L. J., & Snow, R. E. 1977. *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Downey, Douglas B., Paul T. Von Hippel, and Beckett A. Broh. 2004. "Are schools the great equalizer? Cognitive inequality during the summer months and the school year." *American Sociological Review* 69(5): 613-635.

Dougherty, Shaun M., Joshua S. Goodman, Darryl V. Hill, Erica G. Litke, and Lindsay C. Page. 2015. "Middle school math acceleration and equitable access to eighth-grade algebra evidence from the wake county public school system." *Educational Evaluation and Policy Analysis* 37(1 suppl): 80S-101S.

Dreeben, Robert and Rebecca Barr. 1998. "Classroom Composition and the Design of Instruction." *Sociology of Education* 61(3): 129-42.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. *Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya* (No. w14475). National Bureau of Economic Research.

Eccles, Jacquelynne and Robert W. Roeser. 2011. "Schools as developmental contexts during adolescence." *Journal of research on adolescence* 21(1): 225-241.

Epple, Dennis, Elizabeth Newlon, and Richard Romano. 2002. "Ability tracking, school competition, and the distribution of educational benefits." *Journal of Public Economics* 83 (January): 1–48.

Figlio, David N., and Marianne E. Page. 2002. "School choice and the distributional effects of ability tracking: does separation increase inequality?." *Journal of Urban Economics* 51(3): 497-514.

Frank, Kenneth A., Chandra Muller, Kathryn S. Schiller, Catherine Riegle-Crumb, Anna Strassmann Mueller, Robert Crosnoe, and Jennifer Pearson. 2008. "The social dynamics of mathematics coursetaking in high school." *American Journal of Sociology* 113(6): 1645-1696.

Gamoran, Adam. 1992. "The variable effects of high school tracking." *American Sociological Review*: 812-828.

Gamoran, Adam. 2010. "Tracking and inequality." *The Routledge international handbook of the sociology of education*. (Michael W. Apple, Stephen J. Ball, and Luis Armando Gandin, eds.) New York: Routledge.

Gamoran, Adam, and Martin Nystrand. 1994. "Tracking, instruction and achievement." *International Journal of Educational Research* 21(2): 217-231.

Gamoran, Adam, Martin Nystrand, Mark Berends, and Paul C. LePore. 1995. "An organizational analysis of the effects of ability grouping." *American Educational Research Journal* 32(4): 687-715.

Gamoran, Adam, and Robert D. Mare. 1989. "Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?." *American journal of Sociology* 1146-1183.

Hallinan, Maureen T. 1992. The organization of students for instruction in the middle school. *Sociology of Education*, 114-127.

Hallinan, Maureen T. 1994. School differences in tracking effects on achievement. *Social Forces 72*(3), 799-820.

Hallinan, Maureen T. 1994. “Tracking: From theory to practice.” *Sociology of Education* 67(2): 79-84.

Hallinan, Maureen T. 1996. Track mobility in secondary school. *Social Forces* 74(3): 983-1002.

Hallinan, Maureen T., and Aage B. Sørensen. (1985). "Ability grouping and student friendships." *American Educational Research Journal* 22 (4): 485-499.

Hanushek, Eric A., and Ludger Woessmann. 2006 "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries." *Economic Journal* 116(510): C63-C76.

Hoffer, Tomas B. 1992. “Middle School Ability Grouping and Student Achievement in Science and Mathematics. *“Educational Evaluation and Policy Analysis* 14(3): 205-27

Horvath, Hedvig. 2015. “Classroom Assignment Policies and Implications for Teacher Value-Added Estimation” Unpublished manuscript.

Katz, Michael B. (1971). Class, bureaucracy and schools: The American illusion of educational change. *New York: Praeger. Kenway J (1996). The Information Superhighway and Post-Modernity: The Social Promise and the Social Price, Comp. Educ, 32*(2), 217-231.

Kelly, Sean. 2007. “The contours of tracking in North Carolina.” *The High School Journal 90*(4): 15-31.

Kelly, Sean. 2009. “The black-white gap in mathematics course taking.” *Sociology of Education 82*(1): 47-69.

Kelly, Sean, and Heather Price. 2011. "The Correlates of Tracking Policy Opportunity Hoarding, Status Competition, or a Technical-Functional Explanation?." *American Educational Research Journal* 48(3):560-585.

