National Assessment of Educational Progress

A Closer Look at Charter Schools Using Hierarchical Linear Modeling

U.S. Department of Education
NCES 2006-460

IES NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NAEP National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) is a nationally representative and continuing assessment of what America's students know and can do in various subject areas. For over three decades, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other subjects.

NAEP is a congressionally mandated project of the National Center for Education Statistics within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations.

By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement and relevant variables is collected under this program. The privacy of individual students and their families is protected to the fullest extent allowable under the law, and the identities of participating schools are not released.

In 1988, Congress established the National Assessment Governing Board (NAGB) to oversee and set policy for NAEP. The Board is responsible for selecting the subject areas to be assessed; setting appropriate student achievement levels; developing assessment objectives and test specifications; developing a process for the review of the assessment; designing the assessment methodology; developing guidelines for reporting and disseminating NAEP results; developing standards and procedures for interstate, regional, and national comparisons; determining the appropriateness of all assessment items and ensuring the assessment items are free from bias and are secular, neutral, and nonideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of NAEP reports.

U.S. Department of Education
NCES 2006-460

A Closer Look at Charter Schools Using Hierarchical Linear Modeling

August 2006

Henry Braun
Frank Jenkins
Wendy Grigg
Educational Testing Service

William Tirre
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

August 2006

The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch.

Suggested Citation
Braun, H., Jenkins, F., and Grigg, W. (2006). *A Closer Look at Charter Schools Using Hierarchical Linear Modeling* (NCES 2006-460). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office.

For ordering information on this report, write to
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398
or call toll free 1-877-4ED-Pubs or order online at http://www.edpubs.org.

Content Contact
William Tirre
202-502-7361
William.Tirre@ed.gov

The work upon which this publication is based was performed for the National Center for Education Statistics by Educational Testing Service, NAEP Education Statistics Services Institute, Pearson Educational Measurement, and Westat.

Executive Summary

Charter schools are a relatively new, but fast-growing, phenomenon in American public education. As such, they merit the attention of all parties interested in the education of the nation's youth. Accordingly, the National Assessment Governing Board (NAGB), which sets policy for the National Assessment of Educational Progress (NAEP), asked the National Center for Education Statistics (NCES) to conduct a pilot study of charter schools. A special oversample of charter schools, conducted as part of the 2003 fourth-grade NAEP assessments, permitted a comparison of academic achievement for students enrolled in charter schools to that for students enrolled in public noncharter schools. The school sample comprised 150 charter schools and 6,764 public noncharter schools. School participation rates were 100 percent for both charter and public noncharter schools; student participation rates were 92 percent and 94 percent for charter and public noncharter schools, respectively. Initial results employing data from the 2003 NAEP fourth-grade assessments in reading and mathematics were presented in the NCES report *America's Charter Schools: Results From the NAEP 2003 Pilot Study* (NCES 2004).

The present report comprises two separate analyses. The first is a "combined analysis" in which hierarchical linear models (HLMs) were employed to examine differences between the two types of schools when multiple student and/or school characteristics were taken into account. The rationale was that if the student populations enrolled in the two types of schools differed systematically with respect to observed background characteristics related to achievement, then those differences would be confounded with straightforward comparisons between school types.

HLMs were a natural choice for this analysis because such models accommodated the nested structure of the data (i.e., students clustered within schools) and facilitated the inclusion of variables describing student and school characteristics. In the combined analysis, the focus is the average difference in school means between the two types of schools in reading and mathematics. (This difference is similar to but not identical with the average difference between the two student populations.) Parallel analyses were carried out for reading and mathematics. In addition, supplementary analyses were conducted to evaluate the sensitivity of the results to various assumptions.

While the first analysis compares charter and public noncharter schools, the second analysis focuses on charter schools only. HLMs were employed to examine the relationship between mean school achievement and various characteristics of charter schools. Many of these characteristics were derived from a specially designed survey responded to by administrative staff in participating charter schools. Statistical significance was determined at the .05 level.

Results From the Combined Analyses

Reading

In the first phase of the combined analysis, all charter schools were compared to all public noncharter schools. The average charter school mean was 5.2 points lower than the average public noncharter school mean. After adjusting for multiple student characteristics, the difference in means was 4.2 points. Both differences were statistically significant. The adjusted difference corresponds to an effect size of 0.11 standard deviations. (Typically, about two-thirds of scale scores fall within one standard deviation of the mean.)

In the second phase, charter schools were classified into two categories based on whether or not they were affiliated with a public school district (PSD). Each category was compared separately with public noncharter schools. On average, the mean scores for charter schools affiliated with a PSD were not significantly different from those of public noncharter schools. However, on average, the means of charter schools not affiliated with a PSD were significantly lower than the means for public noncharter schools, both with and without adjustment. The effect size of the adjusted difference was 0.17 standard deviations.

In the third phase, the comparison between school types was restricted to schools having a central city location and serving a high-minority population, as there has been particular interest in those students who have traditionally not fared well in public schools. For this subset of 61 charter schools, there were no significant differences (for any fitted model) between the average charter school mean and the average public noncharter school mean.

Mathematics

In the first phase of the combined analysis for mathematics, all charter schools were compared to all public noncharter schools. The average charter school mean was 5.8 points lower than the average public noncharter school mean. After adjusting for student characteristics, the difference in means was 4.7 points. Both differences were statistically significant. The adjusted difference corresponds to an effect size of 0.17 standard deviations.

