Determinants of Student Achievement: New Evidence from San Diego

• • •

Julian R. Betts
Andrew C. Zau
Lorien A. Rice

2003

PUBLIC POLICY INSTITUTE OF CALIFORNIA

Library of Congress Cataloging-in-Publication Data
Betts, Julian R.
Determinants of student achievement : new evidence from San Diego / Julian R. Betts, Andrew C. Zau, Lorien A. Rice.
p. cm.
Includes bibliographical references.
ISBN: 1-58213-044-2
1. Students—California—San Diego—Social conditions—20th century. 2. Academic achievement—Social aspects—California—San Diego. 3. Educational indicators—California—San Diego. 4. San Diego City Schools—Evaluation. I. Zau, Andrew. II. Rice, Lorien, 1968- III. Public Policy Institute of California. IV. Title.

LC205.5.C2B48 2003
371.8'09794'98—dc22 2003015087

Foreword

In the 2001 school year, the San Diego Unified School District (SDUSD) launched a program of reform known as the "Blueprint for Student Success." This ambitious and controversial reform calls for a districtwide intervention program to help students who are falling behind in their grade. The blueprint includes multiple interventions, including peer coaches, extended-length English classes, supplemental class options, reduced class sizes, summer school, and grade retention. Part of the controversy over the program lies in the broad sweep of its approach, both in the number of children affected and the number of options available. Many studies of student achievement in California have used state-level data. But few have used student-level data that link student performance to the resources available in the classroom. In 2000, PPIC entered into an agreement with SDUSD to provide the research and financial support necessary to format, collect, and analyze the student, teacher, and classroom data necessary to provide an accurate portrait of what affects student achievement in San Diego.

This report by Julian Betts, Andrew Zau, and Lorien Rice is the first product stemming from this collaboration. It examines in unprecedented detail the determinants of individual student gains in achievement in SDUSD between fall 1997 and spring 2000. This research also provides an important baseline against which to compare student achievement after the blueprint's implementation in fall 2000. Future PPIC reports will directly assess the effect of the blueprint by extending the database from spring 2000 to the date of the study then undertaken. These data will provide new insights into the detailed effects of the blueprint to all who are interested in the success of this program of reform.

This report provides some important new baseline findings about the interaction between students, their peers, and their teachers that will prove critical for understanding those effects. First, teacher education, credentials, experience, and subject authorization can make a difference

in student outcomes on tests, but the effects are neither as systematic nor as large as some may believe. Second, an individual student's rate of learning appears to be strongly and positively influenced by the initial achievement of students in his or her grade. And third, as has been found in earlier studies, the daunting achievement gaps between students do not appear to be created primarily by the schools as they now exist. Taking everything else into account, income and socioeconomic status still matter, and they matter a great deal.

PPIC has made a significant and long-term investment in working with SDUSD because it was clear from the beginning of this project that the blueprint presented a rare, if not unique, opportunity to look carefully at a major educational reform effort close up, without the usual critique that the reform is either too narrow or poorly implemented. Future reports will build on the current "pre-blueprint" work to provide an analysis of the reform effort itself. Second, it was also clear that the lack of systematic student-level data in California is a serious obstacle to an objective understanding of what is happening at the classroom level in schools throughout the state. K–12 education is the largest item in the state budget, but we have no simple statistic to tell us what works and what does not. Finally, our intent, in working with the district, is to ensure a strong, factual underpinning to the various debates that will doubtless emerge over school quality and the achievement gap in California's public schools. PPIC is committed to the collection and dissemination of such facts and to the hope that the presentation of the facts at the student, teacher, and classroom level will make a real contribution to the reform of education in California and the rest of the nation.

We are grateful to SDUSD for the opportunity to make this contribution and to the spirit of collaboration that has been present over the past three years. We are optimistic that these findings will help San Diego and other school districts throughout the state come to grips with the daunting challenges before them.

David W. Lyon
President and CEO
Public Policy Institute of California

Summary

Statewide surveys by the Public Policy Institute of California (PPIC) have consistently shown that the quality of California's schools tops the list of the public's concerns about problems in California. These surveys have shown that the public is particularly uneasy about the lack of fully qualified teachers in the state's schools. The public wants to see California schools improve and is aware of inequalities across state schools in both student achievement and the resources put into each school.

In light of public concerns about the quality of K–12 education, this report uses a detailed database from San Diego Unified School District (SDUSD) to pursue three goals: to examine the nature of school resource inequalities, to explore trends in student achievement with a focus on the achievement gap among schools and demographic groups, and, most important, to provide detailed statistical estimates of which school and classroom factors have the most influence on the rate at which student achievement increases.

Several organizations, including PPIC, have produced studies that use California's statewide database on school resources and student achievement to explore one or more of the above three questions. These reports have already produced useful policy insights. However, existing statewide datasets have severe limitations, especially when researchers attempt to answer what is probably the most important question from our list: What factors have the greatest effects on rates of student achievement? The central weakness of the state database is that the unit of observation is a grade level in a given school. If test scores rise for grade 5 students at a certain school between 1999 and 2000, we cannot tell whether this gain reflects a true improvement in school quality or merely reflects variations in the background of the two successive cohorts of grade 4 students. A second and related problem is that we cannot directly study the relative effect of class size and teacher qualifications on

student performance, because we do not know which students are assigned to a given teacher or even the actual class size experienced by an individual student. A third drawback of the state database is that it collects only limited information on teacher qualifications.

Given these limitations with the statewide database, we address the policy questions listed above by compiling and analyzing a large student-level dataset from SDUSD, the second-largest school district in California. The resulting database overcomes all three limitations inherent in studies that use the state's databases. Not only do we observe *gains* in student achievement between years, but we can also control for the changing composition of the student body. In addition, we link individual students with their teachers in each subject, allowing us a much more detailed evaluation of the links between teachers' qualifications and their students' progress. We have also developed a much richer characterization of teacher qualifications than is possible using the statewide databases. In addition, by compiling data on each classroom, we know the class size experienced by each student and, perhaps even more important, the characteristics of each student's classroom peers.

The Link Between Poverty and School Resources in San Diego Schools

We begin our analysis by exploring how school resources vary with respect to the affluence of schools' students. By "school resources" we do not mean funding per pupil but rather the actual people and facilities that go into running a school. Class size is one example of a school resource. But arguably the most important school resource is teachers and the many dimensions of their training, including years of teaching experience, their official teacher certifications and subject authorizations, their highest academic degree, and their field(s) of study at college.

When school resources are defined in these ways, there emerges a strong negative link between the level of disadvantage among students and the school resources that they receive (Betts, Rueben, and Danenberg, 2000). To address this issue in San Diego, we divided students into five approximately equal groups, determined by the

percentage of students eligible for free or reduced-price meals in their schools. We found that students in the lowest socioeconomic status (SES) schools were far more likely to be minorities, to have English Learner (EL) status, and to have parents with relatively little education.

The largest inequalities across San Diego schools relate to teacher qualifications in elementary schools. Figure S.1 gives just one example, showing that in the most affluent group of schools (quintile 1), teachers have two and a half times as many years of teaching experience as in the most disadvantaged group of schools (quintile 5). Furthermore, we found that teachers in the most affluent schools are twice as likely to hold a master's degree and are 10 percent more likely to hold a full credential than are teachers at the lower SES schools.

However, at the secondary level we found less strong relationships between student SES and school resources. We found some evidence that in middle and high schools, math and English teachers in low-SES schools are less likely to hold a full authorization to teach in their subject. We also found that these teachers were relatively less likely to hold a master's degree, although the gaps are smaller than in elementary schools.

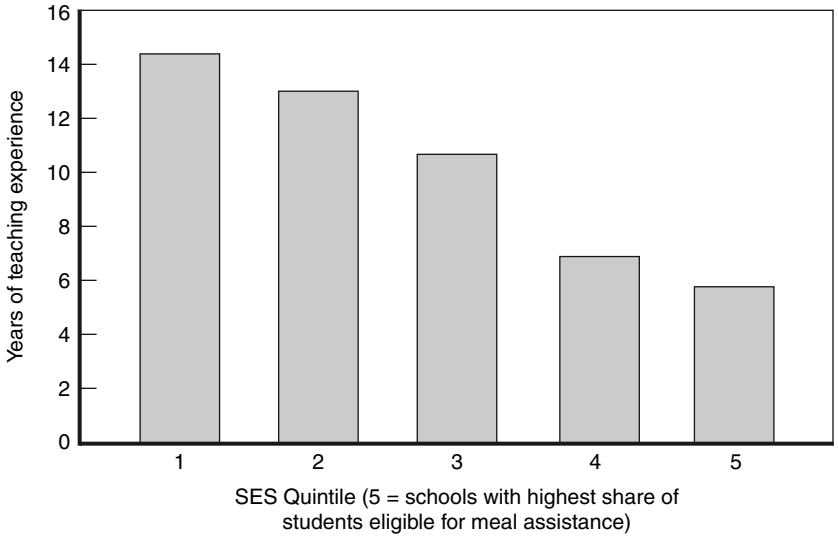


Figure S.1—San Diego Elementary School Teachers' Average Years of Teaching Experience and SES Quintile of School, 1999–2000

On the other hand, we found little link between student disadvantage and class size across schools and, if anything, students in low-SES schools on average had slightly smaller classes in middle and high school.

What Is Happening to Student Achievement in San Diego?

To examine student achievement in more detail, the rest of the report focuses on individual students' scaled scores in California's norm-referenced state test, the Stanford 9. This measure of test scores allows for meaningful comparisons across students at a point in time as well as comparisons of gains in achievement over time.

We examined scores by grade, first for the entire pool of students and then separately by students' demographic groupings. We divided students in numerous ways—by the SES quintiles of the schools that they initially attended in 1997–1998, by race, by EL vs. English-language-fluent status, and by gender.

We divided schools into five SES groups determined by the percentage of students at the school eligible for meal assistance. Figure S.2 shows initial mean reading scores in spring 1998 by the SES quintile of schools. (Results for math are highly similar.) In all grades, the gap in achievement between students in the most and least disadvantaged schools is strikingly large. The bottom and top lines show mean achievement for students in the most disadvantaged and least disadvantaged schools, respectively. A first important observation from this figure is that students, from very early in their educational experiences, appear to exhibit large variations in achievement that are systematically linked to poverty. A second observation from the figure is the extent to which students in the less affluent schools fall behind. For instance, if one traces a series of horizontal lines over the figure, it appears that grade 2 reading achievement in the most affluent schools is not matched by students in the least affluent schools until they reach grade 4. Because the average rate of improvement in test scores decelerates in higher grades, these gaps in "grade equivalents" become even higher in middle and high schools. For instance, the average reading achievement in grade 10 in the most disadvantaged schools lies

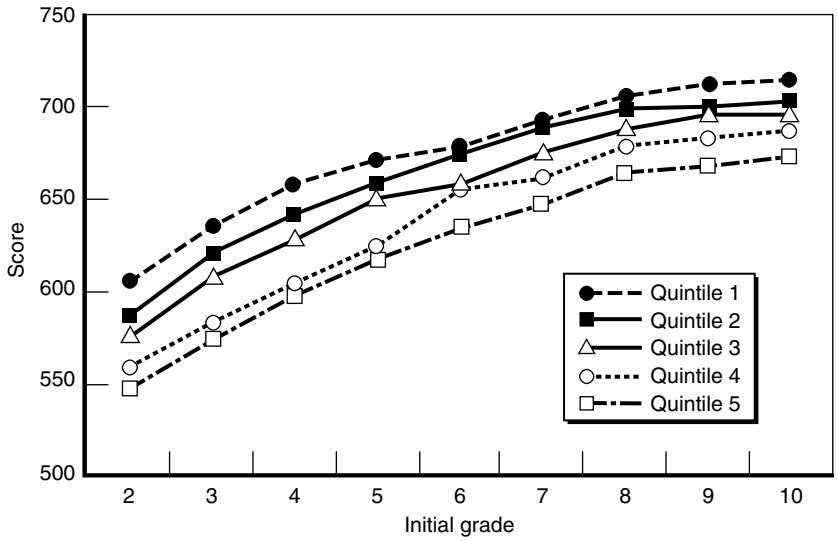


Figure S.2—Spring 1998 Reading Scores, by SES Quintile of School and Grade (5 = schools with highest share of students eligible for meal assistance)

somewhere between mean achievement in grades 5 and 6 in the most affluent schools.

By any standards, these gaps are shocking. However, it is important to realize that other research suggests that the achievement gaps depicted in this figure are not unique to SDUSD, or to California for that matter.

What about *gains* in achievement? If disadvantaged students start their school years less well prepared to learn, it stands to reason that they will only fall further behind as time goes by. To test this idea, we followed the same set of students used for Figure S.2 over the next two years, examining relative gains in test scores between spring 1998 and spring 2000. Figure S.3 shows the results. Achievement among all groups rose substantially but with the largest gains among students who initially were in the lower grades. This may reflect the fact that in the higher grades, teachers devote less attention specifically to reading skills and more to subject matter in diverse subject areas.

Figure S.3 also yields a more subtle, but at least as important, finding. Students who in 1998 were in the lowest SES quintile of schools

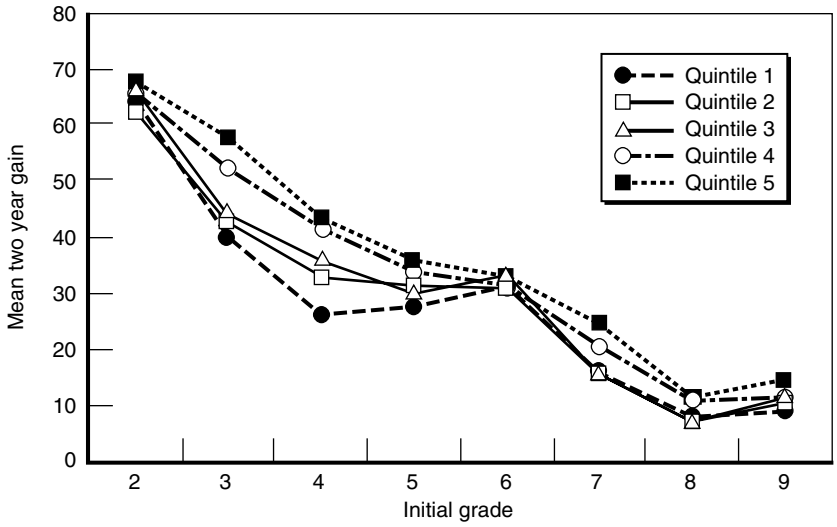


Figure S.3—1998–2000 Gains in Reading Scores, by SES Quintile of School and Grade (5 = schools with highest share of students eligible for meal assistance)

improved their reading performance significantly more than did students in higher SES quintiles of schools. In other words, the achievement gap related to student disadvantage narrowed between 1998 and 2000.

The two dominant patterns that we have described—large gaps between groups that emerge even as early as grade 2 and a narrowing of the gaps over time—are also quite apparent when we divide students not by SES but instead by race or language status. Indeed, no matter how we divide students socioeconomically, by free or reduced-price meal eligibility, by race, or by language status, we always find the same pattern of narrowing in the achievement gap. The only exception is the black-white gap in math achievement, which barely narrowed between the two years.

Overall, the reductions in gaps are substantial. For example, the initial gap in reading achievement between students at the most and least affluent fifths of schools narrowed by 15.2 percent in reading and 11.1 percent in math between 1998 and 2000.

We also examined the gap in achievement between male and female students. In contrast to the gaps we observe by SES, race, and language, any gender gaps in achievement are quite small, and there is no consistent pattern of widening or narrowing of any gaps over time.

Estimating the Determinants of Gains in Student Achievement

The patterns in school resources and student achievement outlined above are both suggestive and confusing. On the one hand, schools in less affluent areas tend to have less experienced, less educated teachers who are less likely to hold full credentials, and these are the schools that have the lowest test scores. However, over time, we found that students in these schools tended to improve their achievement more than did students in more affluent areas. These two ways of looking at the data imply quite different things about whether school resources such as teacher qualifications "matter" for student achievement.

To assess the link between school resources and student learning more rigorously, we estimated a series of models that attempt to explain *gains* in individual students' performance over time, as a function of detailed personal, school, classroom, and teacher characteristics. We conducted separate analyses of the determinants of gains in reading and math achievement in elementary, middle, and high schools in San Diego between the 1997–1998 and 1999–2000 school years.

Several features of the analysis distinguish our approach from the approaches used in statewide studies:

- We model achievement of individual students.
- We examine gains in achievement, not levels, because it is gains that are most likely to be "caused" by the current school year environment.
- A major potential problem in all statistical models of achievement is that the models do not include factors that in reality do determine student achievement. We minimize the potential for this problem in a number of ways:
  - — We take account of a much richer variety of teacher characteristics than is possible using statewide data.

— We take account of the possibility that a student's rate of learning is influenced by the average achievement of those in his or her class or grade.
— We take account of all unobserved factors that are constant during 1998 to 2000 that relate to individual students, their home zip codes, and their schools. One example of unobserved factors is parents' involvement in school activities, at the level of the individual student or the entire school, to the extent that this remains constant across the years.

The Determinants of Gains in Achievement

The regression results can be summarized both in terms of which variables were statistically significant and in terms of the estimated size of the effect of the explanatory variables on gains in reading and math achievement.

One result that appeared meaningful in almost every model that we estimated had to do with the time a student spent at school rather than with school resources themselves. Specifically, the percentage of days a student was absent was a strong negative predictor of each student's gain in achievement in math and reading.

Perhaps the next most consistent finding across all of the models we estimated was that an individual student made much more academic progress in school years in which he or she was surrounded by peers in his or her *grade* who had high scores on the prior spring's test. A strong but less consistent finding was that the average initial test scores of a student's peers in his or her *classroom* also influenced his or her learning. These effects probably work through a number of channels, which can be categorized into the direct effect of a strong peer group (through direct interaction in the classroom and hallway) and indirect effects (such as the increased rigor which a teacher may introduce into a class that is particularly strong). These effects do not merely reflect the student's own prowess, or average school quality, because we statistically control for all unobserved characteristics of both students and schools that are fixed over time, as well as many observable characteristics of each

student, his or her teacher, and school. Rather, the effects are statistically identified by changes from one year to the next in the achievement of individual students' classmates and grade-mates.

Another finding that seems fairly robust is that class size does influence student learning in reading in the elementary grades. But in spite of considerable variations in class size in middle and high schools, we found no evidence that class size matters in these higher grade spans.

Turning to teacher qualifications, our statistical approach involved testing whether a given type of teacher was more or less effective than a teacher in the comparison group, consisting of teachers with a bachelor's degree in education, a full credential, and ten or more years of experience, with no language certification such as a Crosscultural Language and Academic Development (CLAD); and either no university minor, a minor in "other," or a minor in education. At the middle and high school levels, we additionally assigned a full subject authorization in math or English to math and English teachers in the comparison group.

Do these measures of teacher qualifications matter for student learning? Our answer is a qualified yes. We certainly found many instances in which the achievement of students responded positively to higher teacher qualifications. But in most cases, we found no significant difference between less than fully credentialed, relatively inexperienced teachers and teachers in our comparison group.

Overall, teacher qualifications appear to affect gains in student achievement sporadically. However, the effects vary between elementary, middle, and high schools as well as between math and reading achievement.

Comparing results across grade spans, a pattern does emerge: c*lass size appears to matter more in lower grades than in upper grades, whereas teacher qualifications such as experience, level of education, and subject area knowledge appear to matter more in the upper grades*.

Figures S.4 through S.6 give a better idea of the frequency with which key variables were statistically significant predictors of students' gains in reading and math at the elementary and high school levels and the relative size of the predicted effects.

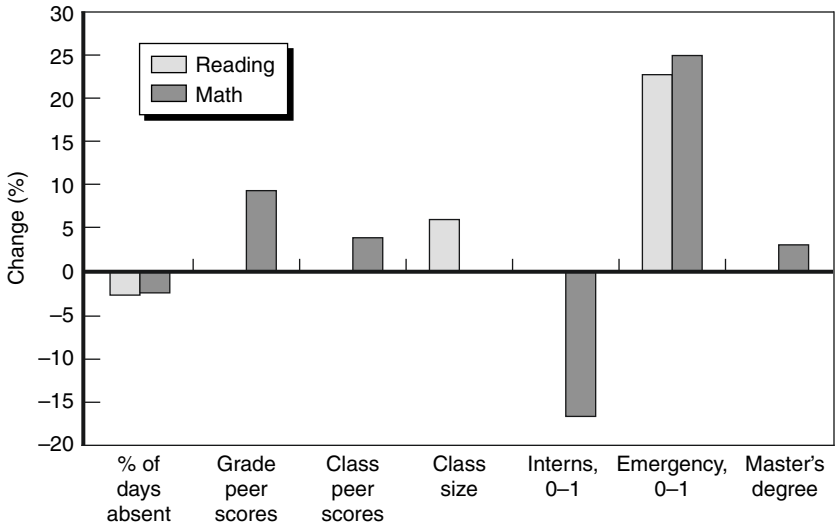


Figure S.4—Predicted Percentage Change in the Rate of Learning Among Elementary School Students

Figure S.4 shows the effect of key variables on students' rate of learning in reading and math in elementary schools. The vertical axis in this figure shows the percentage by which student learning is predicted to change with a given change in classroom or teacher characteristics. In this bar chart, a bar that reached 100 percent would mean that the given intervention was predicted to exactly double the average gain in scaled score points observed in the sample. In many cases, we show the predicted effect of an interquartile change in a given classroom or student characteristic. Suppose that we had 100 observations and ranked them by a classroom characteristic. Then the interquartile change is the change in this characteristic between the 25th and 75th observations, or between the 25th and 75th "percentiles." In cases where we found no statistically significant relation between a factor and a student's gain in achieivement, we omit the corresponding bar in Figures S.4 through S.6.

The first pair of bars shows that an interquartile (25th to 75th percentile) increase in the percentage of days a student is absent is negatively related to gains in students' math and reading achievement.

The next two sets of bars suggest that the initial achievement of peers in the student's classroom and grade appear to be strongly related to student learning in math. The changes simulated in these cases are interquartile changes in peer group test scores, which are likely to be observed if a student switches schools. More conservative estimates that use the median actual year-to-year change in peers' test scores for individual students suggest changes in achievement growth on the order of 1–2 percent.

Class size appears to matter for learning in reading but not math. The predicted effects of an interquartile increase of about 12 students in class size are to reduce gains in reading by about 6 percent.

As shown, a few measures of teacher credentials/teacher experience are similarly, and in a few cases much more strongly, related to student learning. But these results are sporadic—most of our measures of teacher credentials and teacher experience are not statistically significant and the results vary between reading and math. Math achievement seems to suffer when elementary school students are taught by interns with 0-1 years of experience. It is puzzling to note that student gains in both math and reading are predicted to be higher when students are taught by a teacher with an emergency credential and 0-1 years of experience instead of by a fully credentialed teacher with ten or more years of experience. Finally, we find evidence that teachers with a master's degree are marginally more effective in promoting gains in math achievement.

All in all, the elementary school results seem much more powerful with regard to the effect of student absences, peer effects, and class size than they are with regard to teacher qualifications.

Figures S.5 and S.6 show the predicted effects of changing various aspects of the high school student's environment. Figure S.5 considers factors apart from teacher qualifications. Again, student absences are a factor in determining math achievement. Interquartile changes in the grade-level peer scores are predicted to have large effects on gains in math achievement. Overall, in middle and high schools, we found that the size of the grade-level peer effects is much larger than what we found in elementary schools. At the same time, we found that classroom peer

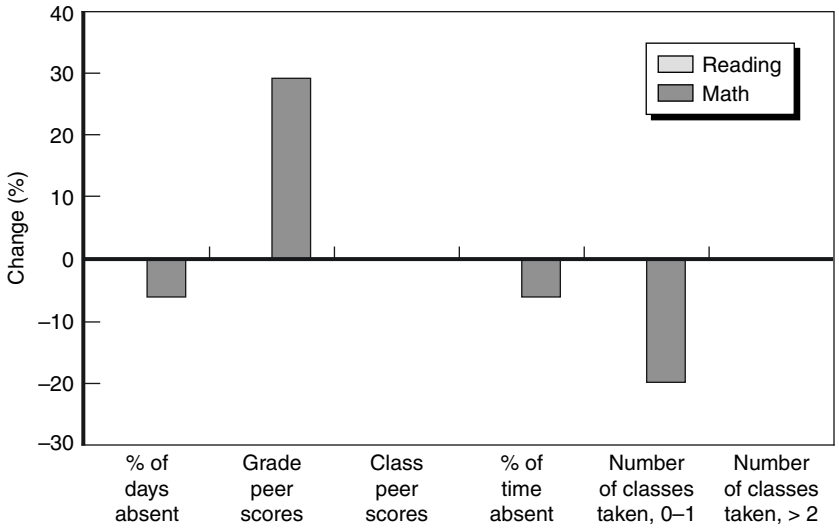


Figure S.5—Predicted Percentage Change in the Rate of Learning Among High School Students, by Absenteeism, Peer Scores, and Courses Taken

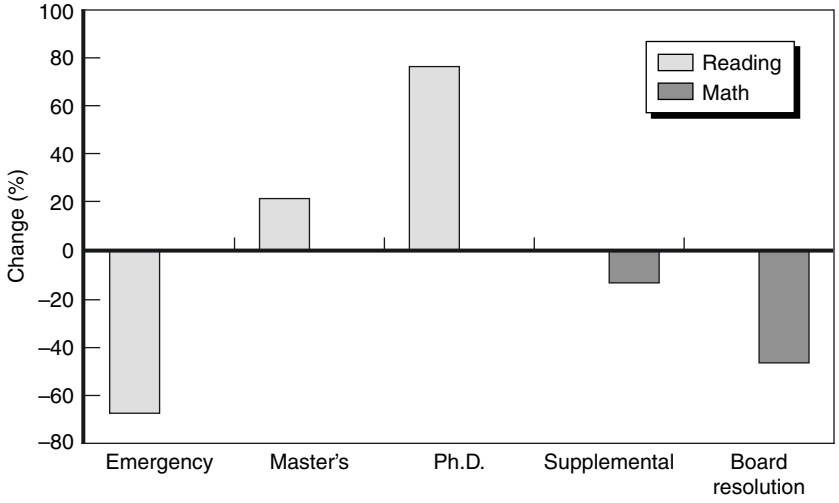


Figure S.6—Predicted Percentage Change in the Rate of Learning Among High School Students, by Teacher Credentials, Education, and Authorization

scores were less likely to be significant predictors of student learning in middle and high schools than in elementary schools. One explanation for these patterns could well be that in middle and high school, students typically switch classrooms during the day, changing their peers from one class to the next. Perhaps in this environment it is less the achievement of peers in the math class that affects a student's improvement in math ability than it is the average achievement of peers in all of his or her classes in the grade.

Figure S.5 also shows evidence that high school students who take 0-1 math classes per year increase their math scores about 20 percent less than do students who take two or more (semester-long) math courses.

Figure S.6 shows the predicted effects of changing teacher qualifications at the high school level. What immediately jumps to the reader's attention is that what matters for math achievement and reading achievement are quite different. For reading achievement, students appear to gain if their English teacher holds a master's or Ph.D. in any field and to lose if their teacher holds an emergency credential. For math achievement, what appears to matter most is the level of math authorization that the math teacher holds. In these simulations our default math teacher holds a full math authorization, signifying that he or she has taken all of the college math courses recommended by the California Commission on Teacher Credentialing (CCTC). The next two highest levels of subject authorization are the CCTC's supplementary authorization followed by the board resolution. It appears that for math at the high school level, the level of subject authorization is important.

In sum, at the high school level, two patterns stand out with regard to teacher qualifications. First, the effects of specific types of teacher qualifications are quite variable across subjects. Second, when a given teacher qualification does matter at the high school level, the predicted effects are very large. Although we do not show them in this summary, middle school results are very similar in both regards.

Although space limitations prevent us from presenting results for English Learners, several important points emerged from our analysis. First, we should not assume that given aspects of the classroom environment affect EL and other students in the same way. Typically,

the patterns were quite different. For example, at the elementary school level the effects of changing class size or peer group achievement appear to be twice as large for EL students as for students taken as a whole. Second, across all three grade spans we found little evidence that teachers with CLAD, Bilingual-CLAD (BCLAD), or equivalent certifications that are designed to help teachers instruct English Learners were associated with faster gains in achievement among these students.

Although we have learned a great deal about the San Diego Unified School District, do these lessons hold any value for districts elsewhere or for policymakers in Sacramento? We believe that the answer is yes. We performed a detailed analysis of test scores, class size, teacher qualifications, and student demographics that compared San Diego with the other largest school districts in California as well as schools in California taken as a whole. Differences exist among the large districts, but overall they bear strong similarities in terms of demographics, teacher qualifications, class size, and student achievement. Perhaps most important, all districts in California operate under the same set of ground rules and financing formulas established by the state government. This increases our confidence that at the very least the broad lessons learned from San Diego will hold relevance for other districts around the state.

Policy Lessons

The seeming paradox that in San Diego the least advantaged students improved their scores by the greatest amount between 1998 and 2000 in spite of having less qualified teachers than average has been partly resolved by our regression results. In essence, teacher education, credentials, experience, and subject authorizations can make a difference, but the effects are neither as systematic nor as big as some might believe.

In some respects, administrators should be reassured to learn that a less than fully credentialed teacher sometimes appears to be as effective as a fully credentialed teacher. California spends roughly $100 million a year on the Beginning Teacher Support and Assessment (BTSA) program, which aims to provide assistance to teachers in their first and second years of teaching. This and related programs might successfully integrate inexperienced teachers into the classroom. In addition, SDUSD has adopted a peer coach program to train teachers in the latest

instructional techniques, which may be particularly helpful for novice teachers.

Similarly, the result that middle and high school English and math teachers with less than a full subject authorization often are just as effective as fully authorized teachers should come as reassuring news, given that it is difficult for a district to ensure that all of its teachers have exactly the right mix of college courses as mandated by the CCTC. The one major exception to this rule was high school math teachers, in which case subject authorization level appears to matter tremendously.

The evidence that teacher experience and credentials have less effect on gains in student achievement than some may think is particularly important given the grim new financial reality facing most California school districts as a result of California's large budget deficits. In San Diego, the district tackled its budget problem in early 2003 by offering early retirement incentives. These incentives led approximately one in ten teachers to opt for retirement. The district plans to replace these teachers with less highly experienced teachers who will be paid less. It seems likely that the short-run effect of this mass retirement will be to make schools less effective simply because of the loss of institutional memory. However, our results suggest that after one or two years, many of the relatively inexperienced recruits may be far more effective teachers than some would believe.

Although the measured effect of teacher qualifications varies substantially by subject and grade span, overall we did find sporadic evidence that in certain cases teacher qualifications matter significantly, especially in the higher grades. In light of these findings, what can be done to equalize teacher qualifications between schools in disadvantaged and more affluent areas?

The strong relationship between student poverty and teacher qualifications appears to be related to clauses in the district's collective bargaining agreement. The agreement requires that schools with teaching vacancies limit their choice from the pool of qualified applicants to the five candidates with the most district seniority. This contract clause, in conjunction with the apparent preferences of teachers to move to schools in relatively affluent areas, generates some relatively severe inequalities in teacher qualifications across San Diego's schools. The

reason is simply that more affluent schools currently have no option but to hire the most highly experienced teachers who apply for one of their coveted openings.

Therefore, one possibility would be to relax first-right-of-transfer clauses in the district's collective bargaining agreement, because these restrictions militate against the need for inner-city schools to retain highly experienced and qualified teachers.

A related possibility would be to redesign the wage schedule for teachers to allow for salary bonuses to teachers with certain skill sets who agree to teach in schools with a shortage of qualified teachers in certain areas. This innovation would represent a major reform to the structure of teacher pay in California. To succeed politically such a reform would probably have to be presented as a pay increase for many teachers that would not decrease the pay of any teacher. Clearly, the current budget situation in California suggests that this reform cannot be implemented in a major way until California has solved its budget problems.

