19

Tracking and inequality
New directions for research and practice

Adam Gamoran[1]

For more than a century, educators and researchers have debated the merits of separating students for instruction into different tracks, classes, and groups, according to their purported interests and abilities (for historical perspectives, see Loveless, 1998, 1999; Oakes, 2005; Oakes *et al.*, 1992; Powell *et al.*, 1985). The practice, known as "tracking" and "ability grouping" in the US and "streaming" and "setting" in the UK, is intended to create conditions in which teachers can efficiently target instruction to students' needs.[2] Despite this intended benefit, tracking has been widely criticized as inegalitarian, because students in high tracks tend to widen their achievement advantages over their low-track peers, and because measures of school performance commonly used to assign students to tracks typically coincide with the broader bases of social disadvantage such as race/ethnicity and social class, leading to economically and/or ethnically segregated classrooms. Yet tracking has been highly resistant to lasting change and remains in wide use in various forms in the US, the UK, and in school systems around the world.

Although struggles over tracking involve instructional and political challenges that play out in schools and classrooms, the persisting debate reflects not only local concerns but also broader tensions inherent in education systems (Oakes *et al.*, 1992). On the one hand, schools are charged with providing all students with a common framework of cognitive and social skills essential for full participation in the civic and economic activities of adult society. On the other hand, schools are structured to sort and select students for different trajectories aligned with their varied orientations and capacities. This ongoing tension between commonality and differentiation is at the heart of the tracking debate: Is the purpose of schooling to provide all students with a common socialization? Or is it to differentiate students for varied futures? The former aim is consistent with mixed-ability teaching, whereas the latter is consistent with tracking, and the debate has no simple resolution because school systems embody both goals.

Building on past research, recent work on tracking has advanced in three areas that indicate promising new directions for research and practice. First, new international scholarship has extended knowledge about the consequences of tracking for student achievement to contexts beyond the US and UK, where most prior research had been conducted. Second, recent studies of attempts to reduce or eliminate tracking and ability grouping have yielded important insights about why tracking is resistant to change and how some of the obstacles to detracking may be

213

surmounted. Third, a new wave of research on classroom assignment and instruction has pointed towards approaches that, while not resolving the tension between commonality and differentiation, may capture the benefits of differentiation for meeting students' varied needs without giving rise to the consequences for inequality that commonly accompany tracking and ability grouping. These findings in turn call for new research and experimentation in practice.

Before turning to these latest findings, I summarize the earlier literature on the effects of grouping and tracking on student achievement. This research has been well covered in prior reviews (e.g. Gamoran, 2004; Gamoran and Berends, 1987; Hallam, 2002; Harlen and Malcolm, 1997; Kulik and Kulik, 1982; Oakes *et al.*, 1992; Slavin, 1987, 1990), but I begin with it here because it sets the stage for the promising work of the present and the new directions for the future. Thus, the remainder of this chapter is divided into four sections: a review of findings about tracking and achievement that links work from the 1970s, 1980s, and 1990s to updated studies in the same vein; a discussion of recent international research on tracking, both between and within schools; an analysis of new studies of efforts to reduce or eliminate tracking; and a conclusion calling for new research and practice based on the latest findings.

Tracking and achievement: increased inequality without benefits to productivity

Following Gamoran and Mare (1989), one may distinguish between two possible consequences of tracking for achievement: it may affect *productivity*, that is, the overall level of achievement in the school or class; and it may affect *inequality*, that is, the distribution of achievement across the different tracks, classes, or groups. Although not all studies have reached the same conclusions about these outcomes, the weight of the evidence indicates that tracking tends to exacerbate inequality with little or no contribution to overall productivity. This occurs because gains for high achievers are offset by losses for low achievers. A compelling example of this pattern comes from Kerckhoff's (1986) study of ability grouping between and within schools in England and Wales. Kerckhoff used data from the National Child Development Study, which followed for more than thirty years all children born in the UK in the first week of March in 1958. He examined secondary school achievement in reading and mathematics among students enrolled in schools for high achievers (grammar schools), low achievers (secondary modern schools) and those of widely varying achievement levels (comprehensive schools). He also compared students assigned to high, middle, low, and mixed-ability classes within the different types of school. Comparisons between and within schools told a consistent story: There were no overall benefits to average achievement in contexts that differentiated students for instruction as compared with mixed-ability contexts. However, sorting students into selective schools and classes was associated with increasing gaps between high and low achievers over time (see also Kerckhoff, 1993). The comparison of tracking to mixed-ability teaching has received less attention in the US because tracking has been nearly universal at the secondary level (Loveless, 1998), but comparisons of ability-grouped with mixed-ability classes in middle school mathematics and science (Hoffer, 1992) and English (Gamoran and Nystrand, 1994) have yielded the same pattern. National survey analyses in the US also demonstrated that, over the course of high school, students assigned to high and low tracks grow farther and farther apart in achievement (e.g. Heyns, 1974; Alexander, Cook and McDill, 1978; Gamoran, 1987a, 1992; Gamoran and Mare, 1989; Lucas and Gamoran, 2002).

Because track location is correlated with traditional bases of socio-economic disadvantage, tracking not only widens achievement gaps but also reinforces social inequality (Lucas and Berends, 2002; Oakes *et al.*, 1992). In contrast to socio-economic status, which has direct effects on track assignment, race and ethnicity affect track assignment indirectly: Minority students whose test scores and socio-economic background match those of Whites are no less likely to be placed in high tracks (Gamoran and Mare, 1989; Lucas and Gamoran, 2002; Tach and Farkas, 2006). However, because minority students tend to reach high school with lower test scores and less advantaged socio-economic circumstances, tracking works to the disadvantage of minority students and contributes to achievement gaps.