In the second phase, charter schools were classified into two categories based on whether or not they were affiliated with a PSD. Each category was compared separately with public noncharter schools. On average, the mean scores for charter schools affiliated with a PSD were not significantly different from those for public noncharter schools. However, on average, the means of charter schools not affiliated with a PSD were significantly lower than the means for public noncharter schools, both with and without adjustment. The effect size of the adjusted difference was 0.23 standard deviations.

In the third phase, the comparison between school types was restricted to schools having a central city location and also serving a high-minority population. There was a significant difference between the average of all charter school means and the average of public noncharter school means, as well as between charter school means not affiliated with a PSD and public noncharter school means. In both cases, the difference favored public noncharter schools, and the effect size of the adjusted difference was 0.17 standard deviations. However, there were no significant differences between the average of public noncharter school means and the means of charter schools affiliated with a PSD.

Sensitivity analyses

Since most charter schools are located in a relatively small number of jurisdictions, the distribution of charter schools across jurisdictions is not proportional to the distribution of all public schools. It is possible, therefore, that a national comparison between school types could be confounded with average differences in achievement among states. Accordingly, a set of parallel analyses for reading and mathematics was conducted for which the criterion was the difference between the standard student outcome and the mean NAEP score for the state. The results of the second set of analyses were very similar to those from the first set, with the effect size in the second set typically being a little smaller. While there appeared to be some confounding, it was not sufficient to alter the conclusions materially.

NAEP data are derived from a complex survey, and reported NAEP statistics are based on appropriately weighted student data. The HLM results were also based on the use of both student-specific and school-specific weights. Since there is no consensus on how to apply weights in a multilevel regression context (Pfefferman, et al. 1998), HLM analyses were rerun with different combinations of weights. Again, the results were quite similar to those obtained in the primary analysis.

Results From the Charter-School-Only Analysis

In addition to background data about the school, the charter school survey collected information about a number of areas related to school functioning, including policies from which the school had waivers or exemptions, areas in which the school was monitored, entities to which the school was required to report, student population served, and program content. For each area, a number of variables were constructed to represent the responses to the questions. All of these factors, together with student and school background variables, were incorporated in a series of HLMs in order to identify those characteristics that best accounted for differences in mean achievement across charter schools. The variation among school means for reading was nearly twice as large as it was for mathematics. Moreover, the number and nature of characteristics retained differed for reading and mathematics.

Reading

Nearly two-thirds of the variation among all students can be attributed to the variation between students within schools. Differences among schools on student variables (such as gender, race/ethnicity, disability status, status as an English language learner, and eligibility for free/reduced price lunch) accounted for 57 percent of the variance among school means. A reduced set of 10 school characteristics (such as teacher experience, region of the country, areas in which charter schools are monitored, and whether or not a charter school was part of another public school district) accounted for a further 27 percent of the variance. Thus, overall, student and school characteristics accounted for about five-sixths of the variance among school means. Of the 10 school characteristics, 3 were derived from the charter school survey (state monitoring of student achievement, monitoring for compliance with state/federal regulations, and charter school type), and 1 of the 3 (charter school type) was not statistically significant.

Mathematics

Approximately two-thirds of the variance among all students can be attributed to the variation between students within schools. Differences among schools on student variables accounted for 55 percent of the variance among school means. A reduced set of seven school characteristics (such as waivers for certain requirements, areas monitored, and charter granting agency) accounted for a further 11 percent of the variance. Thus, overall, student and school characteristics accounted for about two-thirds of the variance among school means. All seven school characteristics were derived from the charter school survey, and three (waiver for curriculum requirements, waiver for assessment requirements, and state agency granted charter) were statistically significant.

Cautions in Interpretation

There are a number of caveats to bear in mind in interpreting these results. First, the conclusions presented pertain to national estimates. Results based on a census of public schools in a particular jurisdiction may differ. Second, the data are obtained from an observational study rather than a randomized experiment, so the estimated effects should not be interpreted in terms of causal relationships. In particular, charter schools are "schools of choice." Parents may have been attracted to charter schools because they felt that their children were not well-served by public schools, and these children may have lagged behind their classmates. On the other hand, the parents of these children may be more involved in their children's schooling and provide greater support and encouragement. Without further information, such as measures of prior achievement, there is no way to determine how patterns of self-selection may have affected the estimates presented. That is, the estimates of the average difference in school means are confounded with average differences in the student populations, which are not adequately captured by the student characteristics employed in the analysis. It is also the case that students currently enrolled in charter schools have spent different amounts of time in one or more such schools. Consequently, the contributions of charter schools to students' learning vary across students both because of the differential effectiveness of the programs and the different amounts of exposure students have had to these programs.

Summary

After adjusting for student characteristics, charter school mean scores in reading and mathematics were lower, on average, than those for public noncharter schools. The size of these differences was smaller in reading than in mathematics.

Charter schools differ from one another in many ways. Some characteristics pertain to all public schools. Other characteristics—such as policies from which the school had waivers or exemptions, areas in which the school was monitored, entities to which the school was required to report, student population served, and program content—pertain only to charter schools. Such characteristics accounted for some of the observed variation in mean school performance.