One of the many "achievement gaps" identified by this report is the one between EL and English-language-fluent students. In separate models for EL students, we discovered that at the elementary school level, class size reduction appeared to be twice as effective at improving reading achievement for EL students relative to students overall. Although we did not find evidence that teachers with CLAD or BCLAD certificates were unusually effective with EL students, a number of different measures of teacher qualifications, such as whether the teacher held a master's degree, in some cases were associated with higher gains in EL student achievement. Although our limited sample size of EL students has limited the precision of our estimates, all of these resource issues related to English Learners deserve continued attention.

Policymakers also should be interested in one of the most consistent findings in this study—that an individual student's rate of learning appears to be strongly and positively influenced by the initial achievement of students in his or her grade, and with somewhat less consistency by that of students in his or her classroom. This finding holds great policy relevance. Obviously, ability grouping within the school will affect each student's peers. Similarly, students who volunteer

for busing in the district are likely to alter their peer group in substantial ways. Both of these issues are worthy of more detailed study.

In fall 2000, SDUSD implemented its Blueprint for Student Success. This reform is designed to accelerate the learning of students who lag far behind grade level. The reform has attracted favorable national attention and generated intense local controversy. Although some elements of the blueprint such as those related to peer coaching of teachers were implemented toward the end of our period of study (the school years 1997–1998 through 1999–2000), the main parts of the reform were put in place in fall 2000 after the period we study. Therefore, our report cannot speak to the extent to which the blueprint will boost student achievement. However, our results do allow us to comment on the general approach taken by the blueprint. First, our finding that in the late 1990s reading achievement in the district lagged behind national norms to a greater extent than did math performance suggests that the initial focus of the blueprint on reading may make good policy sense. Second, our analysis of the large achievement gap in the district between more and less affluent students, between white students and students of other ethnicities, especially Hispanics and blacks, and between English Learners and fluent speakers of English suggests that the blueprint is on the right track in its central tenet that more resources must be devoted to students who lag behind academically. Third, we found that increasing teacher credentials and education, although suggestive of better teaching, are not a panacea. The variable effects of mainstream teacher qualifications certainly provide some rationale for the heavy investments that the district is currently making on new teacher professional development programs.

Thus, our findings provide considerable support for the idea that the district would do well to overhaul its interventions both for students who are struggling and for the assistance it provides to teachers. The blueprint is moving in exactly these directions. Of course, none of our analyses can predict the extent to which the blueprint itself will increase academic achievement.

Finally, we note that the daunting achievement gaps between students do not appear to be *created* primarily by the schools as they now exist. These gaps, related to income and socioeconomic status more

generally, emerge by the time young children reach school age. One implication is that at the federal and state level, policymakers may want to examine the value of Head Start and similar preschool programs as a way to reduce the achievement gap of disadvantaged students before they begin their formal schooling.

As for the schools themselves, in San Diego Unified, at least, schools appear to have been working effectively to reduce these gaps between 1997–1998 and 1999–2000. We should not use this sign of success as an excuse to ignore the large achievement gaps that remain. But it should give us some perspective. Schools are not a part of the problem; they are a part of the solution. The goal of this report, and ensuing reports, has been and will be to shed some light on the most promising ways to devote limited financial resources to making schools more effective solutions than they already are today.

Contents

Foreword iii
Summary v
Figures xxvii
Tables xxxi
Acknowledgments xxxiii

1. INTRODUCTION 1

2. CHALLENGES IN ANALYZING THE RELATION BETWEEN SCHOOL INPUTS AND STUDENT ACHIEVEMENT 7
   Evidence on the Relation Between School Resources and Student Outcomes 8
      School Resources and Student Achievement 8
      Research from California 10
      Research from Texas 12
      National Studies Using School Resources Aggregated to the State Level 12
      School Resources, Educational Attainment, and Earnings . . 13
   How Representative Is San Diego Unified? 14
      Student Demographics and Student Achievement 14
      School and Teacher Inputs 19
   Conclusion 20

3. THE LINK BETWEEN POVERTY AND SCHOOL RESOURCES IN SAN DIEGO SCHOOLS 21
   Dividing Schools on the Basis of Student Demographics 21
   Student Mobility, Student Retention, and Dropout Rates . . . 22
   Class Size 25
   Teacher Characteristics 26
   Conclusion 34

4. TRENDS IN STUDENT ACHIEVEMENT IN SAN DIEGO 37
Introduction 37
Overall Trends in Achievement Gains Between Spring 1998 and Spring 2000 40
Variations in Improvement Across Schools and in Particular Student Groups 42
SES Quintile of the School 42
Student Race and Ethnicity 47
English Learners vs. Non-English Learners 50
Male vs. Female Students 51
Summary 52
Conclusion 52

5. DETERMINANTS OF GAINS IN STUDENT ACHIEVEMENT IN ELEMENTARY SCHOOLS 55
Introduction 55
Overview of the Procedure for Statistically Estimating the Determinants of Gains in Student Achievement 55
Variables Included in Models of Gains in Test Scores 56
Results 61
The Effect of Demographics of the Student Body and Peers' Initial Test Score 61
The Effect of Class Size and Teacher Credentials, Experience, and Education 63
Conclusion 78

6. DETERMINANTS OF GAINS IN STUDENT ACHIEVEMENT IN MIDDLE AND HIGH SCHOOLS 81
Introduction 81
Findings for Middle and High Schools 84
Patterns of Statistical Significance 84
The Predicted Effect of Explanatory Variables on Students' Rate of Learning 90
Robustness Checks 104
Conclusion 105

7. POLICY CONCLUSIONS 107
Overview of Central Findings 107

Efficiency 108
Equity 108
The Determinants of Student Learning 110
Policy Implications 113

Appendix
A. Methods Used to Take Account of Unobserved Factors Affecting Student Learning 119

B. Details on the Regression Models for Elementary School Students 127

Web-Only Appendix (www.ppic.org/main/publication.asp?i=321)
C. Additional Information About Student Demographics, School Resources, and Test Scores in San Diego and Other Districts

D. Student SES and School Resources in San Diego

E. Additional Information About Achievement Gaps Between Various Student Subgroups in San Diego

F. Details on the Regression Models for Elementary School Students

G. Details on the Regression Models for Middle School and High School Students

Bibliography 131

About the Authors 137

Related PPIC Publications 139

Figures

S.1. San Diego Elementary School Teachers' Average Years of Teaching Experience and SES Quintile of School, 1999–2000 vii
S.2. Spring 1998 Reading Scores, by SES Quintile of School and Grade ix
S.3. 1998–2000 Gains in Reading Scores, by SES Quintile of School and Grade x
S.4. Predicted Percentage Change in the Rate of Learning Among Elementary School Students xiv
S.5. Predicted Percentage Change in the Rate of Learning Among High School Students, by Absenteeism, Peer Scores, and Courses Taken xvi
S.6. Predicted Percentage Change in the Rate of Learning Among High School Students, by Teacher Credentials, Education, and Authorization xvi
2.1. Student Percentages, by Race, 1999–2000 15
2.2. Percentage of Students Who Are English Learners and Who Are Eligible for Free or Reduced-Price Meals, 1999–2000 16
2.3. Student Performance Against National Norms in Reading, 1999–2000 17
3.1. Percentage of San Diego High School Students Whose More Educated Parent Has a Bachelor's Degree or Higher and SES Quintile, 1999–2000 23
3.2. Percentage of San Diego High School Students Who Are Unexpected Transfers, Retained a Grade, or Drop Out of School, and SES Quintile, 1998–1999 24
3.3. Class Size in San Diego, by Grade Span and SES Quintile, 1999–2000 25
3.4. Percentage of San Diego Elementary School Teachers with a Master's Degree or Higher and SES Quintile, 1999–2000 27

3.5. Percentage of San Diego High School Math and English Students Whose Teacher Majored in Math or English and SES Quintile, 1999–2000 27
3.6. San Diego Elementary School Teachers' Average Years of Teaching Experience and SES Quintile, 1999–2000 28
3.7. Percentage of San Diego High School Math and English Students Whose Teacher Held a Full Authorization in Math or English and SES Quintile, 1999–2000 31
3.8. Percentage of San Diego Middle School Students Whose Teacher Held a Full or Supplemental Authorization and SES Quintile, 1999–2000 33
4.1. Spring 1998 Reading Scores, by SES Quintile of School and Grade 43
4.2. Spring 1998 Math Scores, by SES Quintile of School and Grade 44
4.3. 1998–2000 Gains in Reading Scores, by SES Quintile of School and Grade 46
4.4. 1998–2000 Gains in Math Scores, by SES Quintile of School and Grade 47
4.5. Spring 1998 Reading Scores, by Ethnicity and Grade . . . 48
4.6. Spring 1998 Math Scores, by Ethnicity and Grade 48
4.7. 1998–2000 Gains in Reading Scores, by Ethnicity and Grade 49
4.8. 1998–2000 Gains in Math Scores, by Ethnicity and Grade 50
5.1. Predicted Percentage Change in the Rate of Learning Among Elementary School Students 77
6.1. Predicted Percentage Change in the Rate of Learning Among Middle School Students, by Absenteeism, Peer Scores, and Courses Taken 101
6.2. Predicted Percentage Change in the Rate of Learning Among High School Students, by Absenteeism, Peer Scores, and Courses Taken 101
6.3. Predicted Percentage Change in the Rate of Learning Among Middle School Students, by Teacher Credentials, Experience, and Authorization 103

6.4. Predicted Percentage Change in the Rate of Learning Among High School Students, by Teacher Credentials, Education, and Authorization 103
A.1. Identical Students with Different Quality Classrooms . . . 120
A.2. Hypothetical Example of Incorrect Inferences About the Value of Teacher Experience for Student Learning, Caused by Unobserved Variations in School Quality 121
A.3. Hypothetical Example of Correct Inferences About the Value of Teacher Experience for Student Learning, After Taking Account of Unobserved Differences in School Quality 122
A.4. Hypothetical Example of Incorrect Inferences About the Value of Teacher Experience for Student Learning, Caused by Unobserved Variations in Student Ability . . . 124

Tables

2.1. Stanford 9 Test Score Distribution: Unwighted Average Across Grades, All Students 18
4.1. Mean Scaled Scores by Year for All Students, Reading . . . 40
4.2. Mean Scaled Scores by Year for All Students, Math 41
4.3. Percentage Reduction in Test Score Gaps, 1998–2000 . . 51
5.1. Student, Family, and Neighborhood Controls Used in the Statistical Models 57
5.2. School, Classroom, and Student Body Controls Used in the Statistical Models That Include Both EL and Non-EL Students 59
5.3. Statistical Significance of Demographics of the Student Body and Average Initial Test Scores in the Student's Classroom and Grade in Elementary School Models 62
5.4. Statistical Significance of Class Size and Teacher Qualifications in Elementary School Models 64
5.5. Statistical Significance of Teacher's CLAD, BCLAD, and Alternative Certifications in Elementary School Models 66
5.6. Statistical Significance of Teacher's College Major and Minor in Elementary School Models 68
5.7. Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for Elementary School Students . . . 70
5.8. Predicted Effect of Stated Changes in Personal, School, Classroom and Teacher Characteristics on the Rate of Learning in Math for Elementary School Students 72
6.1. Statistical Significance of Demographics of the Student Body and Average Initial Test Scores in the Student's Classroom and Grade in Middle and High School Models 85

6.2. Statistical Significance of Class Size and Teacher Credentials, Experience, Education Level, and Subject Authorization in Middle and High School Models 87
6.3. Statistical Significance of Teacher's CLAD, BCLAD, and Alternative Certifications in Middle and High School Models 90
6.4. Statistical Significance of Teacher's College Major and Minor in Middle and High School Models 91
6.5. Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for Middle School Students 92
6.6. Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Math for Middle School Students 94
6.7. Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for High School Students 96
6.8. Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Math for High School Students 98

Acknowledgments

We are greatly indebted to many people at the San Diego Unified School District, without whom the collaborative working relationship upon which this report is based would not have been possible. We thank Superintendent Alan Bersin and Chancellor of Instruction Anthony Alvarado for inviting us to work with the district. We are particularly indebted to Karen Bachofer, who has worked tirelessly as the chief district liaison to this project. In addition, we would like to acknowledge the following for their interest in this project and their generous help on many occasions: Deberie Gomez, Peter Bell, Gary Knowles, Jeff Jones, Susie Millett, Barbara Jarrold, Teresa Walter, and David Lee. Karen Bachofer, Eric Hanushek, Christopher Jepsen, Marianne Page, Susanna Cooper, Joyce Peterson, and Christopher Weare devoted considerable time to reading all or parts of the manuscript; their suggestions have led to some major improvements. We also thank Kevin King for research assistance. Last but not least, we would like to acknowledge the financial support of the Public Policy Institute of California. Without this support, the multiyear effort needed to develop and analyze the massive database upon which this research is based would have been impossible.

1. Introduction

Over the past few years, Californians have consistently placed the state of K–12 education at or near the top of their list of political concerns. For instance, the Public Policy Institute of California (PPIC) Statewide Survey conducted in October 2002 found that education was the issue likely voters most wanted to hear gubernatorial candidates discuss in the upcoming election—at 21 percent—compared to 14 percent for the next most frequently mentioned issue—jobs and the economy (Baldassare, 2002). Indeed, the PPIC Statewide Surveys, apart from a brief interlude at the peak of the electricity crisis, have consistently reported over the past four years that education ranks as the number one priority of the public.

Although multifaceted, these concerns about education boil down to two central issues: efficiency and equity. Californians want to see more efficient schools that spend money wisely. They seem particularly concerned about the need to increase the percentage of teachers in the classroom who are fully qualified. For instance, the February 2002 PPIC survey found that majorities of both likely voters and parents of public schoolchildren listed teacher quality, including recruitment and training, and overall spending as areas in which they were dissatisfied with recent education reform efforts in California.

As for equity, it is apparent that California's schools vary radically in the resources they receive. Betts, Rueben, and Danenberg (2000) show that teacher qualifications, along a number of dimensions, tend to be much lower in schools in relatively disadvantaged areas than in affluent areas. Given California's new educational accountability system that offers both financial carrots and sticks to schools that lag behind, inequality in school resources becomes of even greater concern.

Not only do the state's schools vary in resources but they also vary dramatically in student achievement. Results from the first few years of the state's testing program, started in spring 1998, have revealed

considerable gaps in student achievement between the have and have-not schools. Betts, Rueben, and Danenberg (2000) analyzed spring 1998 test scores and found that the most important predictor of the share of a school's students scoring at or above national norms in reading and math was the percentage of students eligible for free or reduced-price meals, even after accounting for differences in resources such as class size and teacher education across schools. In relatively well off elementary schools (which ranked 25th out of 100 in family affluence), typically 55 percent of students scored at or above national norms in math and reading. In contrast, at more disadvantaged schools (which ranked 75th out of 100 in affluence), only about 27 percent of students scored at or above national norms.

Californians appear to be well aware of these inequalities, and in addition they appear to be committed to doing something about it. In the February 2000 PPIC Statewide Survey, 78 percent of respondents answered no to the following question: "Do you think that schools in lower-income areas have the same amount of resources—including good teachers—as schools in wealthier areas?" Perhaps more surprising, when asked the question: "Do you think that school districts with the lowest student test scores in the state should or should not be given more resources than other school districts?" 70 percent of respondents answered that they should (Baldassare, 2000). Clearly, Californians care about unequal resources and unequal outcomes in the public schools.

The startling chasm in achievement across schools together with unequal school resources raise some major policy questions:

- How big are the variations in school resources across schools?
- What are the trends in student achievement? Have the achievement gaps between disadvantaged students and affluent students widened in light of the gap in school resources received by these students? What about racial/ethnic gaps?
- Given the large variations across schools in resources, especially in teacher qualifications, where should the state focus future budget increases? Should it reduce class size or focus on improving teacher qualifications?

- If the state attempts to enhance teacher qualifications, what types of teacher qualifications should it focus on? What matters most for student learning? Do teachers' experience, their credentials, their overall level of education, or their major at a university have the largest effect on student learning?
- Betts, Rueben, and Danenberg (2000) found that two-thirds or more of the achievement gap between California students and those nationwide reflects the relatively large number of English Learner (EL) students in California. If we compare EL students to the entire student body, do we find differential effectiveness of class size, teacher experience, and so on? In particular, how effective is the training provided to teachers to help them teach students whose first language is not English?

The goal of this report is to address these vital policy issues by analyzing in detail the patterns of resource allocation and student achievement in the state's second-largest school district, San Diego Unified School District (SDUSD).

The issues we list above are clearly of statewide importance. So, it is natural to ask why one would want to answer these questions using data from a single district, instead of using statewide or even national data. PPIC has active education research programs that use both statewide data gathered by the state Department of Education and nationwide data. However, nationwide datasets are typically not representative of California's schools, which differ in quite fundamental ways from schools in other parts of the country. To give just one example, the national sample of students used to norm the Stanford 9 test now given throughout California had only about 2 percent EL students, compared to roughly 25 percent in California. Similarly, Sonstelie, Brunner, and Ardon (2000) show divergent trends between California and the rest of the country in overall funding per pupil and specific measures of school resources since 1980. California's relatively anemic school funding per pupil suggests that it may be difficult to extrapolate from national studies to California.

As for statewide research, it is true that a great deal can be learned from sifting through California Department of Education data. Reports

by the CSR Consortium on class size reduction, by the California Center for the Future of Teaching and Learning (2000) on the qualifications of California's teachers, by Jepsen and Rivkin (2002) of PPIC on class size reduction, and by Sonstelie, Brunner, and Ardon (2000) and Betts, Rueben, and Danenberg (2000) provide five examples.[1]

But California's statewide education database is sorely lacking in several dimensions. For example, consider the issue of teacher quality, which PPIC's Statewide Survey has shown to be an issue of key public concern. What do we really know from state data about the effect of teacher qualifications on student learning? The annual state Department of Education survey does reveal much about both the overall level of teacher qualifications in California and variations across schools. But what we really want to know is the extent to which a student learns more quickly if taught by a teacher with ten years of experience and a full teaching credential instead of by a teacher with one year of experience and an emergency credential. At the state level, the state test does not link individual students to individual teachers, so we have no hope of answering such questions using the statewide databases. Even more frustrating for researchers, California's Standardized Testing and Reporting System (STAR) gathers test scores for each student in California in grades 2 through 11 annually, but does not allow one to follow individual students over time. This means that using state data, we cannot hope to answer the key question: What aspects of classrooms, teachers, and schools contribute the most to *gains* in achievement of individual students over time? Analysis of state-level data does not allow much more than a series of annual snapshots that show correlations between school resources and outcomes at the school level.

For this reason, we have entered into a collaborative research arrangement with SDUSD to delve more deeply into both the distribution of school resources within and among schools and the determinants of student learning. We are thus able to obtain detailed information that is simply impossible to obtain at the state level. Most important, we have linked individual student records across years, so that

[1]For an example of research by the first organization, see Stecher and Borhnstedt (2002).

we can examine each student's *gains in achievement*, and at the same time we have linked students to the teachers who teach them each year. This provides a powerful analytical tool for examining the relative effect on student learning of teachers' years of experience, highest academic degree, college major and minor, and type of teaching credentials and subject authorizations. Because we know the identity of every student inside every classroom in the district, we also have a rare opportunity to examine the importance of a student's classroom and grade-level peers on his or her own learning.

In sum, by working with an individual district we are able to look "inside" the classroom to obtain a better picture of the variations in teacher and classroom characteristics and the contributions of these characteristics to student learning.

A natural next question to ask is: Why choose SDUSD? First, the existing research literature on school quality suggests that the relation between school resources and student outcomes is subtle, complex, and, as some researchers have claimed, rather tenuous. To infer the effect of school resources on student learning, we need a large district that provides both large samples of students and schools and significant variation across schools in both resources and achievement. SDUSD, second only in size in California to Los Angeles Unified, meets both these criteria.

A second reason for choosing SDUSD is that we hope to learn policy lessons that are likely to be of interest throughout California. For our research results to hold much relevance elsewhere, we must choose a district that is largely representative of what is observed in other districts. Overall resource levels, student demographics, and the level of student achievement should match the state average reasonably well. SDUSD typically matches California norms as closely if not more closely than do the other large districts.

A third reason for our choice of SDUSD is the national attention the district has recently garnered for its "Blueprint for Student Success." Implemented in the 2000–2001 school year, this ambitious and controversial plan calls for a districtwide intervention program to help students who are identified as falling behind grade level. The blueprint calls for multiple interventions, such as the placement of peer coaches in

schools to assist with teacher training, extended length English classes for students who lag behind in reading, before- and after-school classes, reduced class sizes in certain cases, summer school, and, if necessary, grade retention.

Our report analyzes patterns of achievement and school resources between the school years 1997–1998 and 1999–2000. The district implemented the most far-reaching components of the blueprint in the 2000–2001 school year. Therefore, our current study cannot directly assess the effect of the blueprint. Nevertheless, it provides an important baseline against which to compare future progress.

The structure of the report is as follows. Chapter 2 reviews the knowledge gained on the determinants of student achievement from earlier studies. The chapter then examines whether our San Diego data can solve some of the problems in earlier work. Chapter 3 examines the relation between the level of economic disadvantage of students in San Diego schools and the level of school resources that they receive. Chapter 4 provides a detailed examination of trends in student achievement. Besides examining overall performance, the chapter examines trends in the achievement gap between more and less disadvantaged students, between white students and other racial/ethnic groups, and between EL students and students fluent in English. Chapter 5 summarizes our findings concerning the determinants of math and reading achievement in elementary schools. Chapter 6 presents findings for middle and high schools. Chapter 7 draws conclusions and discusses policy lessons.

2. Challenges in Analyzing the Relation Between School Inputs and Student Achievement

This report, like many other studies of education, focuses on students' test scores as a measure of student success. One may question whether test scores are at all relevant for any of the outcomes that really matter, such as the level of education that students ultimately obtain, or earnings of workers years after they have finished school. Test scores do not explain everything by any means, but test scores on California's state tests are likely to be positively linked to students' scores on the SAT and other college entrance exams. In addition, researchers have shown directly that test scores are positively linked to the probability of college attendance as well as earnings of students later in life, and that this latter linkage appears to have grown in recent decades.[1] This chapter provides a brief summary of what we know about the relation between specific measures of school resources such as class size, teacher education, and teacher experience and student outcomes such as test scores, years of education completed, and earnings years after students have left school.

The review shows that a major obstacle in past research has been the lack of data that follows the progress of individual students over time while measuring the school resources that the student receives at the classroom level. This is particularly true in California, where the state testing system analyzes test scores by school rather than by student. The database we use from San Diego solves most of these problems. However, our focus on a single school district potentially creates a new problem in that the district may not be at all representative of students,

[1]See Grogger and Eide (1995) and Murnane, Willett, and Levy (1995).

teachers, and schools statewide. We consider this issue in the second part of the chapter.

Evidence on the Relation Between School Resources and Student Outcomes

School Resources and Student Achievement

In regard to the determinants of student test scores, a good place to start is the Coleman report, an early landmark in the school quality literature. Coleman (1966) undertook a massive national study that attempted to explain the level of student test scores as a function of students' personal background and the characteristics of their teachers and schools. He found surprisingly little relationship between standard measures of school quality and student achievement. He found that students' socioeconomic status explained a far greater proportion of the variation in test scores than did measures of school resources such as the pupil-teacher ratio and teacher attributes.

The results of the Coleman report might in part stem from the fact that the author attempted to explain levels of achievement, not gains in achievement. Our analysis of test scores in SDUSD will show that students who are in some sense disadvantaged start their school years significantly behind their more advantaged peers. This initial "preschool" gap cannot be attributed primarily to what goes on in schools. A more reasonable test of whether school resources matter might be to test for a link between gains in achievement and school resources such as class size. The Coleman report did not include any data on gains in achievement.

But it is not so easy to dismiss Coleman's results. Numerous studies since that time have modeled gains in achievement, as does this report, to eliminate the problem of unfairly holding a grade 4 teacher responsible for the level of his or her students' achievement. Rather, many of these later studies model one-year gains in achievement instead of the level. Yet many of these more sophisticated studies have found results quite similar to those of the Coleman report.

In a series of influential reviews of the literature, Hanushek (1986, 1996) concludes that a small proportion of studies have found that additional school resources lead to significantly higher achievement.[2] For many measures of school resources, such as class size, Hanushek reports that most studies find no significant link to student achievement. Of the various school resources examined in these studies, the one that he most regularly finds to matter for student achievement is teacher experience. Overall spending per pupil and teacher salary are the school resources that appear to matter the second and third most often. Few studies have found that teacher education affects student achievement.

With regard to teachers, we should emphasize that the research finding that teacher qualifications are only weakly associated with student achievement is not the same thing as stating that teacher quality does not matter. Murnane (1975) tested whether some teachers on average produced better test-score gains among their students than others, even after taking account of variations in the standard measures of teacher qualifications and other factors. He found strong evidence that teachers did vary systematically in the rates at which their students' achievement improved over time. In other words, teacher quality does vary, but these variations are not strongly linked to factors such as teachers' education or experience. Numerous studies since that time have replicated Murnane's finding that teachers do vary in quality in ways that cannot be explained by credentials, education, and the like.[3]

Perhaps the strongest evidence to date in favor of the hypothesis that school resources "matter" comes from Tennessee's class size reduction experiment of the 1980s. Students in kindergarten through third grade

---

[2]Although Hanushek's claims have been influential, they are not universally accepted. See, for instance, the exchange between Hedges, Laine, and Greenwald (1994) and Hanushek (1994).

[3]See Goldhaber (2002) for a review of the more recent literature. Some educators and national professional associations that are involved in the teaching credentialing process have made well-known claims that teacher certification is by far the most important determinant of student learning. These claims have long puzzled many researchers who have been involved in contributing to the quantitative literature on school quality. For a review that is highly critical of the claim that teacher certification is a decisive factor in determining student performance, see Walsh (2002).

were randomly assigned to one of three groups. The first group had class sizes as low as 15 students; the second group had class sizes in the low 20s and one teacher's aide per class; and the third group had class sizes in the low 20s. Since then, numerous studies have compared test scores for the three groups.[4]

The results indicate that students placed in the small classes learned more quickly than other students. Most of the gains accrued to students in the first year they were in smaller classes, and students of low socioeconomic status (SES) gained somewhat more than others. However, these gains largely disappeared after students were returned to regular sized classes (Krueger and Whitmore, 1999). Specifically, students in smaller classes had a 4.5 percentile point advantage over other students at the end of third grade, after which they returned to regular sized classes, but this advantage had diminished to 1 percentile point by the end of eighth grade.

The Tennessee experiment offers the most persuasive evidence to date for reducing class size. Even so, the results suggest that such reductions produce very modest gains, especially if students are placed in larger classes in later grades.

Research from California

A number of recent studies have examined school resources and student achievement in California. For example, Betts, Rueben, and Danenberg (2000) analyze the distribution of resources and test scores at the school level for 1997–1998. Their regression analyses suggest that by far the best predictor of student achievement at each school was the percentage of students eligible for free or reduced-price meals. The predicted effects of changing teacher credentials, experience, education, or class size were minor compared to the effect of student SES. In part, this finding should not be surprising, because the authors were able to use only the first year of results from the Stanford 9. The level of achievement in any given grade will be the cumulative result of experiences not only in that grade but in earlier grades and in the preschool years as well. Notably, however, the report finds equally

[4]See, for instance, Grissmer (1999, p. 2).

strong results in favor of student SES as an explanatory factor in the models of grade 2 achievement as in the models of achievement in higher grades. Betts and Danenberg (2001) use the results of Betts, Rueben, and Danenberg (2000) to estimate the possible effect of partial or full equalization of resources across California's schools and find that even the radical step of fully equalizing teacher preparation across schools would contribute only modestly to eliminating the achievement gap among schools.

The CSR Consortium has also studied the effect of recent class size reductions (CSR) in California (Bohrnstedt and Stecher, 1999, 2002; CSR Research Consortium, 1999, 2000). As the consortium authors note, they cannot draw firm conclusions because of limitations in the state's student data system along with the wholesale implementation of the reform in a way that prevents the availability of a valid control group. The first two reports by the CSR Consortium provide some evidence that third-grade test scores have risen modestly because of class size reductions. In the first year of the study, the CSR Consortium (1999) compared state test scores of students at elementary schools that had implemented class size ceilings of 20 students to students at schools that had not yet adopted the reform. However, the students at schools that did not implement class size reduction in the first year came from lower-SES families, making any simple comparison problematic. The authors attempt to adjust statistically for this problem but express reservations about the reliability of their results. The second CSR Consortium report (2000) uses a more complex comparison technique to estimate the effects of class size reduction. Again, the authors find statistically significant but modest effects of class size reduction and indicate that the lack of a true comparison group prevents them from generalizing their results. Their 2002 report compares patterns of class size reduction across schools with time trends in student achievement by school. They conclude that "the statewide pattern of score increase in the elementary grades does not match the statewide pattern of exposure to CSR, so no strong relationship can be inferred between achievement and CSR" (Bohrnstedt and Stecher, 2002, p. vii).

Jepsen and Rivkin (2002) study trends at the school level in grade 3 test scores in California schools. They conclude that class size reduction

has led to modest improvements in test scores, and that class size appears to be more influential than standard measures of teacher qualifications available in the statewide database in determining student achievement.

A weakness of all of these California studies is that they cannot follow individual students over time, so that measures of class size and teacher characteristics at the grade or school level are only approximate measures of the actual classroom experience of each student.

Research from Texas

Recent research from Texas stands in stark contrast to what has generally been done for California. Unlike California, Texas has built a state testing system that explicitly tracks the test scores of individual students. Particularly relevant for our subsequent analysis are two recent manuscripts—Hanushek et al. (2001) and Hoxby (2000)—that find evidence from Texas that the average achievement of a student's peers in the same grade is related to the student's subsequent rate of achievement growth.

This sort of research is simply not possible statewide in California because no student-level data are released to researchers, and even state contractors are unable to link student achievement over time. Our current San Diego study uses a database that is obviously much smaller than the Texas dataset. But it shares the same advantage of linking student test scores across years. Further, unlike the Texas data system, it provides data on the individual classrooms in which each student studies, allowing for more precise tests of whether one's peers, class size, or the qualifications of one's teacher influence learning.

National Studies Using School Resources Aggregated to the State Level

Some recent analyses of the effect of school resources on achievement have used state-level measures of school resources. The results of these studies are quite divergent but tend to reach much more optimistic conclusions than much of the school- or classroom-level research. For example Grissmer et al. (2000) use data from each state that participated in the National Assessment of Educational Progress (NAEP) between 1990 and 1996. They model average test scores as a function of class

size, teacher education, teacher experience, and several other measures of educational resources. They find that class size variations explained more of the achievement gap than did variations in other measures of school resources, including teacher education and experience. Darling-Hammond (2000) examines NAEP data from 1990 to 1996 and finds that teachers' credentials and experience were the two most important factors explaining interstate variations in test scores, with class size being far less important. Klein et al. (2000) examine NAEP data from a slightly different set of years in the 1990s than do Grissmer et al. (2000) and find that Texas, which the Grissmer report ranks at the top of state school quality rankings, outpaced the national average in only one of four achievement tests they examined.[5] These conflicting results point to the limited value of using solely state-level data on school resources. Small changes in the specifications and time period can lead to very different results. Furthermore, these data do not capture the striking variations in achievement and resources across schools and districts, especially in a state as large and diverse as California.

School Resources, Educational Attainment, and Earnings

In addition to studying test scores, it is useful to examine whether school resources are related to the years of schooling students ultimately attain. Betts (1996) reviews this relatively small body of research and finds weak evidence that school resources affect educational attainment.

A third way to test whether school resources "matter" is to examine the relation between school resources and the earnings of students after they leave school and enter the labor force.

A number of studies have found a relation between adult males' earnings and school resources in their state of birth, but the literature is by no means unanimous (Betts, 1996). Betts (1995), Grogger (1996), and others show that when school resources are measured at the school actually attended, the relationship between school inputs and earnings is not statistically significant. Furthermore, the estimated effect of increased school spending on students' subsequent earnings is extremely

[5]For a critique of the Grissmer et al. (2000) and Klein et al. (2000) studies, see Hanushek (2001a, 2001b).

small. This is true regardless of whether one measures school resources at the school actually attended, in the district attended or whether one instead uses the person's state of birth to create a rough proxy for school resources.

How Representative Is San Diego Unified?