As the demographic make-up of US schools has changed, new patterns of inequality associated with tracking have become more salient. With regard to language minority students, Callahan (2005) argued that schools often conflate limited proficiency in English with limited ability to master academic content. As a result, English language learners are tracked into classes with modified curricula that are less rigorous than those of regular classes, which prevents these students from gaining access to advanced instruction even as their language skills develop. While Callahan supported these assertions with a study of a rural California school, Paul (2005) reached a similar conclusion based on her study of five diverse urban schools. Paul noted that enrollment in algebra 1, the gateway to the college-preparatory curriculum, was stratified by race and ethnicity, with Asian American and White students enrolled in higher proportions, and African American and Hispanic students enrolled in lower proportions. When English language learners enrolled in the same levels of algebra as fluent English speakers, they had similar rates of college-preparatory course work. Foreshadowing this work, Padilla and Gonzales (2001) argued that one reason recent immigrants to the US from Mexico outperform second-generation students is that the immigrants have spent less time in low tracks in US schools.

New forms of tracking in the US have exhibited patterns of inequality comparable with those of earlier forms. Using high-school transcripts from a national sample of students, Lucas (1999) showed that students were grouped on a subject-by-subject basis rather than by broad curricular programs. Nevertheless, students' course levels tended to correlate across subject areas, and this more subtle version of tracking still resulted in achievement inequality. Mitchell and Mitchell (2005) demonstrated that multi-track, year-round schools also tended to stratify students by social origins. Both Lewis and Cheng (2006) and Mickelson and Everett (2008) found that the transformation of vocational education into career and technical education, though accompanied by greater emphasis on academic work within technical courses of study, still resulted in stratified class enrollments.

Generally, elementary and middle schools have witnessed a pattern of increasing inequality similar to that observed at the high school level (e.g. Gamoran *et al.*, 1995; Hoffer, 1992; Rowan and Miracle, 1983). Until recently, national data have been available only at the secondary level, so it was not possible to examine the generalizability of patterns of inequality associated with elementary school ability grouping. However, recent analyses of data from a national sample of children who entered kindergarten in 1998 have confirmed the pattern of widening gaps for within-class reading groups in kindergarten (Tach and Farkas, 2006). Using later waves of the same data, Lleras and Rangel (2009) reported similar findings for between-class ability grouping in Grades 1 and 3. Taking exception to the general pattern, Slavin (1987) reported, based on a synthesis of research on elementary school grouping, that within-class grouping for mathematics had positive effects for students in low-ranked as well as those in high-ranked groups. Slavin also noted that, when students were regrouped for specific subjects, rather than being tracked for the entire school day, ability grouping had positive effects for students at all

achievement levels. On the basis of these findings, Slavin proposed that elementary school ability grouping can have positive effects when assignment is based on criteria relevant to the subject, when students can be moved from one group to another as appropriate to their progress, and when curriculum and instruction are differentiated to meet the needs of students assigned to the different groups.

Slavin's conclusions have recently been reaffirmed by Connor and her colleagues (Connor *et al.*, 2007; Connor *et al.*, 2009). Connor's work shows that small reading groups can be used effectively to tailor reading instruction to students' needs. In a randomized comparison, Connor *et al.* (2007) reported that students taught by teachers who arranged students into reading groups according to carefully assessed student performance levels, and who aimed instruction at students' specific needs, performed much better by the end of first grade than those taught by teachers who did not have access to the systematic approach to assigning students and differentiating instruction. Though based on less precise evidence, Tomlinson *et al.* (2003) advanced similar claims about the value of within-class differentiation of instruction as a strategy for effective teaching of students with varied interests and skills.

Challenges in measuring track effects

Two methodological challenges have confronted researchers studying the impact of tracking and ability grouping on student achievement. One challenge has been to measure accurately students' group and track locations. At the secondary level, research from the 1970s and 1980s often relied on students to report whether their curricular programs could best be described as academic/college-preparatory, vocational, or general. This *social-psychological* measure of tracking was useful as an indicator of students' perceptions, but did not necessarily represent students' actual learning opportunities. Lucas (1999) developed a *structural* measure of track location by using students' transcripts to identify tracks based on the courses students had taken. Lucas and Gamoran (2002) showed that structural and social-psychological dimensions of tracking had independent effects on student achievement, and both contributed to achievement gaps. Other researchers have used network analysis techniques to identify tracks through the configuration of courses in which students enroll (Friedkin and Thomas, 1997; Heck *et al.*, 2004), reaching similar conclusions about tracking and inequality. More recent studies have also uncovered inequality using teacher reports to distinguish among ability groups at high, middle, and low levels (Carbonaro, 2005; Tach and Farkas, 2006).

The second methodological challenge has been to distinguish the effects of track assignment from the effects of pre-existing differences among students assigned to different tracks. Obviously, students in high and low tracks are on different achievement trajectories to begin with; that is how they came to be located in different tracks. All the analyses discussed here have controlled for prior achievement and social background, but owing to unreliability and measurement error, not all pre-existing conditions may have been captured by the controls, and the potential for selectivity bias remains. Researchers have endeavored to respond to this challenge in two ways. First, a few studies, mainly prior to 1970, used random assignment to tracked or untracked settings to rule out selectivity bias (Slavin, 1987, 1990). These studies yielded widely varying estimates of track effects that centered around zero. Because they provided little information on what was going on inside the tracks, it is difficult to assess the generalizability of these small and long-ago experiments. In at least some cases of zero effects, teachers designed instruction and curriculum to be the same across tracks, in contrast to the real world where tracking is typically accompanied by curricular and instructional differentiation. These

findings led Gamoran (1987b) to argue that the effects of tracking depend on how it is implemented, a conclusion later supported by both case study (Gamoran, 1993) and survey analyses (Gamoran, 1992).