For example, charter schools differ on whether or not they are affiliated with a public school district. In reading and mathematics, average performance differences between public noncharter schools and charter schools affiliated with a public school district were not statistically significant, while charter schools not affiliated with a public school district scored significantly lower on average than public noncharter schools.

Acknowledgments

The charter school pilot study and this analysis of the data resulting from it were conducted as a special project of the National Assessment of Educational Progress (NAEP), under the direction of the National Center for Education Statistics (NCES) and overseen by the National Assessment Governing Board. NAEP activities are carried out by Educational Testing Service (ETS), Pearson Educational Measurement, American Institutes for Research, NAEP Education Statistics Service Institute, and Westat. The collaborative and collegial work of many people made this report possible. The authors are grateful for all their efforts.

Data support and analysis activities for this report were carried out by Catherine Trapani and Laura Jerry. The complex analysis reported here would not have been possible without their expertise and guidance.

The design and production of this report were overseen by Loretta Casalaina with the assistance of Susan Mills. Ming Kuang led documentation and data-checking procedures with the assistance of Janice Goodis and Kit Chan. Janice Lukas coordinated the editorial process with the assistance of Mary Daane, Arlene Weiner, and Linda Myers. Rick Hasney coordinated the Web version of the report.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in the final version of this report: Chris Chapman, Young Chun, John Clement, Aaron Douglas, Hilary Forster, Arnold Goldstein, Steven Gorman, Andrew Kolstad, Roslyn Krob, Taslima Rahman, Anne Ricciuti, Marilyn Seastrom, and Alexandra Sedlacek. William Tirre, the project manager, was instrumental in assuring the project's completion.

THIS PAGE INTENTIONALLY LEFT BLANK.

Contents

List of Tables

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 1

Introduction

Charter schools are intended to provide an avenue of choice for parents within the public school framework. Since 1991, as more and more states have passed charter school authorizing legislation, charter schools have been a focus of attention for policymakers, educators, and the public at large, as well as the research community. This interest has led to a large number of studies that have examined different aspects of the charter school movement.

Most published studies have examined student outcomes, either at the state or national level. Comparisons of the achievement or, where possible, the growth in achievement of charter school students to public noncharter school students have been of particular interest. Two reviews of the extant literature are presented in Carnoy, et al. (2005) and in the report of the Charter School Achievement Consensus Panel (2006). The latter, in particular, carries out a detailed analysis of the different methods that have been used to study the achievement of students in charter schools. It points out that experimental studies can have high internal validity (the appropriateness of the inferences to the data at hand) but may lack external validity (the appropriateness of the inferences to other populations) since the sample of schools is usually not representative of a specific population. There are very few experimental studies of charter schools.

On the other hand, nonexperimental or observation studies typically have reasonably high external validity to the extent that the sample of schools is representative of schools in a state or the nation as a whole. However, in such studies, since there is no control over which students attend which schools, comparisons between types of schools must be interpreted cautiously, as they are subject to the confounding effects of various forms of selection bias. Consequently, it is impossible to unambiguously isolate the contributions that schools make to their students' learning.

Investigators have attempted to address the problem of selection bias in observational studies by utilizing auxiliary information about both students and schools in order to generate so-called adjusted comparisons that (it is hoped) are less subject to selection bias. Since there is no "gold standard" to appeal to, they often explore different model specifications to get a sense of the sensitivity of these estimated comparisons to various untestable assumptions. In addition, comparisons are sometimes conducted between categories of schools defined by the type of school, the location of the school, and/or the population served by the school. Unfortunately, it is difficult to determine to what extent variation in findings across studies reflects true differences in these comparisons and to what extent they are due to differential success in correcting for selection bias and/or differences in sample sizes.

At the state level, both Texas and North Carolina have provided fertile ground for research. In addition to consistent and long-standing state testing programs, they have comprehensive, state-wide databases and a relatively large number of charter schools. For example, Hanushek, Kain, and Rivkin (2002) analyzed Texas school data and were able to control for prior student achievement, as well as other student background variables. Overall, they found that there were no significant differences in achievement between students in the two types of schools. However, they did find that students enrolled in schools chartered by school districts had greater gains than those in public noncharter schools, while those in schools chartered by the state had lesser gains than those in public noncharter schools.

Bifulco and Ladd (2004) analyzed North Carolina data, making use of longitudinal data on student achievement in reading and mathematics, as well as background information about the students. They conducted two sets of comparisons. The first comparison was between students in charter schools and students in public noncharter schools. After adjusting for a number of student and family characteristics, they found that students in charter schools made, on average, significantly smaller achievement gains than comparable students in public noncharter schools. The second comparison focused on those students who had attended both public noncharter schools and charter schools during the period in question. Their gains in the two types of schools were compared, in effect using each student as his or her own control. Again, the finding was that gains for students enrolled in charter schools were significantly smaller on average than gains for students enrolled in public noncharter schools.

Hoxby (2004) carried out a multistate analysis, comparing students attending charter schools with those attending the nearest public noncharter school or the nearest public noncharter school with a similar racial composition. Again, the aim of such comparisons was to mitigate the effects of selection bias. Of the 21 states and the District of Columbia considered by Hoxby, in most cases there was a statistically significant advantage in the percent proficient (mathematics or reading) for students attending charter schools. Roy and Mishel (2005) argue that Hoxby's methodology does not adequately control for differences in the relevant characteristics of the students enrolled in the two types of schools. When that is done, almost all of the differences become nonsignificant.