Given the rather mixed current state of knowledge that we have just described, the student-level database that we have built in collaboration with SDUSD offers key advantages for finding out what factors affect student achievement. But it is important to ask whether the San Diego Unified School District is in any way representative of schools statewide. This section addresses this question by examining student demographics, school resources, and test scores in SDUSD and California as a whole. Web Appendix C, available in the web version of this report at www.ppic.org, extends the analysis by comparing SDUSD with other large districts in the state and in addition provides detailed data comparisons. We draw data mainly from the California Basic Education Data System (CBEDS)—a survey of districts, schools, and teachers performed statewide each October.[6]

Student Demographics and Student Achievement

SDUSD is the second-largest district in California, after Los Angeles Unified School District. In 1999–2000 it enrolled 141,000 students.[7]

Figure 2.1 presents the ethnic mix of students in the district in the 1999–2000 school year for the district and for California public schools as a whole. Clearly, the district serves an ethnically diverse set of students. As is true of the other large districts in the state, SDUSD does not exactly match the ethnic and racial mix of students in the state as a whole. San Diego has significantly greater percentages of black and Filipino students than the state, a slightly smaller percentage of Hispanic

[6]Data that are not included in the CBEDS survey, such as percentage of students eligible for meal assistance, can be found at http://www.cde.ca.gov/demographics/ or http://data1.cde.ca.gov/dataquest/.

[7]In our statistical analysis below, our sample will include 123 elementary schools, 24 middle schools, 17 high schools, and five charter schools, the latter of which span various grades.

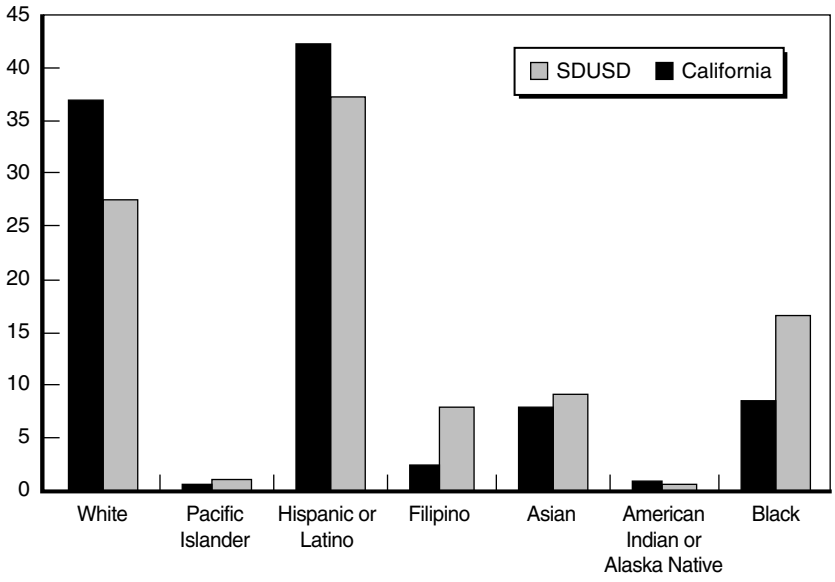


Figure 2.1—Student Percentages, by Race, 1999–2000

students, and in 1999–2000 roughly 9 percent fewer white students. The five largest districts have one thing in common: a far smaller share of students who are white than is found in the state as a whole.

An important measure of diversity within schools is the percentage of students who are English Learners. Another commonly used measure is the percentage of students eligible for free or reduced-price meals. This percentage is a widely used indicator of poverty in school populations. Figure 2.2 shows that in 1999–2000, SDUSD enrolled larger shares of students who were EL or who were eligible for meal assistance than did the state as a whole. This is typical of large urban districts in California.[8] Clearly, many students in SDUSD face significant challenges because

[8]Tafoya (2000) reports that in 2000 nearly 25 percent of all California public school students were English Learners. As Tafoya describes in more detail, schools assess the English language proficiency of students who speak a language other than English at home. Those who do not meet district fluency standards are identified as EL students. These students are tested periodically; once they reach fluency standards they are redesignated as Fluent English Proficient (FEP).

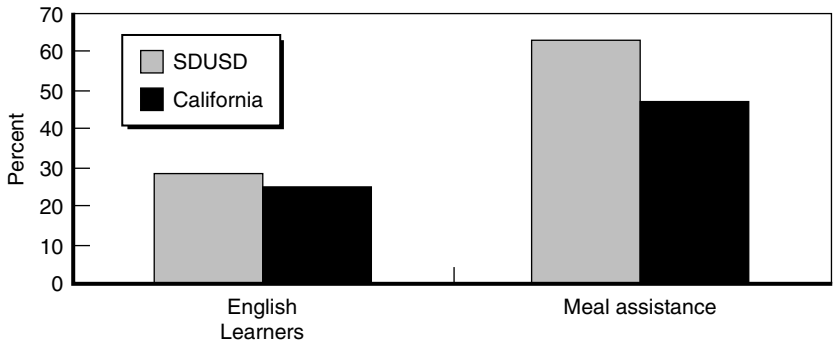


Figure 2.2—Percentage of Students Who Are English Learners and Who Are Eligible for Free or Reduced-Price Meals, 1999–2000

poverty and a lack of English language proficiency among students create barriers to learning.

What about student achievement? Beginning in spring 1998, California initiated a new state test, the Stanford 9, which has been given annually to all students in grades 2 through 11 since that time. The Stanford 9 is a standardized test that has been normed using a national sample of students. This provides a national performance yardstick against which California's students can be compared. Throughout this report we focus on math and reading scores on the Stanford 9. Our reason is simple: Although the Stanford 9 includes additional subject areas in certain grades, the math and reading tests represent the very core of educational achievement.

Figure 2.3 illustrates the reading results for San Diego and the state as a whole in 1999–2000. The figure shows the percentage of students in San Diego and California who exceeded the test scores obtained by the students ranked at the 75th, 50th, and 25th percentiles nationally in reading in 1999–2000. (If district students were identical to students nationwide, then exactly 25 percent, 50 percent, and 75 percent of district students should have exceeded each of these targets.) By these measures, district students were lagging very slightly behind national standards in 1999–2000. The figure also shows that in 1999–2000 reading achievement in the district closely matched that observed in the

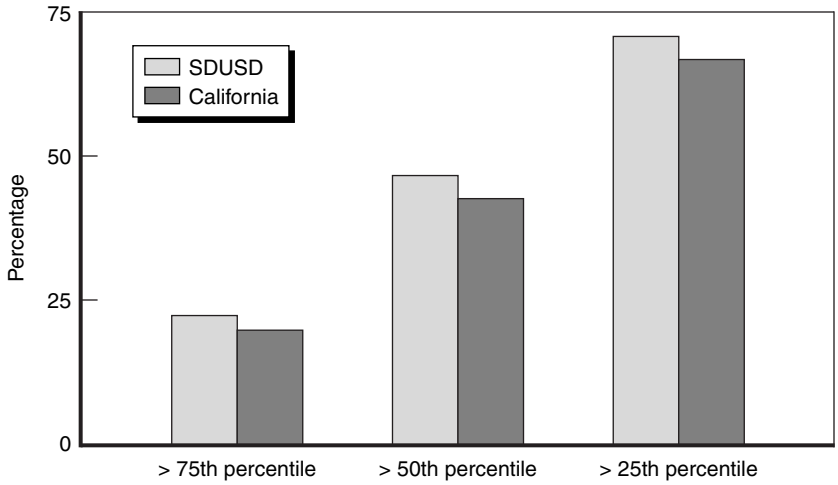


Figure 2.3—Student Performance Against National Norms in Reading, 1999–2000

state as a whole but was slightly higher. In 1997–1998, the first year of the new statewide test, the differences were even smaller, reflecting the fact that SDUSD has improved the reading achievement of its students slightly more quickly than did the state as a whole over this period.

What about trends in math achievement in San Diego? In 1999–2000, after the third year of testing, students in San Diego Unified performed better against national norms in math than in reading and in fact narrowly exceeded national norms in math. This finding is relevant for policy, because in fall 2000 SDUSD implemented an ambitious and controversial "Blueprint for Student Success," which devoted additional resources to students whose achievement lags behind. The blueprint calls for an initial emphasis on reading scores, which seems to be the subject area in greater need of reform.[9]

Table 2.1 provides more detail on test scores. It shows the percentage of students in each district and in California as a whole who exceeded the test scores obtained by the students ranked at the 75th,

[9]Web Appendix C provides a much more detailed analysis of test score trends in San Diego, the other large districts in the state, and the state as a whole.

Table 2.1

Stanford 9 Test Score Distribution: Unweighted Average Across Grades, All Students

	Reading			Math		
Percentile	1997–1998	1999–2000	Change, 1997–1998 to 1999–2000	1997–1998	1999–2000	Change, 1997–1998 to 1999–2000
California						
% > 75th	17.9	19.9	2.0	20.7	27.1	6.4
% > 50th	39.3	42.8	3.5	42.4	50.9	8.5
% > 25th	62.3	66.6	4.3	65.3	72.8	7.5
San Diego Unified						
% > 75th	19.4	22.4	3.0	22.0	29.2	7.2
% > 50th	40.8	46.4	5.6	44.7	54.0	9.3
% > 25th	63.5	70.4	6.9	67.5	75.5	8.0
Fresno Unified						
% > 75th	10.4	11.4	1.0	12.6	16.1	3.5
% > 50th	25.3	27.6	2.3	30.0	36.5	6.5
% > 25th	47.0	51.1	4.1	53.6	61.5	7.9
Long Beach Unified						
% > 75th	12.8	14.1	1.3	15.2	23.1	7.9
% > 50th	30.2	33.7	3.5	35.4	46.1	10.7
% > 25th	53.3	59.2	5.9	59.9	69.6	9.7
Los Angeles Unified						
% > 75th	8.6	9.7	1.1	10.9	14.2	3.3
% > 50th	22.8	25.8	3.0	27.4	33.4	6.0
% > 25th	44.7	50.6	5.9	51.5	59.1	7.6
San Francisco Unified						
% > 75th	21.4	22.3	0.9	32.9	36.9	4.0
% > 50th	43.9	47.0	3.1	55.0	60.4	5.4
% > 25th	67.8	71.8	4.0	73.8	79.0	5.2

50th, and 25th percentiles nationally in reading and math in 1997–1998 and 1999–2000. A comparison of results in the first year of the test, 1997–1998, clearly shows that San Diego Unified students more closely matched statewide averages than did students from any of the other large districts in California.

What all of these districts, and California as a whole, have in common is that in virtually all cases in 1997–1998, students in California lagged behind national norms in both reading and math. (San Francisco Unified remains an exception.) Betts, Rueben, and Danenberg (2000, Chapter 7) demonstrate that in 1997–1998, two-thirds to three-quarters of the achievement gap between California and the country as a whole reflects the preponderance of EL test-takers in California—about 20 percent, compared to about 2 percent in the national norming sample. Similar analysis in Web Appendix C shows that much of the gap between the achievement of students in San Diego and students nationally is related to the much greater than average percentage of EL students in the district.

School and Teacher Inputs

Another way in which San Diego may or may not be representative of schools statewide is in the level of school inputs available to students, measured in terms of class size and teacher qualifications. The pupil-teacher ratio in SDUSD matches the state average very closely. However, each of the five largest districts has a unique pattern of teacher qualifications that distinguishes it from the state average. For instance, SDUSD has a relatively high number of teachers with a master's degree and full credentials, but at the same time, SDUSD's teachers have less teaching experience than teachers elsewhere. Although none of the large urban districts has a mix of school and teacher characteristics that is exactly representative of schools statewide, SDUSD looks quite similar to the average throughout the state.[10]

---

[10]For readers interested in a more detailed comparison, Web Appendix C provides a detailed comparison of school and teacher characteristics among SDUSD, the state as a whole, and other large urban districts in California.

Conclusion

The existing research on school resources and student achievement does not suggest that there are strong and systematic effects of school resources on student learning. The strongest piece of evidence in this regard is the class size reduction experiment in Tennesee, although there the effects are modest to begin with and "wear off" in later grades. Overall, the results suggest a relatively weak relationship between school resources on the one hand and student achievement, educational attainment, and future earnings on the other.

The recent California studies that we reviewed all suffered in one or more regards, the one universal failing being that none followed individual students over time, while linking their gains in test scores to specific characteristics of the classroom and the student's teacher. This approach, although highly desirable, is simply not possible with California's current testing system.

Given the limits of California's education data, it becomes clear why a study of a large district that allows researchers to explore the determinants of achievement at the level of individual students can add much to our knowledge. Overall, San Diego appears to provide a district that is quite representative of patterns and trends statewide. Perhaps the greatest difference between SDUSD and the state's school system in general is that SDUSD has relatively more EL students and more students who are economically disadvantaged. We do not view either of these differences as a disadvantage.

Much of the achievement gap between districts in California reflects differences in students' economic disadvantage and language status. Similarly, most of the achievement gap between California and the rest of the nation reflects the unusually high share of English Learners in California. For both of these reasons, the concentration of economically disadvantaged students and English Learners in San Diego Unified makes it all the more interesting to study.

3. The Link Between Poverty and School Resources in San Diego Schools

As any parent knows, not all schools are equal. Betts, Rueben, and Danenberg (2000) document large variations across California in school resources such as teacher qualifications and the degree of rigor offered in the high school curriculum. In California, schools attended by disadvantaged students receive fewer resources. Teacher mobility appears to drive much of this pattern. That is, as teachers gain experience and enhance their teaching credentials, they tend to move to schools that have relatively advantaged students. This pattern has pivotal importance for education policy, given equity issues and public perceptions that resources make a difference in the quality of schooling that students receive. This chapter shows that SDUSD is no stranger to that pattern.

Dividing Schools on the Basis of Student Demographics

Eligibility for free or reduced-price meals is the most commonly used indicator of SES in education research. To analyze the link between poverty and school resources, for each year we divided students into five approximately equally sized groups or "quintiles," determined by the percentage of students receiving free or reduced-price meals at their schools. For instance, in the 1999–2000 school year, we divided elementary schools into five roughly equally sized groups, based on enrollment. The upper cutoff points for 1999–2000 were 35, 55, 78, 90, and 100 percent.

Although this chapter focuses on variations in San Diego school resources by student SES, to a significant extent the analysis also speaks

to racial gaps in school resources. This is because the percentage of students receiving free or reduced-price meals is strongly related to the racial makeup of the school. For instance, in quintile 1 of elementary schools, which has the lowest share of students eligible for meal assistance, about 48 percent of students in 1997–1998 were nonwhite, compared to almost 96 percent of students in quintile 5. Similar variations in the share of students who are not white appear in middle and high schools and across all years we examined.[1]

San Diego schools with greater percentages of students eligible for meal assistance also tend to have a greater share of students who are English Learners. For instance, in elementary schools in 1999–2000, EL students constituted 12 percent and 65 percent of school populations at the quintiles of schools with the lowest and highest shares of students eligible for meal assistance, respectively. The gaps are slightly less dramatic at middle and high schools because at these levels a smaller percentage of students are English Learners.

As expected, parental education is strongly related to meal assistance. Figure 3.1 shows the share of students at each quintile of high school whose more educated parent holds a bachelor's degree or higher. At the schools with the lowest and highest shares of students eligible for meal assistance (quintiles 1 and 5), 68 percent and 18 percent of students have at least one parent with a bachelor's degree or more, respectively. These percentages are slightly lower in middle and especially elementary schools, perhaps because students with less educated parents are more likely to drop out of high school.

Student Mobility, Student Retention, and Dropout Rates

One challenge for schools serving disadvantaged populations is that these students tend to be more geographically mobile. Students who switch unexpectedly between schools may suffer academically if the two

[1]Readers who are interested in learning more about the details are invited to read Web Appendix D. This appendix provides tables for elementary, middle, and high schools that document the discussion in this chapter and provide the data for the figures presented here.

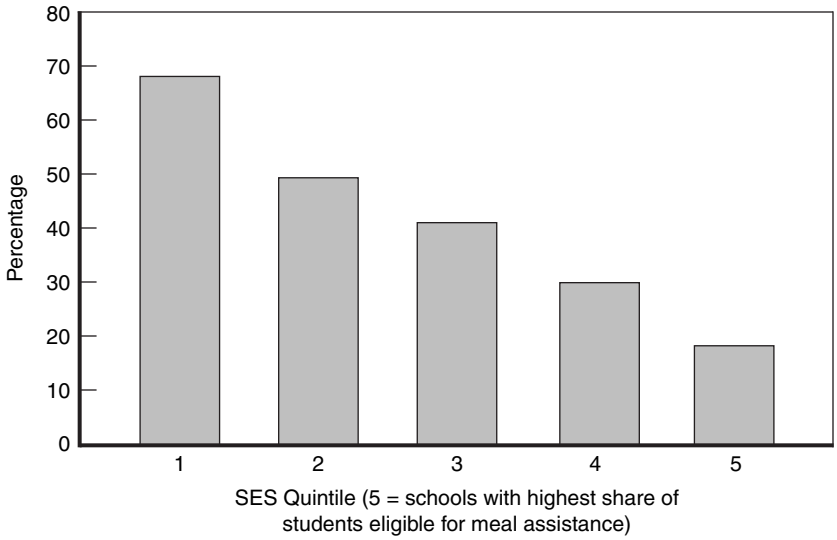


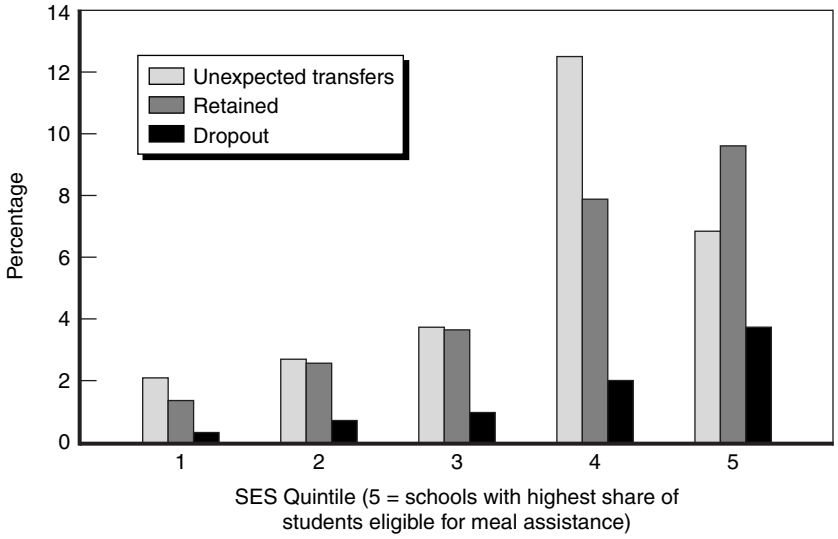
Figure 3.1—Percentage of San Diego High School Students Whose More Educated Parent Has a Bachelor's Degree or Higher and SES Quintile, 1999–2000

schools arrange their curricula differently. The resulting disruption can also affect students who remain at a school for several years but experience influxes of new students in their classrooms. To explore this issue, we developed a measure that indicates whether a student has switched schools unexpectedly. First, we labeled as "unexpected school switches" any midyear move between schools by a student. Second, we looked for unusual types of transitions between schools between the end of one school year and the start of the next school year. Expected school switches include the transitions between elementary and middle school, and middle and high school. We concluded that an unexpected transfer had occurred if: (1) The student was new to the school in the given year, and both (2) the student was not at the entry level grade of the new school, and (3) the student did not graduate from the prior school.

Two other relevant measures that affect student outcomes are the percentage of students who are retained between grades, that is, those who are not promoted to the next grade, and the percentage of high school students who drop out.

Figure 3.2 illustrates all three measures for high schools in each of the five SES quintiles. Some very strong patterns emerge. Schools with higher shares of students eligible for meal assistance in general have far higher rates of unexpected transfers of students into their schools. Schools serving more disadvantaged students also have sharply higher percentages of students who are retained a year or who drop out. For instance, in the high schools in the most affluent areas (quintile 1), fewer than 1 percent of students dropped out in 1998–1999, compared to almost 4 percent in the quintile 5 schools.

In the following section, we examine characteristics of the school that are best thought of as "purchased inputs" provided by the school district.



NOTES: Dropout percentage is calculated as the average percentage of dropouts in grades 10 through 12. Our data for this figure, unlike other figures in this chapter, refer to 1998–1999 because dropout data were not available in time for us to analyze patterns in 1999–2000.

Figure 3.2—Percentage of San Diego High School Students Who Are Unexpected Transfers, Retained a Grade, or Drop Out of School, and SES Quintile, 1998–1999

Class Size

Figure 3.3 shows average class size by SES quintile for elementary, middle, and secondary schools. In addition, because the class size reduction initiative in California reduces class size to 20 students or fewer in kindergarten through grade 3, we separate classes in these grades from those at higher grade levels in elementary schools.

The figure suggests that within grade spans, very little inequality in class size related to SES exists among San Diego schools. The most striking pattern in the figure is that class size rises considerably after the third grade. The figure suggests that in middle and high schools serving disadvantaged populations, class sizes are slightly smaller than average. Betts, Rueben, and Danenberg (2000) find a similar pattern statewide in California.[2]

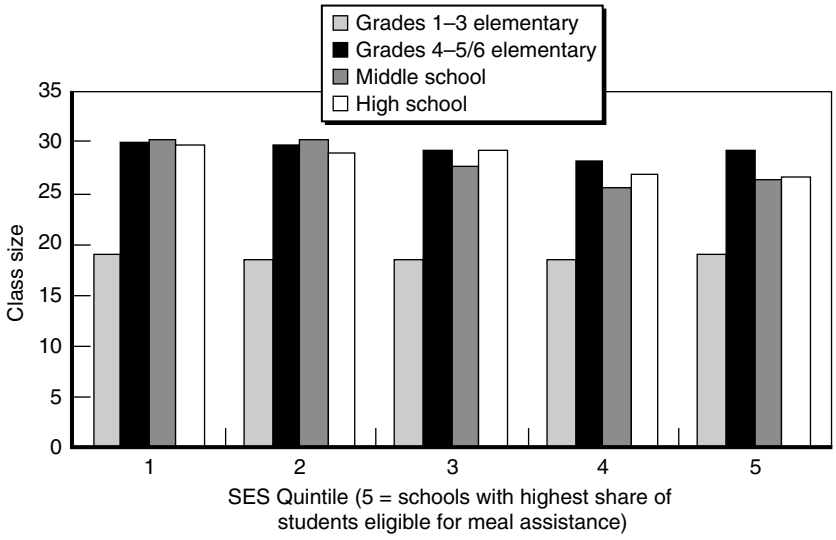


Figure 3.3—Class Size in San Diego, by Grade Span and SES Quintile, 1999–2000

[2]The data for Figure 3.3 take an average of class size across all academic subjects for middle and high schools. Tables D.12 and D.13 in Web Appendix D show that in schools with high percentages of students eligible for meal assistance, both English and math classes are somewhat smaller than in schools in more affluent areas of the city.

Teacher Characteristics

Figure 3.4 shows the percentage of teachers with a master's degree or higher in elementary schools in 1999–2000. The gap between schools is stark, with almost twice as many teachers in the most affluent fifth of schools holding a master's degree relative to teachers in the schools serving the most disadvantaged populations.[3]

Another way of gauging teachers' education is to ask whether they majored or minored in the subject that they teach. As Figure 3.5 shows, there appears to be less of a disparity among math teachers' education when measured this way than when measured by whether the teachers hold a master's degree. There is no clear linear relation between SES and the percentage of math teachers who majored in math. If anything, the middle-SES schools have fewer of these teachers than schools at either extreme of SES.[4]

The disparities in teacher education hint at the possibility that as teachers gain more experience and work toward their master's degree, they also tend to migrate toward schools that serve relatively advantaged students. This issue of teacher mobility can be examined more directly by looking at the distribution of teachers across schools by their level of

---

[3]Web Appendix D, Table D.14, shows the percentage of teachers with a master's degree or higher at the five SES quintiles for all three grade spans and years. The percentage of teachers with a master's degree or higher generally increases with each school level (that is, elementary, middle, and high schools) regardless of SES quintile. But within each grade span, low-SES schools employ a far smaller percentage of highly educated teachers. One interesting trend is a slight decrease in the percentage of teachers with a master's degree or higher at both the middle and high school levels among high-SES schools over a three- year period. This is in contrast to low-SES schools, which maintain roughly the same percentage over time.

[4]As shown in Web Appendix D, an interesting trend is that there appears to be a dropoff recently among high-SES high schools in the percentage of math teachers holding a degree in math and a corresponding increase in the low-SES high schools, to the point where in 1999–2000, math teachers at the low-SES schools actually were slightly more likely to hold a bachelor's degree in math. A second notable pattern, illustrated in Web Appendix D, Table D.16, is that there was quite a disparity between low- and high-SES schools in the percentage of English teachers who held a degree in English in 1997–1998, much more so than observed for math teachers. This is especially true at the high school level. However, over the three-year period, these inequalities have generally narrowed. In 1999–2000, middle schools in the middle-SES quintiles actually had a greater percentage of English teachers with a degree in English than did schools in the other quintiles.

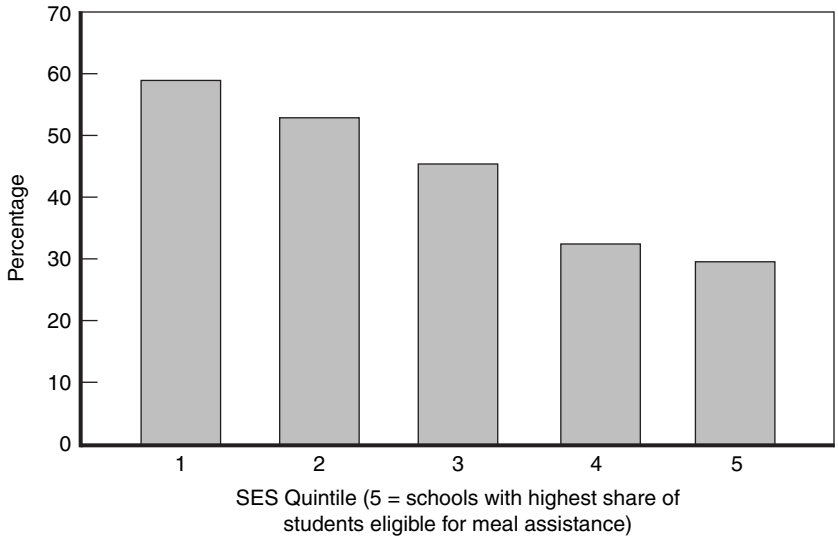


Figure 3.4—Percentage of San Diego Elementary School Teachers with a Master's Degree or Higher and SES Quintile, 1999–2000

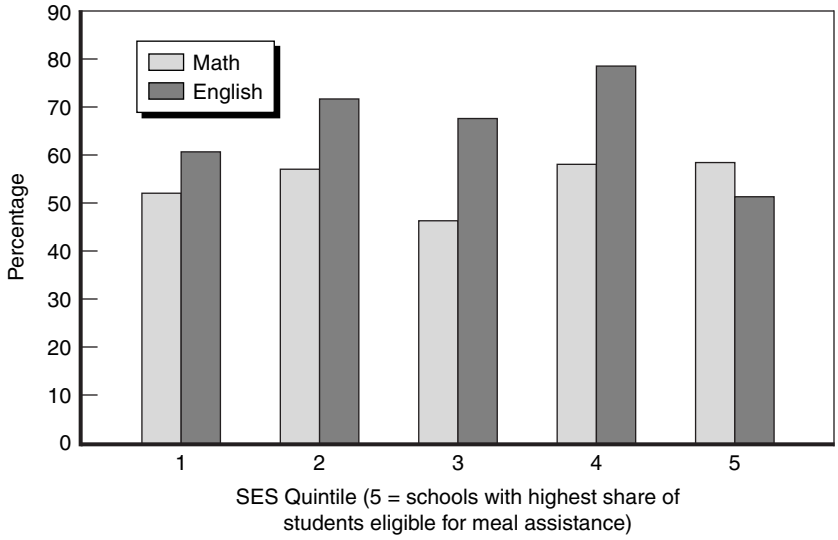


Figure 3.5—Percentage of San Diego High School Math and English Students Whose Teacher Majored in Math or English and SES Quintile, 1999–2000

experience. Figure 3.6 reveals strong relationships between student disadvantage and average teaching experience among elementary school teachers. As shown in Web Appendix D, the link between teacher experience and the SES quintile of the school is almost as strong in middle and high schools as it is in elementary schools. The difference between the highest- and lowest-SES schools can be as many as eight years of teaching experience on average.

These dramatic relations between teaching experience and student SES appear to be largely caused by the transfer of teachers from lower-SES schools once they have gained more experience. The district's collective bargaining agreement with teachers clearly outlines the "post-and-bid" method through which teaching vacancies are filled:[5]

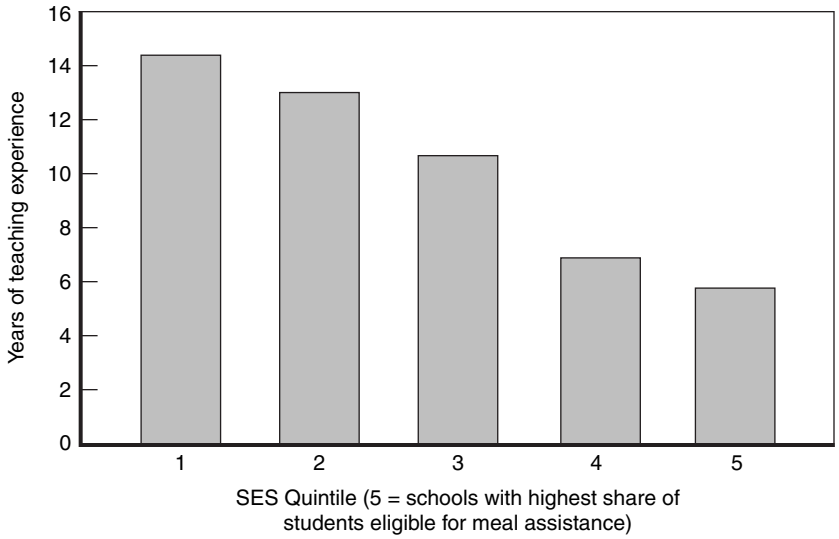


Figure 3.6—San Diego Elementary School Teachers' Average Years of Teaching Experience and SES Quintile, 1999–2000

[5]San Diego Unified School District and San Diego Education Association (2002). Betts, Rueben, and Danenberg (2000) report similar wording in the teachers' contract in force between 1998 and 2001.

> 12.2.5. Awarding of positions will be based upon the criteria specified in the posting. The Personnel Administration Department will certify that the unit member has the required major or minor or has completed the minimum legally-required number of units for majors and minors (currently the equivalent of twenty [20] semester units for a minor and thirty [30] semester units for a major), based on the unit member's transcripts on file with the District at the time of the closing of the posting period.
>
> 12.2.6. The District may interview and will select the unit member to fill the posted vacancy from the five (5) unit members who have the greatest district seniority, have bid for the position and have been deemed qualified by the Personnel Administration Department, Certificated . . . .[6]

The wording makes clear that school administrators must select from among the five most senior applicants whose qualifications match the job description. The priority that the post-and-bid system gives to teachers with seniority, combined with teachers' apparent preference to teach in schools in relatively affluent areas, generate the sharp variations in teacher experience across schools in the district. Although we cannot prove that these inequalities would lessen if the post-and-bid system were changed to allow schools to select freely from among applicants, it certainly seems likely that this is in some cases a binding constraint on schools.

Another measure of teacher preparation is credential status. To some extent, this is related to a teacher's years of teaching experience, and so we should expect to see some of the same disparities in teacher credentials that exist for teacher experience. Teachers with a full credential have taken a series of prescribed university courses and finished

---

[6]Section 12.2.6 also gives some limited preference to teachers who have a minor but not a major in the required field, at least for positions that have not received many applicants: "Unit members with an applicable minor may be considered for vacancies that receive less than five (5) qualified bidders with the appropriate required major under the following conditions:
12.2.6.1. Priority consideration shall not apply.
12.2.6.2. The District shall not be required to select a unit member with a minor even though he/she is included among the top five (5) most senior applicants."

