Tracking students into different classrooms according to their prior academic performance is controversial among both scholars and policymakers. If teachers find it easier to teach a homogeneous group of students, tracking could enhance school effectiveness and raise test scores of both low- and high-ability students. But if students benefit from learning with higher-achieving peers, tracking could disadvantage lower-achieving students, thereby exacerbating inequality.

 Debates over tracking reached their high point in the United States in the 1990s. An influential report published in 1998 by the Thomas B. Fordham Foundation argued that the available research did not support the contention that tracking doomed impoverished students to inferior schooling, nor did it support universal adoption of the practice. Over the last decade, patterns in grouping students have changed markedly in the U.S.; high school students are no longer placed in rigidly defined general-education or noncollege tracks but have the flexibility to move between course levels for different subjects. These changes may have assuaged some critics, but the broader debate over tracking remains unsettled.

The central challenge in measuring the effect of tracking on performance is that schools that track students may be different in many respects from schools that do not. For example, they may attract a different pool of students and possibly a different pool of teachers. The ideal situation to assess the impact of tracking on test scores of different groups of students would be one in which students were assigned to tracking or nontracking schools randomly, and the performance of students could be compared across school types.
We shed light on these issues using data from Kenya. In 2005, each of 140 primary schools in western Kenya received funds from the nongovernmental organization International Child Support (ICS) Africa to hire an extra teacher. One hundred twenty-one of these schools had a single 1st-grade class and used the new teacher to split the students into two classes. In 61 randomly selected schools, students were assigned to classes based on prior achievement as measured by test scores. In the remaining 60 schools, students were randomly assigned to one of the two classes, without regard to their prior academic performance.

The results showed that all students benefited from tracking, including those who started out with low, average, and high achievement. At the tracking schools, the test scores of students who started out in the middle of their class do not seem to be affected by which section (top or bottom) the students were later assigned to. In other words, any negative effects of being with lower-achieving peers were more than offset in tracked settings by the benefit of the teacher being able to better tailor instruction to students’ needs.
Primary Education in Kenya
The Kenyan education system includes eight years of primary school and four years of secondary school. Like many other developing countries, Kenya has recently made rapid progress toward the goal of universal primary education. After the elimination of school fees in 2003, primary school enrollment rose nearly 30 percent, from 5.9 million in 2002 to 7.6 million in 2005. This is typical of what is happening in sub-Saharan Africa overall, where the number of new entrants to primary school increased by more than 30 percent between 1999 and 2004.

This progress creates its own new challenges, however. Pupil-teacher ratios have grown dramatically, particularly in lower grades. In our sample of schools in western Kenya, the median 1st-grade class in 2005 (after the introduction of free primary education, but before the class-size-reduction program we study here) had 74 students and the average class size was 83. These classes are heterogeneous in a number of ways: Students differ vastly in age, school readiness, and support at home. Many of the new students are first-generation learners and have not attended preschools, which are neither free nor compulsory in Kenya. These challenges are not unique to Kenya; they confront many developing countries where school enrollment has risen sharply in recent years. Understanding the roles of tracking and peer effects in this type of environment is thus critically important.

Our results are most likely to be directly applicable to settings where classes are large, the student population is heterogeneous, and few additional resources are available to teachers. It is unclear whether similar results would be obtained in different contexts, such as developed countries, where smaller class sizes may allow more tailored instruction even without tracking, and extra resources, such as remedial education, computer-assisted learning, and special education programs, may already provide tools to help teachers deal with different types of students.

Design of the Experiment
This study takes advantage of a class-size-reduction program and evaluation that involved primary schools in Bungoma and Butere-Mumias in Western Province, Kenya. Of 210 primary schools in these districts, 140 schools were randomly selected to participate in the Extra-Teacher Program. With funding from the World Bank, ICS Africa provided each of the 140 selected schools with funds to hire an additional 1st-grade teacher on a contractual basis starting in May 2005, the beginning of the second term of that school year. Most of the schools (121) had only one 1st-grade class, which was split into two classes when the new teacher was hired. The 19 schools that already had two or more 1st-grade classes added another class.

It is important to note that the incentives facing the newly hired teachers differed from those facing civil-service teachers already working in program schools. The new teachers had clear incentives to work hard to increase their chances of having their short-term contracts renewed and of eventually being hired as civil-service teachers—a desirable outcome in a society where government jobs are highly valued. In contrast, the difficulty of firing civil-service teachers implies that they had weak extrinsic incentives and may be more sensitive to factors affecting their intrinsic motivation.