At the national level, a recent report using data from the National Assessment of Educational Progress (NAEP) compares the achievement in both reading and mathematics of grade 4 students enrolled in charter schools to those enrolled in public noncharter schools (National Center for Education Statistics 2004). Comparisons are made overall and when the data are disaggregated by single student or school characteristics (e.g., race/ethnicity, gender, eligibility for free or reduced-price lunch, school location). In general, either the differences are nonsignificant or, if significant, students in charter schools are found to score lower on average than their counterparts in public noncharter schools. An earlier analysis of essentially the same data (Nelson, Rosenberg, and Van Meter 2004) obtained similar results. It occasioned a spirited debate concerning the proper interpretation of the findings. (See Carnoy et al. 2005 for a review of the issues.)

A critical question with regard to the comparison of charter and public noncharter schools is, "Would the estimates of any of the comparisons based on data from students enrolled in a subgroup of charter schools and data from students enrolled in a subgroup of public noncharter schools be materially changed if they were adjusted simultaneously with respect to several student characteristics (e.g., race/ethnicity, gender, and eligibility for free or reduced-price lunch)?" The present report addresses this question by examining reading and mathematics data for grade 4 from the 2003 NAEP administration. It provides estimates of a number of comparisons between the two types of schools with various classes of adjustments and systematically explores the sensitivity of the results to a number of assumptions.

This report employs a statistical technique termed hierarchical linear modeling. Hierarchical linear models properly reflect the structure of the NAEP sample, with students grouped by school. They allow for the inclusion of multiple explanatory variables in accounting for the variations in achievement within and between schools and, moreover, facilitate the derivation of appropriate standard errors for parameter estimates.

A recent study by Lubienski and Lubienski (2006) also employs hierarchical linear models to compare the achievement on the 2003 NAEP assessment of students enrolled in charter schools with that of students enrolled in public noncharter schools. That study only examines results in mathematics, but does so for both grades 4 and 8. (Note that there was no oversampling of charter schools in grade 8, so the number of charter schools in the NAEP eighth-grade sample is about 40 percent smaller than the number in the NAEP fourth-grade sample. In addition, a charter school survey was not conducted for the eighth grade.) While the present report excludes private schools, Lubienski and Lubienski (2006) carried out an omnibus set of comparisons among students attending charter schools, public noncharter schools, and various types of private schools. In the fourth grade, they find that, after controlling for student and school characteristics, students in charter schools achieve at significantly lower levels than students in public noncharter schools. In the eighth grade, they find no significant differences in achievement.

In addition to the attention paid to the sorts of comparisons previously described, there is considerable interest in studying the charter school movement as an important development in public education in its own right. Some reports (e.g., Finnigan et al. 2004) have documented the characteristics of charter schools, their staffs, their students, and how they have changed over time. They also describe differences among charter school authorizers and their relationships with the charter schools for which they are responsible.

These reports document the enormous variety among charter schools with respect to the circumstances of their founding, their education philosophies, and the degree to which they have been freed from different regulations, as well as the students they serve. In view of the substantial heterogeneity in academic achievement among charter schools, there is a question of whether there is a statistical relationship between various charter school characteristics and the performance of the students enrolled in those schools. The present report addresses this question, employing data from a special survey of charter schools conducted as part of the 2003 NAEP assessment.

Overview of NAEP

Since 1971, NAEP has been an ongoing, nationally representative indicator of what students know and can do in a variety of academic subjects. Over the years, NAEP has measured student achievement in many subjects, including reading, mathematics, science, writing, U.S. history, geography, civics, and the arts. NAEP is administered by the National Center for Education Statistics (NCES), within the U.S. Department of Education's Institute of Education Sciences, and is overseen by the National Assessment Governing Board (NAGB).

NAEP is not designed to provide scores for individual students and schools; instead, it provides results regarding subject-matter achievement, instructional experiences, and school environment for populations of students and groups of students in those populations. Through the use of complex item-sampling designs that present each participating student with only a portion of the total assessment, NAEP is able to produce accurate estimates of the performance of large groups of students, while minimizing the time burden on any individual student or school. In particular, comparisons of the achievement of students attending charter schools to the achievement of students attending public noncharter schools are suitable targets for estimation from NAEP data.

In 2003, NAEP assessments in reading and mathematics were conducted at grades 4 and 8. The content of each assessment was determined by subject-area frameworks developed by NAGB with input from a broad spectrum of educators, parents, and members of the general public. The complete frameworks for the NAEP reading and mathematics assessments are available on the NAGB website (http://nagb.org/pubs/pubs.html). Additional information about the design of the 2003 assessments is provided in appendix A.

NAEP, in its role as the nation's report card, has a responsibility to gauge student progress in America's schools. As a new kind of school, charter schools are an appropriate subject of study. The varied and changing nature of the charter school movement, however, makes such a study a challenge. Initial results from a pilot study of charter schools conducted as part of the 2003 NAEP fourth-grade assessments in reading and mathematics were presented in the NCES report *America's Charter Schools: Results From the NAEP 2003 Pilot Study* (NCES 2004), and on the NAEP website (http://nces.ed.gov/nationsreportcard/studies/charter/). The results presented in this report are based on additional analyses, using statistical modeling techniques that take demographic and other contextual differences into account in estimating differences in performance between students in charter and public noncharter schools. While the initial results presented in the earlier report were intended for a general audience, the results presented in this report are intended for more quantitatively-oriented social scientists and policymakers.