Second, researchers have used econometric techniques to mitigate selectivity bias. Gamoran and Mare (1989) estimated endogenous switching regressions that model track assignment and track effects simultaneously, allowing for correlated errors among unobserved predictors of assignment and outcomes. Their results, which focused on mathematics achievement and high school completion for the high school class of 1982, indicated that the pattern of increasing inequality observed in standard regression analyses with rich controls was upheld in the more complex technique. Lucas and Gamoran (2002) replicated these results for the high school class of 1992, as well as the class of 1982, and with course-based as well as self-reported indicators of track location. Again, the main findings were upheld. However, Betts and Shkolnik (2000), who estimated both propensity models and two-stage least squares regression models of track effects on mathematics achievement, concluded that the differential effects of tracking for students in high and low tracks were much smaller than reported in earlier studies that relied on simple regressions. Figlio and Page (2002) similarly called into question the inequality consequences of tracking on secondary school math achievement, on the basis of two-stage least squares regression models.[3] While it is premature to conclude that tracking is not harmful to low achievers, these studies, combined with the early experimental research, suggest the effects may be smaller than is typically assumed. Since Gamoran and Mare focused on broad curricular tracking, while Betts and Shkolnik and Figlio and Page examined between-class ability grouping, the findings may also indicate that the latter are less consequential for inequality than the former.

Mechanisms of track effects on achievement

With few exceptions, the evidence indicates that tracking tends to magnify inequality. Why is that the case? Conceptually, researchers have identified mechanisms of social comparison as well as differentiated instruction, but empirically it appears that instructional variation across tracks and groups at different levels is the more prominent reason for increases in achievement gaps between tracks. A number of studies have concluded that students in high tracks encounter more challenging curricula, move at a faster pace, and are taught by more experienced teachers with better reputations, while students in low tracks encounter more fragmented, worksheet-oriented, and slower-paced instruction provided by teachers with less experience or clout (for reviews, see Gamoran, 2004; Oakes et al., 1992). These findings have emerged at the elementary, middle, and high school levels. Instructional differences reflect not only what teachers do in classrooms, but also how students respond. A recent finding along these lines comes from the work of Carbonaro (2005), who demonstrated that achievement diverges in part because high-track students put forth more effort on their schoolwork than low-track students. While this finding reflected, in part, low-track students' responses to instruction that was less intellectually stimulating than the instruction given to high-track classes, it also stemmed from differences that students brought with them to class.

Other new examples of instructional mediation of track effects come from both hypothesis-testing and interpretive research. In a study of sixty-four middle and high school English classes, Applebee et al. (2003) reported greater use of discussion-based approaches to literature instruction in high-ability than in low-ability classes, and this difference accounted for just over one third of the effect of ability group assignment on writing performance. Discussion-based

approaches included authentic questions and uptake (questions with no prespecified answer and those that build on prior statements), open discussion, drawing in multiple perspectives (envisionment-building), and conversations that connected different curricular topics. Watanabe (2008) reported parallel instructional differences based on in-depth analyses of 68 hours of classroom observation in two teachers' language arts classes. In high-ability classes, she found more engagement with challenging and meaningful curricula, more writing assignments in more diverse genres, and more feedback from teachers, as contrasted with more emphasis on test preparation in low tracks.

Findings that instructional differentiation accounts for much of the effect of tracking have led some observers to conclude that tracking per se does not generate inequality, but rather inequality has emerged because of the way in which tracking has been implemented (e.g., Hallinan, 1994). If instruction in low tracks could be effectively geared towards students' needs, this argument states, then tracking might mitigate rather than exacerbate inequality. While reasonable in principle, this goal has proven difficult to accomplish in practice, and there are few examples of effective instruction in low-track classes (for exceptions, see Gamoran, 1993, and Gamoran & Weinstein, 1998). At the same time, it is important to acknowledge that most studies of ability grouping and curriculum tracking have found that high-achieving students tend to perform better when assigned to high-level groups than when taught in mixed-ability settings. Proponents of tracking tend to emphasize the benefits of high-level classes for high-achieving students, with little attention to implications for inequality, while critics tend to focus on inequality without acknowledging the effects for high achievers. As a result, proponents and critics are apt to talk past one another with little chance for resolution, and student-assignment policies often lurch from one system to another, without recognition of the strengths and shortcomings of each (Boaler et al., 2000; Gamoran, 2002; Tsuneyoshi, 2004).

New international research on tracking and achievement

An emerging body of international work is largely consistent with the findings from the US and the UK. Perhaps the most revealing results come from new cross-national studies of international achievement data. Analyses from PISA 1999 (Program on International Student Assessment), a study conducted in twenty-eight OECD countries, indicated that countries with more differentiated school systems are characterized by greater inequality by social origins in reading achievement (OECD, 2002). Hanushek and Woessmann (2006) reinforced this conclusion by comparing twenty countries that participated in both PISA and PIRLS (Progress in International Reading Literacy Study), showing that achievement inequality tends to increase more between the primary and secondary grades in countries that practice early tracking than in countries that do not. Similarly, research on twenty-four countries that participated in TIMSS 2003 (Trends in International Mathematics and Science Survey) at Grades 4 and 8 showed that countries that rely on between-class ability grouping for mathematics exhibit more growth in achievement inequality from Grades 4 to 8 than countries that make less use of ability grouping (Huang, in press). These findings are consistent with numerous single-nation studies showing that tracking tends to reinforce inequality.