Average class size was reduced from 84 to 46 students in the 140 schools that received funds for a new teacher. The program continued for 18 months, which included the last two terms of 2005 and the entire 2006 school year, and the same cohort of students remained enrolled in the program.

From the 121 schools that had originally only one 1st-grade class, 60 schools were randomly selected to assign students to one of the two classes by chance. We call these schools the “nontracking schools.” In the remaining 61 schools (the “tracking schools”), the children were divided into two sections according to their scores on exams administered by the school during the first term of the 2005 school year. The 50 percent of the class with the lowest exam scores were assigned to one section (the “bottom class”) and the rest were assigned to the other (the “top class”).

After students were assigned to classes, the contract teacher and the civil-service teacher were also randomly assigned to classes. In the second year of the program, all children not repeating the grade remained assigned to the same group of peers and the same teacher.

Data
Our initial sample consists of approximately 10,000 students enrolled in 1st grade in March 2005 in one of the 121 primary schools participating in the study. The outcome of interest is student academic achievement, as measured by scores on a standardized math and language test first administered in all schools 18 months after the start of the program. Trained proctors administered the test, which was then graded blindly by data processors. In each school, 60 students (30 per class) were drawn from the initial sample to participate in the tests. If a class had more than 30 students, students were randomly sampled.

The test was designed by a cognitive psychologist to measure a range of skills students may master by the end of 2nd grade. One part of the test was written and the other part oral, administered one-on-one. Students answered math and literacy questions ranging from counting and identifying letters to subtracting three-digit numbers and reading and understanding sentences.

To limit attrition from the experiment, proctors were instructed to go to the homes of sampled students who had
dropped out or were absent on the day of the test and to bring them to school for the test. It was not always possible to find the child, however, and the resulting attrition rate on the test was 18 percent. However, there was no difference between tracking and nontracking schools in overall attrition rates. In total, we have postintervention test-score data for 5,796 students.

In addition, each school received unannounced visits several times during the course of the study. During these visits enumerators checked, upon arrival, whether teachers were present in school and whether they were in class and teaching, and then took a roll call of the students.

To measure whether the effects of the program persisted, the children who had been sampled for the first postintervention test were tested again in November 2007, one year after the program ended. During the 2007 school year, these students were overwhelmingly enrolled in grades for which their school had a single class, so tracking was no longer an option. Most of these students had reached 3rd grade by that time, but those repeating an earlier grade were also tested. The attrition rate for this portion of the experiment was 22 percent. Neither the proportion nor the characteristics of children who could not be tested differed between the tracking and nontracking schools.

The Impact of Tracking
We estimate the impact of tracking on student achievement by comparing the postintervention (18 months after the experiment began) test scores of students in the tracking and nontracking schools. Taking the average of students’ scores on math and literacy exams, we find that students in tracking schools scored 0.14 standard deviations higher than students in nontracking schools overall. When we adjust the comparison to take into account minor differences in student characteristics across the two groups of schools, the effect increases to 0.18 standard deviations. There was no significant difference between the impact of the program on math and literacy scores when we examined the subjects separately.

How large were these effects? A typical student with a literacy score one standard deviation above the average could correctly spell 5.5 of 10 words included on the exam, while the average student could spell only 2. Similarly, students with a math score one standard deviation above the average were able to perform single-digit multiplications, whereas those at the mean could not. The average effect of tracking was roughly one-fifth the size of these performance differences.

These gains persisted beyond the duration of the program (see Figure 1). When the program ended, most students had reached 3rd grade, and all but five schools had only one 3rd-grade class. The remaining students had repeated and were in 2nd grade where, once again, most schools had only one large class, since after the program ended they did not have funds for additional teachers. Even so, the test scores of students in tracking schools remained 0.16 standard deviations higher than those of students in nontracking schools overall (and 0.18 standard deviations higher with control variables). The persistence of the benefits of tracking is striking, as many evaluations find that the test-score effects of successful interventions fade over time. It seems that tracking helped students master core skills in 1st and 2nd grade that in turn helped improve their learning later on.

We also examine whether the effect of tracking differs between initially high-scoring students (who are grouped with other strong students in tracking schools) and initially low-scoring students (who are grouped with other low-scoring students

The persistence of the benefits of tracking is striking, as many evaluations find that the test-score effects of successful interventions fade over time. It seems that tracking helped students master core skills in 1st and 2nd grade that in turn helped improve their learning later on.
in tracking schools). We find that both groups of students benefited from tracking, and by approximately the same amount. A year after the intervention ended, the effect persisted for both the top and bottom classes.