What Is a Charter School?

While charter schools are similar to other public schools in many respects, they may differ from each other in some important ways, including management, curriculum focus, student population, and exemptions from certain state or district policies. The unique characteristics of charter schools require additional information to be collected, beyond the information obtained from the regular NAEP questionnaires.

Charter schools, like other public schools, are a very diverse group of institutions—but one thing they have in common is that they are institutions of choice. Some charter schools, for example, are specifically designed to provide an alternative for parents who desire a learning environment different from that of the regular public school available to them. A representative group of charter schools is likely, then, to include schools that look very different from each other, which can make drawing general conclusions about charter schools complicated. Also, because the charter school movement is still new and evolving, the number and types of schools and students attending them are continually changing. As a result, some of the conclusions drawn from this study may apply only to the time period of this study.

The Charter School Pilot Study

As the charter school movement has grown, interest in how charter schools function and how well their students perform academically has increased. Motivated by this interest, NAGB, which sets policy for NAEP, asked NCES to conduct a pilot study of charter schools. The pilot study, conducted as part of the 2003 national assessment of fourth-graders in reading and mathematics, was designed to investigate the feasibility of assessing and reporting on the performance of students attending charter schools.

Charter school students took the same reading and mathematics assessments at the same time as students in all other schools. In addition to the information collected from the standard NAEP questionnaires completed by the students, their teachers, and school administrators, a newly created telephone survey designed to address issues relevant to charter schools was conducted. The respondents were administrative staff in the sampled charter schools who were knowledgeable about the school and its history. Questions and response choices from the charter school survey are available on the NAEP website (http://nces.ed.gov/nationsreportcard/studies/charter/).

Charter school sample

In the 2002–2003 school year (the year in which information for the NAEP charter school study was collected), there were 2,695 charter schools in 36 states (Center for Education Reform 2003). The number of charter schools differs from state to state in part because of state legislation regarding charter schools. This uneven distribution of charter schools across the states posed a sampling challenge for NAEP. In NAEP reading and mathematics assessments, the total number of schools and students sampled from each state is quite similar. Public schools in three states with a large number of charter schools—California, Michigan, and Texas—were oversampled as part of the 2003 NAEP assessments. This ensured that enough charter schools would be sampled and enough charter school students would be assessed to provide reliable national estimates (see the section on Sample Design in appendix A for additional information). Appropriate sample weights were applied to ensure that reported statistics were unbiased estimates of the results for the nation's charter schools.

A number of sources were used to construct the final sample of charter schools. First, the 2000–2001 Common Core of Data (CCD),[1] updated by state departments of education, was used to sample charter schools. Then the NAEP state coordinators independently verified the charter status of these schools. Additional charter schools were identified from the NAEP school questionnaire. Finally, in telephone interviews, a few schools were found not to be charter schools or not to have fourth-grade students eligible for the survey. A total of 150 schools were ultimately identified as charter schools, including 12 additional schools not originally identified on the NAEP website at the time of the 2003 NAEP data release. These schools, most of which did not return a school questionnaire, were discovered through the multiple sources of information just described.

Within each of the 150 participating charter schools, a random sample of students participated in either the reading or mathematics assessment—about one-half participated in reading and about one-half participated in mathematics. There were, however, some schools in which students were only assessed in one subject. Table 1-1 displays the number of charter school students sampled for the pilot study as well as the number of public noncharter school students sampled for the regular reading and mathematics assessments.

Table 1-1. Student sample size by type of public school and subject assessed, grade 4: 2003

	Student sample size by type of public school	
Subject	Charter schools	Public noncharter schools
Reading	3,296	188,148
Mathematics	3,238	188,201

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

School and student participation

The school participation rate for charter schools was 100 percent for the reading and mathematics assessments (which were conducted in the same schools). The school participation rate for public noncharter schools in each assessment was also 100 percent (6,764 schools participating). The student participation rate was 92 percent for charter schools in both the reading and mathematics assessments. For public noncharter schools, the student participation rate was 94 percent in both assessments. These rates were well within the NCES standards for ensuring unbiased samples and reporting data (see appendix A).

Every effort is made to ensure that all sampled students who are capable of participating in the assessment are assessed. A sampled student who is identified by the school as a student with a disability or as an English language learner may be assessed with accommodations allowed by NAEP; students so identified may be excluded from the assessment if they do not meet criteria for inclusion established by NAEP (see the section on participation of students with disabilities and/or English language learners in appendix A). The number of students assessed in the two subjects varied somewhat because more students tend to be excluded from reading assessments than mathematics assessments. In 2003, the exclusion rates for reading were 4 percent in charter schools and 6 percent in public noncharter schools, and the rates for mathematics were 2 percent and 4 percent, respectively (see table A-6 in appendix A).