Kelly, Sean, and William Carbonaro. 2012. “Curriculum tracking and teacher expectations: Evidence from discrepant course taking models.” *Social Psychology of Education 15*(3): 271-294.

Kerckhoff, Alan C. 1986. "Effects of ability grouping in British secondary schools." *American Sociological Review* :842-858.

Kerckhoff, Alan C. 1995. "Institutional arrangements and stratification processes in industrial societies." *Annual review of sociology* 21: 323-347.

Kliebard, Herbert M. 1975. The rise of scientific curriculum making and its aftermath. *Curriculum Theory Network 5*(1), 27-38.

Legewie, Joscha, and Thomas A. DiPrete. 2012. "School context and the gender gap in educational achievement." *American Sociological Review 77*(3): 463-485.

Lewis, Amanda E., and John B. Diamond. 2015. *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.

Long, Mark C., Dylan Conger, and Patrice Iatarola. 2012. "Effects of high school course-taking on secondary and postsecondary success." *American Educational Research Journal* 49(2): 285-322.

Loveless, Tom. (2011). *The tracking wars: State reform meets school policy*. Washington, DC: Brookings Institution Press.

Lucas, Samuel R. 1999. *Tracking Inequality: Stratification and Mobility in American High Schools. Sociology of Education Series*. New York: Teachers College Press.

Lucas, Samuel R., and Mark Berends. 2002. "Sociodemographic diversity, correlated achievement, and de facto tracking." *Sociology of Education* 328-348.

Lucas, Samuel. R., Aaron D. Good. 2001. "Race, class, and tournament track mobility." *Sociology of Education* 139-156.

Mann, Horace. (1848). Twelfth annual report to the Massachusetts Board of Education. *The republic and the school: Horace Mann and the education of free men*.

Marsh, Herbert W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of educational psychology 79*(3), 280.

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

Metz, Mary Haywood.1978. *Classrooms and Corridors: The Crisis of Authority in Desegregated Secondary Schools*. Berkeley: University of California Press.

Metz, Mary Haywood. (2003). *Different by design: The context and character of three magnet schools*. Teachers College Press.

Meyer, John W., Brain Rowan, and M.W. Meyer. 1978. "The Structure of Educational Organization."

Mickelson, Roslyn Arlin. 2001. "Subverting Swann: First-and second-generation segregation in the Charlotte-Mecklenburg schools." *American Educational Research Journal 38*(2): 215-252.

Nomi, Takako, and Elaine M. Allensworth 2012. Sorting and Supporting Why Double-Dose Algebra Led to Better Test Scores but More Course Failures. *American Educational Research Journal* 50(4): 756-788.

Nomi, Takako, Stephen W. Raudenbush. (2016). Making a Success of "Algebra for All" The Impact of Extended Instructional Time and Classroom Peer Skill in Chicago. *Educational Evaluation and Policy Analysis* 38(2), 431-451.

Nystrand, Martin and Adam Gamoran. 1997. "The big picture: Language and learning in hundreds of English lessons." *Opening Dialogue*. New York: Teachers College, Columbia University 30-74.

Oakes, Jeannie. 1985. *Keeping track: How Schools Structure Inequalities*. Yale University Press.

Oakes, Jeannie, and Gretchen Guiton. 1995. "Matchmaking: The dynamics of high school tracking decisions." *American Educational Research Journal 32*(1): 3-33.

Oakes, Jeannie, Amy Stuart Wells, Makeba Jones, and Amanda Datnow. . (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform.Teachers College Record 98; 482–510.

Page, Reba Neukom. 1991. *Lower-Track Classrooms: A Curricular and Cultural Perspective*. New York: Teachers College Press.

Powell, Arthur G., Eleanor Farrar, and David K. Cohen. (1985). The shopping mall high school. Boston: Houghton Mifflin.

Raudenbush, Stephen, and Anthony S. Bryk. 1986. "A hierarchical model for studying school effects." *Sociology of education*: 1-17.

Raudenbush, Stephen W., and Robert D. Eschmann. 2015. "Does Schooling Increase or Reduce Social Inequality?." *Annual Review of Sociology* 41: 443-470.

Rickles, Jordan H. 2011. "Using interviews to understand the assignment mechanism in a nonexperimental study: The case of eighth grade Algebra." *Evaluation review* 35(5); 490-522.