Tracking increases test scores for students taught by contract teachers. In fact, students initially scoring low who were assigned to contract teachers benefited even more from tracking than students who initially scored high. But students who initially scored low showed only a small and statistically insignificant benefit if assigned to a civil-service teacher. In contrast, tracking substantially increased scores for students who initially scored high and were assigned to a civil-service teacher. Below we discuss other evidence that tracking led civil-service teachers to increase effort when they were assigned to high-scoring students but not when assigned to low-scoring students.

Changes in Peer Achievement
Data from the tracking schools allow us to estimate the effect of being taught with a higher-achieving vs. lower-achieving peer group by comparing students with baseline test scores in the middle of the distribution. Because of the way tracking was done (splitting the grade into two classes at the median baseline test score), the two students closest to the median within each school were assigned to classes where the average prior achievement of their classmates was very different.

By comparing pairs of students right around the cutoff, we can estimate the effect of being the lowest-achieving child in the class compared to being the highest-achieving student in the class. We find that, despite the large gap in average peer achievement (1.6 standard deviations in baseline test scores) between the top and bottom classes, the students just below the cutoff have postintervention test scores similar to students just above the cutoff. Moreover, when we compare students around the cutoff at the tracking schools with students of similar ability at the nontracking schools, we find that students at the tracking schools score higher at the end of the intervention than the comparable students in the nontracking schools. These results imply that being the best student in a class of relatively weak students and being the worst student in a class of relatively strong students are both better than being the middle student in a heterogeneous class. This evidence suggests that students benefit from homogeneity because the teacher does not need to spend time addressing the needs of students performing at widely varying levels.

When we compare students around the cutoff at the tracking schools with students of similar ability at the nontracking schools, we find that students at the tracking schools score higher at the end of the intervention than the comparable students in the nontracking schools.

Learning from Peers vs. Learning from Teachers
We took a separate look at students in schools where students were not tracked but instead assigned to classes randomly. The random assignment of students and teachers within these schools made it possible to see whether and how peer achievement affected the performance of individual students when education took place in an untracked setting. We found that it did. If peer achievement was higher—0.10 standard deviations higher, to be exact—students learned 0.04 standard deviations more than they would have otherwise.

These results, taken together with those reported earlier, indicate that peer influence depends on whether or not classes are tracked. In untracked classes, where there is considerable heterogeneity of performance, students learn less if their peers are lower performing. At least in this particular setting, however, the homogeneous classes that are created by tracking seem to allow the teacher to deliver instruction at a level that reaches all students, thus offsetting the effect of having lower-performing peers. Interestingly, combining the direct effect of peer achievement with the fact that
Why Did Tracking Work?

Two additional pieces of evidence shed light on the question of why tracking had such clear benefits. First, we look at teacher presence and effort. Do they spend more time in class and teaching? Then, we examine whether the test-score gains in tracking schools were concentrated among simpler or more complex tasks and whether this varied by students’ initial achievement levels. Our results confirm that students in tracked classes seem to have benefited from more-focused teaching and perhaps also from greater teacher effort.

Teacher absence is a major problem in Kenya, as in many developing countries. Only 59 percent of teachers were in class and teaching during unannounced visits to a comparable sample of schools that did not receive an additional teacher. Overall, teachers in tracking schools were 9.6 percentage points more likely to be found in school and teaching during random spot checks than their counterparts in nontracking schools, who were present and teaching only about half of the time. There were, however, large differences across teachers. The contract teachers were much more likely to be found in school and teaching (74 percent versus 45 percent for the civil-service teachers), and their absence rate was unaffected by tracking (see Figure 2). The civil-service teachers were 10 percentage points more likely to be in schools and teaching during tracking. This difference is statistically significant and amounts to a 25 percent increase in teaching time. However, the difference between tracking and nontracking school types was smaller and statistically insignificant for civil-service teachers assigned to the bottom classes.

These results suggest that teachers may be more motivated to teach a group of students with high initial scores than a group with low initial scores or a heterogeneous group. Recall that students assigned to the top class with a civil-service teacher benefited more from tracking than those assigned to the bottom class with a civil-service teacher. Increased teacher effort may help explain this pattern.

Another hypothesis consistent with both the tracking results and the effects from random peer assignment is that tracking by initial achievement improves student learning because it allows teachers to focus instruction. Teaching a more homogeneous group of students might allow teachers to adjust the material covered and the pace of instruction to students’ needs. For example, a teacher might begin with more basic material and instruct at a slower pace, providing more repetition and reinforcement, when students are initially less prepared. With a group of initially higher-achieving students, the teacher can increase the complexity of the tasks and pupils can learn at a faster pace. With a heterogeneous group, they may be compelled to cover both simple and advanced material, spending less time on each, which would hurt all students.