The reference in the above text to "priority consideration" refers to teachers who have been laid off or otherwise declared "in excess." Such teachers must be interviewed for positions for which they apply if they fully meet the posted description of qualifications. See Section 12.1.9 of the collective bargaining agreement.

a teaching practicum that qualifies them to teach. Every district strives to have every teacher fully credentialed. SDUSD is no exception. At SDUSD, a teacher falls under one of three primary categories: full credential, emergency/waiver, or intern.[7]

At the high school and middle school levels, there is very little difference in the percentage of teachers who are fully credentialed across SES quintiles, although schools in disadvantaged areas do tend to have fewer fully credentialed teachers. The difference is larger at the elementary school level, with a 7 to 9 percentage point gap in the percentage of teachers with a full credential between the lowest- and highest-SES quintiles. For instance, in 1999–2000, in the lowest- and highest-SES quintiles of elementary schools, 91 percent and 99 percent of teachers held a full credential, respectively. This could signal a greater need for teachers overall at the elementary school level and hence the filling of positions through teachers with an emergency credential.[8]

A full credential signifies that a teacher has mastered basic teaching skills but does not guarantee that a teacher has the subject knowledge needed to teach a specific subject in a given grade. In middle schools and particularly high schools, districts aim to place teachers with a full subject authorization in academic classes such as math and English. These subject authorizations are quite distinct from the full credential: The former is awarded based on subject area mastery, and the latter is awarded based on provision of evidence that the teacher has mastered more general teaching skills. To obtain a full authorization, teachers must have completed a set of university courses prescribed by the California Commission on Teacher Credentialing (CCTC) in the relevant subject. Middle school teachers are not required to have formal subject authorizations to teach a specific subject. An alternative path for them is to teach using a multiple subject authorization that allows them to teach multiple subjects to the same group or groups of students.

---

[7]The most common full credential types include multiple subject, single subject, special education, and gifted education. It is quite possible for a teacher to hold both a full credential and an emergency credential. For example, a teacher who has started to teach special education may hold a full multiple subject credential but at the same time hold an emergency credential for teaching special education.

[8]Full details appear in Web Appendix D, Table D.18.

Nonetheless, we should expect that middle school teachers who hold a subject authorization in their subject, even though it is not required, have taken more university courses in their subject than middle school teachers with a multiple subject authorization.

Figure 3.7 shows the percentage of math and English teachers who hold a full authorization in the five SES quintiles of high schools in 1999–2000. The figure does not suggest a strong link between subject authorization and the percentage of students eligible for meal assistance. Middle school data show similarly weak patterns. The main exception was that in 1999–2000 in the most affluent and least affluent middle schools, the percentages of English teachers with a full subject authorization were 37 percent and 27 percent, respectively.

Tables in Web Appendix D also show disparate trends between middle and high schools. Over time, there has been no clear and universal increase or decrease in the percentage of middle school English or math teachers with a full authorization. In contrast, the percentage of high school English and math teachers with a full authorization rose

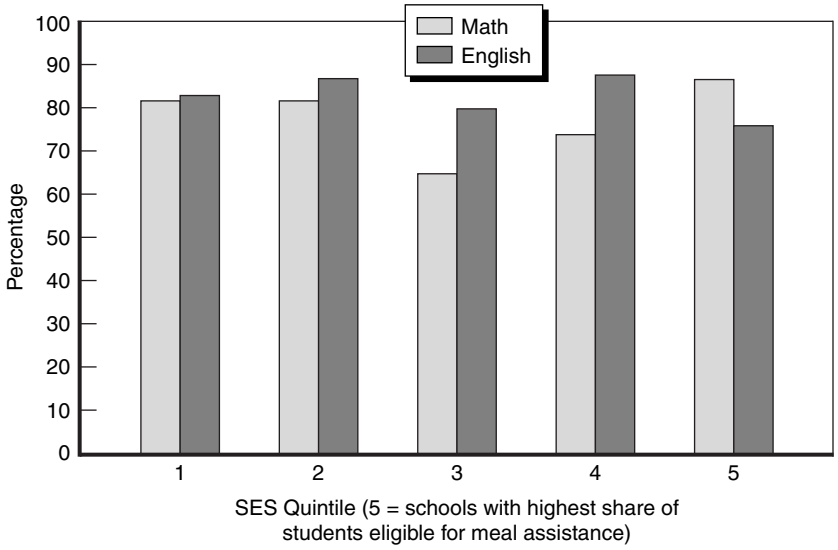


Figure 3.7—Percentage of San Diego High School Math and English Students Whose Teacher Held a Full Authorization in Math or English and SES Quintile, 1999–2000

considerably across the SES spectrum between 1997–1998 and 1999–2000.

Many teachers have taken quite a few university courses in the subject that they teach but not enough to satisfy the CCTC requirements for a full subject authorization. Often these teachers will qualify for a supplemental authorization. When we examine the percentage of teachers holding either full or supplemental authorizations (Web Appendix D, Tables D.21 and D.22), the same broad patterns are evident, with some variations in detail. For instance, there is again greater evidence of a link between student SES and teacher subject authorization in English than in math classes. Perhaps the most important revelation from these tables is that a large percentage of middle school teachers hold a supplemental subject authorization but not a full authorization. So, in middle schools it is important to look at both levels of subject authorization to get a better grasp on the share of teachers who have extensive college preparation in their subject. A second relevant finding, illustrated in Figure 3.8, is that the negative link between English teachers' subject preparation and student eligibility for meal assistance appears to be much stronger in middle schools when we examine the share of teachers who hold a full or supplemental subject authorization than when we look solely at those who hold a full authorization.

Another form of teacher authorization that is particularly important in such a multilingual society as California is the Crosscultural Language and Academic Development (CLAD) certificate, which prepares teachers to teach students who are English Learners. The district is currently encouraging all of its teachers to obtain the CLAD certificate, regardless of teaching assignment. In addition, it is becoming the norm for schools of education to require that teacher trainees obtain it. A closely related credential is the Bilingual CLAD (BCLAD) certificate, which certifies that a teacher is equipped to teach EL students in a language other than English. At SDUSD, almost all holders of the BCLAD certificate hold a Spanish BCLAD.

Because low-SES schools are associated with greater shares of EL students, we would expect to see a higher percentage of teachers who hold a CLAD or BCLAD certificate at low-SES schools. For instance, at the elementary school level, we find that 47 percent of EL students in the

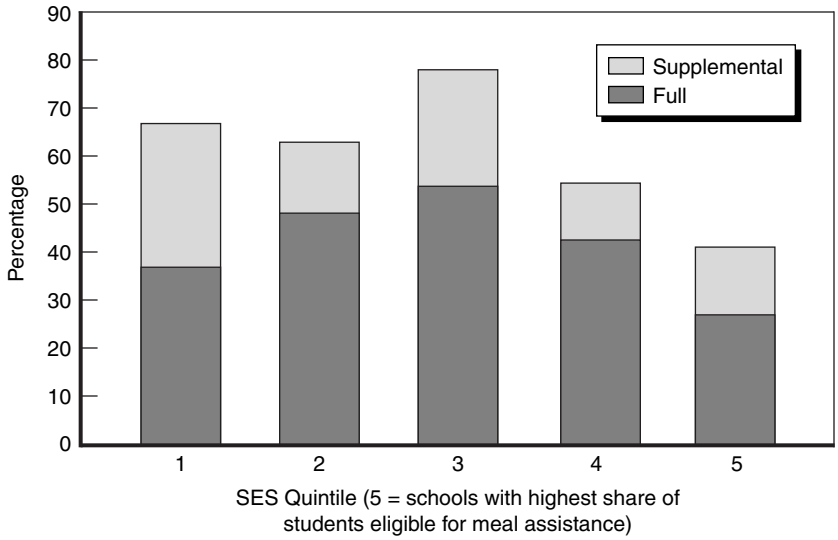


Figure 3.8—Percentage of San Diego Middle School Students Whose Teacher Held a Full or Supplemental Authorization and SES Quintile, 1999–2000

highest-SES fifth of schools and 63 percent of EL students in the lowest-SES fifth of schools had a teacher who had a CLAD or BCLAD.[9] At all grade spans, a higher proportion of teachers hold either a CLAD or BCLAD credential in the lower-SES schools. This appears to signal an appropriate allocation of these teachers, given our evidence that EL students are largely concentrated in the low-SES schools.[10]

---

[9]In Web Appendix D, Table D.23 illustrates the relevant data.

[10]The CLAD and BCLAD program supplements or supersedes earlier programs designed to prepare teachers to teach English Learners. There exist equivalent credentials that are very close in curriculum to the CLAD/BCLAD, but they are not standardized and cannot really be considered exactly equivalent. Web Appendix D Table D.24 replicates Table D.23 but in addition includes teachers who do not hold CLAD or BCLAD certificates but do hold CLAD and BCLAD equivalents. The results show that very few middle or high school teachers hold a CLAD or BCLAD equivalent but that substantial percentages of teachers in high-SES elementary schools hold these equivalents. The concentration of (B)CLAD-equivalent teachers in high-SES elementary schools may reflect the teacher mobility patterns discussed above, whereby more experienced teachers, who are more likely to hold a CLAD equivalent, have migrated over time to schools in more affluent areas. However, the opposite appears to occur in high schools, where teachers with (B)CLAD equivalents are centered at low-SES schools.

Conclusion

This chapter has explored differences in student demographics and school resources between schools with various levels of student eligibility for meal assistance. The results on demographics are straightforward: As one should expect, schools with above-average percentages of students eligible for meal assistance tend in proportional terms to enroll more minorities, more English Learners, more students with relatively less educated parents, more students who transfer unexpectedly from another school, more students who are held back or retained a year, and more students who drop out of school. In many senses, meal assistance serves as a proxy for the many facets of student disadvantage.

The link between student SES and school resources is quite complex. Some measures of school resources, mostly linked to teachers, are highly skewed. For instance, teachers at the higher-SES elementary schools have two and a half times as many years of teaching experience, are twice as likely to hold a master's degree or higher, and are 10 percent more likely to hold a full credential than are teachers at the lower-SES schools.

Other measures of school resources show less strongly positive correlation with students' socioeconomic status. Most notably, class size varies little across schools and, if anything, is slightly smaller in the low-SES schools. The percentage of math and English teachers who hold a full subject authorization in middle and high schools is slightly skewed, usually in favor of the high-SES schools, but the gaps often narrowed between 1997–1998 and 1999–2000. Perhaps the biggest gap in teacher subject knowledge that remained by 1999–2000 was among middle school English teachers. In the high-SES schools, 67 percent of English teachers had taken enough university courses to qualify for either a full or supplemental authorization, compared to only 41 percent in the lowest-SES schools. Finally, schools serving disadvantaged populations appropriately had as large or larger percentages of teachers equipped with CLAD or BCLAD certificates that help to prepare teachers to work with English Learners.

Clearly, it would be inaccurate to claim that students eligible for meal assistance attend schools that receive relatively fewer resources of every type, such as class size and teacher education, certification,

experience, and credentials. But important and dramatic disparities exist in regard to some of the most important indicators of teacher preparation. Teachers who migrate from low-SES to higher-SES schools as they gain experience appear to drive this pattern. The district's collective bargaining agreement, which guarantees open teaching slots to one of the five most senior applicants who meet the job qualifications, probably compounds the inequality in teacher preparation across schools. The gap in teacher qualifications between affluent and less affluent areas is an important factor to bear in mind, given our analysis in the ensuing chapters of the distribution of student achievement across the district's schools.

4. Trends in Student Achievement in San Diego

Introduction

This chapter examines recent trends in student achievement in San Diego Unified, both overall and by individual student subgroups. We disaggregate students using the level of economic disadvantage among the students at their school, student race, language status, and gender.

Some of these ways of disaggregating students bear on the idea that students in some groups begin their schooling at an educational disadvantage. For instance, students who are English Learners may be at a disadvantage learning in an English-only classroom. Similarly, those who live in poorer neighborhoods are likely to have had fewer preschool experiences that would prepare them well for a school environment.

But it is far from clear what trends in achievement differences we should expect over time. Consider, for example, students in low-income areas. Perhaps the most likely hypothesis is that disadvantaged students begin their elementary school years less well prepared and fall further behind as they progress through school. This hypothesis seems particularly likely because of the sharp inequalities in certain school resources between schools with large numbers of disadvantaged students and schools with fewer disadvantaged students that we documented in the previous chapter. Alternatively, disadvantaged students may arrive at elementary school roughly equally prepared as students in higher-income areas, only to fall behind because their homes do not provide the same sort of learning environment. For instance, the homes of students in lower-income families may lack books, magazines, an adequate study space, a computer, and an encyclopedia. Yet another hypothesis is that disadvantaged students start school at a lower level of achievement and neither fall behind nor catch up with their better-off peers. Finally,

disadvantaged students could conceivably have lower achievement initially but gradually catch up over time. This chapter addresses these alternative hypotheses and comes up with some surprising answers.

Although information on average student performance at each school is published annually by the state Department of Education, what we present in this chapter is distinct in a number of important ways from the reports that are publicly available.

First, newspaper accounts on trends in Stanford 9 test scores typically report *average* achievement at each school. This approach misses the fact that the identity of students taking the test each year will change, so that we cannot be sure whether achievement in a school—or the district as a whole—is truly improving. For instance, a common use of the data provided by the Department of Education is to compare grade 2 achievement at a specific school this year and last year. Obviously, we cannot know for sure whether any change has occurred because of changes in school quality or, rather, because of changes in the underlying characteristics of the grade 2 classes in those two successive years, such as the education level of parents of the children, eligibility for meal assistance, and race and ethnicity. A second approach that attempts to solve this problem is to compare achievement of this year's grade 3 class at a given school with the achievement of last year's grade 2 class. Although an improvement, this method is also flawed because it ignores the fact that at most schools some students leave and others arrive during the course of a year. So a drop in achievement between this year's grade 3 class relative to last year's grade 2 class could arise either if some of the highest-achieving grade 2 students left the school this year or if some low-achieving students entered the school in grade 3 this year. These problems are particularly likely to arise in inner-city schools that typically have relatively high rates of student mobility.

As a solution to these problems, this chapter and the subsequent chapters that statistically model test scores will focus not on school averages but instead on *individual* student gains in achievement from one year to the next. In addition, in this chapter we focus on students for whom we have math or reading test scores for spring 1998, spring 1999,

and spring 2000. This allows us to paint a consistent picture of the rate at which students are learning.[1]

A second departure from many of the reports on school performance in the California press is to focus on students' scaled scores rather than their national percentile rankings. The latter measure is a number between 1 and 99 indicating the number of students out of 100 whom the student would beat or match. This way of calculating student scores allows summary comparisons of the percentage of students who are at or above the 50th percentile of national norms. The California Department of Education website provides this information for every school in the state annually; this information was featured in Chapter 2. Although these percentile rankings are useful for comparing school districts against a national yardstick, they are less useful for what really matters: the rate at which students *improve* over time.

Scaled scores provided by the test-maker, Harcourt Brace, provide a measurement system that is specifically designed to measure a student's increase in knowledge from one year to the next. In addition, the test publisher has scaled these scores to ensure that "a difference of 5 points between two students' scores represents the same amount of difference in performance wherever it occurs on the scale" (Harcourt Brace Educational Measurement, 1997, p. 17). The test questions also vary by grade level so that the subject matter gradually becomes more difficult, allowing the test to provide information on student achievement across a range of grades.

For these reasons, this and subsequent chapters will focus almost exclusively on gains in mean scaled scores by individual students over

---

[1]About 70 percent of students have test scores in all three years. This mainly reflects mobility in and out of the district. In addition, the test is offered only in grades 2 through 11, so that in this chapter we do not use test scores for younger students who have had only one or two years of tests by spring 2000 or for students who were in grades 10 or 11 in spring 1998. However, we compared the spring 1998 test scores of those who took the test for three consecutive years and those who had missing test scores in either 1999 or 2000. Initial test scores of these two groups were in all grades within 1 percent of each other, which provides convincing evidence that our sample is quite representative of students in the district as a whole.

time. The analysis below will focus on two-year achievement gains rather than one-year gains, because two-year gains will allow us to be more confident that the trends we observe are due to true changes in student achievement rather than to random events that might have reduced students' performance in any one year, such as a flu bug.

Overall Trends in Achievement Gains Between Spring 1998 and Spring 2000

Tables 4.1 and 4.2 show initial spring 1998 mean scaled scores by grade, and the rise in these scores in spring 1999 and spring 2000, for reading and math, respectively. For example, the first row in Table 4.1 follows the cohort of students who were enrolled in grade 2 during the 1997–1998 school year. Their mean scaled score was 576.76, and over the next two years the mean gain by this set of students was 64.91 points. In our sample, the 25th, 50th, and 75th percentiles of grade 2 reading scores in 1998 were 542, 574, and 608. So, there is tremendous heterogeneity within grade 2 achievement, but the average student improved quite quickly. Within two years the average grade 2 student,

Table 4.1

Mean Scaled Scores by Year for All Students, Reading

Grade	Mean Score, 1998	Mean Score, 1999	Mean Score, 2000	Mean Gain, Year 1	Mean Gain, Year 2	Mean 2 Year Gain
2	576.76	616.48	641.67	39.72	25.20	64.91
3	606.94	636.07	653.15	29.13	17.08	46.21
4	629.65	651.10	664.43	21.45	13.33	34.78
5	648.01	662.88	679.21	14.86	16.34	31.20
6	660.72	676.55	692.41	15.83	15.86	31.69
7	676.59	691.02	694.60	14.43	3.58	18.01
8	689.31	691.58	698.04	2.27	6.46	8.74
9	693.24	697.94	704.58	4.69	6.65	11.34
10	695.32	701.64		6.32		

NOTES: Sample consists of students who had test scores in spring 1998, spring 1999, and spring 2000. For grade 10 students, the sample consists of students who took the reading test in both grades 10 and 11. The grade shows the initial grade of students in spring 1998.

Table 4.2

Mean Scaled Scores by Year for All Students, Math

Grade	Mean Score, 1998	Mean Score, 1999	Mean Score, 2000	Mean Gain, Year 1	Mean Gain, Year 2	Mean 2 Year Gain
2	573.83	609.09	630.20	35.26	21.11	56.37
3	598.99	624.39	649.57	25.40	25.18	50.58
4	617.76	646.27	664.64	28.51	18.38	46.88
5	644.30	663.41	672.88	19.11	9.47	28.58
6	661.49	672.80	681.27	11.31	8.47	19.78
7	673.34	681.43	698.64	8.09	17.21	25.29
8	679.86	695.39	701.27	15.53	5.88	21.41
9	696.35	702.20	709.88	5.85	7.68	13.53
10	698.46	706.16		7.70		

NOTES: Sample consists of students who had test scores in spring 1998, spring 1999, and spring 2000. For grade 10 students, the sample consists of students who took the math test in both grades 10 and 11. The grade shows the initial grade of students in spring 1998.

who by then was in grade 4, scored a 641 in reading, which put him or her considerably above the 75th percentile of achievement in grade 2.[2]

Table 4.1 suggests that, as expected, in spring 1998, students in higher grades on average were more proficient at reading than were students in lower grades. However, the biggest gaps in achievement from one grade to the next occurred among the lower grades. The implication is that students develop their reading skills most quickly in the lower grades. The two-year gains bear this idea out, with students initially in the lower grades typically improving substantially more than those initially in the higher grades. Table 4.1 illustrates the rapid declines in rate of improvement across grades.

Table 4.2, showing results for math scores, tells a similar story: Although students in higher grades had demonstrably higher math proficiency than students in earlier grades, achievement gains were far higher in elementary school than in later years. These patterns may

[2]To put this in further perspective, the minimum possible reading score for the grade 2 form of the Stanford 9 is 359, and the minimum for math is 370.

reflect the fact that in the higher grades, teachers devote less attention specifically to reading and math skills and more to subject matter in diverse subject areas.

We must of course exercise some caution here: Just because math and reading scores rise more slowly in higher grades does not necessarily imply that students are not learning effectively in high school. Rather, the explanation may be that the Stanford 9 test is better matched to the subject matter taught in the lower grades than in middle and especially high school. Still, the overall pattern strongly suggests that most improvement in math and reading skills comes while students are relatively young.[3]

Variations in Improvement Across Schools and in Particular Student Groups

Although the overall patterns and trends in learning are already clear, perhaps of more interest to parents and policymakers alike are the gaps in initial achievement and in learning across schools and groups of students. This section addresses this crucial question in various ways: the level of economic disadvantage at the school attended, student race, language status, and gender.

SES Quintile of the School

The previous chapter showed large differences in resources among the schools in the five quintiles of student eligibility for meal assistance. Just how far behind were students in schools serving the most economically disadvantaged students in 1998, the first year of the new state testing program? And since that time, have students in these schools fallen behind, held their own, or caught up with students in more affluent neighborhoods?

---

[3]As another word of caution we note that it is not appropriate to compare scaled scores or gains in scaled scores between the reading and math tests, as these scales do not measure achievement in the same way. In other words, a gain of 60 scaled points in math compared to a gain of 70 scaled points in reading does not necessarily mean that a student is learning less in math than in reading.

Figures 4.1 through 4.4 provide some startling answers.[4] Figures 4.1 and 4.2 show initial patterns in achievement by school SES quintile in spring 1998 for reading and math, respectively. As one would expect, economic status "matters." There is a clear negative relation between the quintile and initial student performance: Without exception, students who attended schools in a more disadvantaged quintile on average scored lower in 1998 than did those attending schools in a more affluent quintile. For instance, the bottom dotted line in both figures shows the mean scaled scores in 1998 by grade for students attending the fifth of schools with the highest percentage of students eligible for meal assistance.

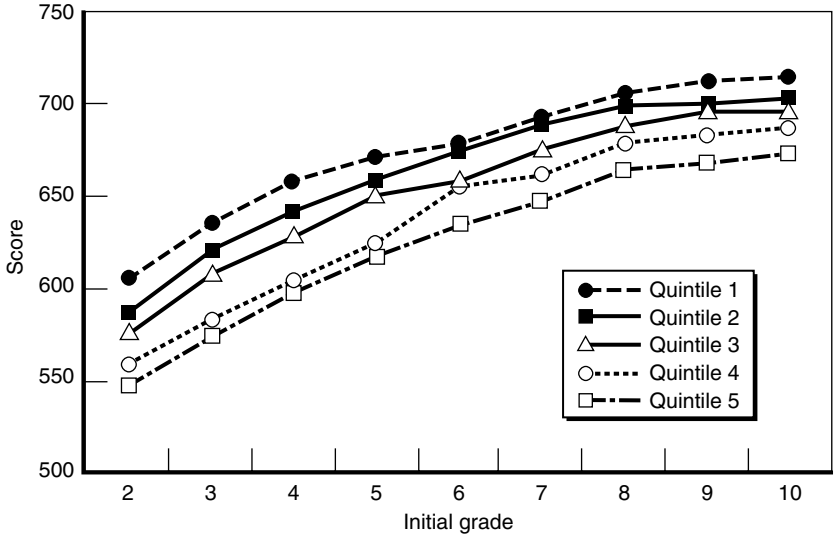


Figure 4.1—Spring 1998 Reading Scores, by SES Quintile of School and Grade (5 = schools with highest share of students eligible for meal assistance)

[4]Throughout the rest of this chapter we will present results graphically. However, the interested reader can find the underlying mean scaled scores and gains for each subgroup discussed in this chapter in Web Appendix E. The tables in Web Appendix E follow the order in which we discuss subgroups in this chapter.

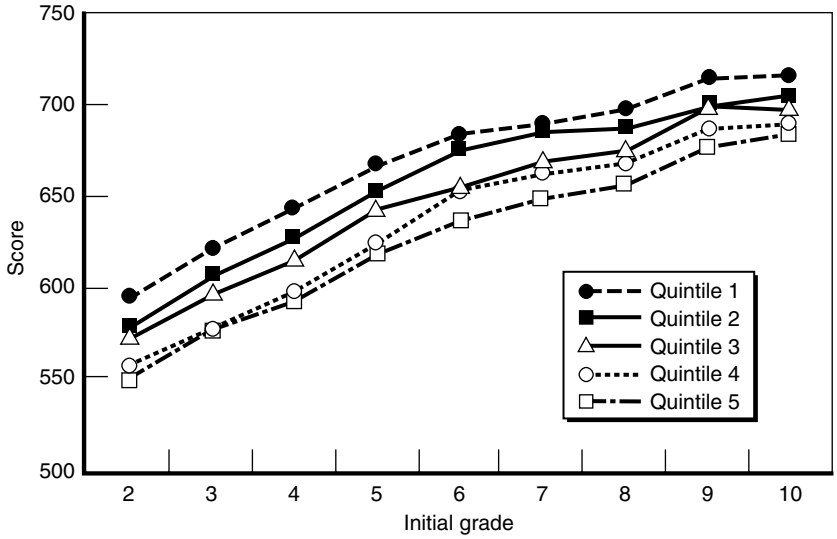


Figure 4.2—Spring 1998 Math Scores, by SES Quintile of School and Grade
(5 = schools with highest share of students eligible for meal assistance)

To some readers, the size of the gaps may appear shocking. For instance, Figure 4.2 shows that the average scaled math score of students in the most affluent fifth of schools in grade 2 was about 600.[5] In the most disadvantaged fifth of schools, depicted by the bottom line in the figure, students in grade 4, two grades more advanced, still had not on average reached this level of achievement.[6]

Because gains in achievement slow down in the higher grades, the same exercise at higher grades suggests that students in disadvantaged schools fall even further back in terms of "number of grades behind" by the middle and high school years. For instance, in quintile 1 schools, the grade 6 mean scaled math score in spring 1998 was about 685. In the most disadvantaged schools, quintile 5, it is not until grade 10 that the

[5]The exact score was 596.08, as shown in Web Appendix E.

[6]For an analysis of the link between poverty and the percentage of students at or above the 25th, 50th, and 75th percentiles of national norms in spring 1998, see Mehan and Grimes (1999). Their analysis of grade 10 performance shows similar patterns to what we describe here.

mean scaled score reaches this level. This implies a gap in math achievement of four years between the quintiles of schools serving the most and least disadvantaged students.

Although these gaps in achievement are extremely large by any objective measure, it is important to realize that this pattern is in no way unique to SDUSD. In nationally representative datasets, it is almost always the case that variation in achievement within a grade dwarfs the average growth across grades. For instance, Betts (1998), using the Longitudinal Study of American Youth, shows that, depending on the grade, from 26 to 40 percent of middle and high school students have not reached the median math test score of students enrolled two grades below, simply because of the huge variation in achievement within any grade level. This heterogeneity among students, combined with the fact that poverty is one of the strongest predictors of student achievement, means that the above results for SDUSD should come as no surprise to those who analyze achievement data on a regular basis. Still, the gaps among schools are large and should be of vital concern to policymakers.

What about trends in achievement gains over time? Figures 4.3 and 4.4 show two-year gains in reading and math scaled scores for students who were initially enrolled in the stated quintiles of schools, respectively. Here, the results suggest a somewhat optimistic interpretation: Students in all SES quintiles of schools show significant gains in achievement and, if anything, the achievement of students enrolled in schools serving the most disadvantaged populations improved the most. For example, Figure 4.3 shows that without exception students initially at the most disadvantaged quintile of schools (quintile 5) exhibited the largest two-year gains in reading achievement between 1998 and 2000, whereas students in the more advantaged quintiles and in particular quintile 1, the most advantaged quintile, generally showed the least improvement. In other words, students at low-SES schools generally narrowed the absolute gap in reading achievement over time. Figure 4.4 reveals a very similar pattern for math scores.

How big was the narrowing in the achievement gap between students attending the most and least disadvantaged schools? For reading, taking a simple average across grades, we find that the average

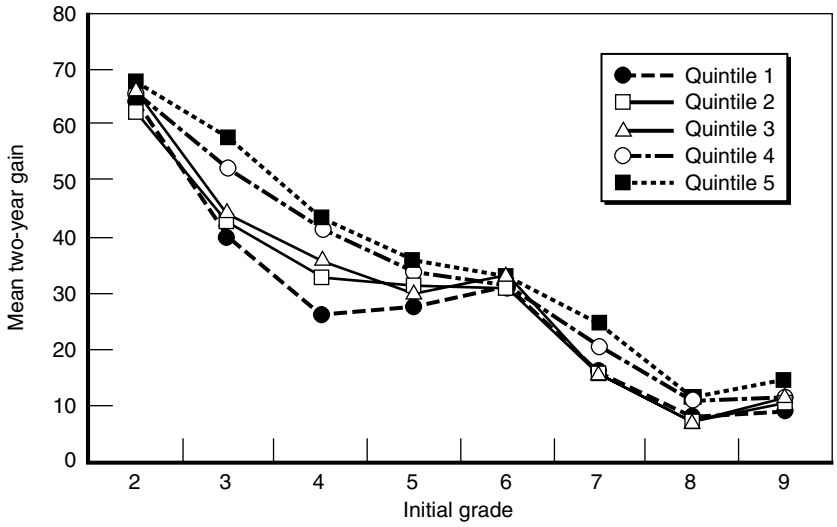


Figure 4.3—1998–2000 Gains in Reading Scores, by SES Quintile of School and Grade (5 = schools with highest share of students eligible for meal assistance)

narrowing of the gap between 1998 and 2000 was 8.2 scaled score points. This translates to an average narrowing of the initial 1998 achievement gap of 15.2 percent. The two cohorts for which the achievement gap narrowed the most were students in grades 3 and 4 in spring 1998. For math, the results are very similar, with the average gap in scaled scores narrowing by 5.1 scaled score points, which represents a narrowing in the initial achievement gap of 11.1 percent. The greatest narrowing occurred for students initially in grades 8, 3, and 4.[7]

In sum, the data point to very large initial gaps in achievement between students at the most and least disadvantaged schools. However,

[7]The finding that students in schools serving disadvantaged students have caught up over time may seem to conflict with the earlier finding that in spring 1998 in higher grades disadvantaged students were more years, or grade-equivalents, behind than in the lower grades. The main explanation for this apparent discrepancy is that rates of gains in student achievement slow down in the higher grades, so that a gap of x scaled score points will translate into more years of learning in the upper grades. Comparing the absolute gap in scaled scores across grades, as we do here, is a more appropriate way to compare achievement across groups, as the scaling of raw test scores is explicitly designed so that a gap of x scaled points anywhere on the distribution represents the same absolute gap in achievement.

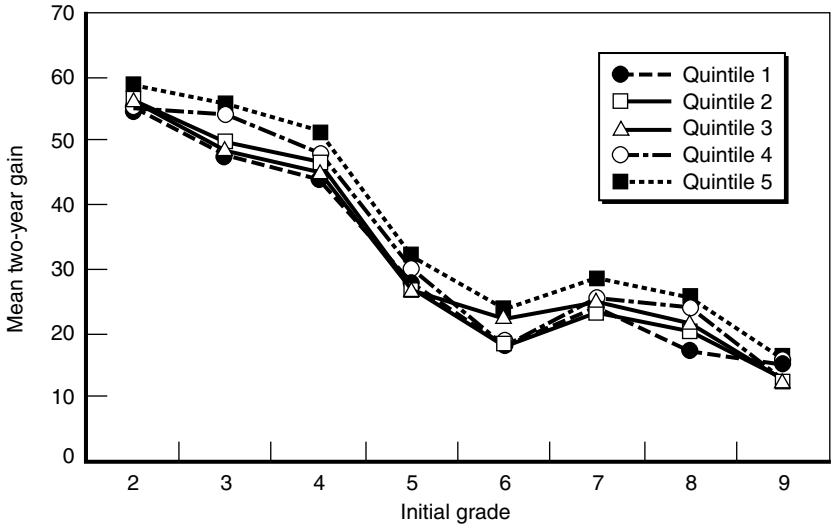


Figure 4.4—1998–2000 Gains in Math Scores, by SES Quintile of School and Grade (5 = schools with highest share of students eligible for meal assistance)

between spring 1998 and spring 2000 this achievement gap narrowed by over 10 percent for both math and reading.

As we are about to show, this pattern of large achievement gaps between more and less disadvantaged groups of students, but with significant reductions in these gaps over time, appears repeatedly when the students of SDUSD are divided in different ways.

