When teachers divide students into separate classes or groups on the basis of prior performance, they do so because they think students are best served by receiving instruction targeted to their particular levels of accomplishment up to that point. Consider the case of first grade reading: Some children enter school without knowing the alphabet, others are familiar with the letter sounds, and still others are already strong readers. To accommodate these differences, teachers typically divide students into reading groups. For another example, consider ninth grade mathematics: Some students enter high school with eighth grade algebra under their belts, others have yet to master arithmetic, and many others are in-between. In response, high schools commonly divide students for ninth grade mathematics into general math, pre-algebra, algebra, and geometry classes. To most teachers, these divisions make sense. They make it possible to think about instruction as organized in a clear sequence, to find each student’s place in the sequence using criteria they consider objective, and to provide instruction intended to move each student along the instructional hierarchy. In short, ability grouping seems like a neutral device for matching instruction to students’ needs.

Despite this sensible logic, there are three reasons why ability grouping cannot be viewed as neutral. First, it leads to divisions that go beyond academic performance.
demic differences. When teachers divide students on the basis of academic performance, they tend to separate students who differ from one another by race, ethnicity, and social class. These inequalities—that have little to do with school—are perpetuated as students move up the ladder of grade levels (Oakes, Gamoran, & Page, 1992). Students from disadvantaged backgrounds tend to score lower on tests for a host of reasons, many of which are unrelated to schooling, and therefore the division of students on the basis of academic performance results in social as well as academic segregation. Second, when teachers create classes that are relatively homogeneous in student performance, they eliminate much of the diversity that might foster rich and productive conversations in classrooms. Although grouping students by performance level may make it possible to sharpen the delivery of instruction to meet students’ levels of skills and knowledge, that sharpening may be double-edged, as it eliminates the very differences that some teachers build on in their instruction. Third, although teachers may intend to provide instruction of equal quality at all levels, in practice, that rarely occurs. Instead, students in lower-ranked classes and groups encounter instruction of lower quality, compared to their peers in higher-ranked classes. Consequently, instead of helping low-achieving students catch up, ability grouping tends to result in wider and wider achievement gaps over time (Gamoran & Berends, 1987; Oakes et al., 1992).

Although we can clearly identify the problems of grouping and tracking, that does not make eliminating these practices an easy solution. On the contrary, detracking is also associated with problems of instructional quality, and successful detracking is rare. Consequently, after we examine the challenges to instruction associated with dividing students by performance level, it will be equally important to consider the instructional challenges associated with mixed-ability grouping.

**TRACKING AND INSTRUCTION**

James Rosenbaum has studied both tracking and detracking. His case study of a working-class high school in the 1970s was one of the first to show how tracking promoted inequality in schools (Rosenbaum, 1976). Courses in which students enrolled—as early as junior high school—often dictated the opportunities they encountered throughout high school. Tracking was like a “tournament”: Students moved down to lower track levels over time, but rarely moved up. Students in high tracks had better academic opportunities than their low-track counterparts, as teachers in high-track classes were more enthusiastic and better prepared for teaching. In the 1980s, Jeannie Oakes (1985) used data from a national study
of secondary schools to confirm and extend Rosenbaum’s case study findings. She found that teachers in low-track classes spent more time managing students’ behavior, whereas high-track teachers could devote relatively more time to instruction. This instructional difference corresponded to differences in student behavior: rates of off-task behavior were higher in low-track classes, although the magnitude of the difference seems small: 4% compared to 1% off-task in low- and high-track mathematics classes, respectively, and 4% compared to 2% of students off-task in low- and high-track English classes, respectively (p.101).

Oakes also learned, from observations of classes and interviews with teachers and students, that the content of high-track classes was more academically rigorous, compared to low-track classes. For example, students were more likely to write essays in high-track than in low-track English classes, and more likely to engage in problem solving in high-track than in low-track mathematics classes. Students in high-track classes had more exposure to high-status content, such as classic and modern literature in English, as opposed to juvenile fiction which was more common in low tracks, and advanced concepts in mathematics as opposed to a focus on facts and computation. When students were asked about the most important thing they learned during the year, a response that typified a high-track science class was, “We have learned the basics of the laws of relativity, and basics in electronics. The teacher applies these lessons to practical situations;” whereas a student in a low-track responded, “I can distinguish one type rock from another” (Oakes, 1985, p. 69, 71).