One way to examine this is to see whether children with different initial achievement levels gained from tracking.

Tracking Gains (Figure 1)

The benefits of tracking persisted even one year after the intervention ended and students returned to regular classrooms. Among students assigned to civil-service teachers, the gains from tracking were statistically significant only for students who had been assigned to the top class.
Our results show that grouping students by preparedness or prior achievement and focusing the teaching material at the most appropriate level could potentially have large positive effects with little or no additional resource cost.

differentially in terms of the difficulty of the material that they learned. While the results for language are mixed, the estimates for math suggest that, although the total effect of tracking on children in the bottom class is significantly positive for all levels of difficulty, these children gained from tracking more than other students on the easier questions and less on the more-difficult questions. Conversely, students assigned to the top class benefited less on the easier questions, and more on the more-difficult questions. In fact, they did not significantly benefit from tracking for the easier questions, but they did significantly benefit from it for the more-difficult questions. These results suggest that tracking helped by giving teachers the opportunity to focus on the competencies that children were not mastering.

Conclusion
A central challenge of education systems in developing countries—the context for which our results are most relevant—is that students in the same grades and classrooms are extremely diverse. Our results show that grouping students by preparedness or prior achievement and focusing the teaching material at the most appropriate level could potentially have large positive effects with little or no additional resource cost. One could also target more resources to the weaker group, further helping them to catch up with their more-advanced counterparts. It is often suggested that there is a trade-off between the value of targeting resources to weaker students, and the costs imposed on them by separating them from stronger students. We find no evidence for such a trade-off in this context.

Our results may also have implications for debates over school choice and voucher systems. A common criticism of such programs is that they may hurt some students if they lead to increased sorting of students by initial achievement and if all students benefit from having peers with higher initial achievement. If tracking is indeed beneficial, this is less of a concern.

Esther Duflo is professor of economics at the Massachusetts Institute of Technology. Pascaline Dupas is assistant professor of economics at University of California, Los Angeles. Michael Kremer is professor of economics at Harvard University.
EDUCATIONAL ENTREPRENEURSHIP
REALITIES, CHALLENGES, POSSIBILITIES
Harvard Education Press, 2006

Though entrepreneurs in K-12 education are gradually remaking the structure of American schooling, the topic has received surprisingly little serious or systematic attention. Most accounts of their work either celebrate successes or bemoan excesses, and seldom do observers stop to examine the challenges and opportunities in store. This collection of essays fills this gap, addressing a number of central questions: What is educational entrepreneurship? Who are the educational entrepreneurs and what motivates them? What policies or practices enable or impede entrepreneurship? A range of contributors offer pieces on the nature of educational entrepreneurship; the political, policy, and legal landscapes facing entrepreneurs; various models of entrepreneurial activity; the role of for-profit organizations; and possible future directions for educational entrepreneurs. The result is a lively, provocative book that introduces this expanding field to readers concerned with K-12 education—and with efforts to reform and improve it.

THE FUTURE OF EDUCATIONAL ENTREPRENEURSHIP
POSSIBILITIES FOR SCHOOL REFORM
Harvard Education Press, 2008

This volume examines the challenge of creating innovative and productive entrepreneurial activity in American education. In the course of exploring these challenges, the book considers a number of crucial issues and circumstances: existing “barriers to entry” that prohibit or obstruct entrepreneurial efforts; the availability—and frequent lack—of venture capital for fueling entrepreneurial activities; the effort to sponsor and create a sufficiently large population of talented educational entrepreneurs; and questions about research, development, and quality control in the burgeoning entrepreneurial sector. A field that is likely to grow in size and importance in the years to come, educational entrepreneurship receives much-needed attention, analysis, and elucidation in this lively, wide-ranging book.

EDUCATIONAL INNOVATION AND PHILADELPHIA’S SCHOOL OF THE FUTURE
Forthcoming from Harvard Education Press, 2010

Grand claims about the transformative power of technology in education are common, but decades of school redesign have yielded mixed results. One of the most widely acclaimed efforts has been the School of the Future in the Philadelphia School District, created in 2006 through a partnership with the Microsoft Corporation. This $63 million school incorporated new technologies in areas such as data and curriculum management, virtual simulations and mentoring, and personalized instruction intended to fundamentally change how students learn. A team of analysts will explore how this pioneering effort is unfolding on the ground and the lessons it holds for the use of technology and high school redesign. Research will examine key elements of the School of the Future—including its genesis, design, and flaws—and broader questions of educational innovation, public-private partnerships, and the integration of technology in the classroom.

Check out these and other publications by Rick Hess at www.aei.org/hess.