Rosenbaum, James E. 1976. *Making Inequality; the Hidden Curriculum of High School Tracking*. New York: John Wiley & Sons.

Rosenbaum, James E. 1999. "If Tracking Is Bad, Is Detracking Better?" *American Educator 23*(4): 24-29.

Shavit, Yossi, and Hans-Peter Blossfeld. 1993. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries. Social Inequality Series*. Westview Press, 5500 Central Avenue, Boulder, CO 80301-2847

Slavin, Robert E. 1988. "Synthesis of research on grouping in elementary and secondary schools." *Educational Leadership* 46(1): 67-77.

Snow, Richard. 1989. Aptitude-treatment interaction as a framework for research on individual differences in learning. In *Learning and individual differences*, eds. P. Ackerman, R.J. Sternberg, and R. Glaser, 13–59. New York: W.H. Freeman.

Sørensen, Aage Bøttger. 1970. Organizational differentiation of students and educational opportunity. *Sociology of Education*: 355-376.

Stainback, Kevin, Donald Tomaskovic-Devey, and Sheryl Skaggs. 2010. "Organizational approaches to inequality: Inertia, relative power, and environments." *Annual Review of Sociology 36*(1): 225-247.

Stein, Mary Kay, Julia Heath Kaufman, Milan Sherman, and Amy F. Hillen. 2011. "Algebra: A challenge at the crossroads of policy and practice." *Review of Educational Research* 81: 453–492.

Tyson, Karolyn. (2011). *Integration interrupted: Tracking, Black students, and acting White after Brown*. Oxford University Press.

Van de Werfhorst, Herman G., and Jonathan JB Mijs. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual review of sociology 36*: 407-428.

Van Houtte, Mieke. 2004. "Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff." *American Journal of Education* 110(4): 354-388.

von Hippel, Paul T., Joseph Workman, and Douglas B. Downey. (2018). Inequality in Reading and Math Skills Forms Mainly before Kindergarten: A Replication, and Partial Correction, of "Are Schools the Great Equalizer?". *Sociology of Education* 91(4): 323-357.

Watanabe, Maike. (2006). " Some People Think This School Is Tracked and Some People Don't": Using Inquiry Groups to Unpack Teachers' Perspectives on Detracking. *Theory Into Practice 45*(1): 24-31.

Wells, Amy Stuart, and Jeannie Oakes. 1996. "Potential pitfalls of systemic reform: Early lessons from research on detracking." *Sociology of Education*: 135-143.

Zimmer, Ron. 2003. A new twist in the educational tracking debate. *Economics of Education Review 22*(3): 307-315.

Tables and Figures

Table 1: Five dimensions of tracking and strategies to measure them using educational administrative data

Dimension	Definition	School-level Measure
Degree of curricular differentiation	The number of distinct curricular positions in the organization	# of different courses available to 8th graders (math and ELA)
Classroom skills homogeneity (ability grouping)	Degree to which organizations assign individuals to different settings based on salient observed characteristics	7th grade test score homogeneity within 8th grade classrooms (math and ELA)
Track exclusiveness	Extent to which access to high-status positions is restricted	% of students in low-status courses (math and ELA)
Track stability	Extent to which organizational positions persist over time	% of students in same track level in 8th and 9th grade (math and ELA)
Track scope	Extent to which organizational position in one domain predicts position in other domains	Correlation between math and ELA track location

Table 2: Descriptive statistics, 8^{th} grade students in 3 Southern California public school districts, 2009-10—2011-12 school years

	District A	District B	District C
District administrative information			
Total 8^{th} grade student enrollment, 2010-2012	12,212	7,913	3,714
N traditional schools enrolling 8^{th} graders	9	10	4
N 8^{th} grade mathematics classrooms[19]	116	103	41
N 8^{th} grade ELA classrooms	165	80	35
Student demographics (averaged over available cohorts)			
% Female	50.6	50.7	47.1
Socioeconomic disadvantage scale (z-scored)	0.07	0.03	-0.40
% African American	2.5	0.5	0.9
% Asian	18.1	37.0	6.5
% Hispanic or Latino	67.1	51.4	44.9
% White	12.3	11.1	47.6
% Free- or Reduced-Price Lunch	70.7	69.9	50.4
% English Language Learners	20.3	28.0	16.9
% Reclassified Fluent English Speakers	43.8	47.1	22.1
% Special Education	6.2	2.4	12.5
7^{th} grade ELA CST	-0.13 (0.97)	0.18 (0.93)	0.06 (1.15)
7^{th} grade Mathematics CST	-0.16 (0.98)	0.22 (0.91)	0.08 (1.13)

[19] Classroom counts average over the 3 study years.