**The Link to Student Achievement**

Although the instructional differences are pervasive, they are relatively small, and Oakes (1985) acknowledged that instruction in all classes emphasized teacher lectures and student recitation and seatwork:

The most significant thing we found is that generally our entire sample of classes turned out to be pretty noninvolving places. As we expected, passive activities—listening to the teacher, writing answers to questions, and taking tests—were dominant at all track levels. And, also not unexpected, the opportunities students had in any group of classes to answer open-ended questions to work in cooperative learning groups, to direct the classroom activity, or to make decisions about what happened in class were extremely limited. In most classes these things just did not happen at all. Any statements that can be made about differences between tracks in this respect must be seen in this context. (p. 129)

If instructional differences are small but persistent, are they large enough to affect student achievement? Research in the 1990s by Gamoran and his
colleagues provided quantitative evidence of the connection between unequal instructional opportunities and unequal achievement across tracks. In a study of 25 Midwestern middle and high schools, the authors noted that differences in the quality of instruction in English classes accounted for much of the widening gap in student achievement on a literature test (Gamoran, Nystrand, Berends, & LePore, 1995). Students in low-track classes were more often off-task, less consistent in completing their assignments, and spent less time in class discussion, compared with their high-track counterparts, and these differences accounted for part of the achievement gap. Teachers in high- and low-track classes were about equally likely to ask open-ended questions, but in high-track classes, the open-ended questions focused on the literature students were studying, and this provided an additional advantage on the literature test. In low-track classes, teacher questions about literature generally focused on reporting plot summaries rather than on more interpretive concerns.

Other quantitative studies, ranging from elementary to high school, support the contention that instructional differences across groups and tracks contribute to achievement differences. At the high school level, students in college-preparatory programs enroll in more academic courses, and particularly more advanced mathematics and science courses, and this contributes to their achievement advantages in those subjects (Gamoran, 1987). In elementary school, students in higher-ranked reading groups cover more new words and read more stories over the course of a school year, making the reading gap between high- and low-ranked groups wider at the end of the year than it was at the beginning of the year (Gamoran, 1986; Rowan & Miracle, 1983). Two students who start the school year at similar reading levels, but who are assigned to different reading groups, end up with different reading achievement at the end of the year depending on whether they were assigned to a higher or a lower group. Thus, although providing different instruction to different groups seems like it would help low-achieving students catch up, usually that is not what happens; instead they fall further and further behind.

Can Tracking Help Instead of Harm Low Achievers?

Not all uses of grouping and tracking are as damaging to the prospects of low-achieving students as typically occurs. Catholic high schools, for example, produce less inequality between tracks than do public high schools (Bryk, Lee, & Holland, 1993; Gamoran, 1992). This occurs because Catholic schools place more academic demands on students who are not enrolled in the college-preparatory program, than do public high schools. Case studies also suggests that a school climate of effort and car-
In the study of English classes in 25 middle and high schools, Gamoran found only two schools in which low-track classes were successful in preventing their students from falling further behind (Gamoran, 1993). Both happened to be Catholic schools; one was a high school and one was a junior high. Gamoran uncovered three commonalities among the two cases of relatively successful low-track classes. First, the schools had no system of assigning weak or inexperienced teachers to low-track classes. This contrasted with other research showing that teachers compete for the privilege of teaching the honors class (Finley, 1984). In these two schools, the same teachers taught high- and low-level classes. Second, the teachers worked hard to conduct oral discourse with students, instead of relying on worksheets to control student behavior as often occurs in low tracks. Third, the teachers exhibited high expectations for student performance, manifested in their refusal to relinquish the academic curriculum. As one teacher exhorted her students, “I know it’s not easy, you guys, I know it’s not easy, but we’re not going to read a Weekly Reader in this class. All right. You deserve to have this information. So stick with it” (Gamoran, 1993, p. 15).