Student Race and Ethnicity

Another way to analyze gaps in achievement is to plot trends separately for students of different races and ethnicities. This approach is of great policy relevance, because San Diego hosts important and well-organized parent groups, representing both black and Hispanic communities, among others, that take a great interest in disparities in learning across schools and racial groups.

Figures 4.5 and 4.6 show initial 1998 test scores in reading and math by ethnicity. White students had by far the highest achievement in all grades, followed by Asian/Pacific Islander students. Black and Hispanic

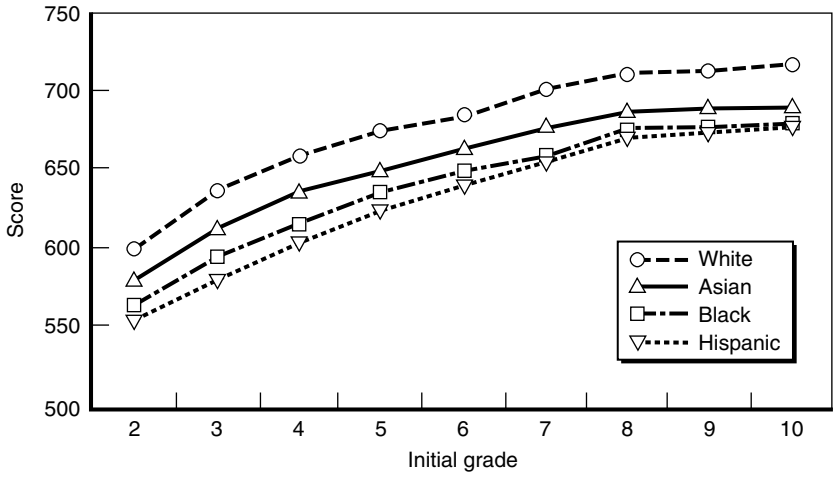


Figure 4.5—Spring 1998 Reading Scores, by Ethnicity and Grade

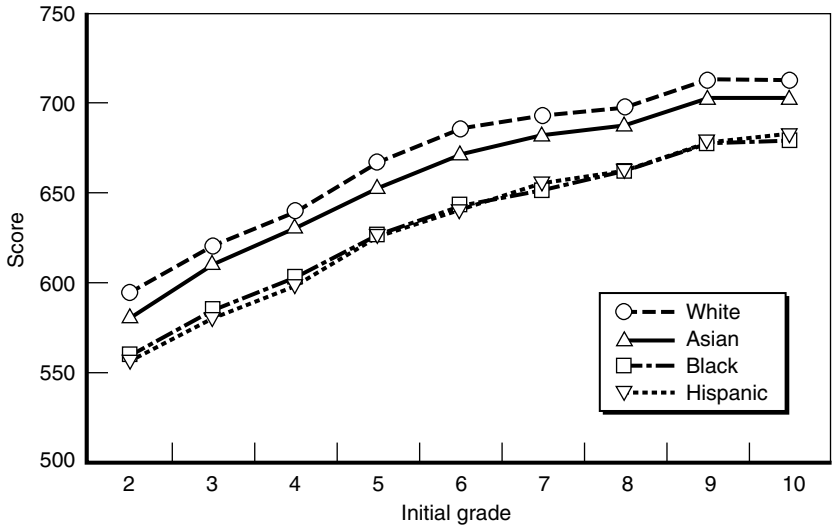


Figure 4.6—Spring 1998 Math Scores, by Ethnicity and Grade

students had the lowest achievement in all grades and were on average quite similar to each other. Notably, the gaps in test scores in each grade between whites on the one hand and blacks and Hispanics on the other are roughly as large as the gaps in average achievement between students attending the lowest- and highest-SES schools, as shown above.

Figures 4.7 and 4.8 show the two-year gains in scaled scores in reading and math by racial/ethnic group. The most obvious point told by these graphs is that rates of growth vary sharply by grade, and all ethnic and racial groups show similar overall growth and variations by grade. But looking more closely, one sees that in general white students showed the smallest two-year gain in test scores, with nonwhites generally but not always increasing their test scores to a greater extent over the two-year period.

When we calculate the percentage reduction in the gap between white test scores and scores of each other group and then average these across grades, we find that between 1998 and 2000 the ethnic reading achievement gap dropped by an average of 13.9 percent for Hispanic students, 13.1 percent for Asian students, and 6.7 percent for black

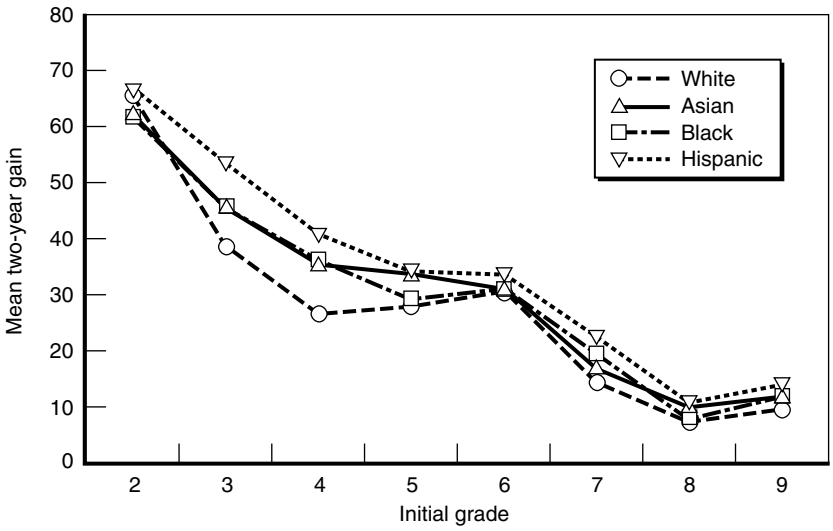


Figure 4.7—1998–2000 Gains in Reading Scores, by Ethnicity and Grade

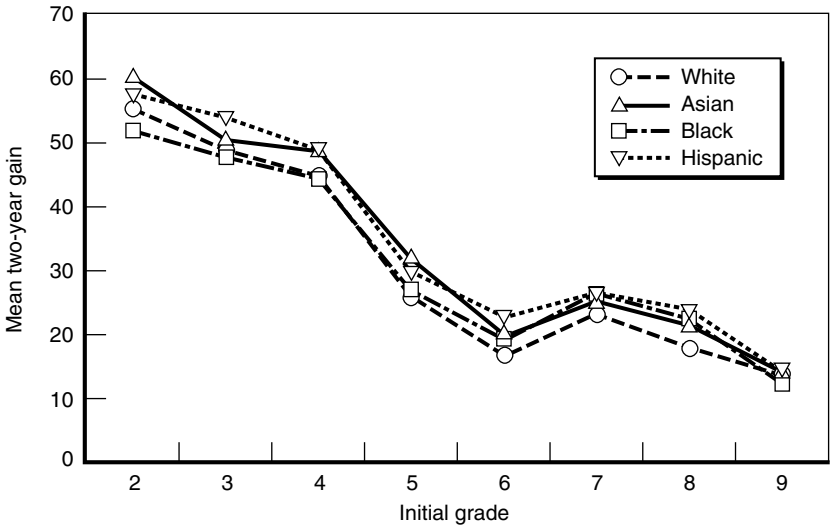


Figure 4.8—1998–2000 Gains in Math Scores, by Ethnicity and Grade

students. There were only two cases where the gain in scaled reading scores was higher for white students than for a minority in a given grade. This occurred in grade 2 for both blacks and Hispanics.

Turning to the racial/ethnic gaps in math achievement, a similar but slightly different story emerges. Averaging across grades, the Hispanic-white math test-score gap narrowed by 9.2 percent, the Asian gap narrowed by 24.8 percent, and the black gap narrowed by only 0.9 percent. For Asians and Hispanics, the gap narrowed in all grades. However, the black-white test-score gap narrowed in only four of eight grades and overall hardly changed.

We conclude that in 1998 large gaps in test scores existed between whites and other racial/ethnic groups, and that between 1998 and 2000 these gaps declined. However, the black-white gap decreased by smaller amounts and less uniformly across grades than did the Hispanic-white and Asian-white achievement gaps.

English Learners vs. Non-English Learners

Another way to view disparities in achievement is to compare EL to non-EL students. Again, a picture emerges of large initial gaps that

become smaller as we follow individual students' progress between 1998 and 2000. Table 4.3 ilustrates the reduction in test score gaps. When we calculated the simple average across grades, we found that between 1998 and 2000, the average EL/non-EL achievement gap in reading and math shrank by 15.5 percent and 15.4 percent, respectively.

We conclude that the narrowing achievement gap between Hispanics and whites is in part related to the fact that EL student scores have risen more quickly, but the story is more complex than that—the racial narrowing also occurs among English-proficient Hispanic students and white students.[8]

Table 4.3

Percentage Reduction in Test Score Gaps, 1998–2000

Groups Being Compared	Reading	Math
SES quintiles 1 and 5	15.2	11.1
Hispanic-white	13.9	9.2
Asian-white	13.1	24.8
Black-white	6.7	0.9
EL/non-EL	15.5	15.4

NOTES: These percentage reductions are based on test scores of individual students who had test scores in spring 1998, spring 1999, and spring 2000. The numbers represent a simple average across grades.

Male vs. Female Students

A long series of research reports have examined male-female differences in learning. See, for instance, Stumpf and Stanley (1996), Allred (1990), and Nowell and Hedges (1998). Probably the most robust findings from this literature have been that girls' math and science achievement sometimes trails that of boys whereas the opposite sometimes arises in language arts.

We found far less evidence of either the existence of an achievement gap or its narrowing between genders than there is among racial, language, and socioeconomic groupings.[9]

---

[8]Refer to Web Appendix E for further discussion and figures.

[9]Refer to Web Appendix E for further discussion and figures.

Summary

Table 4.3 summarizes our results on the extent to which achievement gaps between various groups have changed over time. The table reports the percentage reductions over the 1998–2000 period in achievement gaps between students at the bottom and top SES quintiles of schools, between whites and other racial/ethnic groups, and between EL and non-EL students, averaged across grades. In almost all cases, the reductions are sizeable.

Conclusion

For the students of SDUSD taken as a whole, trends and patterns are quite simple to summarize. Student achievement increases between grades, but as we switch our attention from elementary school to higher grades we find that the achievement gains between one grade and the next become much smaller. This may reflect the fact that in higher grades, teachers devote less attention specifically to reading and basic math skills and more to subject matter in diverse subject areas. As for time trends, between spring 1998 and spring 2000, test scores for individual students rose considerably.

These are interesting findings, but perhaps the most relevant policy question concerns the achievement gaps among various groups of students and whether these gaps have widened or narrowed. We found that in 1998, students who were attending schools with higher than average shares of students eligible for meal assistance had markedly lower reading and math achievement than did students attending schools in more affluent areas. Similarly, we found large achievement gaps between Hispanics and blacks on the one hand and whites on the other, with Asians/Pacific Islanders in-between but in general scoring much closer to whites. A similarly large achievement gap exists between EL and non-EL students. Perhaps understandably, the achievement gap between English Learners and other students is slightly larger in reading than math. In contrast, we found relatively little evidence of achievement gaps between boys and girls. Similar patterns appear to exist in other school districts around the country.

We hypothesized at the start of this chapter that the achievement gap between disadvantaged and less disadvantaged students could well have grown over time, given the evidence in Chapter 3 that teachers at low-SES schools typically have less than average experience and other qualifications. Somewhat surprisingly, we found the opposite to be true: Between 1998 and 2000 the reading and math achievement of students attending low-SES schools improved more quickly than did that of students in the highest-SES schools in every grade. We found similar evidence of a narrowing achievement gap when we compared trends in the achievement of white students and students from other races and ethnicities. The main exception was the black-white achievement gap, which did narrow, but much more weakly than for other minority-white comparisons. We also observed this same pattern of narrowing achievement gaps when comparing English Learners to other students in the district.

By all of these measures, inequality in student achievement narrowed in SDUSD between spring 1998 and spring 2000. What makes this finding more notable is that it came despite robust growth in achievement for even the top-achieving groups.

We have purposely avoided attempting to explain the underlying cause for the apparent reduction in inequality in test scores in the district. It is impossible to know from the simple calculations that we have performed thus far. Indeed, some readers may find it puzzling that achievement has grown most in the lowest-SES schools—the very ones that tend to receive the least qualified teachers. We are fairly certain that this pattern is not unique to SDUSD. Betts and Danenberg (2002) analyze trends in the Stanford 9 test at schools in California and find that the schools with the most students eligible for meal assistance have witnessed the largest increases in the share of students performing in the top half or top three-quarters of national norms, even though statewide these schools tend to employ relatively inexperienced and less educated teachers than other schools, as is the case in San Diego.

The remaining chapters of this report examine paradoxes such as these, by statistically estimating the effect of highly specific measures of

school and classroom characteristics on the achievement gains of individual students in the district. The main task that we address in the ensuing chapters is deceptively straightforward: What determines individual students' rate of learning in San Diego Unified?

5. Determinants of Gains in Student Achievement in Elementary Schools

Introduction

This chapter presents results from our statistical analyses of the gains in individual elementary school students' reading and math achievement between spring 1998 and spring 2000. We postpone discussion of the corresponding results for middle and high schools to the next chapter, primarily because the number of classroom characteristics that we need to consider in these higher grade spans is significantly larger. The elementary school analysis here will provide a good introduction to the analysis that follows for middle and high schools. We estimate separate models for all students (including EL students) and EL students by themselves. The latter models are useful, given the large number of English Learners in San Diego and throughout California and the large gap in achievement between EL students and students fluent in English, as documented above.

Overview of the Procedure for Statistically Estimating the Determinants of Gains in Student Achievement

The richness of the data available for this study provides an unusual opportunity to estimate the relative importance of class size and teacher characteristics in determining the rates at which student test scores rise.

A first highlight of the procedure is that we model *changes* in individual students' test scores over time, rather than *levels* of achievement. This approach is extremely useful, because the level of a student's score in a given grade reflects not only the quality of instruction

he or she received in that grade, but in addition the quality of education he or she received in earlier grades, not to mention learning experiences provided in the home from the time the student was very young. By modeling gains in test scores between grades, we can credibly link improvement in a student's achievement in a given year with the educational environment of the student in that same year.

A second advantage of our estimation method is that it allows us to take account of unobserved but fixed characteristics of students, schools, and neighborhoods that might confound the analysis. To give just one simple example, suppose that some students innately learn more quickly than other students, and that these "fast learners" typically get placed into larger classes than other students. If we lacked a way to control for these unobserved variations across students, we might incorrectly infer that larger class size "causes" faster rates of learning. Our solution to this genre of problem is to include fixed effects for each student, each school, and each student's home zip code as well as for the grade level and the year in which the test was given. This in effect removes all of the inter-student, inter-school and inter-zip-code variation from our data. The inclusion of student fixed effects is particularly important. In practice what it means is that we are inferring the effect of a given variable, such as class size, based on year-to-year variations in class size experienced by the individual student, instead of relying on variations in class size among students. Appendix A presents in more detail the general approach that we take in estimating the determinants of student achievement, focusing on a nontechnical description of the statistical precautions taken and the reasons why they are so valuable. A more technical description of the estimation technique is presented in Appendix B.

Variables Included in Models of Gains in Test Scores

We list the set of explanatory variables that are used in all of our models to explain gains in test scores below. We provide this list in stages to convey the information more clearly.

Table 5.1 summarizes the set of student, family, and neighborhood variables incorporated in the models. When we include fixed effects for

Table 5.1

Student, Family, and Neighborhood Controls Used in the Statistical Models

Student characteristics
Fixed effects for each student to control for all characteristics of a student that are fixed over time, such as race; controls for the student's test score in the given subject last year; for students who changed schools that year, switched schools unexpectedly, or were new to the district; for age and grade level; for language status (EL, FEP, non-Spanish EL, non-Spanish FEP); for special education; for students who skipped a grade that year and retained a grade that year; and for percentage of days absent. Indicator variables are also included for students who skipped a grade that year, unexpectedly, or were new to the district.
Family characteristic
Controls for the level of education of the more highly educated parent.
Neighborhood characteristic
Fixed effects for student's home zip code.

NOTE: FEP = Fluent Engish Proficient.

each student, we in effect subtract the mean of a variable for a given student from the observed values for the student in each year. For this reason, characteristics of students that are fixed over time, such as gender and race, drop out of our models. We included those characteristics in the appendix regression results that do not include student fixed effects, and readers interested in finding out what additional variables we included in those models, such as student race, can find the answers in Web Appendix F, which shows the regression results for elementary school students.[1]

Most of the variables listed in this table are self-explanatory. We included the student's lagged test score because we found strong evidence

[1]Some readers may be surprised to see that we can include controls for the education level of the more highly educated parent in these regressions, even though we include student fixed effects. We can do this because parental education actually varies for many students between 1998 and 2000, and so is not completely fixed. Parts of these changes probably reflect genuine increases in parents' education over this period. However, much of the variation for each student reflects measurement error. (Parental education level is gathered each year during the administration of the state test and is either provided by students or listed by teachers in lower grades. District officials repeatedly warned us that these data are subject to measurement error. For this reason we do not emphasize parental education in any part of this report.)

that a student's gain in achievement tended to be large if his or her prior spring's score was unusually low, and vice versa. This might reflect in part "regression to the mean," where some random occurrence, such as a flu bug, depresses test scores one spring, virtually guaranteeing an unusually large gain in achievement for the student next year as he or she rebounds.

The dummy variables for school changers are of two types. Students who "changed schools that year" quite literally moved from one school to another in the middle of the school year. A student who "switched schools unexpectedly" moved from one school last year to a new school this year in a way that does not conform to the normal exit and entrance grades for the schools. (See Chapter 3 for the exact definition used.) We include these measures to test the hypothesis that switching schools in one of these ways can be disruptive for the student. We also include the percentage of days that a student is absent during the school year, to test the idea that "time on task" is positively related to achievement gains.

Table 5.2 provides a list of characteristics of the school, classroom, and student body that all of our regression models take into account. The student body characteristics include the percentage of students eligible for meal assistance, percentage breakdowns by race and ethnicity, and controls for student mobility similar to those we defined for individual students.

As shown in the table we control for many characteristics of the student's teacher(s), including highest degree obtained, the subject in which the teacher majored (English, math, social science, science, foreign language, and other majors, with education being the omitted or comparison category), and the teacher's minor, if any. Because of a lack of teachers with a minor in education, for the minor our omitted or comparison category consists of those with a minor in education, "other minor," or no minor. In some cases, teacher major and minor were not available. We also include information on whether the teacher was an intern or held an emergency credential to test whether these teachers are more or less effective than those who hold a full credential. We also compare the effectiveness of teachers who have either 0–1, 2–5, or 6–9 years of teaching experience with that of more experienced teachers. Because the effect of experience could depend on the type of credential,

Table 5.2

School, Classroom, and Student Body Controls Used in the Statistical Models That Include Both EL and Non-EL Students

School characteristics
Fixed effects for each school to control for all fixed characteristics of the school. Controls for whether the school was a year-round school.
Student characteristics at the school level
Percentage eligible for free or reduced-price meals; separate controls for percentage of students who are Hispanic, black, Asian, Pacific Islander, native American; percentage of students who are EL and percentage FEP; controls for student mobility: percentage who changed schools that year, who switched schools unexpectedly, and who were new to the district.
Student characteristics at the grade level
Mean test scores in previous spring's test of all students in the student's current grade, standardized to district average.
Classroom and teacher characteristics
Class size; controls for teacher characteristics: interactions of credentials (intern, emergency credential, full credential) with indicators of years of teaching experience (e.g., 0–1, 2–5, 6–9); master's degree, Ph.D.; bachelor's in math, English, social science, science, language, other major (except education) (separate variables for each major); corresponding controls for minors by field except that the omitted group is teachers with a minor in education or other; CLAD, (Spanish) BCLAD, CLAD alternative credential, BCLAD alternative credential, interactions for the last four dummy variables with two student indicators for language status (EL or FEP); controls for teachers who are black, Asian, Hispanic, other nonwhite, and female.
Average student characteristics in the classroom
Mean test scores of students in the class in previous spring's test, standardized to district averages.

we interact the experience variables with the full, emergency, and intern variables. The omitted or comparison group is teachers with a full credential and more than nine years of teaching experience.

To summarize, we control for many characteristics of teachers. A nonzero coefficient on any of these variables indicates a difference in the rate of learning between students with a teacher who has the stated characteristic and the "comparison teacher," that is, the omitted type of teacher. The comparison group is teachers with a bachelor's degree in education, a full credential and ten or more years of experience, with no language certification such as a CLAD, and either no university minor or a minor in "other" or education.

As mentioned in Chapter 3, the CLAD and BCLAD credentials prepare teachers to instruct students whose first language is not English, in a setting that is either immersion or Sheltered English on the one hand or bilingual on the other. We also include controls for teachers who hold alternative credentials to the CLAD and BCLAD.[2] Because English Learners and perhaps Fluent English Proficient students are likely to gain more from having a teacher in the classroom who holds a CLAD or BCLAD, we also interacted each of these teacher credential measures with both our EL indicator and our indicator for FEP students.[3]

Of particular importance are our measures of the average achievement of students in the classroom and the entire grade at the student's school. We include these variables to test for a possible influence of academic peers on the individual student's rate of learning. It certainly seems possible that the average achievement of a student's colleagues in the classroom, or more broadly in his or her grade at the school, could set the tone of the learning environment. These measures were standardized so that an increase of one unit in either academic peer group measure would represent a one-standard-deviation increase in test scores relative to the district average. Appendix B provides more details on these two measures. As shorthand, we will refer to these variables as classroom peer achievement and grade-level peer group achievement, respectively.

Given the large number of explanatory variables in our models, we will focus mainly on those that appear to be related to gains in student achievement in a statistically significant fashion. For example, suppose that we found that the coefficient on the dummy variable to indicate

---

[2]Before the CLAD and BCLAD certificates were introduced, Senate Bills 1969 and 395 provided for alternative language certification procedures for teachers, and some earlier programs existed as well. We have learned from district staff that these earlier certification methods typically did not require as many college courses as do the CLAD and BCLAD certificates. Nonetheless, the district does employ teachers who possess some of these precursor language qualifications, and it is important to account for these alternatives.

[3]In rare cases where a variable is missing, we set the variable to zero. But we also include in all of the regressions dummy variables set to one if the corresponding variable, such as a teacher's major in college, is missing.

teachers with a master's degree is 0.05. Does this reliably tell us that students' test scores grow by 0.05 points more during that grade if they are taught by a teacher with a master's degree rather than a teacher with a bachelor's degree (the "omitted" or comparison group)? The answer is that it depends. It is always possible that the true coefficient is exactly zero but that because of randomness in the model, occasionally the coefficient that we estimate could be as high as 0.05. Generally, statisticians claim that this coefficient is "significantly different from zero" if there is only a small probability, given our estimate of the coefficient, that the true coefficient is zero. Accordingly, we list cases in which key coefficients are statistically significant at either the 5 percent or 1 percent levels. This means that there is only a 5 percent or 1 percent probability, respectively, that the coefficient is truly zero.

Results

Most of the student-specific measures such as race and gender are fixed over time and so are removed from the model by the student fixed effect. However, it is worth noting that we find evidence that student absences are negatively and significantly related to gains in both reading and math achievement. Similarly, the student's own lagged test score is negatively and strongly significantly related to the student's current-year gains, suggesting the presence of regression to the mean.

We now turn to the external environment faced by the student. We present tables summarizing the statistical significance of all the key variables and then examine the size of the estimated effects on student learning. Those readers interested in seeing the regression results directly can find them in Web Appendix F.

The Effect of Demographics of the Student Body and Peers' Initial Test Score

Table 5.3 begins by summarizing the statistical significance of key variables describing the demographic characteristics of the student body at the school, and the average achievement levels of students in the specific student's grade or classroom. For instance, a "++" indicates that the coefficient appears to be positively related to gains in test scores, at

Table 5.3

Statistical Significance of Demographics of the Student Body and Average Initial Test Scores in the Student's Classroom and Grade in Elementary School Models

	Gains in Reading		Gains in Math	
Variable	All	EL	All	EL
% of students eligible for meal assistance				
% of school black				
% of school Asian			-	-
% of school Hispanic				
% of school Pacific Islander			-	
% of school Native American				
% of school EL			+	
% of school FEP	+		++	
Grade-level peer achievement			++	++
Classroom peer achievement			+	++

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

the 1 percent level. A "-" indicates a negative relationship that is significant but only at the 5 percent level, whereas a blank entry indicates that the listed variable is not statistically significant. The table shows results from the reading and math regressions for all students and for English Learners separately.

The first finding of note from this table is that schools with lower SES (with large percentages of students eligible for meal assistance), those with large numbers of nonwhite students, or those with large numbers of

English Learners are in no case associated with lower individual achievement gains in reading. The models of gains in math achievement are similar, although the percentage of Asian or Pacific Islander students is negatively associated with individuals' gains in math achievement, and schools with more EL and FEP students are associated with larger gains for the individual student.

By far the most striking pattern in the table, though, appears in the final two rows of the table: The individual student's classroom peer achievement and the grade-level peer achievement are positively related to the individual student's test score gains at the 5 percent level or better for math, for all students, and also for the subsample of EL students. The implication is that an individual student's progress in math is very much influenced by the initial achievement of students around him or her in both the classroom and his or her grade. This influence could work through the direct effect a student's peers have on his or her own effort or through the quality of help that classmates can give. In addition, these effects could be mediated indirectly through the choices that teachers make about *how* to teach based on the initial level of subject mastery of students in the grade and the classroom within the grade. For the reading models, neither peer score variable is statistically significant although, as shown in Web Appendix F, in all cases the coefficients are positive.

The Effect of Class Size and Teacher Credentials, Experience, and Education

Table 5.4 summarizes the extent to which class size and detailed measures of teachers' qualifications are significantly related to student learning.

The first row shows that class size appears to be significantly negatively related to gains in reading achievement for all students as well as for the sample of EL students. This is what we might intuitively expect, as larger classes may be harder to teach. We did not find that class size was statistically significant in the math models, although the coefficient on class size was negative in the models for all students and EL students, as shown in Web Appendix F.

Table 5.4

Statistical Significance of Class Size and Teacher Qualifications in Elementary School Models

	Gains in Reading		Gains in Math	
Variable	All	EL	All	EL
Class size	--	--		
Interns with 0–1 years of experience			--	--
Interns with 2–5 years of experience				
Teachers with emergency credential and 0–1 year of experience	++		++	
Teachers with emergency credential and 2–5 years of experience				
Teachers with full credential and 0–1 year of experience				
Teachers with full credential and 2–5 years of experience				
Teachers with full credential and 6–9 years of experience		--		
Master's degree			+	
Ph.D. degree				

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

The statistical import of teacher credentials and experience is not particularly strong. Consider first whether the teacher holds a full credential, is an intern, or holds only an emergency credential. Because these certification levels vary systematically with teachers' experience level, we interacted these credentials with the total years of teaching experience as shown in the table. In each case, we are comparing teachers of a given credential and experience level to teachers with a full

credential with ten or more years of experience.[4] *In very few cases did we find any statistically significant difference between the effectiveness of fully credentialed teachers with ten or more years of experience and teachers with less experience, regardless of whether they held a full or emergency credential or an internship.*

There are exceptions. In both the reading and math models for all students, teachers with an emergency credential and 0–1 year of experience are associated with larger gains in achievement than the comparison group of teachers (fully credentialed with ten or more years of experience). It is not clear why this might be the case. One possibility is that relatively inexperienced teachers might have been better positioned to design their teaching protocols around the state test than more experienced teachers who have devoted years to fine-tuning their teaching methods.

There are two cases with the more intuitive result in which less experienced teachers are associated with lower gains in achievement. First, students who had intern teachers with 0–1 year of experience improved their math scores significantly more slowly than did students taught by a fully credentialed teacher with ten or more years of experience. This was true for the sample of all students and EL students. Second, reading score gains were significantly and negatively related to whether the teacher held a full credential with 6–9 years of experience, again with the comparison group being fully credentialed teachers with ten or more years of experience.

What about teacher education? For math achievement gains, we found evidence that teachers with a master's degree are more effective than teachers with a bachelor's degree. We found no significant link for reading, although for both the samples of all students and EL students the coefficient on the master's degree variable was positive.

Another type of teacher certification apart from the full/emergency/intern categorization is the various credentials designed to prepare

---

[4]In regressions not included in this report, we repeated the elementary, middle, and high school models using fully credentialed teachers with six or more years of experience as the omitted comparison group. The results were little changed.

teachers to instruct students who are English Learners. Table 5.5 summarizes our findings in this regard. This table differs slightly from Tables 5.3 and 5.4 in that we present results separately for non-EL, non-FEP students and for EL students, rather than for *all* students and EL students as in the earlier tables.[5]

Table 5.5

Statistical Significance of Teacher's CLAD, BCLAD, and Alternative Certifications in Elementary School Models

	Gains in Reading		Gains in Math	
Variable	Non-EL, Non-FEP	EL	Non-EL, Non-FEP	EL
CLAD credential			--	
CLAD-equivalent credential				
Spanish BCLAD credential		-	--	
Spanish BCLAD-equivalent credential		N/A		N/A

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being non-EL, non-FEP students and English Learners. Because the sample of all students included interactions between the teacher credentials listed above and indicators for whether the student was EL or FEP, in the second and fourth columns above we are able to report the effect of these credentials on all students who were neither EL nor FEP.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

N/A = the coefficient could not be estimated because no students in the sample had a teacher with a Spanish BCLAD-equivalent credential in one year and a teacher without this credential in another year.

A similar lack of observations cannot explain why CLAD and CLAD-equivalent credentials appear to have no effect on EL students. 29 percent and 14 percent of observations in the EL sample had a teacher with a CLAD credential or a CLAD-equivalent credential.

[5]The results for non-EL, non-FEP students come from the models that include all students. Because those models interact the language credentials with students' EL and FEP status, the coefficients on the credentials without any interactions measure the effect on students who are neither EL nor FEP, in other words native English speakers and others who were never identified by schools as needing accommodation in English. The EL results come from models run on the EL subsample only.

EL students' learning in both reading and math appears to be unrelated to whether their teachers held a CLAD or equivalent credential. We found some weak evidence that EL students improved their reading scores less if their teacher held a Spanish BCLAD.

There is no obvious direct reason why any of these credentials would necessarily affect teachers' ability to improve non-EL, non-FEP students' rate of learning. In most cases we found no link, although teachers with a CLAD or Spanish BCLAD credential were associated with weaker gains in math for these students.

Tables 5.4 and 5.5 omit many of the characteristics of teachers that we included in the model, such as the teachers' gender, race, and major and minor in college (with majors in education and minors in education or "other" and those with no minor as the comparison group for majors and minors, respectively). Table 5.6 shows the results for teachers' major and minor. In general, these variables were not linked to student learning in reading in a statistically significant way. But a few important exceptions arise. In the model of reading achievement gains for all students, the variable indicating whether the teacher minored in English during his or her bachelor's program is positively related to reading gains at the 5 percent level, whereas the opposite was true for graduates in science. We found in the math model that numerous teacher degrees were associated with better gains for students than the comparison teacher degree (in education). Surprisingly, a bachelor's degree in math was not among these apparently more effective degrees. The separate models we estimated for EL students show somewhat weaker links between teacher degree and gains in student achievement, perhaps because of smaller sample size. There are two cases in which the models for EL students tell the same story as the models for all students: In both cases, teachers with a bachelor's degree in science are associated with smaller gains in math achievement whereas teachers with a degree in any other major than the ones listed and education, which is the comparison group, are associated with stronger gains in math achievement.

It is important to state that these findings do not tell us how effective an average person with a certain college degree would be if, say, government randomly assigned people to teach in the classroom. People

Table 5.6

Statistical Significance of Teacher's College Major and Minor in Elementary School Models

	Gains in Reading		Gains in Math	
Variable	All	EL	All	EL
Bachelor's degree in English				
Bachelor's degree in science (biology, chemistry, physics)	--		-	-
Bachelor's degree in social science			++	
Bachelor's degree in foreign language				
Bachelor's degree in math				
Bachelor's degree in other major		++	++	++
Minor in English	+		+	
Minor in science (biology, chemistry, physics)		-		
Minor in social science				
Minor in foreign language				++
Minor in math				

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/ other minor/no minor.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

who enter teaching are a self-selected group, who may not be representative of all adults with the same college degree. For example, recall the result from the math achievement model that teachers with a bachelor's degree in math do not vary significantly from teachers with a major in education. This pattern may reflect self-selection. College graduates with a major in math ostensibly have many career possibilities. Given the rigid way in which teacher salaries are set, and the low salaries of California teachers relative to salaries in other occupations that require a college degree, it could well be that the most promising math graduates find far more remunerative jobs than teaching elementary school.