A recurring theme in the international work is that grouping and tracking come in many forms, a point that is easily missed when one focuses on a single nation. For example, countries differ on whether tracking occurs largely between schools (e.g. Japan, Germany), within schools (Australia, Belgium, Israel, US), or both (Taiwan, UK). In these different tracking systems, the

scope of tracking may be wide (covering many subjects) or narrow (implemented on a subject-by-subject basis). Countries also differ on whether differentiation is introduced early or late, and whether or not the system is flexible enough to allow mobility between tracks. These structural differences were anticipated by Sørensen (1970), but have been greatly elaborated as international differences have become evident (LeTendre *et al.*, 2003). What is striking about the variation in the *forms* of tracking, however, is that the *results* are broadly similar: where tracking systems are present, achievement tends to diverge, and to reinforce initial differences by social class. New studies from Japan (Ono, 2001), Korea (Park, 2009), South Africa (Hoadley, 2008), Israel (Ayalon, 2006), Germany (Cheng *et al.*, 2007), Belgium (Van de Gaer *et al.*, 2006; Van Houtte, 2004), and the UK (Boaler *et al.*, 2000; Ivinson and Duveen, 2005; Ireson *et al.*, 2002) all identify aspects of increasing inequality associated with grouping between or within schools. Moreover, as ethnic minority groups increase in size, and ethnic inequality is increasingly recognized in nations that were formerly relatively homogeneous (such as European countries with new populations of guest workers), researchers are finding that tracking reinforces ethnic inequalities (Cheng *et al.*, 2007). Ivinson and Duveen (2005) in the UK and Ayalon (2006) in Israel also demonstrated that horizontal differentiation (i.e. divisions between subjects) tend to stratify students by social origins, just as does vertical differentiation (divisions between levels). Finally, Van Houtte (2004) presented findings from Belgium that supported the conclusion from US research that track effects are driven by instructional differences to an important degree.

Within this common framework, interesting differences also emerge. For example, in countries with well-articulated standards tied to curriculum and assessment, the harmful effects of tracking may be mitigated by incentives for success in lower level classes. Broaded (1997) reported that high-stakes exams targeted at different achievement levels in Taiwan led all students, including those in low tracks, to work hard at their studies, and, as a result, tracking contributed to *smaller* achievement inequalities. Similarly in the case of Israel, Ayalon and Gamoran (2000) found that schools with multiple ability levels within college-preparatory mathematics programs tended to have *less* inequality by social origin than schools with only a single level. They attributed this result to meaningful incentives attached to lower level mathematics courses that, like higher level courses, led to high stakes assessments at the end of high school. Likewise, a secondary curriculum reform in Scotland that raised standards for lower level students resulted in declining inequality of achievement over time (Gamoran, 1996), and in Australia, a reform in secondary English that reduced the number of tracks and simultaneously raised standards in low tracks may have boosted test scores overall (Stanley and McCann, 2005). In the US, a parallel finding is that Catholic schools, which place more academic demands on students in lower tracks than public schools, tend to exhibit less achievement inequality between tracks than public schools (Gamoran, 1992). These findings reinforce Broaded's (1997) conclusion that the impact of tracking is context-dependent and suggest that, in principle, tracking's pernicious effects on low achievers can be reduced or eliminated. Thus far, however, attempts to use ability grouping to raise achievement in the context of high standards in US public schools have met with limited success (Lewis and Cheng, 2006; Mickelson and Everett, 2008; Sandholtz *et al.*, 2004).

New insights from US research on detracking

More than fifteen years ago, Oakes (1992) insightfully identified three challenges to detracking: normative challenges, based on long-standing beliefs that young persons differ by ability and

that schools should be structured to meet those differences; political challenges, reflecting the difficulty of overcoming vested interests in tracking such as those held by parents of high-achieving students and by teachers who enjoy teaching honors classes; and technical challenges, reflecting the difficulty of instructing students of widely varying levels of performance, a task for which few teachers are prepared. Most of the emphasis in Oakes' subsequent work (see especially the 2005 edition of her classic book, *Keeping track*) and that of her colleagues and students (e.g. Oakes and Wells, 1998; Wells and Serna, 1996; Welner, 2001; Yonezawa *et al.*, 2002) has been on the normative and political challenges, reasoning that if these challenges could be met, the technical difficulties could be overcome. Recent evidence, however, suggests the opposite: failure to solve the technical problems of mixed-ability teaching is a major impediment to addressing the normative and political challenges. While the technical challenges have defied easy solution, recent work has identified conditions under which effective teaching in mixed-ability contexts may be more successful than in the past.

Challenges of detracking

Loveless's (1999) analysis of detracking reforms in California and Massachusetts revealed substantial resistance from teachers who believed that they were not equipped to succeed in instructing students at widely varying performance levels within the same classrooms. Teachers' attitudes towards detracking tended to differ by subject matter, with mathematics and foreign language teachers more resistant than teachers in other subjects, owing to beliefs about the sequential nature of knowledge in these disciplines (see also Ball, 1987; Gamoran and Weinstein, 1998). Even in social studies, however, a subject area that might be viewed as particularly conducive to mixed-ability teaching because of the potential for discussion of topics from diverse viewpoints, detracking efforts have run into technical difficulties. One case study found that teachers struggled to engage students in classes with widely varying achievement levels: low-achieving students had difficulty with assignments, while high-achieving students were bored (Rosenbaum, 1999). In another study, Rubin (2008) found that detracking in social studies seemed to work well in a middle-class suburban school with a relatively homogeneous population, as teachers emphasized active learning and differentiated assignments for students at different performance levels. However, detracked social studies classes appeared less effective in a more diverse school, where teachers aimed more for relevance than for high standards; and in an inner-city school with a low-income population, detracking resulted in a highly routinized curriculum with little challenge for students. Rubin's observations in the inner-city school mirrored earlier findings by Gamoran and Weinstein (1998) from an urban school in which tracking in mathematics was eliminated by diluting the curriculum in mixed-ability classes to a level that all students could follow, with the result that teachers complained students were not being prepared to move to more advanced mathematics.