Figure 1: Distribution of standardized 7th grade achievement scores by eighth grade course track, math and ELA in 3 California school districts 2010-2012.

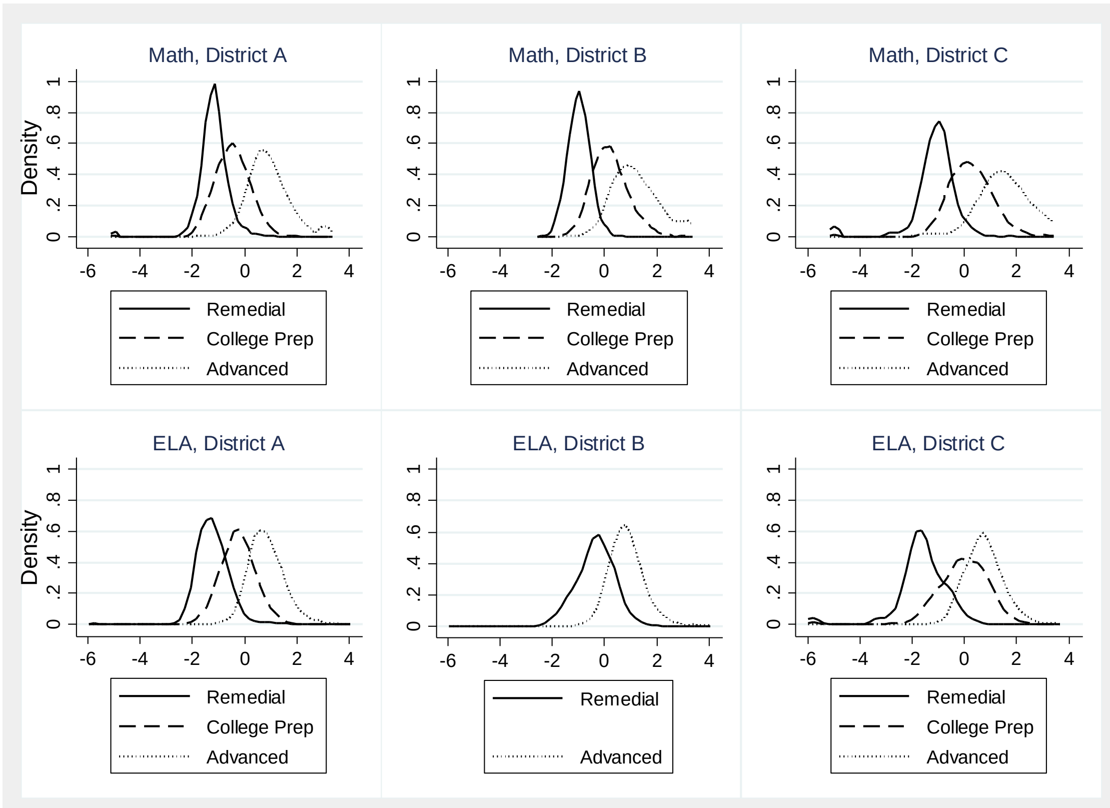

Table 3: Descriptive statistics for measures of dimensions of organizational differentiation in 3 Southern California public school districts, 2009-10—2011-12 school years

		Mean	SD	% variance between schools	% variance within schools (over time)
Degree of curricular differentiation					
# distinct courses					
	Math	4.06	1.06	67.7	32.3
	ELA	3.26	1.31	56.6	43.4
Classroom Skills homogeneity					
8^{th} grade classroom ICC, 7^{th} grade scores					
	Math	0.52	0.17	77.3	22.7
	ELA	0.50	0.18	52.0	48.0
Exclusiveness					
% in lower-track courses					
	Math	0.15	0.12	60.2	39.8
	ELA	0.37	0.27	49.8	50.2
Stability					
% in same 8^{th} and 9^{th} grade track					
	Math	0.59	0.16	51.2	48.8
	ELA	0.66	0.27	49.2	50.8
Scope					
Correlation: Math to ELA track		0.67	0.16	53.3	46.7