A study of highly restructured schools also revealed a case in which high standards helped prevent divisions among students from resulting in instructional inequality (Gamoran & Weinstein, 1998). At Red Lake Middle School, classroom observers found high-quality instruction in both “regular” eighth grade mathematics and in the high-track algebra classes. At both levels, instruction was characterized by rigorous content, higher-order thinking, and serious discussions of academic material. One of the two algebra classes ranked at the highest level on these criteria of any of the 72 mathematics classes that were observed across 24 restructured schools. The other algebra class and the regular class both averaged higher on these criteria than the typical high-track class in the study. High-quality instruction at Red Lake was supported by a pervasive ideology of teaching to one’s “passions”—teachers were encouraged to develop courses and curricula that energized them, yet the courses were linked to rigorous disciplinary content. Within a differentiated structure, teachers at Red Lake focused on establishing high standards for all students, and by the year after the observations took place, all students were minimally prepared to enroll in algebra when they graduated from Red Lake and entered high school.
In another instance, Andrew Porter and his colleagues examined programs in New York and California designed to improve the quality of high school mathematics instruction for low-achieving, low-income youth (Gamoran, Porter, Smithson, & White, 1997; White, Gamoran, Smithson, & Porter, 1996; White, Porter, Gamoran, & Smithson, 1997). Students were still sorted into separate classes, but teachers attempted to provide instruction that would bridge the gap between elementary and college-preparatory mathematics, so that students could make a transition into college-preparatory courses. The "transition" courses were partially successful in meeting their goals. Students in the transition courses were more likely to complete a college-preparatory sequence than others who initially enrolled in general math (the traditional low-track class). However, students had the best chance of completing the college-preparatory sequence if they skipped the transition course and enrolled immediately in the college-preparatory track. Moreover, student achievement growth in the transition courses fell in between that of general math and that of college-bound courses such as algebra. More rigorous course content accounted for the achievement benefits of the transition courses over the general math classes that they were replacing.

These cases show that, although it is not what usually happens, it may be possible to group students in a way that promotes equity instead of inequity. Maintaining high standards seems a key to success: Providing a rigorous curriculum, communicating expectations, teaching with passion, and avoiding a system in which less experienced teachers are relegated to lower-level classes all played important roles in these rare success stories. Evidence from outside the United States also supports the notion that high standards are essential for using grouping more effectively. Hanna Ayalon and Adam Gamoran (2000) compared course work divisions within academic programs in the United States and Israel. In American high schools, the more students differed from one another in their mathematics course taking, the more inequality in achievement among students from different social backgrounds, and the lower achievement overall. In Israel, by contrast, course work differences were associated with no drop in average achievement, and with less inequality by social background. The difference between the two systems is that in the United States, where students in low-track classes have little incentive to do more than receive a passing grade, course content tends to be diluted and students exert minimal effort. In Israel, however, students at all levels of academic courses are preparing for important national examinations that occur at the end of high school. Schools that offer different levels of such courses (which correspond to different levels on the examinations) give low-achieving students better chances for success, while maintaining
strong incentives for both teachers and students to cover meaningful course material.

DETRACKING AND INSTRUCTION

Even the most successful uses of grouping still encounter the problem of separating students of different social backgrounds. Social and economic inequalities outside schools contribute to substantial differences in test scores inside schools, so when educators divide students by achievement level, the result is classes that differ by social background. For this reason, many educators would prefer to avoid the practice altogether rather than trying to use grouping more effectively.

However, detracking offers its own set of challenges to those who wish to provide high-quality instruction. Although tracking often results in poor instruction for low achievers, it also tends to sustain high-quality instruction for high achievers. Thus, efforts to detrack seem to confront the classic tension between excellence and equity. Can this tension be surmounted? Ultimately, it may come down to a question of values: Is it worth sacrificing some opportunities for the highest achievers for the sake of more equitable opportunities for all students? Research can still contribute much to resolving this dilemma by showing what the tradeoffs are and what it might take to provide equitable opportunities without sacrificing high standards of excellence. Can we provide the same high-quality instruction in mixed-ability classes that is now typical in high-track classes?