Ironically, findings from all three of these case studies (Gamoran and Weinstein, 1998; Rosenbaum, 1999; Rubin, 2008) suggest that high-achieving minority students may have the most to lose when detracking is unsuccessful. These students are often found in urban schools where detracking has not resulted in challenging instruction in mixed-ability classes, and they may lack the support outside of school to succeed in the absence of a challenging curriculum. Rubin (2003) brought this problem to life based on interviews and observations of a high-achieving minority student in a detracked school, who socialized with a small group consisting of less academically oriented peers, to the detriment of her academic work.

Some schools have attempted to reduce the use of tracking by allowing students to select their own track assignments. Recent case studies suggest, however, that student choice is not an effective detracking mechanism, because students tend to sort themselves into classes in much the same way as a traditional tracking system, and with the corresponding results for social class and race/ethnic divisions (Watanabe, 2007). Yonezawa *et al.* (2002) proposed that differential access to information and varied aspirations among students contributed to this pattern. In addition, they noted that minority students preferred classes in which they were not racially isolated and in which their cultural backgrounds were valued. These findings reflect the familiar tension between commonality and differentiation: while there may be benefits to students' academic performance from pursuing a common curriculum, students are motivated by their interests and social concerns, which may result in ethnic as well as academic divisions among students.

Boaler and Staples (2008) uncovered mixed success in another detracking case study. Initially, achievement gains appeared in one detracked school compared with two others that did not detrack. However, the gains were not sustained over the three years of the study. Moreover, the achievement benefits were not evident on the high stakes state standardized test, and it is difficult at any rate to attribute achievement trends to any single reform in a sample of three schools. Nonetheless, the study is enticing in its call for further examination of instruction in detracked schools.

Addressing the technical challenge: differentiated instruction in mixed-ability classes

Not all cases of mixed-ability teaching have met with frustration. In the same research project that uncovered a case of diluted curriculum in a detracked school (discussed in the last section), Gamoran and Weinstein (1998) also identified a successful instance of detracking in secondary school mathematics. In this urban, east-coast high school, in which half the students were eligible to receive free or reduced-price lunch, student performance on authentic assessments was the highest of all the twenty-five highly restructured schools from which this case was drawn. In this school, mathematics and science instruction were integrated in the same class, and student work was project-oriented; for example, researchers observed students applying principles of mathematics and physics in completing an assignment to design rides in an amusement park. Students were assessed based on portfolios of work in a variety of subjects, and expectations for students took into account their progress as well as the levels of excellence they had attained. Moreover, students were expected to have mastered elementary mathematics, and, if they had not, a Saturday tutoring program was available to help them along. Key elements that supported a rigorous curriculum in a mixed-ability setting in this school were small classes (limited to fifteen students), the supplemental tutoring program, a visionary leader who had selected a staff with congruent attitudes, and the opportunity to interview students prior to students' admission to the school.

More recently, Burris and her colleagues (Burris *et al.*, 2006, 2008) also identified cases of high achievement in mathematics that resulted from a move to mixed-ability teaching. The authors used an interrupted time series design to assess the impact of the reform, comparing the achievement trajectories of schools before the reform with their trajectories afterwards, as well as with the trajectories of other schools that did not undergo the reform over the same time period. At the middle school level in this New York school district (Burris *et al.*, 2006), teachers implemented an accelerated curriculum for all students paired with a supplemental

221

workshop to support students who had trouble keeping up. They also introduced common preparation time for teachers and increased the use of calculators in class. At the high school level, the low-track non-Regents class was eliminated, and all students were placed in mathematics classes that led to the Regents diploma. Students who struggled with this class had available to them a supplementary class that met three times each week. At both levels, student achievement rose following the introduction of the reform. Achievement gaps narrowed as low achievers gained more than high achievers, but there was no evidence that high achievers suffered in their performance as a result of the reform. Achievement gains did not reflect increasing high-school dropout rates; on the contrary, dropout rates declined over the period of the reform. It should be noted that this case involved an economically advantaged school district with relatively few high-needs students compared with other New York school districts. The supplemental class also provided about 50 percent more mathematics instruction to low-achieving students.

The new research by Burris and colleagues is extremely important because it demonstrates that detracking *can* result in gains for low achievers *without* the losses for high achievers observed in earlier attempts. As in the case study reported by Gamoran and Weinstein (1998), however, success was based in part on favorable circumstances, particularly the resources that enabled the school to offer extra mathematics instruction for struggling students. This accomplishment calls for replication in other contexts to assess its broader viability.

Conclusion: new directions for research and practice

While definitive solutions remain elusive, the present time is witness to exciting new prospects for balancing the aims of commonality and differentiation in arranging students for instruction. Recent findings lend support to two approaches that merit further experimentation in research and practice: raising standards for low achievers in differentiated classrooms; and providing differentiated learning opportunities in mixed-ability classrooms. The key to evaluating both approaches will be careful monitoring of the nature and quality of instruction and the relation between instruction and achievement, however students are arranged for class.

Raising standards for low-achieving students

The practical conclusion from years of tracking research that low-level, dead-end courses should be eliminated is no longer seriously debated. High-school courses such as general math and business English do not prepare students for post-secondary opportunities and are less effective than regular courses such as algebra and college-preparatory English, even for students with low skill levels in these areas. This conclusion still leaves open the possibility, however, that meaningful instruction at all skill levels could make differentiated classes an effective way to organize students for learning.