Cautions in Interpretation

NAEP data are collected as part of an observational study rather than as a randomized experiment. Families choose to enroll their children in charter schools, and it is possible that there are systematic differences between those families and their children and the general population of families and their children that are not captured by the student characteristics available for analysis. If such differences are correlated with student achievement, then the estimated average difference in achievement between charter school students and public

[1] The Common Core of Data (CCD) is a program of the National Center for Education Statistics that annually compiles information about the nation's public schools and school districts and makes this information available through a public database. For more information, see http://nces.ed.gov/ccd/.

noncharter school students (even after adjusting for observed student characteristics) will be confounded to some degree with the unobserved differences between the families of the children in the two school types. This is usually termed "selection bias."

The methodological review presented in the report of the Charter School Achievement Consensus Panel (2006) makes it clear that selection bias is a serious threat to validity and that the estimated effects obtained should not be interpreted in terms of causal relationships. Although some studies employ a rich set of student characteristics, it cannot be assumed that selection bias has been eliminated. The concern is all the greater with the present study in which a number of relevant variables are not available. Perhaps the most critical is a measure of prior achievement, which is often employed as a covariate in such studies. Apparent differences in average achievement between charter school students and public noncharter school students may simply reflect average differences in achievement between their respective student populations (at entry into the fourth grade) that are not adequately captured by observed student characteristics.

Other relevant unobserved variables may include (but are not limited to) the following:

- The length of time that students in the charter school sample will have spent in the charter school system.
- The possible attraction of parents to charter schools because they felt that their children were not well served by public noncharter schools.
- The extent to which parents provide differential amounts of support and encouragement for academic achievement.

Inasmuch as NAEP draws samples of schools and students, estimates of the differences in achievement between school types are subject to uncertainty. In particular, the number of charter schools in the sample (150) is an order of magnitude smaller than the number of public noncharter schools (6,764). Consequently, the (estimated) standard errors of the difference estimates will tend to be higher than expected, given the total number of schools in the sample, because they are strongly influenced by the size of the smaller sample.

Finally, estimates of the school-type contrast are based on a nationally representative sample of charter schools and target the average difference in adjusted school means between school types across all jurisdictions. The sample of charter schools in any one jurisdiction is too small to make meaningful comparisons within that jurisdiction. Consequently, there may be jurisdictions in which the true average difference between charter schools and public noncharter schools deviates substantially from the national average. Detecting such deviations reliably, however, would require much larger samples.

Many of these cautions pertain equally to the analyses of charter schools only, where the focal research question is which charter school characteristics are correlated with student achievement after adjusting for other general school characteristics. With at most 150 charter schools in a study sample and more than 60 school characteristics to be examined, there is a danger of overinterpreting the results of any single analysis. These 60 characteristics include both general school characteristics and characteristics specific to charter schools. Since many pairs of characteristics are moderately correlated with each other, there is a concern that multicollinearity could mask the association between some of these characteristics and student achievement.

To mitigate the difficulty, a phased approach has been adopted, whereby small groups of substantively related characteristics are separately introduced into the regression. One group, for example, comprises indicators of the kinds of waivers obtained by each charter school. In the first phase, the "strongest" indicators in each group (if any) are called out and carried forward to the second phase, where they are pitted against the indicators of other groups to account for the variance in achievement across schools. Although there is no optimal strategy to deal with multicollinearity, the phased approach offers a reasonable way of identifying school characteristics of possible interest. Nonetheless, in view of the relatively small number of charter schools in the sample, the results obtained can only be suggestive of what might be found with a much larger sample.

Overview of Study Design and Application of Hierarchical Linear Modeling

Hierarchical linear modeling (HLM) is a class of techniques for analyzing data having a hierarchical or nested structure. For example, a database may consist of students who are nested within the schools they attend. Analyzing such data structures poses special problems. Conventional regression techniques either treat the school as the unit of analysis (ignoring the variation among students within schools) or treat the student as the unit of analysis (ignoring the nesting within schools). Neither approach is satisfactory.

In the former case, valuable information is lost, and the fitted school-level model can misrepresent the relationships among variables at the student level. In the latter case, it is assumed that if the model is correctly specified, then all the observations (e.g., student outcomes) are independent of one another. However, students attending the same school share many common, educationally relevant experiences that affect academic performance. As a result, scores on academic measures for students in the same school will not be independent, even after adjusting for student characteristics. Violation of the independence assumption means that, typically, estimates of standard errors of means and regression weights related to academic performance will be biased. Such bias, in turn, leads to situations in which statements of significance can occur too often or not often enough; that is, the actual Type I or Type II error rates can be quite different from the nominal ones.[2]

With HLM, on the other hand, the nested structure is represented explicitly in a multilevel model, with different variances assumed for each level. This ameliorates the above-mentioned problems with single-level models. Moreover, it is possible to postulate a separate student-level regression for each school. Both student and school characteristics can be included, and standard errors of means and regression coefficients can be estimated without bias. Consequently, the corresponding significance tests have the proper Type I error rate. At present, the use of HLM is strongly recommended for nested data. For further discussion, see Raudenbush and Bryk, chapter 1 (2002).

Hierarchical linear models are very flexible. They consist of two or more sets of linear regression equations that can incorporate predictor variables at each level of the data structure. In the example above, at the lower level (level 1) there is a regression equation for each school relating a student's outcome to one or more student characteristics (e.g., gender, race, socioeconomic status). The relationship between test scores and students' characteristics, represented by a set of regression coefficients, can differ from one school to another. At the higher level (level 2), each school's set of regression coefficients is predicted by one or more school characteristics (e.g., school type, school size, racial composition).