Successful detracking is rare, and Oakes (1992) has explained that three conditions make detracking difficult:

- **Normative barriers:** Widely shared beliefs hold that people differ from one another, and that it is appropriate to create educational structures that correspond to those differences;
- **Political barriers:** Certain individuals and groups have vested interests in maintaining the current tracking system, such as parents of children who would be placed in honors classes, and teachers who seek the opportunity to teach honors classes;
- **Technical barriers:** Few curricula and teaching methods are currently available that have been designed to meet the challenge of teaching students at widely varying levels of academic performance.

Although much of the research literature focuses on the first two barriers, the third barrier is also formidable. Many teachers are reluctant to teach
mixed-ability classes, fearing that weak students will not be able to keep up, or that strong students will be held back, or both (Loveless, 1998; White et al., 1996). Indeed, it seems likely that if the technical barriers to detracking could be overcome, the normative and political barriers would begin to weaken.

Resistance to detracking among teachers seems strongest in subjects where they perceive the curriculum as rigidly sequential, so that students must master one topic before they proceed to the next. Mathematics and foreign languages are examples of such subject areas (Gamoran & Weinstein, 1998; Loveless, 1994). It is difficult to know whether these perceptions are inherent in the subject matter, or if they reflect ingrained beliefs that might be successfully challenged if teachers could be shown that high-quality instruction in mixed-ability classes is possible.

Instructional Problems Associated with Detracking

What is the evidence about detracking and instruction? Most studies of detracking are unable to answer this question because they do not look inside classrooms. Instead, they focus on the political battles that embroil detracking reformers (Oakes & Wells, 1998; Oakes, Wells, Yonezawa, & Ray, 1997; Wells & Serna, 1996). Indeed, most detracking reforms are not fully implemented, so their implications for classroom activities have been limited.

A small number of cases of fully detracked schools provide evidence about instruction. Rosenbaum’s recent study of detracking in a high school social studies department raises important questions about the viability of detracking to create high-quality instructional opportunities for all students (1999). In social studies, one might think that increasing the academic and social diversity of the classroom would increase the richness of instructional activities because teachers could capitalize on differences among students to create varied roles, topics, and activities that would facilitate dialogue and debate. However, high school social studies relies heavily on reading and writing assignments, and the teachers in this school struggled to find the right level of demands. As one teacher explained:

Piquing the interest of the brighter kids would require extra readings, extra writing assignments, and extra discussions that we would have to schedule outside of class. It’s too hard to do all of this. I really don’t do enough for them. There’s not enough time. (Rosenbaum, 1999, p. 26)

High-achieving students were, perhaps understandably, resistant to the notion of doing more work for no additional reward. At the same time,
teachers felt that weaker students were struggling to keep up with the reading and writing assignments.

Because tracking generally is disproportionately harmful to minority students, it is paradoxical that high-achieving minority students are among those who may have most to lose from detracking. In the school Rosenbaum studied, a group of high-achieving Hispanic students was “slowed down and bored, just like the other bright kids,” according to one of the teachers, and in the detracked classroom, “the brighter Hispanic students seem to face a lot of peer pressure in the class from other Hispanic kids who aren’t doing well in the class” (Rosenbaum, 1999, p. 27). High-achieving minority students are especially vulnerable to changes that reduce academic demands. According to Rosenbaum (1999),

These minority students come from working- and lower-class families; their parents do not have strong educational backgrounds. If these students do not find academic challenge at school, they may not find it at all. (p. 27)

In the study of highly restructured schools discussed earlier, three high schools were successful in implementing a structure that did not have any tracks or ability groups (Gamoran & Weinstein, 1998). Analysis of these cases revealed that two of them, Wallingford and Marble Canyon, experienced instructional problems similar to the ones Rosenbaum reported. In sharp contrast, the third case, Cibola High, provided evidence of high-quality instruction in mixed-ability classes.