Critics of tracking such as Oakes (2005) argue that, because tracking is inherently stratifying, it is just not possible to offer effective instruction to low-achieving students in ability-grouped classes. Indeed, examples of high-quality instruction in low-ability classes are rare. Yet recent international research shows that differentiated class settings for low achievers can be effective when they are tied to meaningful outcomes, such as assessments that are aligned to the curriculum, and provide access to jobs and further education. Studies from Taiwan (Broaded, 1997) and Israel (Ayalon and Gamoran, 2000) demonstrated that differentiation within academic

programs in which meaningful instruction and valued incentives are present at all ability levels can result in less inequality than systems of fewer levels in which low-achieving students lack access to meaningful incentives. Other research from Scotland (Gamoran, 1996) and Australia (Stanley and McCann, 2005) observed that the negative effects of tracking for low achievers diminished when the degree of tracking was reduced and when academic standards in the lower level classes were elevated. The common ingredient in all four cases was a meaningful assessment that had value for students in lower-level as well as higher-level classes.

Do these findings have any bearing on the US, where classes for low achievers typically lack meaningful incentives for effort or performance? The finding that Catholic schools obtain smaller achievement gaps between tracks than public schools by providing more rigorous instruction in low tracks, and cases of successful low-track instruction in Catholic schools (Gamoran, 1993) and restructured public schools (Gamoran and Weinstein, 1998) merely show that exceptions are possible, not that making low-track instruction more effective by raising standards overall is a viable reform strategy for the US. The current emphasis on test-based accountability in the US might, in principle, lead schools to create effective low-ability classes in order to meet accountability requirements. However, the evidence so far suggests that accountability-driven tracking is no more effective for low achievers than other forms of tracking (Sandholtz et al., 2004). Based on insights from the international work, one can identify at least three elements that would need to change to make low-track classes more effective: First, the assessments towards which students were striving would need to be tied to futures that were more visibly meaningful to students than is currently the case. At present, students are prodded to perform on multiple-choice tests whose underlying standards are not evident to students and which demand fragmented knowledge rather than coherent mastery of subject matter that has relevance beyond the test itself. Second, the assessments would need to offer incentives for students as well as schools; at present, schools are held accountable for student performance, but the students themselves are not. Positive incentives such as access to jobs and/or post-secondary education would need to be offered, not merely negative sanctions such as denial of a high school diploma. Third, the alignment between the course curriculum and the assessment would need to be tighter than has typically been the case in the US.

Differentiating instruction in mixed-ability settings

Although detracking remains a challenging solution, with more examples of failure than success, the findings of recent studies are positive enough to warrant further efforts. An examination of reports of effective instruction in mixed-ability classes yields several common ingredients. First, the success stories all recognize that students differ in the skills and interests they bring with them to class. Successful cases reported by Burris, Gamoran, and Connor and their colleagues are not instances in which teachers acted as if students were all alike. Instead, teachers responded to variation among students in their teaching. Second, and correspondingly, all the successful cases involved differentiated instruction within the mixed-ability setting. In the secondary school cases reported by Burris, Gamoran, and their colleagues, differentiation involved supplemental instruction that was available for students who struggled with class materials. In Connor's elementary school research, differentiation meant carefully analyzing students' skill levels, matching skills to particular instructional strategies, and arranging students for instruction within classes in such a way as to match the skill levels with instructional approaches. Third, teachers in each of these cases had access to important resources that allowed them to supplement instruction and tailor it to students' needs. Future efforts would do well to keep these elements in mind.

Combining research on tracking with research on teaching

After a century of research on tracking and ability grouping, one might expect to see a definitive answer to the question of how best to organize students for instruction. Yet the dilemma persists, because the goals of commonality and differentiation lie in uneasy proximity to one another, because every approach has disadvantages as well as advantages, and because the consequences of different solutions vary by context. Research in the last decade has made important progress, however, by focusing on the instruction provided to students assigned to class in different ways. Ultimately, how students are arranged matters less than the instruction they encounter, so bringing together research on tracking with research on teaching offers the most useful way to continue to shed light on this topic of continuing interest.

Notes

1 The author is grateful for helpful research assistance from Michelle Robinson and exceptional editing from Cathy Loeb.
2 US writers often use the terms "tracking" and "ability grouping" interchangeably. For brevity I use the single term "tracking" to capture all the various forms of structural differentiation for instruction. When distinguishing among different forms, I use the term "tracking" to refer to the practice of dividing students into separate classes (or clusters of classes) for all of their academic subjects, and the term "ability grouping" to mean the division of students into classes on a subject-by-subject basis. This use parallels the different meanings of the terms "streaming" and "setting" used in the UK. I use the terms "within-class ability grouping" to refer to the use of instructional groups within class for a particular subject, and "between-school grouping" to refer to systems in which students are assigned to separate schools targeted to different futures on the basis of varied academic performance.
3 Models estimated by Betts and Shkolnik (2000) and Figlio and Page (2002) rely on very strong assumptions, so their results should be interpreted with particular caution. Betts and Shkolnik's conclusions rest on comparisons of classes at similar ability levels as reported by teachers but located in schools that differed on whether the principal reported that tracking was used for mathematics. Yet teacher reports of class ability levels may reflect between-class ability grouping irrespective of the principal's report. Figlio and Page (2002) used as instruments for track assignment indicators that, on the face of it, seem far-fetched: two- and three-way interactions between the number of courses required for graduation, the number of schools in the county, and the fraction of voters in the county who voted for President Reagan in 1984. Weak instruments would undermine the estimates of track effects and could bias them towards zero.

References

Alexander, K.L., Cook, M.A. and McDill, E.L. (1978) "Curriculum tracking and educational stratification," *American Sociological Review* 43: 47–66.

Applebee, A.N., Langer, J., Nystrand, M. and Gamoran, A. (2003) "Discussion-based approaches to developing understanding: classroom instruction and student performance in middle and high school English," *American Educational Research Journal* 40: 685–730.

Ayalon, H. (2006) "Nonhierarchical curriculum differentiation and inequality in achievement: a different story or more of the same?" *Teachers College Record* 108: 1186–1213.

——— and Gamoran, A. (2000) "Stratification in academic secondary programs and educational inequality: comparison of Israel and the United States," *Comparative Education Review* 44: 54–80.