An analysis based on HLM yields a decomposition of the total variance into a between-student, within-school component and a between-school component. In addition, the output of the level 1 regression tells us how much of the variation in test scores between students within schools (i.e., the first component) can be accounted for by differences in student characteristics. Similarly, the output of a particular level 2 regression tells how much of the variation in school means, or adjusted school means (i.e., the second component), can be accounted for by differences in school characteristics such as school type.

Because the NAEP database conforms to a hierarchical structure—students nested within schools—HLM is well suited for carrying out an investigation that can help to elucidate the differences between charter and public noncharter schools. Previously published descriptive data indicate that the average mathematics score of students enrolled in charter schools is lower than the average score of students enrolled in public

[2] The Type I error rate is the probability that a statistical test will (incorrectly) reject a null hypothesis of no difference when the null hypothesis is true. The Type I error rate is set in advance of the analysis, and .05 is a typical value. The Type II error rate is the probability that a statistical test will (incorrectly) accept a null hypothesis when the null hypothesis is false. The Type II error rate is determined by the Type I error rate, the statistical test used, and the extent of the departure from the null hypothesis.

noncharter schools (NCES 2004). Although the average reading score appeared lower in charter schools than in public noncharter schools, the apparent difference was not found to be statistically significant. This relationship between averages generally persists when the data are disaggregated by student characteristics (such as gender or socioeconomic status) taken one at a time. The exception is that the differences are not statistically significant when the data are disaggregated by race.

Ideally, to ascertain the difference between the two types of schools, an experiment would be conducted in which students are assigned (by an appropriate random mechanism) to either charter or public noncharter schools. With a sufficiently large sample, such a procedure would guarantee that, on average, there are no initial differences between students attending charter and public noncharter schools, and would facilitate a fair comparison of the two types of schools. However, students are not randomly assigned to schools; families choose to seek admission for their children to charter schools. Thus, it is possible that students enrolled in the two types of schools differ on key characteristics that are associated with achievement. To the extent that is true, estimates of the average difference in achievement between school types will be confounded with initial differences between their student populations. This is of special concern if measures of prior academic achievement are unavailable, as is the case here.

The most common method to reduce the impact of confounding is adjustment by regression. Consequently, for this report, primary interest centers on how the inclusion of multiple predictor variables at the student level affects the estimated average difference in school means between charter and public noncharter schools. Secondary interest focuses on the impact of the inclusion of school covariates in level 2 of the model on the estimated average difference, as well as on the proportion of variation at each level that the predictor variables can account for.[3]

Note that the average difference in school means is, in general, not the same as the average difference in student outcomes (see appendix B for more details). Furthermore, the proper interpretation of the results of an analysis based on HLM must consider the substantive nature of the variables included in the model, as well as their statistical properties. This is addressed in the section on the specifics of the HLM analyses presented later in this chapter.

Combined and charter-school-only analyses

For both reading and mathematics, this report presents two sets of analyses. In the first set, HLM is used to estimate the size of the average difference in school means between charter and public noncharter schools. This set, referred to as the "combined analyses," comprises three phases:

Phase 1. Charter schools are compared to all public noncharter schools, using a variety of models that incorporate different combinations of student and school characteristics. There is substantive interest in the estimated average difference in school means between school types for each of the models.[4]

Phase 2. Charter schools are classified into two groups based on whether or not they are affiliated with a public school district (PSD). About one-half of charter school students nationally attend schools that are part of a public school district, and this characteristic is associated with differential achievement among charter schools (NCES 2004). The models in phase 1 are then rerun, enabling the estimation of the average difference in school means between each type of charter school and all public noncharter schools.

Phase 3. A subset of public schools (charter and noncharter) that have both a central city location[5] and also serve a high-minority student population[6] is selected. The analyses described in phase 1 and phase 2 are replicated for this subset.

[3] For some purposes, there may also be some interest in the magnitude and sign of the regression coefficients associated with the predictor variables.
[4] The interpretation of the estimated average difference depends on what characteristics are included in the model.
[5] Central city is defined in appendix A in the section on school-level variables.
[6] High-minority student population schools were defined as schools in which at least 50 percent of the students were Black or Hispanic.

The analyses in these three phases are carried out twice, using two different versions of test scores as the criterion in the level 1 regressions. First, the standard student outcomes produced by NAEP are used. Second, student outcomes modified to account for differences in mean achievement across states are used. For further details, see the section describing the model sequence for the combined analyses later in this chapter.

In the second set of analyses, the focus is on charter schools only. Here, HLM is used to investigate whether certain characteristics of charter schools are related to differences in the achievement of their students. As before, student characteristics are introduced as predictors of achievement at level 1, and various school characteristics are introduced at level 2. However, at level 2, variables specific to charter schools, such as type of governance, years of operation, and types of waivers, are also included. Interest centers on which of the regression coefficients associated with the charter-school-specific variables are statistically significant, as well as on the proportion of variation among school means (adjusted for student characteristics) they account for.[7]

Data Preparation

Summary data

For this study, the software program HLM6, which carries out the complex calculations associated with fitting HLMs, is used.[8] This program is designed to handle the NAEP data structure, which incorporates five plausible values for each assessed student.[9] The analysis procedure for each model is run five times, once for each set of plausible values. That is, in each run the plausible values play the role of the criterion in the regression. The final estimates are the averages of the results from the five analyses (Mislevy, Johnson, and Muraki 1992). These steps are automated in the HLM program.