Overall, one of the main messages from this table is that there is not an automatic link between a student's rate of learning and the number of college courses the student's teacher completed in a given subject. The

general lack of significance of these variables suggests to us that at the elementary school level, the teacher's subject major and minor are only weakly related to student learning. Interested readers can find the full results in Web Appendix F.

What about the size of the estimated effect of each variable on student learning? Tables 5.7 and 5.8 show the predicted changes in the rate of student learning that result from simulated changes in given explanatory variables for reading and math, respectively. We omit variables from the previous tables that are not significant at the 5 percent level. (Blank entries in the tables indicate that the variable was not statistically significant for the given sample of students.) In addition, we add one student characteristic that we consistently found to matter: the percentage of days the student was absent during the school year.

For many of the variables, we simulate the effect of changing the variable from the 25th to the 75th percentile observed in our data. In other words, we simulate the effect of an "interquartile change." For instance, the first row in Table 5.7 indicates that the interquartile range in the percentage of days absent was 3.89. We calculate the predicted change in the average gain in reading that results from such an increase in days absent. We multiply the coefficient on this variable by 3.89 to obtain the predicted change in the gain in the student test score, which is 3.89*(–0.2179) = –0.85 mean scaled points. Finally, we express this predicted drop in learning as a percentage of the average gain in mean scaled score in the sample (28.1 points), to arrive at a final estimate that the rate of reading learning will fall by 100%*(–0.85/28.1) = –3.02%.

This is the approach we take for many of the variables. In other cases, typically related to variables indicating whether the teacher held a given credential and had a given range of experience, our simulation was instead to consider what would happen if the student's teacher switched from the comparison type of teacher to one with the stated combination of credentials and experience. (Recall that the comparison category for teachers is a teacher with a bachelor's degree in education, either no minor or a minor in "other"/education, a full credential, ten or more years of teaching experience, and no language credential). In the tables, we label this sort of simulation as "nta" indicating a change from having

Table 5.7

Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for Elementary School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq	3.89	–3.02	iq		
School and peer group characteristics						
% of school FEP	iq	3.27	2.71	iq		
Grade-level peer achievement	iq			iq		
Classroom peer achievement	iq			iq		
Grade-level peer achievement, p25 to p50	oth			oth		
Classroom peer achievement, p25 to p50	oth			oth		
Grade-level peer achievement, p50 to p75	oth			oth		
Classroom peer achievement, p50 to p75	oth			oth		
Grade-level peer achievement, median absolute change	oth			oth		
Classroom peer achievement, median absolute change	oth			oth		
Class size and teacher characteristics						
Elementary class size	iq	–11.667	5.91	iq	–11.333	12.15
Teachers with emergency credential and 0–1 year of experience	nta	1	22.53	nta		
Teachers with full credential and 6–9 years of experience	nta			nta	1	–9.11
Spanish BCLAD credential	nta			nta	100	–7.09

Table 5.7 (continued)

Variable	All Students			EL Students		
	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Teacher major/minor subjects						
B.A. in science	nta	100	–10.83	nta		
B.A. in other major	nta			nta	100	6.80
Minor in English	nta	100	5.52	nta		
Minor in science	nta			nta	100	–39.74

NOTES: iq = interquartile change, nta = "none-to-all" (that is, the results from changing from a teacher without the given characteristic to having a teacher with this characteristic for the entire school year), oth = "other, see footnotes." Blank entries mean that the effect is not statistically significant. The "other" simulations apply to the peer group estimates, where we break down the effect of an interquartile change in peer group into the effect from changing from the 25th to 50th percentile and from the 50th percentile to the 75th percentile; in addition, we simulate changing the median of the absolute value of the changes in peer group test scores for individual students. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 28.06 points for the sample of all students and 31.18 for the sample of EL students.

Table 5.8

Predicted Effect of Stated Changes in Personal, School, Classroom and Teacher Characteristics on the Rate of Learning in Math for Elementary School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq	3.89	–2.70	iq	3.89	–3.70
School and peer group characteristics						
% of school Asian	iq	16.45	–8.62	iq	16.26	–19.54
% of school Pacific Islander	iq	0.82	–3.32	iq		
% of school EL	iq	27.5	12.93	iq		
% of school FEP	iq	3.23	4.28	iq		
Grade-level peer achievement	iq	0.782	9.29	iq	0.672	16.02
Classroom peer achievement	iq	0.924	3.74	iq	0.745	6.56
Grade-level peer achievement, p25 to p50	oth	0.407	4.84	oth	0.234	5.58
Classroom peer achievement, p25 to p50	oth	0.454	1.84	oth	0.335	2.95
Grade-level peer achievement, p50 to p75	oth	0.375	4.46	oth	0.438	10.44
Classroom peer achievement, p50 to p75	oth	0.47	1.90	oth	0.41	3.61
Grade-level peer achievement, median absolute change	Oth	0.148	1.76	oth	0.146	3.48
Classroom peer achievement, median absolute change	oth	0.296	1.20	oth	0.299	2.63
Teacher characteristics						
Interns with 0–1 year experience	nta	1	–16.64	nta	1	–16.64
Teachers with emerg. credential and 0–1 year of experience	nta	1	24.73	nta		
Master's degree	nta	100	3.10	nta		
CLAD credential	nta	100	–4.24	nta		
Spanish BCLAD-equivalent credential	nta	1	–58.29	nta		

Table 5.8 (continued)

Variable	All Students: Simulation Type	All Students: Simulated Change	All Students: % Predicted Change in Student Learning	EL Students: Simulation Type	EL Students: Simulated Change	EL Students: % Predicted Change in Student Learning
Teacher major/minor subjects						
B.A. in social science	nta	100	3.57	nta		
B.A. in science	nta	100	–8.20	nta	100	–13.28
B.A. in other major	nta	100	5.96	nta	100	10.83
Minor in English	nta	100	5.14	nta		
Minor in foreign language	nta			nta	100	17.22

NOTES: iq = interquartile change, nta = "none-to-all" (that is, the results from changing from a teacher without the given characteristic to having a teacher with this characteristic for the entire school year), oth = "other, see footnotes." Blank entries mean that the effect is not statistically significant. The "other" simulations apply to the peer group estimates, where we break down the effect of an interquartile change in peer group into the effect from changing from the 25th to 50th percentile, from the 50th percentile to the 75th percentile; in addition, we simulate changing the median of the absolute value of the changes in peer group test scores for individual students. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 28.04 points for the sample of all students and 30.20 for the sample of EL students.

the given type of teacher none of the time to all of the time, or "none to all."

Table 5.8 suggests that both classroom and grade-level peer group achievement have a quantitatively important relationship with individual students' rate of learning in math. An interquartile change in classroom peer achievement is associated with a 3.7 percent increase in the rate of learning in math; the corresponding number for the grade-level peer achievement level is 9.3 percent. In some senses, it is surprising that the predicted effect of interquartile change in grade-level peer group scores is greater than the effect of an interquartile change in classroom peer scores. But the overall achievement in the grade level could have important effects on the social norms of students—that is, their attitudes toward school—and in turn influence the extent to which teachers give challenging material to students in all classes in the grade.

It might seem doubtful that any individual student is likely to experience a change in peer achievement equal to the districtwide interquartile range in peer achievement. Such a change might require that a student be bused from the inner-city to a high-achieving suburban district, for instance. But approximately one in four students in the district is in a school choice program of some sort, showing that most students really do have the ability to change peers meaningfully by changing schools. Alternatively, in schools that group students by ability, such a large increase in classroom peers might entail a radical reassignment between ability groups. We certainly do observe such large variations in peer achievement for some students in the data.

Still, we can only wonder how many students could experience such a large change. Accordingly, the next two rows in Tables 5.8 break down an interquartile change into simulations of what would happen if the achievement of a student's peers in the classroom and the grade changed from the 25th to the 50th percentile and from the 50th to 75th percentile. The resulting changes, which mechanically should add up to the predicted change from the original interquartile simulation, are about half as large, with variations in how those changes are divided between the two smaller changes in peer achievement. These changes, which are still large, might correspond to being promoted from one ability group to

another within a school, or to being bused from a good to a very good school.

Finally, we offer another simulation, in which the changes in classroom and grade-level peer group achievement are set to the median of the absolute value of the *actual changes in peers' past achievement* for students in the sample. These changes are in fact reasonably large for math—0.15 standard deviation at the grade level and 0.30 standard deviation at the classroom level. Even with these relatively conservative changes in peer achievement at either the classroom or the grade level, we obtain meaningful predicted gains in learning of 1.8 percent and 1.2 percent, respectively.

Clearly, the initial achievement of one's peers in the classroom and the grade significantly affect individuals' rate of learning in a positive way. *In short, peers matter tremendously at the elementary level, for math. For reading, the coefficients on the peer variables were positive but were not statistically significant so that we can be less confident of peer effects in reading.*

Next, we consider the estimated effect of class size reduction on student learning. We found no statistically significant effect of class size on math achievement. But an interquartile reduction in class size, which amounts to a substantial reduction (about 12 students), is predicted to boost the rate of learning in reading by 5.9 percent overall and 12.2 percent for EL students (Table 5.7). These are significant increases in the rate of learning, although brought about by a very large investment in reducing class size.

Teachers' credentials and years of teaching experience appeared to matter for student learning in only a few cases. Tables 5.7 and 5.8 and the earlier tables suggest that the predicted effects of switching to less fully credentialed and experienced teachers are usually not statistically significant, but in the rare instance when they are, the effects can be meaningful. For instance, switching from the comparison group of teachers (with a full credential and ten or more years of experience) to a teacher with a full credential but only six to nine years of experience is predicted to lower the rate of reading gains by 9.1 percent for EL students. In contrast, novice teachers with an emergency credential are predicted to increase rates of gain in reading achievement by 22.5

percent, relative to teachers with a full credential and ten or more years of experience.

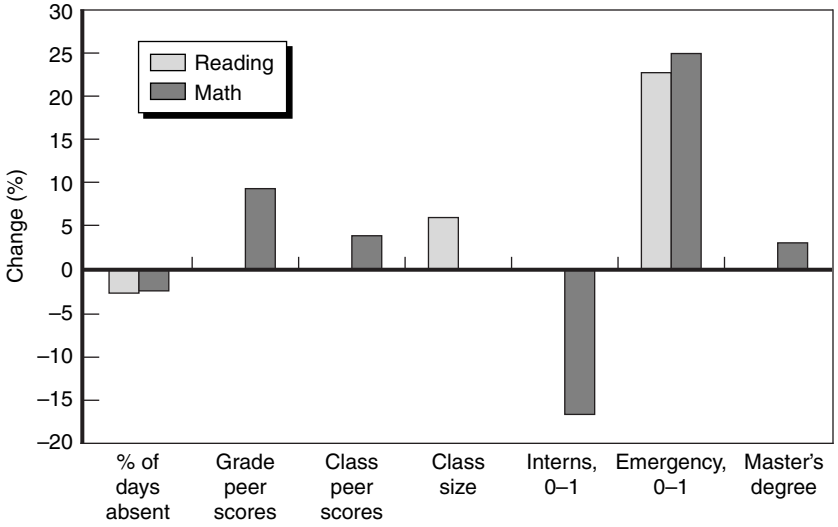
We find similarly mixed evidence on teacher education. In only one case do we find that a teacher's highest degree is associated with higher gains in student achievement. For all students, math scores are predicted to grow 3 percent more quickly if a student switches from a teacher with a only bachelor's to a teacher with a master's degree.

Although some of these teacher qualification effects are fairly large, they are not completely persuasive, given the lack of significance of most of the other combinations of credentials, experience, and education for which we controlled; the lack of corresponding findings for both the entire sample and for EL students; and the lack of corresponding findings for both math and reading gains. It seems that sometimes a teacher's level of experience, credential, and degree can matter, but in general this is not the case.

We did find that a teacher's language credentials were related to learning in different ways for EL and non-EL students. The only significant finding for EL students was that teachers with a BCLAD credential were associated with 7.1 percent lower rates of gain in reading achievement. Because the district does not have readily available data on whether specific classes are bilingual, Sheltered English immersion, or mainstreamed classes, it is hard to know whether this effect has to do with teacher training, the structure of the class, or other unmeasured characteristics of the teacher or class.

Tables 5.7 and 5.8 show that the predicted effects of switching from a teacher with a bachelor's degree in education to one with a degree in other fields are quite variable. One of the most consistent findings is that teachers with a bachelor's degree or a minor in science tend to be associated with lower gains in math and reading achievement, typically with a predicted drop in gains in achievement of 5–15 percent. The largest predicted change is a 39.7 percent drop in gains in EL students' reading achievement when taught by a teacher with a minor in science. However, this result is not mirrored in the sample of all students and so may reflect something idiosyncratic about the relevant EL students or teachers in the sample.

Finally, Figure 5.1 summarizes some of the most dramatic findings by illustrating the relative importance of absenteeism, peer group test scores, class size, teacher education, and two measures of teacher credentials and experience. The initial achievement of peers in the student's classroom and grade level appear to be among the variables strongly related to student learning in math. A few measures of teacher credentials/teacher experience are as strongly or more strongly related to student learning. But these results are sporadic—most of our measures of teacher credentials and teacher experience are not statistically significant, and in one case it appears that less highly qualified teachers are more effective than more highly qualified teachers. Increasing class



NOTES: The percentage of days absent, peer scores, and class size simulations reflect interquartile changes in the listed variables. For example, the bar for the estimated effect of class size on reading scores reflects a drop in class size of 11.7 students. The predicted effects for switching to a teacher with an emergency credential or intern credential with 0–1 years of experience are based on a comparison with a fully credentialed teacher with ten or more years of experience; the simulation for a teacher with a master's degree uses as a comparison a teacher with a bachelor's degree. The missing bars for certain variables indicate that these measures of teacher qualification were not related significantly to gains in achievement in the given type of test.

Figure 5.1—Predicted Percentage Change in the Rate of Learning Among Elementary School Students

size appears to influence student learning in reading, but the effects are dwarfed by some of the other effects.[6]

Conclusion

This chapter presents a complex picture of "what matters" for student learning. Perhaps the most consistent findings are that a student's absence rate and, at least in math, the initial academic achievement of students in the given student's classroom and grade are strongly related to the student's own rate of learning.

The result that a student who is absent particularly often will learn relatively less seems intuitive. What is less obvious is the mechanism through which the peer effects work. The effect of these classroom and grade peer test score measures could be capturing the direct learning effect that results from being surrounded by high-achieving peers. In addition, teachers may alter their teaching methods and curriculum in reaction to changes in the composition of the classroom and the grade. Either way, it appears that it is not just the teacher who matters for a student's learning but also the aptitude of other students.

It is important to remember that these and other findings cannot simply be caused by mere correlation whereby quick learners attend schools with other quick learners. Because we control for unobserved but fixed characteristics of both schools and students, we are distilling these peer group achievement effects from *changes* in the student's classroom and grade level peers from one year to the next.

Teacher qualifications do appear to affect student learning, but not in a strong or consistent fashion. Teacher education seems to matter weakly. In a very few cases, teacher credentials/experience seem to matter as well, but these effects are inconsistent. Class size appears to be a much stronger predictor of elementary students' rate of learning in reading than are the detailed measures of teacher qualifications that we include. Conversely, for math achievement we did not find that class size

[6]The next chapter, on middle and high school results, will briefly discuss some robustness checks that we performed on the models for elementary, middle, and high schools.

"matters." We also found that teachers with master's degrees are associated with higher gains in math achievement.

Finally, we recall our finding from Chapter 4 that nonwhite students have been catching up with white students. This result seems to be something of a paradox, because schools with more nonwhite students typically have the fewest fully credentialed, highly experienced, and highly educated teachers, and yet these schools appear to have shown particularly sharp gains in achievement. Part of the answer, clearly, is that teacher qualifications do matter for student learning but far less than many appear to believe.

In the following chapter, we examine our findings for middle and high school students. An important reason for doing separate analyses by grade span is to test which patterns we have just outlined in elementary schools are corroborated by results in middle and high schools. A second and more important reason to study middle and high schools separately is that education in these higher grades is a more complex process. Additional measures of teacher qualifications, in the form of subject authorizations, become relevant at these higher grades. At the same time, students in middle and high school begin to vary in the number and type of courses taken. We address these issues specifically.

6. Determinants of Gains in Student Achievement in Middle and High Schools

Introduction

Linking student learning to classroom and school characteristics is more difficult at the middle and high school levels than at the elementary school level, primarily because there are more factors that we need to take into account. For example, students—especially in the upper grades—vary in the number of English and math courses they take. A second and more important main complication in the higher grade spans is that we need to consider not only whether a teacher has a full teaching credential or an emergency credential but in addition whether he or she has the appropriate authorization(s). The *credential* refers to the teacher's overall level of qualification to teach. *Subject authorizations* are less a measure of a teacher's overall readiness to teach. Rather, they indicate the degree of mastery of the subject matter at hand.

For single-subject teachers in middle and high schools, there is in fact an entire spectrum of subject authorizations that determine the course level they may teach *in specific subjects*. These are full authorization, supplementary, board resolution, and limited assignment emergency (LAE). Full and supplementary subject authorizations are official authorizations mandated by the California Commission on Teacher Credentialing. Board resolutions refer to decisions by the San Diego School Board to authorize a teacher to teach a specific subject, if the teacher has taken relevant college courses. These teachers may lack one or two courses required for a supplementary authorization or may have enough in the general subject area but not the exact set of courses required by the CCTC. LAE authorizations are short-term

authorizations for teachers with less subject knowledge. These should not be confused with an emergency credential, because LAE credentials are given to fully credentialed teachers teaching outside their normal assignment. After consultation with district officials and documents from the CCTC, we interpret the teachers' level of knowledge of the given subject as descending in the order listed above. Some high school teachers may not hold any of the above subject authorizations, because they are not yet fully credentialed.

None of the above subject authorizations is required for subjects taught at roughly the middle school level. A subsection of the Education Code (44258.1) allows teachers who hold a multiple-subject credential to teach multiple single-subject courses at the middle school level, provided they teach separate classes to the same group of students in blocks. This fulfills the multiple-subject requirement of teaching to the same group of students as opposed to single-subject teachers who may teach to different groups of students each class period. Still, a middle school teacher with a full or supplementary authorization in the subject taught can be assumed to have taken more university courses in that subject. It is important to test whether middle school students who have a teacher with a subject authorization in the given subject learn more quickly than other similar middle school students.

Accordingly, the set of explanatory variables that we use to model students' gains in math and reading in middle and high schools will be more extensive than the set we used in the previous chapter to model elementary students' rate of learning. Table 5.1 in the preceding chapter accurately portrays the set of personal, family, and school characteristics that we will control for in the middle and high school regressions, but we need to modify the teacher characteristics. First, unlike in elementary schools, where students spend most of their day with one teacher, students in middle and especially high schools often have different teachers in various subjects. We therefore redefine the classroom peer group test score to refer to the math or English classroom, depending on whether we are modeling gains in math or reading achievement. Similarly, we model gains in math scores as a function of the math teacher(s) and math classroom(s) that the student had in a given year, and likewise we focus on English classes when modeling gains in reading

achievement. We continue to control for teachers' education, but we add indicators for whether the teacher holds anything less than a full subject authorization. Thus we add controls for a supplementary, board resolution, or LAE subject authorization.

Because middle and high school teachers are less likely than elementary school teachers both to lack a full credential and to be in their first year or two of teaching, we modify the controls used for teacher credentials and experience. To ensure that we model a reasonable number of teachers of each qualification level, the lowest range of experience we use for fully credentialed teachers is 0–2 years, rather than 0–1 years as in the previous models for elementary schools. In addition, there are far fewer interns and teachers with only an emergency credential in middle and high schools than in elementary schools. Therefore, we control for whether the teacher was an intern or had an emergency credential but we do not distinguish between interns and emergency-credentialed teachers with high vs. low levels of experience.

Finally, because students vary in the number of courses in math and English that they take each year, we also included controls for the number of courses taken. In addition, we assigned math classes in high school to one of four levels of difficulty and added dummy variables for whether the most advanced math class taken in a given year belonged to one of the three more advanced categories.[1]

---

[1]We categorized courses into the eight categories listed below, following a classification system developed by the National Center for Education Statistics, and then combined categories 2 and 3 into the omitted category (low-level math), labeled levels 4 and 5 as midlevel 1 and 2, respectively, and combined the sparsely populated categories 6 to 8 into a single category we labeled as advanced.

1. **no mathematics**

2. **nonacademic** (general 1, general 2, basic 1, basic 2, basic 3, consumer,technical, vocational, review)

3. l**ow academic** (prealgebra, algebra 1 part 1, algebra 1 part 2, geometry informal)

4. **middle academic I** (algebra 1, geometry of planes, geometry of planes-solids, unified 1, unified 2, other)

5. **middle academic II** (algebra 2, unified 3)

6. **advanced I** (algebra 3, algebra-trigonometry, algebra-analytic geometry, trigonometry, trigonometry-solid geometry, analytical geometry, linear algebra, probability, probability-statistics, statistics, other, independent study)

As in the preceding chapter, our summary of regression results focuses on a tabular and graphical presentation of the estimated effects of school and classroom characteristics that "matter" in a statistical sense. In addition, we focus on the models that include fixed effects not only for schools and students' home zip codes but also for the students themselves. The regression results are provided in Web Appendix G. This appendix also shows results when the student fixed effects are not included. Although we consider these latter models less reliable because of possible contamination from unobserved student characteristics, the models do allow identification of the effect of fixed variables such as student race and gender. Readers interested in these results can consult Web Appendix G.

Findings for Middle and High Schools

We spare the reader a specific analysis of "what matters" and what does not; instead, we summarize the broad patterns in the results. We also highlight similarities among the elementary, middle, and high school results.

Patterns of Statistical Significance

Table 6.1 shows the patterns of statistical significance of the key school demographic variables as well as the class-level and grade-level peer group test scores. The share of students at the school who are nonwhite or EL in some cases is positively related to individual students' rate of learning. This is more often the case in the middle school models than in the high school models. The patterns in Chapters 3 and 4 provide context for these findings. Low-SES schools, which also tend to have concentrations of nonwhite and EL students, do indeed have lower test scores; but as shown in Chapter 4, students in these schools, if anything, appear to be improving more quickly than those in other schools. These patterns appear to hold up in our regression analysis,

---

7. **advanced II—precalculus** (introductory analysis)

8. **advanced III—calculus** (Advanced Placement calculus, calculus-analytical geometry, calculus)

Table 6.1

Statistical Significance of Demographics of the Student Body and Average Initial Test Scores in the Student's Classroom and Grade in Middle and High School Models

	Gains in Math		Gains in Reading	
Variable	All	EL	All	EL
Middle school results				
% of students eligible for meal assistance				
% of school black	++		++	
% of school Asian			+	
% of school Hispanic			+	
% of school Pacific Islander				
% of school Native American	++		++	
% of school EL	+			
% of school FEP	+		++	
Grade-level peer achievement	++		++	
Classroom peer achievement	++	++		
High school results				
% of students eligible for meal assistance				
% of school black	++			
% of school Asian				
% of school Hispanic				
% of school Pacific Islander				-
% of school Native American	++			
% of school EL				
% of school FEP				
Grade-level peer achievement	++			
Classroom peer achievement				

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

even though we are controlling for a host of personal, family, and school characteristics as well as for unobserved but fixed characteristics of each student, school, and zip code neighborhood.

We cannot say for sure why these indicators are often positively linked to learning. One possibility is that the district, by focusing its

policy on low achievers, had already started to reap some benefits by spring 2000. Another possibility is that teachers are able and likely to tailor their teaching styles to students in these groups when the groups, such as low-SES, black, and Hispanic, are large and growing, in a way that is not practical when these students make up a smaller proportion of a class. Clearly, there are many other possibilities.

As we found for elementary schools, in middle schools the peer test scores in the classroom and the student's grade are strongly positively related to the student's own gains in math scores. At the high school level, we find that grade-level peer scores are positively linked to gains in math achievement, but the classroom peer score is not statistically significant. The math peer group results for the EL subsample are slightly weaker in this regard, with classroom but not grade-level peer scores mattering in middle schools and having no statistically significant effects at the high school level.

For reading achievement in middle schools, we found strong evidence that peer scores at the grade level are positively linked to student learning. This contrasts with both the elementary and high school results, where we found no significant effects.

Table 6.2 summarizes the extent to which class size, courses taken, and our measures of teachers' credentials are significantly related to student learning. The class size results are considerably weaker in the higher grade spans than in elementary schools: A significant relationship emerges only for all students' gains in math scores, and the coefficient is perverse, suggesting that larger classes are more effective than smaller classes.

Rose and Betts (2001), using national data, find that students who take more advanced high school courses, especially in math, earn significantly more than average later in life. On a similar note, we examined whether the type and number of math courses taken in middle and high school affect gains in math achievement. To test this hypothesis, we added controls for number of courses taken, and at the high school level, we also added controls for the level of difficulty of the math course taken. The math models at both the middle and high school levels suggest that students who take more math courses during the year improve their math achievement by significantly greater

Table 6.2

Statistical Significance of Class Size and Teacher Credentials, Experience, Education Level, and Subject Authorization in Middle and High School Models

	Gains in Math		Gains in Reading	
Variable	All	EL	All	EL
Middle school results				
Class size (in math or English)	+			
Number of courses is 0 or 1 (in math or English)				
Number of courses is more than 2 (in math or English)	++	+		
Teacher characteristics (in math or English)				
Intern				
Emergency credential				
Teachers with full credential and 0–2 years of experience	-			
Teachers with full credential and 3–5 years of experience				
Teachers with full credential and 6–9 years of experience	-			
Supplemental subject authorization		+		
Board resolution subject authorization	++			
Limited Assignment Emergency subject authorization				
Any master's degree		+		
Any Ph.D.				
High school results				
Class size				
Number of courses is 0 or 1 (in math or English)	-			
Number of courses is more than 2 (in math or English)				
Teacher characteristics (in math or English)				
Intern				
Emergency credential		++	-	
Teachers with full credential and 0–2 years of experience				
Teachers with full credential and 3–5 years of experience				
Teachers with full credential and 6–9 years of experience				
Supplemental subject authorization	-			
Board resolution subject authorization	--			
Limited Assignment Emergency subject authorization				
Any master's degree			++	+
Any Ph.D.			+	

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

amounts than average. However, we did not find evidence that students who took more advanced courses increased their math scores by greater amounts.[2]

Our measures of teacher experience and credentials in general provide only weak evidence that either is linked to student learning. The most consistent results in this regard are for math achievement among middle school students, where students' test scores rise significantly more slowly when they are taught by teachers with 0–2 or 6–9 years of experience instead of by teachers with ten or more years of experience. In contrast, for middle school reading achievement, and both math and reading achievement at the high school level, teacher experience does not enter significantly.

We found some evidence that teachers' highest degree matters positively for student learning, with a math teacher holding a master's degree entering positively in the middle school math model for EL students. We find similar results for English teachers in the models for the samples of all students and EL students in reading at the high school level. In addition, English teachers with a Ph.D. are associated with larger gains in reading achievement at the high school level.

We found mixed evidence regarding interns and teachers with an emergency credential. These variables did not enter significantly in middle school. Emergency credentials were associated with lower reading gains for the sample of all students in high school. Curiously, teachers with emergency credentials were associated with larger gains in math among EL students in high school.

---

[2]Indeed, as shown in Web Appendix G, the only significant result was that students taking the second-highest level of math, "midlevel 2," increased their math achievement by *less* than did students who took the least-demanding math classes. This counterintuitive result could have two explanations. First, it seems quite plausible that at the high school level, students who take advanced math improve their math abilities in ways that are not at all well represented by the Stanford 9 test. Second, recall our description above of the peer test score effect as possibly working through a direct effect of one's peers on one's own rate of learning, and an indirect effect mediated through the difficulty of the curriculum that teachers choose. Because we have simultaneously controlled for class and grade-level peer scores as well as the type of course taken, there could be a problem with collinearity.

What about the teacher characteristics that are unique to the middle and high school models—that is, teachers' subject authorizations? It is important to note that in high school, math achievement appears to grow significantly more slowly if students are taught by a teacher holding a supplementary or board resolution math authorization rather than by a teacher with a full authorization. We did not find any other significant effects for English or math authorization at the high school level. At the middle school level, teachers' subject authorizations in English are not significantly linked to students' gains in reading. For math, we find two cases where a subject authorization matters. It is surprising to see that students' math scores appear to grow significantly more quickly when their math teacher holds a board resolution math authorization than if the teacher holds a full authorization. As surprising, EL students' math score gains tend to be higher when their teacher holds a supplemental authorization instead of a full authorization.

In a sense, these mixed results are good news for the district, in that teachers who hold supplementary, board resolutions, or LAE authorizations are apparently holding their own in terms of improving math and reading achievement, with the major exception of math achievement at the high school level.

Table 6.3 shows results for the various credentials related to assisting EL students. We find that the CLAD, BCLAD, and their equivalents are only occasionally significant in middle and high schools. For EL students, the only case in which one of these credentials becomes significant is for reading gains in high school, where teachers with a CLAD are associated with lower gains in achievement.

Table 6.4 shows results for teachers' majors and minors. As shown, a teachers' major or minor is only rarely a significant predictor of student outcomes. In middle schools, students' math achievement grows more quickly if their teacher has a major or minor in English. One possible explanation for this puzzling result is that such teachers have excellent communication skills that improve their ability to teach a different subject.

Table 6.3

Statistical Significance of Teacher's CLAD, BCLAD, and Alternative Certifications in Middle and High School Models

Variable	Gains in Math		Gains in Reading	
	Non-EL, Non-FEP	EL	Non-EL, Non-FEP	EL
Middle school results				
CLAD credential	+			
CLAD-equivalent credential				
Spanish BCLAD credential				
Spanish BCLAD-equivalent credential			--	
High school results				
CLAD credential				--
CLAD-equivalent credential	+			
Spanish BCLAD credential				
Spanish BCLAD-equivalent credential	N/A	N/A	N/A	N/A

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being non-EL, non-FEP students and English Learners. Because the sample of all students included interactions between the teacher credentials listed above and indicators for whether the student was EL or FEP, in the second and fourth columns above we are able to report the effect of these credentials on all students who were neither EL nor FEP.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

N/A = the coefficient could not be estimated because no students in the sample had a teacher with a Spanish BCLAD-equivalent credential in one year and a teacher without this credential in another year.

The Predicted Effect of Explanatory Variables on Students' Rate of Learning

Tables 6.5 through 6.8 show the predicted effect of various changes in the explanatory variables that we have found to influence gains in reading or math achievement in a statistically significant way. As in the last chapter, the numbers in these tables report the predicted percentage change in the annual average gain in achievement that results from changing given explanatory variables such as teacher education. For

Table 6.4

Statistical Significance of Teacher's College Major and Minor in Middle and High School Models

Variable	Gains in Math		Gains in Reading	
	All	EL	All	EL
Middle school results				
Bachelor's degree in subject taught				+
Bachelor's degree in science (biology, chemistry, physics)				
Bachelor's degree in social science				
Bachelor's degree in foreign language				
Bachelor's degree in English (math courses)/math (English courses)	++			
Bachelor's degree in other major				
Minor in subject taught				
Minor in science (biology, chemistry, physics)				
Minor in social science			++	
Minor in foreign language				
Minor in English (math courses)/math (English courses)	+			
High school results				
Bachelor's degree in subject taught				
Bachelor's degree in science (biology, chemistry, physics)				
Bachelor's degree in social science				
Bachelor's degree in foreign language				
Bachelor's degree in English (math courses)/math (English courses)				
Bachelor's degree in other major				
Minor in subject taught		-		
Minor in science (biology, chemistry, physics)				
Minor in social science			-	
Minor in foreign language				
Minor in English (math courses)/math (English courses)				

NOTES: Each column refers to a separate model, with the dependent variable being gains in math or reading achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/no minor.