Ball, S.J. (1987) *The micro-politics of the school: Towards a theory of school organization*, London: Methuen.

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

Boaler, J. and Staples, M. (2008) "Creating mathematical futures through an equitable teaching approach: the case of Railside School," *Teachers College Record* 110: 608–645.

—— Wiliam, D. and Brown, M. (2000) "Students' experiences of ability grouping: disaffection, polarisation, and the construction of failure," *British Educational Research Journal* 26: 631–648.

Broaded, C.M. (1997) "The limits and possibilities of tracking: some evidence from Taiwan," *Sociology of Education* 70: 36–53.

Burris, C.C., Heubert, J.P. and Levin, H.M. (2006) "Accelerating mathematics achievement using heterogeneous grouping," *American Educational Research Journal* 43: 105–136.

—— Wiley, E., Welner, K. and Murphy, J. (2008) "Accountability, rigor, and detracking: achievement effects of embracing a challenging curriculum as a universal good for all students," *Teachers College Record* 110: 571–607.

Callahan, R. (2005) "Tracking and high school English learners: limiting opportunity to learn," *American Educational Research Journal* 42: 305–328.

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

Cheng, S., Martin, L. and Werum, R.E. (2007) "Adult social capital and track placement of ethnic groups in Germany," *American Journal of Education* 114: 41–74.

Connor, C.M., Morrison, F.J., Fishman, B.J., Schatschneider, C. and Underwood, P. (2007) "The early years: algorithm-guided individualized reading instruction," *Science* 315: 464–465.

—— Piasta, S.B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P. and Morrison, F.J. (2009) "Individualizing student instruction precisely: effects of child by instruction interactions on first graders' literacy development," *Child Development* 80: 77–100.

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

Friedkin, N.E. and Thomas, S.L. (1997) "Social positions in schooling," *Sociology of Education* 70: 239–255.

Gamoran, A. (1987a) "The stratification of high school learning opportunities," *Sociology of Education* 60: 135–155.

—— (1987b) "Organization, instruction, and the effects of ability grouping: comment on Slavin's 'best-evidence synthesis'," *Review of Educational Research* 57: 341–345.

—— (1992) "The variable effects of high school tracking," *American Sociological Review* 57: 812–828.

—— (1993) "Alternative uses of ability grouping in secondary schools: can we bring high-quality instruction to low-ability classes?" *American Journal of Education* 101: 1–22.

—— (1996) "Curriculum standardization and equality of opportunity in Scottish secondary education, 1984–1990," *Sociology of Education* 29: 1–21.

—— (2002) *CES briefing: Standards, inequality, and ability grouping*, Edinburgh, Scotland: Centre for Educational Sociology.

—— (2004) "Classroom organization and instructional quality," in M.C. Wang and H.J. Walberg (eds) *Can unlike students learn together? Grade retention, tracking, and grouping*, Greenwich, CT: Information Age Publishing, pp. 141–155.

—— and Berends, M. (1987) "The effects of stratification in secondary schools: synthesis of survey and ethnographic research," *Review of Educational Research* 57: 415–435.

—— and Mare, R.D. (1989) "Secondary school tracking and educational inequality: compensation, reinforcement, or neutrality?" *American Journal of Sociology* 94: 1146–1183.

—— and Nystrand, M. (1994) "Tracking, instruction, and achievement," *International Journal of Educational Research* 21: 217–231.

—— , ——, Berends, M. and LePore, P.C. (1995) "An organizational analysis of the effects of ability grouping," *American Educational Research Journal* 32: 687–715.

—— and Weinstein, M. (1998) "Differentiation and opportunity in restructured schools," *American Journal of Education* 106: 385–415.

Hallam, S. (2002) *Ability grouping in schools*, London: University of London Institute of Education.

Hallinan, M.T. (1994) "Tracking: from theory to practice," *Sociology of Education* 67: 79–84.

Hanushek, E.A. and Woessmann, L. (2006) "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries," *The Economic Journal* 116: C63–C76.

Harlen, W. and Malcolm, H. (1997) *Setting and streaming: a research review*, Edinburgh: Scottish Council for Research in Education.

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

Heyns, B. (1974) "Selection and stratification within schools," *American Journal of Sociology* 79: 1434–1451.

Hoadley, U. (2008) "Social class and pedagogy: a model for the investigation of pedagogic variation," *British Journal of Sociology of Education* 29: 63–78.

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

Huang, M.-H. (in press) "Classroom homogeneity and the distribution of student math performance: a country-level fixed-effects analysis," *Social Science Research*.

Ireson, J., Hallam, S., Hack, S., Clark, H. and Plewis, I. (2002) "Ability grouping in English secondary schools: effects on attainment in English, mathematics, and science," *Educational Research and Evaluation* 8: 299–318.

Ivinson, G. and Duveen, G. (2005) "Classroom structuration and the development of social representations of the curriculum," *British Journal of Sociology of Education* 26: 627–642.

Kerckhoff, A. (1986) "Effects of ability grouping in British secondary schools," *American Sociological Review* 51: 842–858.

—— (1993) *Diverging pathways: social structure and career deflections*, Cambridge, England: Cambridge University Press.

Kulik, C.-L. and Kulik, J.A. (1982) "Effects of ability grouping on secondary school students: a meta-analysis of evaluation findings," *American Educational Research Journal* 19: 415–428.

LeTendre, G.K., Hofer, B.K. and Shimizu, H. (2003) "What is tracking? Cultural expectations in the United States, Germany, and Japan," *American Educational Research Journal* 40: 43–89.

Lewis, T. and Cheng, S.-Y. (2006) "Tracking, expectations, and the transformation of vocational education," *American Journal of Education* 113: 67–99.