Figure 2: Illustrating classroom skills homogeneity: Distribution of 7th grade math CST scores by 8th grade classrooms in 1 school, 2010 and 2012

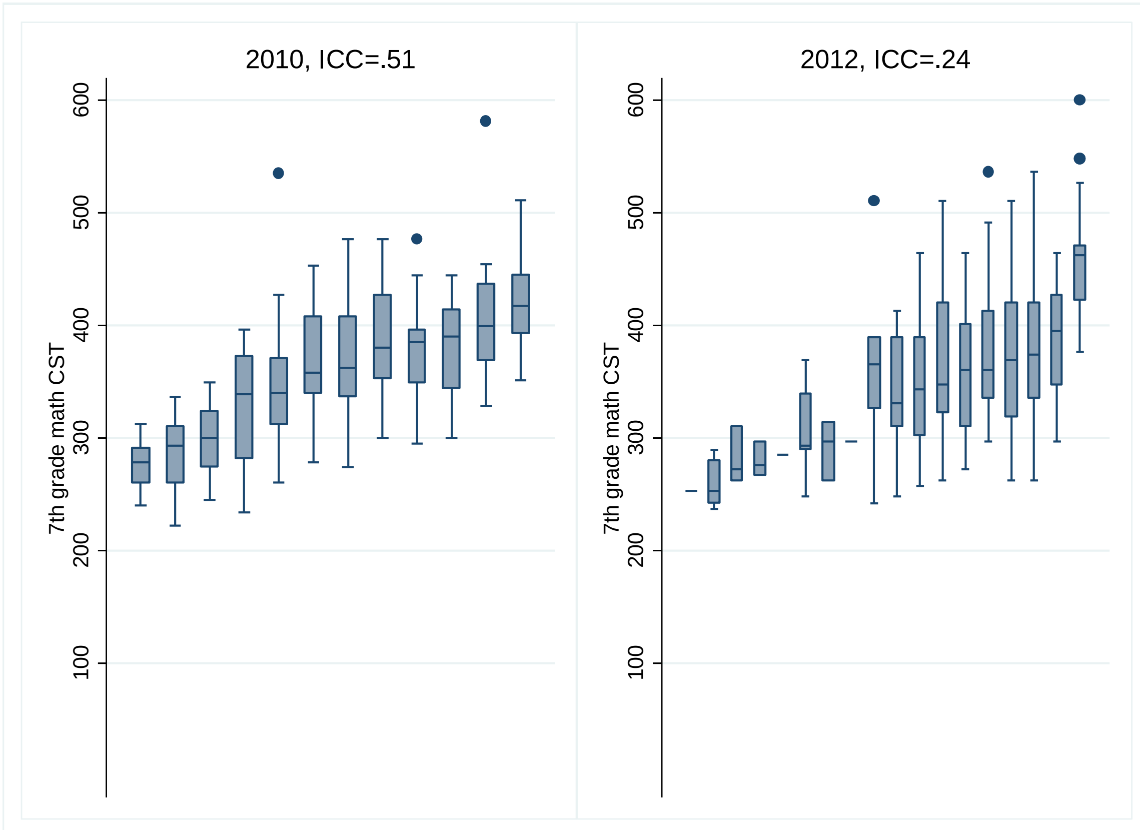

Table 4: Correlation of school-level measures of dimensions of organizational differentiation in 3 Southern California public school districts, 2009-10—2011-12 school years

	# courses (Math)	# courses (ELA)	Skills homogeneity (Math)	Skills homogeneity (ELA)	% in lower tracks (Math)	% in lower tracks (ELA)	% in same track 8-9 (Math)	% in same track 8-9 (ELA)	Correlation: Math to ELA
Differentiation									
# courses (Math)	1.00								
# courses (ELA)	0.00	1.00							
Skills homogeneity									
Skills homogeneity (Math)	0.36	0.08	1.00						
Skills homogeneity (ELA)	0.20	0.44	0.30	1.00					
Exclusivity									
% in lower tracks (Math)	0.42	-0.30	0.18	0.23	1.00				
% in lower tracks (ELA)	0.07	-0.16	-0.39	0.09	0.23	1.00			
Stability									
% in same track 8-9 (Math)	0.45	-0.55	0.17	-0.02	0.71	0.35	1.00		
% in same track 8-9 (ELA)	0.04	-0.31	-0.40	0.01	0.25	0.95	0.46	1.00	
Scope									
Correlation: Math to ELA	0.03	0.05	-0.12	-0.20	-0.56	0.13	-0.16	0.17	1.00