Blank entry = not statistically significant.

++ = positive relationship and statistically significant at the 1 percent level.

+ = positive relationship and statistically significant at the 5 percent level.

-- = negative relationship and statistically significant at the 1 percent level.

- = negative relationship and statistically significant at the 5 percent level.

Table 6.5

Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for Middle School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq	5	–4.08	iq	5.56	–10.21
School and peer group characteristics						
% of school Asian	iq	23.41	28.32	iq		
% of school black	iq	11.46	25.14	iq		
% of school Hispanic	iq	17.39	34.91	iq		
% of school Native American	iq	0.64	18.72	iq		
% of school FEP	iq	3.72	10.45	iq		
Grade-level peer achievement	iq	0.867	45.73	iq		
Grade-level peer achievement, p25 to p50	oth	0.272	14.35	oth		
Grade-level peer achievement, p50 to p75	oth	0.595	31.38	oth		
Grade-level peer achievement, median absolute change	oth	0.057	3.01	oth		
Teacher characteristics						
Spanish BCLAD-equivalent credential	nta	100	–94.94	nta		
Hispanic teacher	nta				100	–31.18
Teacher major/minor subjects						
B.A. in English	nta			nta	100	31.66
Minor in social science	nta	100	–12.46	nta		

Table 6.5 (continued)

NOTES: Each column refers to a separate model, with the dependent variable being gains in reading achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/other minor/no minor. Some regressors range from 0–1, they are full credential, 0–2 years experience; full credential, 3–5 years experience; full credential, 6–9 years experience; supplemental authorization; board resolution; Limited Assignment Emergency; and interactions of CLAD, BCLAD, CLAD-equivalent, and BCLAD-equivalent with EL and FEP students. Other variables related to teacher qualifications range from 0–100. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 13.64 points for the sample of all students and 16.74 for the sample of EL students.

Blank entry = not statistically significant.

Table 6.6

Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Math for Middle School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq	5	–5.94	iq	5.56	–7.20
School and peer group characteristics						
% of school black	iq	11.34	18.18	iq		
% of school Native American	iq	0.7	15.98	iq		
% of school EL	iq	19.66	28.02	iq		
% of school FEP	iq	4.64	7.32	iq		
Grade-level peer achievement	iq	0.602	52.65	iq		
Classroom peer achievement	iq	0.048	0.92	iq	0.724	20.94
Grade-level peer achievement, p25 to p50	oth	0.298	26.06	oth		
Classroom peer achievement, p25 to p50	oth	0.048	0.92	oth	0.724	20.94
Grade-level peer achievement, p50 to p75	oth	0.304	26.59	oth		
Classroom peer achievement, p50 to p75	oth	0	0.00	oth	0	0.00
Grade-level peer achievement, median absolute change	oth	0.094	8.22	oth		
Classroom peer achievement, median absolute change	oth	0.392	7.54	oth	0.4375	12.65
Class size and teacher characteristics						
Number of math classes taken, >2	nta	1	30.12	nta	1	18.53
Average math class size	iq	–8.5	–3.61	iq		
Teachers with full credential and 0–2 years of experience	nta	1	–7.09	nta		
Teachers with full credential and 6–9 years of experience	nta	1	–8.61	nta		

Table 6.6 (continued)

	All Students			EL Students		
	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Teachers with a supplementary authorization	nta	1	–0.14	nta	1	–11.86
Teachers with a board resolution	nta	1	13.07	nta		
Master's degree	nta			nta	100	10.99
CLAD credential	nta					
Female teacher	nta	100	4.86	nta		
Asian teacher	nta	100	10.48	nta		
Teacher major/minor subjects						
B.A. in English	nta	100	31.86	nta		
Minor in English	nta	100	29.64	nta		

NOTES: Each column refers to a separate model, with the dependent variable being gains in math achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/other minor/no minor. Some regressors range from 0–1, they are full credential, 0–2 years experience; full credential, 3–5 years experience; full credential, 6–9 years experience; supplemental authorization; board resolution; Limited Assignment Emergency; and interactions of CLAD, BCLAD, CLAD-equivalent, and BCLAD-equivalent with EL and FEP students. Other variables related to teacher qualifications range from 0–100. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 11.74 points for the sample of all students and 15.01 for the sample of EL students.

Blank entry = not statistically significant.

Table 6.7

Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Reading for High School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq			iq	6.11	–33.71
School and peer group characteristic						
% of school Pacific Islander	iq			iq	0.55	–58.46
Teacher characteristics						
Teachers with emergency credential	nta	100	–68.15	nta		
Master's degree	nta	100	20.97	nta	100	40.61
Ph.D.	nta	100	75.40	nta		
CLAD credential	nta			nta	100	–54.72
Teacher major/minor subject						
Minor in social science	nta	100	–26.41			

NOTES: Each column refers to a separate model, with the dependent variable being gains in reading achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/other minor/no minor. Some regressors range from 0–1, they are full credential, 0–2 years experience; full credential, 3–5 years experience; full credential, 6–9 years experience; supplemental authorization; board resolution; Limited Assignment Emergency; and interactions of CLAD, BCLAD, CLAD-equivalent, and BCLAD-equivalent with EL and FEP students. Other variables related to teacher qualifications range from 0–100. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 4.96 points for the sample of all students and 8.15 for the sample of EL students.

Blank entry = not statistically significant.

instance, a predicted change of 50 percent means that students would on average improve their achievement 50 percent faster if they received the listed change in resources.

We will not go through these tables line by line. However, comparing these tables to the analogous tables in Chapter 5, we find that in some cases the predicted effects of increasing a given variable are much larger in middle and high schools than in elementary schools. Part of this pattern stems from the fact that average test score gains are smaller in the higher grades, so that a predicted increase in learning of, say, 10 points implies a much bigger effect in percentage terms.

As noted above, in middle school, especially for reading, the racial composition of the school is often associated with test score gains. Tables 6.5 and 6.6 show that the size of some of these gains is quite large. For example, an interquartile increase in the percentage of students who are Hispanic in middle schools is 17.4 percent. Table 6.5 shows that this increase is associated with a 34.9 percent increase in the average rate of reading score achievement. These patterns are not nearly as prevalent in elementary and high schools.

The grade-level peer group test score variable, which is quite consistently significant, continues to have large and in some cases far larger predicted effects in middle and high schools than it did in elementary schools. In our most conservative simulations, we calculate the absolute value of the actual changes in peer test scores by individual students between grades and then take the median of these. In this case, we find predicted effects on reading gains of about 3 percent in middle schools, and predicted gains in math achievement growth in middle and high schools of about 8 percent and 4 percent. The predicted effects of an interquartile (25th to 75th percentile) change in classroom and grade-level peer test scores are substantially higher.

The lone case where the classroom peer group appears to matter in the sample of all students is for math achievement in middle school. Here, the predicted increase in test score growth from changing peer scores by the median of the absolute observed change is about 8 percent. This was the same case in which classroom peers appeared to matter for EL students in middle school math. Again, the predicted effects are quite large. The predicted effect from changing peer scores by the

Table 6.8

Predicted Effect of Stated Changes in Personal, School, Classroom, and Teacher Characteristics on the Rate of Learning in Math for High School Students

	All Students			EL Students		
Variable	Simulation Type	Simulated Change	% Predicted Change in Student Learning	Simulation Type	Simulated Change	% Predicted Change in Student Learning
Student characteristic						
% of days absent	iq	5	–6.31	iq		
School and peer group characteristics						
% of school black	iq	15.04	38.04	iq		
% of school Native American	iq	0.54	20.95	iq		
Grade-level peer achievement	iq	0.535	28.86	iq		
Grade-level peer achievement, p25 to p50	oth	0.249	13.43	oth		
Grade-level peer achievement, p50 to p75	oth	0.286	15.43	oth		
Grade-level peer achievement, median absolute change	oth	0.065	3.51	oth		
Teacher characteristics						
Number of math classes taken annually, 0–1	nta	1	–19.81			
Highest annual math level, midmath 2	nta	1	–28.61			
Teachers with emergency credential	nta			nta	100	62.54
Teachers with a supplemental authorization	nta	1	–13.49			
Teachers with a board resolution	nta	1	–46.55			
CLAD-equivalent credential	nta	100	13.70			
Asian teacher	nta	100	–22.70			
Hispanic teacher	nta	100	16.74			
Teacher major/minor subject						
Minor in math				nta	100	–46.48

Table 6.8 (continued)

NOTES: Each column refers to a separate model, with the dependent variable being gains in math achievement, and the sample being all students or English Learners. The omitted college major is education and the omitted minor is education/other minor/no minor. Some regressors range from 0–1, they are full credential, 0–2 years experience; full credential, 3–5 years experience; full credential, 6–9 years experience; supplemental authorization; board resolution; Limited Assignment Emergency; and interactions of CLAD, BCLAD, CLAD-equivalent, and BCLAD-equivalent with EL and FEP students. Other variables related to teacher qualifications range from 0–100. The predicted percentage gains in student learning are calculated by dividing the predicted gain in the test score by the average one-year gains in mean scaled scores. These average gains were 9.56 points for the sample of all students and 11.64 for the sample of EL students.

Blank entry = not statistically significant.

median of the absolute observed change is an increase in the average annual gain for EL students of about 13 percent.

For all of these peer group simulations, the predicted effects from changing the student's peer score by an interquartile change are much larger. These sorts of changes are most likely to occur when a student switches schools.

To facilitate comparison of the size of the peer group effects with the predicted effects from changing various other measures, Figures 6.1 through 6.4 show the predicted effect of changing given variables by the interquartile range observed in the data. (For teacher characteristics, the simulation instead changes the teacher from the comparison group of teachers (teachers having a bachelor's degree in education, a full credential, ten or more years of experience, and a full subject authorization in the subject taught) to a teacher with the given credential, experience, or education.)

Figure 6.1 shows for middle schools the effect of interquartile changes in the percentage of days absent and the peer group measures, as well as the predicted effect of changing the number of courses taken (in English for the reading score models and in math for the math score models). Figure 6.2 shows the same comparisons for the high school models. As in Chapter 5, when a bar is missing from a graph, it indicates that the given variable was not a statistically significant determinant of the given test score.

As for elementary students, students in the higher grade spans who are absent about 5 percent of days experience roughly 5 percent lower achievement growth. This is not an example of a "school resource" but provides an easily understandable baseline against which to compare some of the other simulated changes.

Interquartile changes in the peer scores at the student's grade level in his or her school appear to be very strongly related to the student's own rate of achievement gain, with the notable exception of reading gains in high school. The size of these effects, often approximating a 50 percent boost in the annual gain in achievement, is much larger than what we found in elementary schools. At the same time, we found that classroom peer scores were less likely to be significant predictors of student learning in middle and high schools than in elementary schools. As noted above,

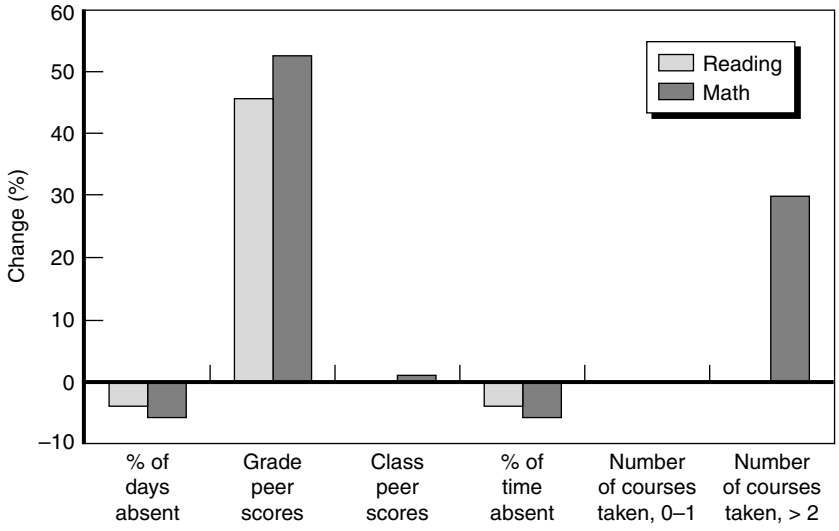


Figure 6.1—Predicted Percentage Change in the Rate of Learning Among Middle School Students, by Absenteeism, Peer Scores, and Courses Taken

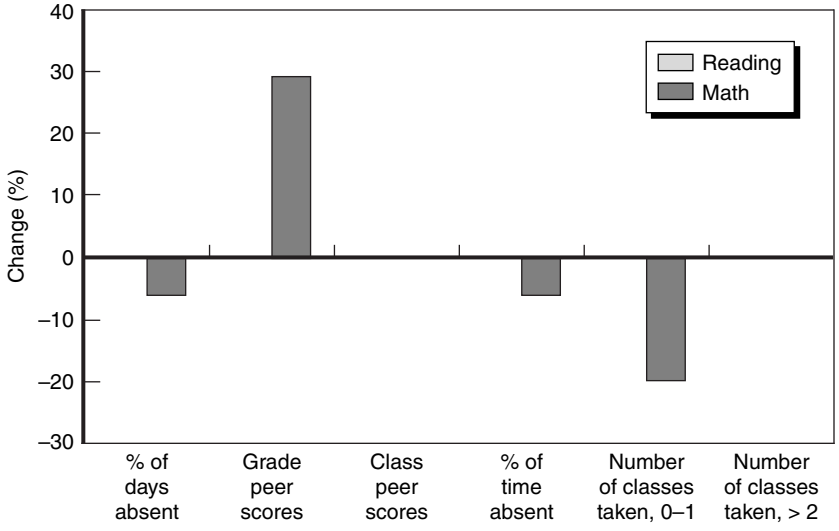


Figure 6.2—Predicted Percentage Change in the Rate of Learning Among High School Students, by Absenteeism, Peer Scores, and Courses Taken

one explanation for these patterns could well be that in middle and high school, students typically switch classrooms during the day, changing their peers from one class to the next. Perhaps in this environment, it is less the achievement of peers in the English class that affects a student's improvement in reading ability than it is the average achievement of peers in all of his or her classes in the grade. Similar arguments may apply to math classes and gains in math achievement.

Figures 6.1 and 6.2 also suggest that sizeable variations in the rate of achievement growth in math appear to result from variations in the number of math courses taken. In the figures, our comparison group is students who take exactly two math and English courses. In middle schools, we find that if a student takes two math courses one year and more than two the next, his achievement growth is roughly one-third higher in the second year. Similarly, in high school a student who takes two math courses one year and 0–1 course the next year learns about 20 percent less in the later year.

These figures are also notable in that they exclude class size. In no case did we find significant evidence that smaller classes led to higher gains in either math or reading in middle and high schools.

Figures 6.3 and 6.4 continue the comparison of middle and high school results, showing the predicted effect of changing various measures of teacher qualifications. The graphs show slightly different simulations, because different aspects of teacher qualifications seem to matter in middle and high schools. Figure 6.3 shows that in middle school, students who have teachers with less experience than our comparison group (teachers who hold ten or more years of teaching experience) in two out of three cases have reduced gains in math. On the other hand, teacher subject authorizations, which are optional for middle school teachers, do not appear to matter much for student learning in middle schools. Indeed, the strongest effect we found was that math teachers with a board resolution in math, meaning that they have taken relatively few of the math courses needed for a full math authorization, are associated with much higher gains in student math achievement than are teachers with a full authorization.

We obtain quite different results when we examine high school teachers' qualifications. We find only limited evidence that years of

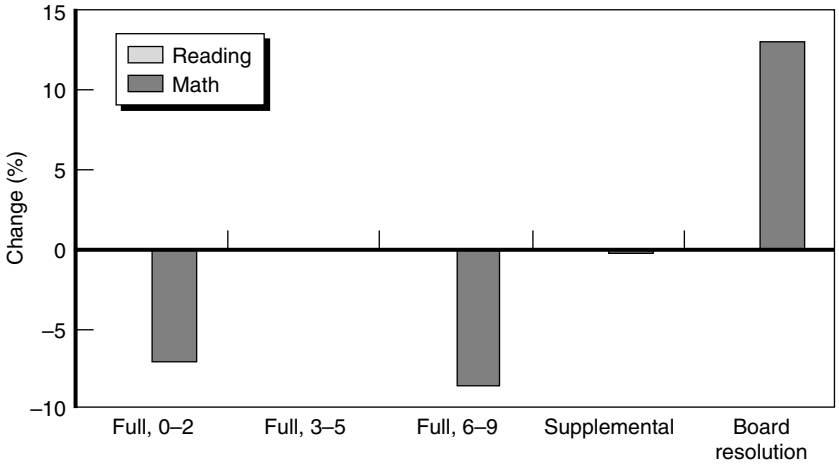
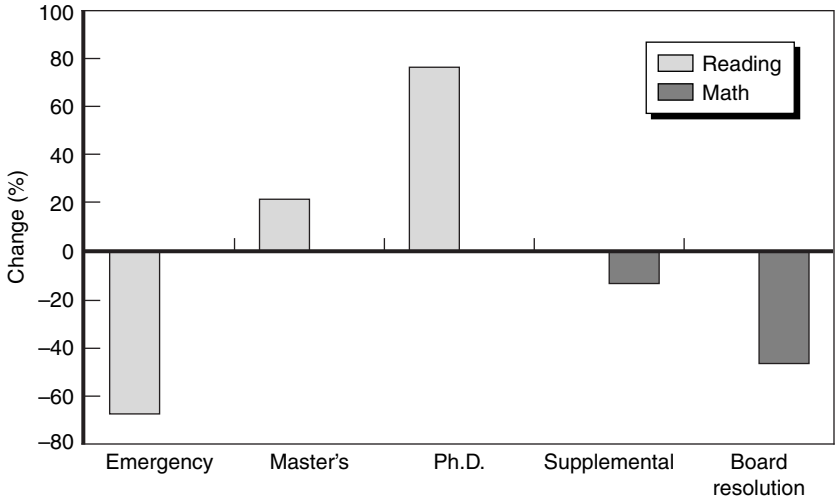


Figure 6.3—Predicted Percentage Change in the Rate of Learning Among Middle School Students, by Teacher Credentials, Experience, and Authorization



NOTE: The figure shows the predicted effect of switching from a teacher with a full credential, a full subject authorization, a bachelor's degree, and ten or more years of experience to a teacher with the stated credential or education but who is otherwise identical.

Figure 6.4—Predicted Percentage Change in the Rate of Learning Among High School Students, by Teacher Credentials, Education, and Authorization

experience matter, and we find that education and subject authorizations do matter in important ways. Although years of experience do not enter significantly, the few high school English teachers who hold an emergency credential are associated with rates of reading achievement gain that are almost two-thirds below those of teachers with a full credential. Similarly, English teachers with a master's degree or Ph.D. are associated with increased student rates of improvement in reading on the order of 20 percent and 75 percent, respectively. Turning to math achievement, we find that math teachers who hold a supplementary or board resolution authorization in math are associated with 13 percent and 47 percent slower rates of growth in math achievement for their students than teachers with a full authorization. By any measure, these effects are large. However, they are sporadic in that what matters in one subject does not matter in the other.

Robustness Checks

We undertook some robustness checks on our test score models for elementary, middle, and high schools. In brief, our checks included the following. First, we removed students who were in charter or atypical schools to make sure that our main conclusions do not derive from some idiosyncrasy of these schools. Very little changed in the sense that coefficients on key variables did not change substantially, no key variable became significant or insignificant in the subsample models, and no coefficient that was statistically significant changed signs. Additionally, we removed controls for students switching schools and the percentage of days absent, in case these variables were endogenous. Again, little changed. We also tried adding a separate dummy for "expected" school switchers to include controls for all students who had changed schools, as a partial check on whether particularly large gains or drops in peer scores were really capturing unobserved differences related to a move between schools. The peer coefficients changed very little.

Finally, and most interesting, we tested for an asymmetry in peer effects. Specifically we asked whether the test-score gain of a student whose peer group improves equals, in absolute terms, the test-score drop of a student whose peer group deteriorates from one year to the next. For both the grade-level and classroom peer effects, we found evidence

that typically suggested that losses were greater than gains. This analysis could have implications for busing and ability grouping, because it suggests that attempts to make all classes look alike by grouping heterogeneous students together may harm high achievers slightly more than it helps low achievers. This analysis is very preliminary, and we intend to follow this up in further research, using alternative measures to test for asymmetric effects.

Finally, we examined the issue of whether our inclusion of fixed effects for zip codes, years, schools, and, especially, students left sufficient variation in the data to identify the effect of measures of class size, teacher characteristics, classes taken, and peer achievement at the class and grade level. There was substantial underlying variation in these variables, and so our judgment was that as long as these fixed effects together could not explain more than 95 percent of the variation in these variables, there would be sufficient variation left to identify effects that were large. In most cases the data easily met this requirement. We found in most cases in each of the three grade spans that the fixed effects could account for about 50–85 percent of the variation in the explanatory variables, with a few exceptions that were higher. The most consistent exception to this rule was the class and grade peer achievement variables. Once we added student fixed effects we found that 86–98 percent of the variation in these explanatory variables was removed. (The two highest cases of 98 percent occurred in the case of the middle and high school reading models, for grade peer achievement.) This makes it all the more remarkable that we find a statistically significant effect of peer achievement in our models.

Conclusion

This chapter has studied the effect of changes in the environment of middle and high school students from one grade to the next on students' rate of improvement in math and reading. We find that class size does not seem to "matter" in these grade spans, and that measures of the number of courses taken, peers' achievement, and teacher qualifications are related sporadically to gains in student achievement. Some of these effects are quite large. The most compelling results appear to be that the initial achievement of students in a given student's grade is positively

related to the student's own subsequent gains in achievement, especially in math.

Similarly, math and English teachers' qualifications in some cases do appear to be related positively to student gains in math and reading, but the results are variable and inconsistent between the two grade spans and between math and reading. This should not be interpreted to mean that teacher qualifications are irrelevant. Indeed, some of the effects that we found were extremely large. To give just one example, high school English teachers with an emergency credential are associated with student achievement gains in reading that are about two-thirds below those associated with teachers with a full credential. However, it would be highly misleading to conclude that teacher credentials, subject authorizations, education, and experience always matter in important ways. This is well illustrated by the finding that high school math teachers with less than a full math authorization appear to produce far smaller gains in math achievement than do those with a full authorization. But we found no such evidence in middle schools and no similar evidence for English teachers in either middle or high schools. This may simply indicate that high school math is one of the few areas in which teachers truly do excel by having successfully completed a rigorous university curriculum in the subject. Less technical types of teaching may not depend as heavily on taking the "right" mix of relevant university courses.

7. Policy Conclusions

Overview of Central Findings

This report has examined the link between school resources and student achievement within the context of San Diego Unified School District. Although this research will be of particular interest to readers in San Diego, we believe that it also conveys findings of interest to policymakers, school administrators, and parents throughout California. SDUSD is quite representative of the state as a whole: It enrolls a demographically diverse set of students, taught by teachers who vary considerably in their education, experience, and credentials. In terms of student demography, student test scores, and school resources such as class size and teacher characteristics, SDUSD looks like other major urban districts and also resembles the state as a whole quite closely. As is typical of other districts statewide, SDUSD's distribution of teachers across schools is far from random. Teachers in schools serving economically disadvantaged students are far more likely to lack a full credential, to be in their first few years of teaching, and to lack a master's degree. In part, this inequality probably reflects teachers' own preferences and teacher mobility among schools. All of these patterns in resource allocation and demographic diversity are shared by other large districts around the state.

Given that SDUSD appears to be quite representative of the state as a whole, it is a good testing ground for learning more about the determinants of overall student achievement and inequalities in achievement in the state.

We began this report by citing survey evidence indicating that the California public is deeply concerned about public schooling in California. The public wants to see better schools for all students and also seems prepared to devote additional resources to schools where student achievement is low. In short, the California public desires

greater efficiency *and* greater equity in the state's schools. And so what have we learned on these two counts?

Efficiency

Our analysis of the test score data suggests, tentatively, that SDUSD schools on average may be increasing their effectiveness. Our evidence of greater efficiency is simply that test scores rose considerably in math and reading between spring 1998 and spring 2000. Of course, some of this must be attributed to the fact that the state test was introduced in spring 1998. Several studies in other states have found that test scores almost always rise in the first few years after the introduction of a new test, as students and teachers become more familiar with the test format and questions.[1] This is particularly an issue with the Stanford 9, because California chose not to alternate among test forms between one year and the next. Still, the point gains are large; and over this period, gains in SDUSD outstripped gains in the state as a whole. We cannot know for certain, but both of these facts lend credence to the notion that the district has experienced some genuine improvement in average student achievement.

Equity

What has our study revealed about the inequalities in student achievement, the associated trends, and the causes? It is widely known that in California, test scores in schools serving economically disadvantaged students tend to lag far behind national norms. In SDUSD, as elsewhere, the gaps in achievement between EL and non-EL students, between Hispanic and white students, between black and white students, and between students in schools in affluent and disadvantaged areas can only be described as huge. This immediately raises some pivotal questions. Are schools in some sense to be blamed for lagging test scores in schools in disadvantaged areas? In particular, does the lower level of school resources in these schools, especially related to teacher qualifications, contribute to the achievement gap?

---

[1] See for example Koretz (1996).

Our results suggest some surprising answers. It is certainly true that schools serving disadvantaged students tend to have far lower student achievement. But we found that the achievement gap in reading and math typically is at its *largest* in grade 2—the first grade in which students are tested statewide. In other words, disadvantaged students start their schooling years with levels of achievement that seriously lag behind those of their more advantaged peers. This statement is true regardless of whether we define "disadvantage" in terms of our fairly crude proxy—the percentage of students at the school who are eligible for meal assistance, or instead in terms of race or language disadvantage.

This is an important finding: Inequality in achievement appears to arise well before students are old enough to enter public school. We conclude that it would be simplistic and unfair to hold schools accountable for preexisting variations in achievement. That said, our finding should not be cause for complacency. Although schools should not be blamed for preexisting inequalities in achievement among entering students, perhaps they should be held accountable for improving achievement among all groups of students. So, have the initial gaps in achievement between various groups of students widened or narrowed over time?

We followed individual students over three years of testing, and we found strong evidence of increasing equity within SDUSD. We divided students in several ways—by the percentage of students at their initial school who were eligible for meal assistance and by the students' race and language status. *By any of these criteria, intergroup gaps in achievement declined dramatically over the two-year period, typically with drops in the achievement gap of over 10 percent.* To give just two examples, between spring 1998 and spring 2000 the initial gap in achievement between white and Hispanic students fell by 13.9 percent in reading and 9.7 percent in math, and the gap between students attending the quintiles of schools serving the most and the least economically disadvantaged fell by 15.2 percent and 11.1 percent in reading and math, respectively. The main exception to the rule was the black-white achievement gap, which on average did fall, but by far smaller amounts: 6.7 percent and 0.9 percent in reading and math.

The welcome news that minorities, English learners, and those attending schools in less-affluent areas are catching up does raise a further question. One might think that the reason why the achievement gap has narrowed is that additional resources have been devoted to disadvantaged students. Yet our analysis shows that schools serving less-affluent areas have distinctively fewer resources, especially when resources are defined in terms of the qualifications of the teachers. If less-affluent schools have less highly qualified teachers, how could it possibly be that students in these schools have caught up over time?

The Determinants of Student Learning

To address this paradox, we statistically modeled the determinants of students' gains in reading and math achievement over the three-year period. We took full advantage of the fact that we have repeated observations for individual students and schools. This rich "longitudinal" nature of the data enabled us to control fully for any unobserved but fixed characteristics of the students, their schools, and the characteristics of the environment in the students' home zip codes. To some extent, our findings may be overly conservative because of our extensive set of controls for these unobserved factors. But it increases our confidence that when we find that something matters for student learning, it truly does matter.

Our results are striking not only because of which factors matter for student learning but also because of the factors that apparently do not matter, or matter only sporadically. In short, our findings partially resolve the paradox that the achievement of disadvantaged students has improved the most even though on average these students attend schools with less highly qualified teachers. The resolution comes from the general result that in many cases, teachers who have less education and experience and fewer credentials are not necessarily less-effective teachers than their more qualified counterparts.

Indeed, perhaps the most consistent finding in this research has been that factors apart from teachers themselves appear to influence students' rate of learning. A principal finding that applies across the three grade spans is that an individual student's achievement gains are strongly positively related to the initial achievement level of students in his or her

grade level and occasionally the achievement level within his or her classroom. One might think that this effect is not causal and merely reflects ability grouping that occurs within schools. But because we control for unobserved ability and motivation of each student, at least to the extent that these remain constant over the three-year time span of this study, this objection seems moot. In effect, we are identifying the grade level and classroom peer test score effect by small variations in each student's peers between grades. Typically, the change in a student's peer group from one year to the next is predicted to change the student's rate of learning by 3 to 8 percent, although in many cases, especially at the classroom level, the effect is not statistically significant. When we instead simulated the effect of more radical changes in classroom peer test scores, the predicted effects were even larger. Such changes could result, for example, from busing of students between neighborhoods. This is a relevant simulation because approximately one in four students in the district participates in one of several forms of school choice program.

In 1996, California implemented an ambitious and expensive program to reduce class size to 20 students in kindergarten through grade 3. We found solid evidence at the elementary school level that smaller classes promote learning in reading but not math. For instance, a reduction of class size from 32 to 20 is predicted to increase elementary students' rate of growth in reading achievement by 6 percent overall and about 12 percent for English Learners. However, at the middle and high school levels, we could not find any evidence that class size mattered for student learning. Although larger samples in future work might overturn this finding, it seems quite plausible that class size matters most during children's earliest school years.

We also examined whether there is a link between students' rate of learning and an exceptionally rich portrait of teacher characteristics. These teacher characteristics include highest degree earned, college major and minor, basic credential level, teaching experience, language certification to teach English Learners, and at the middle and high school levels, subject authorization. To what extent do these qualifications matter? We certainly found some evidence that each of these measures of teacher preparedness can contribute to faster student learning. But it

would overstate our findings tremendously to claim that these aspects of teacher qualifications *always* matter.

One example of this is the perennial debate over the merits of the teacher credentialing system and the closely related debate about the importance of placing an experienced teacher in every classroom. We found quite contradictory evidence on this question in elementary schools, but in middle and high schools, we did occasionally find that less experienced teachers or teachers with an emergency credential appeared to be relatively less effective than more highly qualified teachers.

This same pattern, in which teacher qualifications matter in some cases but not others, replicates itself in other regards. In some cases, such as high school reading achievement, students seem to learn more quickly if they have a teacher with a master's instead of a bachelor's degree, but the evidence is weaker in middle and elementary schools. Similarly, a teacher who holds a full subject authorization in his or her field of teaching does not appear to do any better in middle school, but in high school we find that math teachers' level of authorization in math is extremely important.

Although complex, these results might be characterized in the following way. C*lass size appears to matter more in lower grades than upper grades, whereas teacher qualifications such as experience, level of education, and subject area knowledge appear to matter more in the upper grades*. There is some intuition supporting both of these conclusions. For instance, a careful reading of Krueger and Whitmore (1999) suggests that the main gains to class size reduction occur in the first year that a student is in a small class. This finding is consonant with the hypothesis that class size matters more in the early grades. Similarly, one can imagine that teachers' knowledge of subject matter, as expressed through his or her level of subject authorization and overall level of education, might matter more in the higher grades as the curriculum becomes more difficult to master.

In cases where teacher qualifications do appear to matter significantly, the size of the effect on student learning can be quite large. For example, in high schools, students whose teacher holds only an emergency credential appear to increase their reading achievement by

only about one-third of the norm in which teachers hold a full credential. Similarly, high school math teachers with a board resolution authorization in math appear to produce gains in math achievement only about half as large as do teachers with full math authorizations. But our reaction to the impressive size of these differences must be tempered by the fact that most often a positive finding in math is not replicated in reading, and by the related fact that the effect of a selected teacher characteristic matters in one grade span but not another.