Wallingford High School is located in an eastern city and all of its students come from minority backgrounds. As part of a restructuring program, all divisions of students within grades and subjects were eliminated. Thus, all students were enrolled in the same ninth grade mathematics course. Teachers reported a variety of problems with this approach. According to Gamoran and Weinstein (1998), “two teachers acknowledged that they had lowered their standards for heterogeneous classes, and one said he had given up trying to cover all the intended material. This resulted in students being promoted without being prepared for the next math course” (p. 393). Consistent with these difficulties, observers reported that the quality of instruction was poor:

Almost none of the observed classes exhibited more than a minimal amount of thoughtfulness and depth. In math, students’ tasks consisted mainly of applying algorithms to routine computations. We did not observe students discussing problems with one another; all the lessons used traditional lecture formats. (p. 393)

Similarly, at another urban high school in the east, teachers held low standards for student performance. This school, Marble Canyon, served
students who came from a wider range of ethnic backgrounds than Wallingford. Teachers focused on getting students to attend and remain in school, and did not make rigorous academic demands. One teacher explained that, “Because we’re aiming at keeping at-risk kids in school, the more motivated kids aren’t being stretched enough” (Gamoran & Weinstein, 1998, p. 393).

Detracking and High-Quality Instruction

In contrast to the problems associated with detracking at Wallingford and Marble Canyon High Schools, observers reported extraordinarily high levels of instructional depth, thoughtfulness, and substantive conversation at Cibola High School, the third detracked high school (Gamoran & Weinstein, 1998). Cibola is also located in the east, and it serves an ethnically and economically diverse population: about half its students come from minority backgrounds, and about half are on free or reduced-price lunch (but the minority students are not necessarily the ones on free lunch). Observers rated the quality of instruction at Cibola as the highest of any of the highly restructured schools. How was this accomplished in a context of heterogeneous classes? The finding held in both mathematics and social studies, and it was particularly notable in mathematics where teachers in other schools most strongly resisted detracking. The absence of tracking seemed to enhance learning opportunities in this school, in contrast to other detracked schools in which classes have been observed. What made the difference?

Gamoran and Weinstein (1998) noted several conditions in Cibola High School which supported high-quality instruction in mixed-ability classes. First, the school had a visionary leader who was able to select her staff and, to some degree, her students. Students had to apply to be admitted and those who did not come from the feeder elementary school were interviewed before being admitted. Choosing staff and students made it possible to establish a distinctive school culture that supported instructional innovation. At Cibola, mathematics and science instruction were integrated, and students worked in small, heterogeneous groups on long-term projects that involved mathematical and scientific principles. At the time the researchers observed instruction, students were working on plans for designing an amusement park. Teachers rejected the notion that mathematics had to be taught in a strict sequence. They did not ignore elementary mathematics, but neither did prevent a student whose arithmetic was weak from working on algebraic problems. One teacher told the students,
If you need work on your basic math skills, I don’t slow down for you. You’re supposed to be putting in extra time. We have somebody here on Saturdays. You can stay after school. On Saturday the library is open. There are two teachers and a resource teacher there to tutor. (p. 402)

On the one hand, the teachers’ approach of setting forth complex problems and not slowing down meant that all students were challenged. On the other hand, students with academic deficiencies were well supported. Classes were small—limited to 20 students. In addition, the school had won outside grants that supported a Saturday tutoring program. To graduate from Cibola, students had to prepare “portfolios” in 14 subjects and submit 3 of them to be judged by a panel of school insiders and outsiders. Although all students were required to demonstrate competencies through the portfolios, expectations were also somewhat differentiated according to the capacities of the student. Interviews with educators revealed that “grading is highly individualized … in a context of serious academic content” (p. 403). Thus, the culture of Cibola supported the approach that the teachers Rosenbaum studied were unable to implement: more work for the highest-achieving students, and differential assessment of student performance, though always with a high minimum threshold.