Table 5: Multilevel models of school-level predictors of dimensions of organizational differentiation measures for all District A, B, and C middle schools 2010-2012 (School-year level data, with school-level random effects. Outcomes as well as % Female, Disadv, x CST and, Enrollment are z-scored)

	# courses (Math)	Skills homogeneity (Math)	% in lower tracks (Math)	% in same track 8-9 (Math)	# courses (ELA)	Skills homogeneity (ELA)	% in lower tracks (ELA)	% in same track 8-9 (ELA)	Correlation: Math to ELA
% Female	-0.09	-0.02	-0.1	-0.03	-0.05	-0.05	-0.01	0.02	0.00
Disadv	-0.26	-0.31	0.25	0.10	0.32*	0.14	0.22*	0.42***	-0.63*
$\bar{\bar{x}}$ CST	-0.33	-0.48	-0.27	0.67***	-0.38*	-0.61**	-0.08	0.50***	-0.23
Enrollment	0.34	0.26	-0.49**	-0.26*	0.18	0.06	0.00	-0.16**	0.69***
SD CST (Math)	1.04	1.52	1.91	0.86	1.67	1.37	-0.49	-0.40	-0.38
SD CST (ELA)	0.3	1.05	0.78	0.71	-1.29	-0.99	0.01	0.37	0.73
2011	0.34	-0.03	0.26	0.11	-0.16	0.12	-0.06	-0.05	-0.27*
2012	0.71**	0.21	0.74***	0.58***	-0.32	0.23*	-0.18***	-0.14***	0.02
District B	0.81*	0.19	0.18	-0.06	-0.37	0.58	1.69***	0.97***	1.33*
District C	-0.15	0.09	0.01	0.06	-1.12***	-0.61	-0.34	-0.60***	-0.88*
Constant	-1.89	-2.47*	2.81**	-1.66*	0.17	-0.6	-0.25	-0.24	-0.73
N=	69	69	69	68	69	69	69	66	69

Table 6: Selected coefficients from multilevel models of the relationship between dimensions of 8th grade school tracking system and 10th grade mathematics achievement, for students in District A, B, and C middle schools 2010-2012

	Model 1 (Fully Unconditional)	Model 2 (Dimensions only)	Model 3 (Controls)	Model 4 (Controls, school-mean centered)	Model 5 (Controls, school-mean centered, interaction homogeneneity*7th grade scores)
Differentiation (# courses)	0.03***	0.01	-0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Classroom skills homogeneity	-0.03***	-0.05***	-0.03	-0.02*	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Exclusivity (% in lower tracks)	0.04***	0.04	0.00	0.11*	0.06**
	(0.01)	(0.02)	(0.01)	(0.05)	(0.02)
Stability (% in same track 8-9)	0.04	0.00	0.02	0.04	0.00
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
Scope (Correlation: Math to ELA)	0.02	0.05	0.02	-0.01	-0.00
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Interaction: Classroom skill	--	--	--	--	0.02***
homogeneity*7th grade scores	--	--	--	--	(0.01)
Demographic controls	No	No	Yes	Yes	Yes
Prior achievement	No	No	Yes	Yes	Yes
School-mean centered	No	No	No	Yes	Yes
District FE	No	No	Yes	Yes	Yes
Cohort FE	No	No	Yes	Yes	Yes
School RE	Yes	Yes	Yes	Yes	Yes
Classroom RE	Yes	Yes	Yes	Yes	Yes
N=	20,545	20,545	20,545	20,545	20,545

NOTE: Each of the independent variables as well as the dependent variable in these models are z-score standardized. Demographic controls include indicators for student gender race/ethnicity, free or reduced-price lunch status, English-language learner status, and special education enrollment. Prior achievement controls are continuous measures of student 7th grade math and ELA test scores.