Policy Implications

From a policy perspective, what are we to make of these findings? Is the glass half full or half empty? In some respects, administrators should be reassured to learn that a less than fully credentialed teacher sometimes appears to be as effective as a fully credentialed teacher. This reassurance is particularly important at the current time. In spring 2003, SDUSD responded to the dire state budget situation by instituting an early retirement incentive plan for its staff. As a result, approximately one in ten of the district's teachers opted to retire in summer 2003 (Moran, 2003). The result will probably be that in fall 2003, newly hired recruits will replace some of the most experienced teachers in unprecedented numbers. The results in this report cannot be used to predict the effect of such a large shock to the system. Further, it is easy to imagine that the loss of institutional memory created by this mass retirement will reduce the effectiveness of San Diego schools in fall 2003. But a year or two into this new regime, many observers may be pleasantly surprised to find that the relatively inexperienced teachers may be faring better in the classroom than they would have predicted.

Why is it that in many but not all cases less-experienced teachers appear to be equally effective as more experienced teachers? California spends roughly $100 million a year on the Beginning Teacher Support and Assessment (BTSA) program, which aims to provide assistance to teachers in their first and second years of teaching. It could be that this and related programs successfully integrate inexperienced teachers into the classroom. In addition, SDUSD has adopted a peer coach program to train teachers in the latest instructional techniques, which may be particularly helpful for novice teachers.

Similarly, the news that middle and high school English and math teachers with less than a full subject authorization often are just as effective as fully authorized teachers should come as reassuring news given that it is virtually impossible for a district to ensure that all of its teachers have exactly the right mix of college courses as mandated by the CCTC.

Still, the preponderance of evidence is that teachers who on the surface appear more qualified to teach math or English in some but not all cases are somewhat more effective. This brings us back to the findings of Chapter 3. There, we showed that teachers at schools serving economically disadvantaged students on average are significantly less qualified along a number of dimensions than their counterparts at schools in more affluent areas.

What policy reforms might the district enact to equalize these differences in teacher qualifications between the have and have-not schools? Above, we discussed stipulations in the district's collective bargaining agreement guaranteeing that open teacher positions will go to one of the five qualified teachers with the most district seniority. Betts, Rueben, and Danenberg (2000) report similar first-right-of-transfer clauses in teacher contracts in other large California districts. The observed tendency of teachers to transfer to schools in more affluent areas as they gain more experience can only be compounded by these contract stipulations, which make it automatic that an affluent school must choose from among the most highly experienced teachers on its applicant list. Clearly, it is more than just these contract stipulations that cause the observed inequalities in teacher preparation between affluent and disadvantaged areas. But they certainly exacerbate patterns created by teachers' exhibited preferences. One can imagine a mutual agreement between union and district to relax these stipulations on the grounds that they work against the interests of some of the most needy students in the district. However, such reforms cannot be mandated by administrators alone and may entail what labor economists refer to as a "compensating differential"—in other words, an increase in average salary for more senior teachers, to compensate them for any loss or reduction in first-right-of-transfer privileges.

A related aspect of teachers' contracts throughout California that militates against equalization of teacher qualifications among schools is the teacher pay schedule. As is true in other districts, teachers' pay rates in San Diego are largely determined by their highest degree and teaching experience. One possibility would be for the district and the teacher's union, the San Diego Education Association, to agree to salary bonuses designed to attract highly qualified and experienced teachers to the schools that are currently lacking them. Obviously, such negotiations are more likely to succeed in a time of budget plenty, so that teachers in some schools would receive bonuses without reducing the pay of teachers at other schools. At the time of this writing, SDUSD's budget is quite tight. However, if in some more prosperous year all parties agree that a high priority is to boost the share of teachers at inner-city schools who are highly qualified, then they should pursue this possibility in a way that would leave no teacher worse off, and many students better off.

Another aspect of this report that bears upon policy is the attempt that we have made to model separately the determinants of learning for all students and the subsample of English learners in the district. In this initial report, with only three years of test-score data, we fear that the relatively small size of the EL sample may have prevented us from discovering all of the ways that class size, classroom and grade-level peer group achievement, curriculum, and teacher preparation influence learning among EL students. This makes it all the more impressive when we find that a given classroom or teacher characteristic appears to influence learning among these students. Perhaps the most powerful finding is that in the elementary grades the effect of changing class size is about twice as strong among EL students as it is in the general student population. At the high school level, we found distinctly mixed messages regarding teacher qualifications and EL students.

Perhaps the most consistent finding in the report is that an individual student's rate of learning appears to be strongly positively influenced by the initial achievement of students in his or her grade level, and with somewhat less consistency that of students in his or her classroom. This finding is obviously of great policy relevance but is very hard to translate into a specific policy prescription. Obviously, ability

grouping within the school will affect each student's peers. Similarly, students who volunteer for busing in the district are likely to alter their peer group in substantial ways. Our research falls far short of providing specific ideas on whether or how either of these practices should evolve over time. Both of these issues are worthy of more detailed study.

It seems fitting to end this report by touching upon SDUSD's Blueprint for Student Success. Implemented in fall 2000, this reform is designed to accelerate the learning of students who lag far behind grade level. The reform has at the same time attracted favorable national attention and generated intense local controversy. Although some elements of the blueprint, such as those related to peer coaching of teachers, were implemented toward the end of our period of study (school years 1997–1998 through 1999–2000), the main parts of the reform were put in place in fall 2000, after the period we study.

Thus, our report cannot speak to the extent to which the blueprint will succeed; but our results do allow us to comment on the general approach taken by the blueprint. First, the initial years of the blueprint have placed greater emphasis on reading improvement than on math improvement. The general notion of starting with reading as a foundation skill before expanding the scope to include math and other subjects garners support from our analysis of test scores in Chapter 2. There, we found that reading achievement in the district lagged behind national norms to a greater extent than did math performance.

More fundamentally, is there a solid empirical basis for the central thesis of the blueprint—that additional resources need to be devoted to students who lag behind? Certainly our analysis has found that in spring 1998 exceedingly large gaps in achievement existed between more- and less-affluent students, between white students and students of other ethnicities, especially Hispanics and blacks, and between English Learners and fluent speakers of English. Over the next two school years, these gaps all narrowed, but troublingly large achievement gaps still exist. These facts argue in favor of devoting additional help in some form to the many students in the district who lag the furthest behind national norms.

The blueprint also calls for intensive teacher training and professional development. As a survey of teachers conducted by the

American Institutes of Research for the SDUSD School Board has shown, teachers in the district disagree with many aspects of the district's new professional development program. Again, our research cannot provide direct insight on the design of this element of the reform. However, our findings indicate that the traditional measures of teacher qualifications, such as education, credentials, experience, and subject authorizations, are not as strongly or as consistently related to student learning as some might think. The general concept that districts should look "outside the box" for additional ways to help teachers improve their teaching receives strong support from our findings.

A final comment relevant to the reform is simply this. All who are involved in making the public schools more effective and equitable—teachers, parents, administrators, and outside parties—must bear in mind that the daunting achievement gaps between students do not appear to be *created* by the schools as they now exist. These gaps, related to income and socioeconomic status more generally, emerge by the time young children reach school age.

One implication is that at the federal and state level, policymakers may want to examine the value of Head Start and similar preschool programs as a way of reducing the achievement gap of disadvantaged students before they begin their formal schooling. Notably, a working group of California's Joint Committee to Develop a Master Plan for Education—Kindergarten Through University, in its final report in early 2002, proposed an expansion of preschool funding to prepare the state's young children better for regular school.[2]

As for K–12 school systems themselves, in San Diego Unified, at least, schools appear to have been working effectively to reduce inequalities in achievement between 1997–1998 and 1999–2000. We should not use this sign of success as an excuse to ignore the large achievement gaps that remain. But it should give us some perspective. Schools are not a part of the problem; they are part of the solution. The goal of this report, and ensuing reports, has been and will be to shed

---

[2]See Joint Committee to Develop a Master Plan for Education—Kindergarten Through University (2002).

some light on the most promising ways to devote limited financial resources to making schools more effective solutions than they already are today.

Appendix A

Methods Used to Take Account of Unobserved Factors Affecting Student Learning

This appendix provides a nontechnical summary of the advantages of the statistical method used to infer the determinants of student achievement gains.

Gains versus Levels of Achievement

Most of the early research on the determinants of students' test scores in the 1970s and 1980s attempted to explain the *levels* of student achievement in a given grade. But this approach has limitations. It is surely the case that a student's test score in grade 5 reflects not only the quality of instruction he or she received in that grade but also the quality of education he or she received in earlier grades, not to mention learning experiences provided in the home since the student was very young. It is extremely uncommon for researchers to have information describing the classroom experience of students from kindergarten through the current grade. It is even more uncommon for researchers to know much about students' early childhood educational experiences in the home.

It is quite easy to imagine situations where researchers could attempt to "model," or explain, the *level* of a student's test score in a given year as a function of classroom characteristics that year, and arrive at quite incorrect conclusions. Figure A.1 provides a hypothetical example. Suppose two otherwise identical students are placed in different classrooms in each grade from grade 2 through 5. The figure shows the test scores at the end of each grade for the two students. By the end of grade 5, student A has a higher test score. But the quality of his or her classroom environment appears to be markedly worse in grade 5 than it is for student B, whose test score rises much more than does student A's

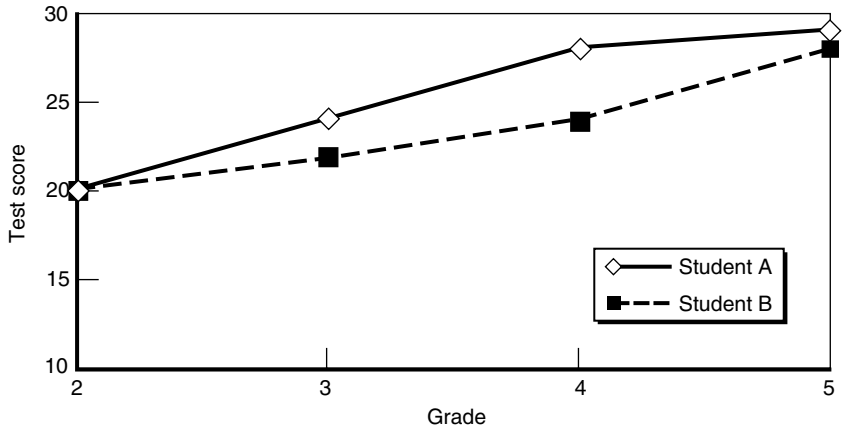


Figure A.1—Identical Students with Different Quality Classrooms

score between the end of grade 4 and the end of grade 5. If we naively attempted to explain the grade 5 test scores of these two students on the basis of, for example, class size in grade 5, we would make exactly the wrong inference—that student A had a better grade 5 experience than did student B.

The SDUSD data allow a solution to this problem. Because we have up to three years of test score data for each student, we model the *gain* in student test scores between spring of one grade and spring of the next grade as a function of the classroom characteristics in the latter grade. This comes far closer to allowing us to estimate the causal effect of classroom characteristics on student learning.

We should note that many studies over the past two decades have used a similar "value added" approach that estimates the added achievement that results from a student spending an additional year in school. Unfortunately, it is not possible to analyze gains in student achievement at the state level in California, given the state decision not to link student test scores between years. The approach we use here provides a useful check on earlier California research that has modeled levels of student test scores as a function of school resources, such as Betts, Rueben, and Danenberg (2000).

Taking Account of Unobserved Characteristics of Each School

No dataset can hope to capture all of the characteristics of a school environment that might influence student learning. Attitudes of students, teachers, and administrators, subtle differences in teaching styles, and so on could lead to some schools consistently outperforming others. The danger of this for our analysis is that without an attempt to take account of these unobserved variations among schools, we may incorrectly attribute some of these gains to variations across schools in some of the characteristics that we do have in our model.

Consider the following hypothetical example. Suppose that for every additional year of experience that a teacher has, a student's gain in test score rises by 1 point. We have data on one year of test score gains for two students at each school. The two solid lines in Figure A.2 show the gains in test scores for the pair of students at each school—at each school, the students with the more experienced teachers learn more quickly than the students with the less-experienced teachers. But as shown in Figure A.2, there are quite big unobserved differences between the schools attended by these four students. For reasons that we do not observe, students at School A, the school with teachers having far less experience, on average improve by a far greater margin in a given year.

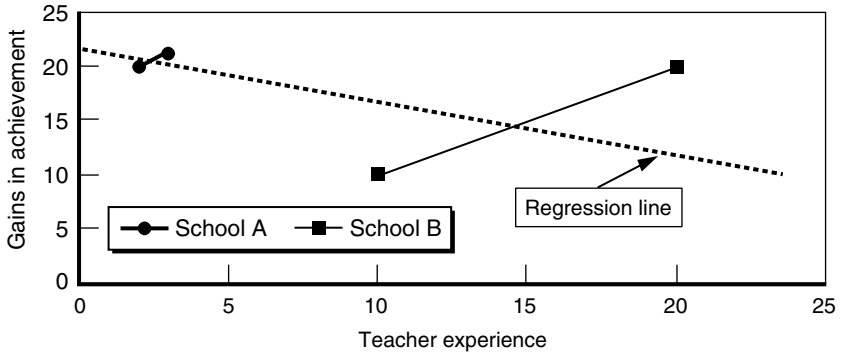


Figure A.2—Hypothetical Example of Incorrect Inferences About the Value of Teacher Experience for Student Learning, Caused by Unobserved Variations in School Quality

If we attempted to fit a regression line to these data, we would incorrectly infer that teachers with greater experience are associated with *lower* gains in student achievement. The regression line is shown by the dotted line in the figure. The position of this regression line is chosen in a certain sense to minimize the "distance" between data points and the line.

To avoid making such incorrect inferences, in all of our models we include dummy variables for every school in the sample. These indicator variables, equal to zero or one, take account of all unobserved aspects of a school's quality that were constant or fixed over the 1998 through 2000 period. It can be shown that inclusion of these fixed effects is equivalent to first calculating the average of each variable for a given school, then subtracting this mean from the value observed for all observations from that school, and then fitting the best line through the adjusted data points. In other words, we remove all of the variations among schools, which leaves only the variation within the school.

Figure A.3 shows what happens after we subtract the school averages from both test score gains and teacher experience for each of the four students who were shown in Figure A.2. The four observations now line up perfectly on a positively sloped line. A linear regression will now accurately calculate that test score gains rise by 1 point with every one-year increase in teacher experience.

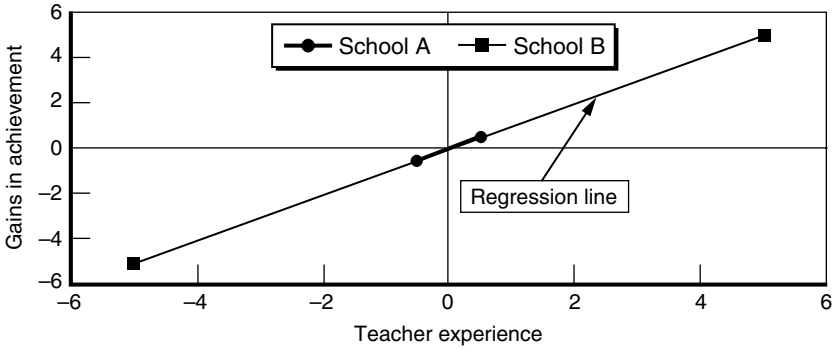


Figure A.3—Hypothetical Example of Correct Inferences About the Value of Teacher Experience for Student Learning, After Taking Account of Unobserved Differences in School Quality

All of our regression models will incorporate school fixed effects to remove any unobserved variations across schools that are fixed over time.

Taking Account of Unobserved Characteristics of Each Student's Neighborhood

A serious risk in all analysis of student achievement is that unobserved characteristics of the neighborhood that influence student achievement may be wrongly attributed to the quality of the school attended. For example, it seems quite evident from Chapter 4 that disadvantaged students begin elementary school less prepared than other students. It would be wrong to blame schools for low initial achievement.

We have partly taken account of such issues already by modeling gains in student achievement rather than levels. This will account for most of the large gap in initial achievement in grade 2 between disadvantaged and more-affluent students. It makes sense to remove these gaps as they appear to have more to do with preschool influences perhaps related to family or neighborhood environment than with the schools themselves.

But the risk remains that the gains in student achievement might still be higher in some schools than others because of unobserved variations in neighborhood characteristics that influence gains in achievement. For this reason, all of our models include indicator variables that indicate the zip code in which the student lives.

Taking Account of Unobserved Variations in Each Student's Rate of Learning

Finally, we need to take account of the fact that some students, irrespective of their academic environment, improve their academic achievement more quickly than others, either because of differences in innate ability, motivation, or unobserved characteristics of their home environment. This can lead to serious errors in our attempt to estimate the effect of various school resources on learning if there is a nonzero correlation between students' average rate of learning (or ability) and classroom characteristics.

Our solution to this problem, again afforded by the unusually rich dataset at hand, is to include fixed effects for each student. The advantages of this approach can be explained in the same way that we explained the need for school fixed effects. For instance, suppose that we have a pair of observations for two students, one of whom habitually learns more quickly but whom, through chance, has less experienced teachers than does the other student. Figure A.4 illustrates this, with the pair of observations for Student A, who naturally learns more quickly, illustrated in the upper left-hand corner of the graph. The dotted line shows that without taking account of the variations in ability between the students, we would incorrectly infer that students learn more slowly when they are placed in a class with a more-experienced teacher. Inclusion of student fixed effects solves the problem by subtracting the mean of each variable for each student, leading to the "correct" regression line, similar to what we showed in Figure A.3 in the explanation of school fixed effects. The use of student fixed effects is likely to be of great importance, given that schools do tend to steer students of a given achievement level toward certain types of classrooms. With the student fixed effects, we get around this problem by instead identifying the effect of school and classroom characteristics on learning by using variations from one year to the next in the environment faced by a student.

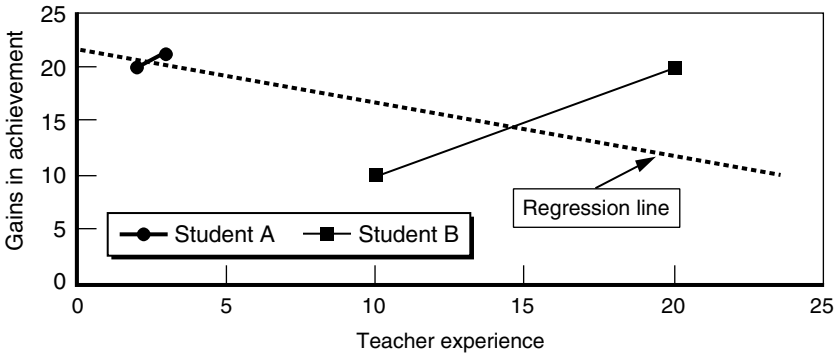


Figure A.4—Hypothetical Example of Incorrect Inferences About the Value of Teacher Experience for Student Learning, Caused by Unobserved Variations in Student Ability

Conclusion

Attempts to gauge the relative importance of various school characteristics on student achievement are hardly new. But the fact that we can follow individual students over time while linking them to teacher, classroom, and school characteristics provides us with some opportunities to take account of confounding influences. Specifically, we attempt to explain gains in individual achievement, not levels of achievement, because the latter likely reflect an entire lifetime's influences on each student. In addition we control for unobserved but fixed variations related to students' home zip code, their school, and their own ability. Chapters 5 and 6 focus on models that include fixed effects for home zip codes, schools, and students.

Appendix B

Details on the Regression Models for Elementary School Students

As outlined in the text, we model gains in test scores, or $\Delta Score_{icgst}$ for student i in classroom c in grade g in school s in year t as a function of school, family, personal, and classroom characteristics. (Classroom characteristics include teacher characteristics, class size, and classroom peer test scores.) Our regression model is

$$\begin{aligned}\Delta Score_{icgst} = {} & \alpha_s + \beta_{Zipcode_{it}} + \gamma_i + Score_{icgs,t-1}\omega \\ & + \mathbf{FAMILY}_{it}E + \mathbf{PERSONAL}_{it}\Phi + \mathbf{CLASS}_{icgst}\Gamma \\ & + \mathbf{SCHOOL}_{ist}\Lambda + \varepsilon_{it}\end{aligned}$$

where the first three variables on the right-hand-side of the equation represent fixed effects for the student's school, home zip code, and the student him or herself. $Score_{icgs,t-1}$ is the student's prior year score, added as a control for regression to the mean,. Items in bold face indicate vectors of time-varying family, personal, classroom, and school characteristics. The corresponding Greek letters are vectors of coefficients, and ε_{it} is an error term.

Chapter 5 outlines the list of right-hand-side variables in the above equation, which we use to "explain" the variation in gains in test scores. Two explanatory variables that deserve further explanation are the average test scores in a student's classroom and in his or her grade at the school. Suppose student i is in a class of n students. Define

$$\overline{Score}_{g,t-1}$$

as the average score in grade g in period t – 1 for all students in the district, with $\sigma_{g,t-1}$ representing the standard deviation across all students in the district of the score in grade g in period t – 1. Then in period t, we define

$$Peer_{icgs,t} = \frac{\frac{\sum_{j \neq i} Score_{j,g-1,t-1}}{n-1} - \overline{Score}_{g-1,t-1}}{\sigma_{g-1,t-1}}$$

In other words, for student i in class c in grade g in school s in year t, the average classroom peer achievement variable is set to the average test score in the previous year for all of the other (n – 1) students in the classroom, minus the district average test score last year in the previous grade, and all of this divided by the standard deviation of test scores last year in the previous grade districtwide. So, a value of 1.0 for this variable means that the student's classroom peers this year on average last year scored one standard deviation above the district mean. A value of –2.5 means that the student's classroom peers last year scored 2.5 standard deviations below the district average.

The other measure of a student's peers' achievement is analogous to the above but is defined as the average test scores last year of all the other students who this year are in that student's grade g at school s. Again, we subtract the district average and divide by the district standard deviation to standardize the measure.

Chapters 5 and 6 focus on results from the models that include student fixed effects, but in Web Appendices F and G we also present results from models that do not include student fixed effects but only the school and home zip code fixed effects. It is important to understand the tradeoffs between these two models. We argued in Appendix A that it is all too easy to obtain biased coefficients in models of test scores because of unobserved characteristics of the student that are correlated with some of the right-hand-side variables. This will bias the regression coefficients. The inclusion of the student fixed effects in the above model removes all unobserved but fixed influences on gains in test scores for the individual students. We believe that these models provide the most reliable estimates of the effect of classroom and other factors on student learning. However, these models "throw out" all of the variation among students in the data and so may be overly conservative. We provide the models without student fixed effects in Web Appendices F and G but limit our

references to these results in the main text, because variables that seem to "matter" in these models may only matter because of bias caused by unobserved student heterogeneity.

Bibliography

Allred, R. A., "Gender Differences in Spelling Achievement in Grades 1 Through 6," *Journal of Educational Research*, Vol. 83, No. 4, March–April 1990, pp. 187–193.

Baldassare, Mark, *PPIC Statewide Survey: Californians and Their Government*, Public Policy Institute of California, San Francisco, California, February 2000.

Baldassare, Mark, *PPIC Statewide Survey: Californians and Their Government*, Public Policy Institute of California, San Francisco, California, October 2002.

Betts, Julian R., "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics,* Vol. 77, 1995, pp. 231–250.

Betts, Julian R., "Is There a Link Between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings Institution, Washington, D.C., 1996.

Betts, Julian R., "The Two-Legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools," *Economic Policy Review*, Vol. 4, No. 1, 1998, pp. 97–116.

Betts, Julian R., Kim S. Rueben, and Anne Danenberg, *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California*, Public Policy Institute of California, San Francisco, California, 2000.

Betts, Julian R., and Anne Danenberg, "Resources and Student Achievement: An Assessment," in Jon Sonstelie and Peter Richardson, eds., *School Finance and California's Master Plan for Education*, Public Policy Institute of California, San Francisco, California, 2001, pp. 47–79.

Betts, Julian R., and Anne Danenberg, "School Accountability in California: An Early Evaluation," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2002,* Brookings Institution, Washington, D.C., 2002, pp. 123–197.

Bohrnstedt, George W., and Brian M. Stecher, eds., *Class Size Reduction in California: Early Evaluation Findings, 1996–1998* (CSR Research Consortium, Year 1 Evaluation Report), American Institutes for Research, Palo Alto, California, 1999.

Bohrnstedt, George W., and Brian M. Stecher, eds., *Class Size Reduction in California: Findings from 1999–00 and 2000-01,* California Department of Education, Sacramento, California, 2002.

California Department of Education, *The 1999 Base Year Academic Performance Index (API)*, available at http://www.cde.ca.gov/psaa/api/base/baseapi.htm, 2000.

California Center for the Future of Teaching and Learning, *The Status of the Teaching Profession 2000*, California Center for the Future of Teaching and Learning, Santa Cruz, California, 2000.

Coleman, James S., *Equality of Educational Opportunity*, Government Printing Office, 1966.

CSR Research Consortium, *Class Size Reduction in California 1996–98: Early Findings Signal Promise and Concerns*, American Institutes for Research, Palo Alto, California, 1999.

CSR Research Consortium, *Class Size Reduction in California: The 1998–99 Evaluation Findings*, American Institutes for Research, Palo Alto, California, 2000.

Darling-Hammond, Linda, "Teacher Quality and Student Achievement: A Review of State Policy Evidence," *Education Policy Analysis Archives,* Vol. 8, No. 1, January 2000.

Goldhaber, Dan, "The Mystery of Good Teaching," *Education Next*, Spring 2002, pp. 50–55.

Grissmer, David W., ed., Special Issue of *Educational Evaluation and Policy Analysis,* Vol. 20, Issue 2, Summer 1999.

Grissmer, David W., Ann Flanagan, Jennifer Kawata, and Stephanie Williamson, "Improving Student Achievement: What NAEP State Test Scores Tell Us," RAND, Santa Monica, California, 2000.

Grogger, Jeff, "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics,* Vol. 14, 1996, pp. 231–253.

Grogger, Jeff, and Eric Eide, "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources,* Vol. 30, Spring 1995, pp. 280–310.

Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature,* Vol. 24, 1986, pp. 1141–1177.

Hanushek, Eric A., "Money Might Matter Somewhere: A Response to Hedges, Laine and Greenwald," *Educational Researcher,* Vol. 23, 1994, pp. 5–8.

Hanushek, Eric A., "School Resources and Student Performance," in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings Institution, Washington, D.C., 1996, pp. 43–73.

Hanushek, Eric A., "Deconstructing RAND," *Education Matters,* Vol. 1, No.1, January 2001a.

Hanushek, Eric A., "RAND vs. RAND," *Education Matters,* Vol. 1, No.1, January 2001b.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin, "Does Peer Ability Affect Student Achievement?" National Bureau of Economic Research Working Paper 8502, Cambridge, Massachusetts, 2001.

Harcourt Brace Educational Measurement, *Stanford Achievement Test Series Spring Norms Book*, Ninth Edition, Harcourt Brace and Company, San Antonio, Texas, 1997.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential

School Inputs on Student Outcomes," *Educational Researcher,* Vol. 23, 1994, pp. 5–14.

Hoxby, Caroline, "Peer Effects in the Classroom: Learning from Gender and Race Variation," National Bureau of Economic Research Working Paper 7867, Cambridge, Massachusetts, 2000.

Jepsen, Christopher, and Steven Rivkin, *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools*, Public Policy Institute of California, San Francisco, California, 2002.

Joint Committee to Develop a Master Plan for Education—Kindergarten Through University, "School Readiness Working Group Final Report," Sacramento, California, 2002.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, and Brian Stecher, "What Do Test Scores in Texas Tell Us?" RAND, Issue Paper 202, 2000, available at http://www.rand.org/publications/IP/IP202/.

Koretz, Daniel, "Using Student Assessments for Educational Accountability," in Eric A. Hanushek and Dale W. Jorgenson, eds., *Improving America's Schools: The Role of Incentives*, National Academy Press, Washington, D.C., 1996.

Krueger, Alan B., and Diane M. Whitmore, "The Effect of Attending a Small Class in the Early Grades on College Test-Taking and Middle School Test Results: Evidence from Project STAR," Princeton University Industrial Relations Section, Working Paper #427, Princeton, New Jersey, 1999.

Mehan, Hugh, and Scott Grimes, *Measuring the Achievement Gap in San Diego City Schools*, San Diego Dialogue, San Diego, California, 1999.

Moran, Chris, "S.D. City Schools Retirement Rate Is 3 Times Normal," *San Diego Union-Tribune*, May 14, 2003, pp. B1 and B4.

Murnane, Richard J., *Effect of School Resources on the Learning of Inner City Children.* Ballinger, Cambridge, Massachusetts, 1975.

Murnane, Richard J., John B. Willett, and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determination," *Review of Economics and Statistics,* Vol. 77, May 1995, pp. 251–266.

Nowell, A., and L. V. Hedges, "Trends in Gender Differences in Academic Achievement from 1960 to 1994: An Analysis of Differences in Mean, Variance, and Extreme Scores," *Sex Roles*, Vol. 39, Nos. 1–2, July 1998, pp. 21–43.

Rose, Heather, and Julian R. Betts, *Math Matters: The Links between High School Curriculum, College Graduation, and Earnings*, Public Policy Institute of California, San Francisco, California, 2001.

San Diego Unified School District and San Diego Education Association, "Reformation (Extension) of the Term of the Current Collective Negotiations Contract to July 1, 2000, through June 30, 2003," 2002, available at ww.sdea.net.

Sonstelie, Jon, Eric Brunner, and Kenneth Ardon, *For Better or Worse? School Finance Reform in California,* Public Policy Institute of California, San Francisco, California, 2000.

Stecher, Brian M., and George W. Borhnstedt, eds., *Class Size Reduction in California: Findings from 1999–00 and 2000–01*, California Department of Education, Sacramento, California, 2002.

Stumpf, H., and J. C. Stanley, "Gender-Related Differences on the College Board's Advanced Placement and Achievement Tests, 1982–1992," *Journal of Educational Psychology*, Vol. 88, No.2, June 1996, pp. 353–364.

Tafoya, Sonya, "Linguistic Landscape of California Schools," *California Counts*, Vol. 4, No. 3, Public Policy Institute of California, San Francisco, California, February 2000.

Tully Tapia, Sarah, and Maria Sacchetti, "Testing Suggests Greater Fluency," *Orange County Register*, March 15, 2002.

Walsh, Kate, "Positive Spin: The Evidence for Traditional Teacher Certification, Reexamined," *Education Next*, Spring 2002, pp. 79–84.

About the Authors

JULIAN R. BETTS

Julian Betts is a senior fellow at the Public Policy Institute of California (PPIC) and a professor of economics at the University of California, San Diego. Much of his research has focused on the economic analysis of public schools. He has written extensively on the link between student outcomes and measures of school spending and the role that standards and expectations play in student achievement. He serves on The National Working Commission on Choice in K–12 Education and is a member of both the national advisory board of the Center for Research on Education Outcomes, Stanford University, and the San Diego Achievement Forum. He holds a Ph.D. in economics from Queen's University, Kingston, Ontario, Canada.

LORIEN A. RICE

Lorien Rice is a research fellow at PPIC. Her primary areas of interest include poverty, labor economics, education, and the distribution of income and wealth. She is also researching the role of transportation and geographic access in determining individuals' employment and educational outcomes. Before joining PPIC, she was a research and teaching assistant in the Department of Economics at the University of California, San Diego. She holds an M.A. in economics from the University of California, San Diego, and is currently completing work on her Ph.D. in economics at U.C. San Diego.

ANDREW C. ZAU

Andrew Zau is a research associate at PPIC. His current research focuses on the determinants of student achievement in the San Diego City School District. Before joining PPIC, he was an SAS programmer and research assistant at the Naval Health Research Center in San Diego, where he investigated the health consequences of military service during Operation Desert Shield/Desert Storm. He holds a B.S. in bioengineering from the University of California, San Diego, and a master of public health in epidemiology from San Diego State University.

Related PPIC Publications

Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California
Julian R. Betts, Kim S. Rueben, and Anne Danenberg

Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools
Christopher Jepsen and Steven Rivkin

For Better or For Worse? School Finance Reform in California
Jon Sonstelie, Eric Brunner, and Kenneth Ardon

Student and School Indicators for Youth in California's Central Valley
Anne Danenberg, Christopher Jepsen, and Pedro Cerdán

PPIC publications may be ordered by phone or from our website
(800) 232-5343 [mainland U.S.]
(415) 291-4400 [Canada, Hawaii, overseas]
www.ppic.org