Lleras, C. and Rangel, C. (2009) "Ability grouping practices in elementary school and African American/Hispanic achievement," *American Journal of Education* 115: 279–304.

Loveless, T. (1998) *The tracking and ability grouping debate*, Washington, DC: Fordham Institute.

—— (1999) *The tracking wars: state reform meets school policy*, Washington, DC: Brookings Institution Press.

Lucas, S.R. (1999) *Tracking inequality*, New York: Teachers College Press.

—— and Berends, M. (2002) "Sociodemographic diversity, correlated achievement, and de facto tracking," *Sociology of Education* 75: 328–348.

—— and Gamoran, A. (2002) "Track assignment and the black–white test score gap: divergent and convergent evidence from 1980 and 1990 sophomores," in T. Loveless (ed.) *Closing the gap: promising strategies for reducing the achievement gap*, Washington, DC: Brookings, pp. 171–198.

Mickelson, R.A. and Everett, B.J. (2008) "Neotracking in North Carolina: how high school courses of study reproduce race and class-based stratification," *Teachers College Record* 110: 535–570.

Mitchell, R.E. and Mitchell, D.E. (2005) "Student segregation and achievement tracking in year-round schools," *Teachers College Record* 107: 529–562.

Oakes, J. (1992) "Can tracking research inform practice? Technical, normative, and political considerations," *Educational Researcher* 21(4): 12–22.

—— (2005) *Keeping track: how schools structure inequality*, 2nd edn, New Haven, CT: Yale University Press.

—— Gamoran, A. and Page, R.N. (1992) "Curriculum differentiation: opportunities, outcomes, and meanings," in P.W. Jackson (ed.) *Handbook of research on curriculum*, New York: Macmillan, pp. 570–608.

—— and Wells, A.S. (1998) "Detracking for high student achievement," *Educational Leadership* 55(6): 38–41.

OECD (Organization for Economic Cooperation and Development) (2002) *Education policy analysis*, Paris: OECD.

Ono, H. (2001) "Who goes to college? Features of institutional tracking in Japanese higher education," *American Journal of Education* 109: 161–195.

Padilla, A.M. and Gonzales, R. (2001) "Academic performance of immigrant and U.S.-born Mexican heritage students: effects of schooling in Mexico and bilingual/English language instruction," *American Educational Research Journal* 38: 727–742.

Park, H. (2009) "Growing curriculum differentiation and its implications for inequality in Korea," Paper presented at the Annual Meeting of the Association of Asian Studies, Chicago, March.

Paul, F.G. (2005) "Grouping within Algebra I: a structural sieve with powerful effects for low-income, minority, and immigrant students," *Educational Policy* 19: 262–282.

Powell, A.G., Farrar, E. and Cohen, D.K. (1985) *The shopping mall high school*, Boston, MA: Houghton Mifflin.

Rosenbaum, J.E. (1999) "If tracking is bad, is detracking better?" *American Educator* 47: 24–29.

Rowan, B. and Miracle, A.W. Jr. (1983) "Systems of ability grouping and the stratification of achievement in elementary schools," *Sociology of Education* 56: 133–144.

Rubin, B.C. (2003) "Unpacking detracking: when progressive pedagogy meets students' social worlds," *American Educational Research Journal* 40: 539–573.

—— (2008) "Detracking in context: how local constructions of ability complicate equity-geared reform," *Teachers College Record* 110: 646–699.

Sandholtz, J.H., Ogawa, R.T. and Scribner, S.P. (2004) "Standards gap: unintended consequences of local standards-based reform," *Teachers College Record* 106: 1177–1202.

Slavin, R.E. (1987) "Ability grouping and achievement in elementary schools: a best-evidence synthesis," *Review of Educational Research* 57: 293–336.

—— (1990) "Achievement effects of ability grouping in secondary schools: a best-evidence synthesis," *Review of Educational Research* 60: 471–499.

Sørensen, A.B. (1970) "Organizational differentiation of students and educational opportunity," *Sociology of Education* 43: 355–376.

Stanley, G. and MacCann, R.G. (2005) "Removing incentives for 'dumbing down' through curriculum re-structure and additional study time," *Educational Policy Analysis Archives* 13: 1–10.

Tach, L.M. and Farkas, G. (2006) "Learning-related behaviors, cognitive skills, and ability grouping when school begins," *Social Science Research* 35: 1048–1079.

Tomlinson, C.A., Brighton, C., Hertberg, H., Callahan, C.M., Moon, T.R., Brimijoin, K., Conover, L.A. and Reynolds, T. (2003) "Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: a review of literature," *Journal for the Education of the Gifted* 27: 119–145.

Tsuneyoshi, R. (2004) "The new Japanese educational reforms and the achievement 'crisis' debates," *Educational Policy* 18: 364–394.

Van de gaer, E., Pustjens, H., Van Damme, J. and De Munter, A. (2006) "Tracking and the effects of school-related attitudes on the language achievement of boys and girls," *British Journal of Sociology of Education* 27: 293–309.

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

Watanabe, M. (2007) "Lessons from a teacher inquiry group about tracking: perceived student choice in course-taking and its implications for detracking reform," *Teachers College Record* 109: 2136–2170.

—— (2008) "Tracking in the era of high-stakes state accountability reform: case studies of classroom instruction in North Carolina," *Teachers College Record* 110: 489–533.

Wells, A.S. and Serna, I. (1996) "The politics of culture: understanding local political resistance to detracking in racially mixed schools," *Harvard Educational Review* 66: 93–118.

Welner, K.G. (2001) *Legal rights, local wrongs: when community control collides with educational equality*, New York: SUNY Press.

Yonezawa, S., Wells, A.S. and Serna, I. (2002) "Choosing tracks: 'Freedom of choice' in detracked schools," *American Educational Research Journal* 39: 37–67.