Table 7: Selected coefficients from multilevel models of the relationship between dimensions of 8th grade school tracking system and 10th grade ELA achievement, for students in District A, B, and C middle schools 2010-2012

	Model 1 (Fully Unconditional)	Model 2 (Dimensions only)	Model 3 (Controls)	Model 4 (Controls, school-mean centered)	Model 5 (Controls, school-mean centered, interaction homogeneity*7th grade scores)
Differentiation (# courses)	-0.01	-0.02	0.00	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Classroom skills homogeneity	0.00	0.00	0.02	0.00	0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Exclusivity (% in lower tracks)	0.09***	-0.03	-0.10***	0.04	0.05
	(0.02)	(0.06)	(0.02)	(0.03)	(0.03)
Stability (% in same track 8-9)	0.12***	0.15**	0.11**	0.00	0.02
	(0.03)	(0.06)	(0.03)	(0.05)	(0.05)
Scope (Correlation: Math to ELA)	0.03	0.02	0.01	0.02	0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Interaction: Classroom skill	--	--	--	--	0.06***
homogeneity*7th grade scores	--	--	--	--	(0.01)
Demographic controls	No	No	Yes	Yes	Yes
Prior achievement	No	No	Yes	Yes	Yes
School-mean centered	No	No	No	Yes	Yes
District FE	No	No	Yes	Yes	Yes
Cohort FE	No	No	Yes	Yes	Yes
School RE	Yes	Yes	Yes	Yes	Yes
Classroom RE	Yes	Yes	Yes	Yes	Yes
N=	20,545	20,545	20,545	20,545	

NOTE: Each of the independent variables as well as the dependent variable in these models are z-score standardized. Demographic controls include indicators for student gender race/ethnicity, free or reduced-price lunch status, English-language learner status, and special education enrollment. Prior achievement controls are continuous measures of student 7th grade math and ELA test scores.

Figure 3: Predicted 10th grade mathematics achievement scores for students in schools with high and low levels of skills-homogeneous assignment in 8th grade mathematics classrooms