Perhaps the greatest instructional challenge for detracking is teaching secondary school mathematics to students who differ widely in their math skills. Teachers at Cibola were successful in this aim because they rejected the notion that mathematical knowledge has a rigid sequence. Can this approach be adopted more widely? For the past decade, the College Board’s Equity 2000 program has attempted to enroll all students in algebra (or a more advanced course) as soon as they enter high school. In their approach, algebraic concepts are introduced throughout the elementary years to increase students’ readiness for an algebra course. Working with six pilot sites, the College Board has reported increases not only in the proportion of students enrolling in college-preparatory courses, but increases in rates of passing courses, even when the proportion of the cohort enrolled has increased. For example in 1991, 31% of ninth graders were enrolled in algebra (or higher), and only 25% passed the course. By 1997, 99% were enrolled and 55% passed (College Board, 1999). In one sense that is a remarkable success, but at the same time almost half the ninth graders are not passing their first mathematics class.

An important question about Equity 2000 is whether the rigor of instruction in algebra, and thus the benefits of algebra for student learning, are maintained when detracking occurs. This question has not been addressed by the program, but an analysis of national survey data has indicated that all students benefit from taking algebra, and such benefits
are not reduced when the population of algebra-takers in a school is more heterogeneous than when it is more homogeneous (Gamoran & Hannigan, 2000).

One reason teachers at Wallingford and Marble Canyon taught a diluted curriculum to mixed-ability classes is that they did not know what else to offer besides the standard college-preparatory curriculum. This is precisely the technical problem: teachers lack pedagogical strategies and curricula designed for mixed-ability classes—nowhere is this more true than in secondary-school mathematics—and most do not believe such a curriculum is possible. One group of curriculum developers has responded to that challenge by creating the “Interactive Mathematics Program” (IMP), a curriculum that tries to build on the differences among students in the course of problem-oriented, hands-on problems that center on core mathematical issues (Alper, Fendel, Fraser, & Resek, 1997). Small-scale evaluations of IMP are promising, showing that students gained more on average in achievement from IMP than from the traditional course sequence, and high-achieving students gained no less than in traditional classes (Webb, 2003). IMP, along with Equity 2000 and the experience of Cibola High School, suggest that solutions to the technical problems of detracking are possible, even in the most challenging subject areas.

**IMPLICATIONS FOR RESEARCH, POLICY AND PRACTICE**

Both practitioners and researchers can respond to the research findings about classroom organization and instructional quality. The first reaction from practitioners may be to strengthen their inclinations to reduce reliance on grouping and tracking because these practices are associated with unequal classroom instruction and ultimately with unequal achievement. The research is clear that some forms of tracking should be eliminated because better alternatives are available. For example, the research points towards elimination of general math classes in high school. This course simply repeats the arithmetic curriculum of elementary and junior high school and is a dead end for students, blocking access to college-bound curricula and college enrollment. At the elementary level, the research also indicates that the practice of rigid tracking of students in elementary school for the entire school day on the basis of a sole criterion should cease.

Second, practitioners can consider whether to maintain less extreme versions of tracking, or to eliminate all divisions among students. The research is inconclusive as to which alternative is better, and practitioners must consider their own unique circumstances in deciding which
approach best fits their school. One view holds that, as has occurred in some Catholic schools and in a restructured public school, some divisions among students for particular subjects are appropriate, as long as teachers hold students at all levels to high standards of accomplishment. Another view holds that all such divisions should be eliminated. To adopt this approach, it will be necessary to develop curricula and pedagogies that are suited to mixed-ability classes, since the cases of Wallingford and Marble Canyon illustrate the futility of simply following a traditional course with students of widely varying performance levels. In these cases, teachers taught to a lowest common denominator and were concerned that strong students were not challenged. Cibola High School offers a model for successful detracking, but its program was supported by special circumstances including selection of staff and some students, as well as a tutoring program for students who were falling behind.

The lack of conclusive evidence on alternatives to traditional tracking structures also shows where researchers must direct their attention. It is most essential to examine a broader range of schools engaged in responding to the tracking problem so that we can move beyond “existence proofs” to a more generalizable conclusion about the advantages and disadvantages of each policy choice.

REFERENCES