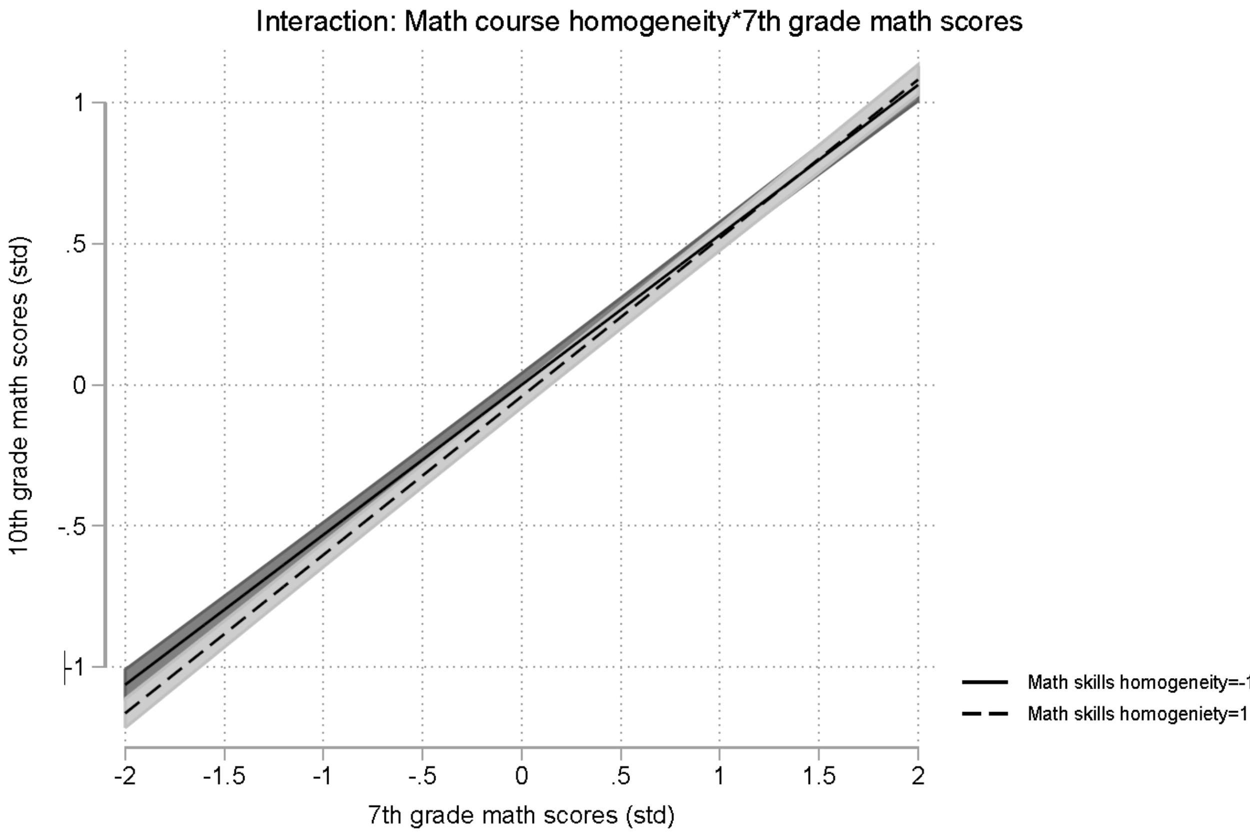

Figure 4: Predicted 10th grade ELA achievement scores for students in schools with high and low levels of skills-homogeneous assignment in 8th grade ELA classrooms

Appendices

Appendix Table 1: Selected coefficients from multilevel models of the relationship between dimensions of 8th grade school tracking system and 8th grade ELA achievement, for students in District A, B, and C middle schools 2010-2012

	Model 1 (Fully Unconditional)	Model 2 (Dimensions only)	Model 3 (Controls)	Model 4 (Controls, school-mean centered)
# courses	-0.05***	-0.05***	-0.01	0
	(0.01)	(0.01)	(0.01)	(0.01)
Classroom skills homogeneity	-0.02	0.00	-0.04***	-0.06***
	(0.02)	(0.02)	(0.01)	(0.01)
% in lower tracks	-0.06*	-0.23***	-0.07**	0.03
	(0.02)	(0.05)	(0.02)	(0.03)
% in same track 8-9	0.14***	0.38***	0.08*	-0.06
	(0.02)	(0.05)	(0.03)	(0.04)
Correlation: Math to ELA	0.00	-0.03	-0.03**	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Demographic controls	No	No	Yes	Yes
Prior achievement	No	No	Yes	Yes
School-mean centered	No	No	No	Yes
District FE	No	No	Yes	Yes
Cohort FE	No	No	Yes	Yes
School RE	Yes	Yes	Yes	Yes
Classroom RE	Yes	Yes	Yes	Yes
N=	20,771	20,771	20,771	20,771

NOTE: Each of the independent variables as well as the dependent variable in these models are z-score standardized. Demographic controls include indicators for student gender race/ethnicity, free or reduced-price lunch status, English-language learner status, and special education enrollment. Prior achievement controls are continuous measures of student 7th grade math and ELA test scores.

Appendix Table 2: Selected coefficients from school mean centered multilevel models of the relationship between dimensions of 8th grade school tracking system and 8th grade Math and ELA achievement, for students in District A, B, and C middle schools who moved on to modal and non-modal high schools 2010-2012

	Outcome: 10th grade Math score		Outcome: 10th grade ELA score	
	Students who enrolled in modal HS	Students who enrolled in non-modal HS	Students who enrolled in modal HS	Students who enrolled in non-modal HS
# courses	-0.01	0.01	0.01	0.05
	(0.01)	(0.03)	(0.01)	(0.03)
Classroom skills homogeneity	-0.03**	0.00	-0.01	-0.02
	(0.01)	(0.03)	(0.01)	(0.04)
% in lower tracks	0.10***	0.07	0.08*	0.03
	(0.03)	(0.07)	(0.04)	(0.10)
% in same track 8-9	0.04*	-0.03	0.04	0.08
	(0.02)	(0.05)	(0.05)	(0.14)
Correlation: Math to ELA	0.00	-0.01	0.02	0.03
	(0.01)	(0.03)	(0.01)	(0.03)
Demographic controls	Yes	Yes	Yes	Yes
Prior achievement	Yes	Yes	Yes	Yes
School-mean centered	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
School RE	Yes	Yes	Yes	Yes
Classroom RE	Yes	Yes	Yes	Yes
N=	16,914	3,210	16,914	3,210

NOTE: Each of the independent variables as well as the dependent variable in these models are z-score standardized. Demographic controls include indicators for student gender race/ethnicity, free or reduced-price lunch status, English-language learner status, and special education enrollment. Prior achievement controls are continuous measures of student 7th grade math and ELA test scores.