Determining appropriate weights to be employed at the different levels in an HLM analysis is a complex matter. The general recommendation (Pfeffermann et al. 1998) is to split the standard NAEP weight into two components: a component applied to students within a school and a component applied to schools. This is the procedure followed in this report. However, alternative weighting schemes are possible, and the sensitivity of the reported results was, in fact, investigated. For details, consult appendix B.

The HLM program requires that the input data be organized in a summary data file. Appropriate summary data files were generated separately for the reading and mathematics assessments for the following samples:

- combined charter and public noncharter schools, with unadjusted test scores;
- combined charter and public noncharter schools, with state-mean-deviated test scores; and
- charter schools only, with unadjusted test scores.

This results in six data sets for analysis. Each data set can be used to fit a variety of models for the particular combination of school sample and test scores.

The first step of the procedure is to create a “flat file” (an ordinary text file) with one record per student, containing all of the corresponding student- and school-level variables. This flat file is then read into the HLM program, along with identification codes for students and schools and a set of student and school weights. This data-definition run establishes appropriate missing-value definitions (for student-level data), as well as variable labels. The HLM program reads this file and creates a multivariate data matrix, incorporating student and school data, that is used in all subsequent analyses. Once a model is specified and the weights selected, the program generates the appropriate likelihood function and obtains maximum likelihood estimates of the model parameters.

[7] In view of the qualitatively similar results obtained from the parallel analyses comparing all charter schools to all public noncharter schools, the charter-school-only analyses were conducted using only the unadjusted outcomes.

[8] For information regarding this program, consult Raudenbush, Bryk, Cheong, and Congdon (2004).

[9] Plausible values are random draws from the posterior distribution of scale scores for each student. The use of plausible values facilitates the unbiased estimation of group statistics and their associated standard errors. See Mislevy, Johnson, and Muraki (1992) for additional information.

Variable definitions

A listing of the student- and school-level variables used in the analyses is presented in figure 1-1. A number of these variables are recoded from their original form. For student-level variables, categories were combined if some categories had few responses or if a simpler categorization yielded adequate predictions. For school-level variables, categories were combined for similar reasons, particularly if a category had no responses. (One limitation of the HLM program is that it cannot handle missing data in the level 2 regression.) For some variables, a small amount of missing data was imputed from the means of similar schools. More detailed descriptions are given in appendix A.

Specifics of HLM Analyses

Centering

When a predictor variable is introduced at the student level, it is centered at the grand mean for that variable, that is, at the mean over all students in the population. This is consistent with standard practice in the analysis of covariance and has implications for the interpretation of the regression coefficients in the model. In particular, it means that, for each school, the intercept of the level 1 model is adjusted for the linear regression of the test scores on that variable. In a sense, that puts all school means on an equal footing with respect to that variable. In the HLM setting, the adjusted intercepts can be described as "adjusted school means." The variation among adjusted means will almost always be less, and usually much less, than the variation among the unadjusted means. For further discussion, see Raudenbush and Bryk, chapter 5 (2002).

Figure 1-1. Student-, school-, and charter-school-level variables

Student-level variables	School-level variables	Charter-school-level variables
Gender	Years of teaching experience	Waivers or exemptions from state/district policies
Race/ethnicity	Teacher certification	Areas monitored by state or chartering agency
Disability status	Student absenteeism	Groups requiring reporting on school's progress
Status as an English language learner	Percentage of students excluded	Charter-granting agency
Eligibility for free/reduced-price school lunch	Percentage of students in racial/ethnic groups	Student population served (e.g., at-risk students, gifted/talented students)
Computer in the home	Student mobility	New or pre-existing school
Number of books in the home	School location	Primary focus of program content
Number of absences	Region of the country	School managed by an organization or company managing other schools
	Percentage of students eligible for free/reduced-price school lunch	Part of another public school district or a local education agency
	Percentage of English language learners	State with strong chartering laws
	Percentage of students with a disability	

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Combined analyses

As noted above, the initial results (NCES 2004) indicated that the average score of charter school students was lower than that of public noncharter school students.[10] A natural follow-up question is: How large is the average difference in achievement between the two types of schools, after adjusting for differences in student characteristics? To answer the question, school means adjusted for student characteristics are estimated through a standard linear regression. This is referred to as the level 1 model. The adjusted school means are then regressed on an indicator of school type (i.e., charter or public noncharter). This is referred to as the level 2 model. The fitted coefficient of the school-type indicator is the desired estimate of the average difference in (adjusted) school means between the two school types. It is also possible to extend the previous analysis by incorporating school characteristics in the level 2 model.

To make these ideas more concrete, consider the following model:

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \ldots + \beta_{pj}X_{pij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + u_{0j}$

$$\beta_{1j} = \gamma_{10}$$

$$\vdots$$

$$\beta_{pj} = \gamma_{p0}$$

where i indexes students within schools, j indexes schools;

y_{ij} is the outcome for student i in school j ;

$X_1, \ldots, X_p$ are p student characteristics, centered at their grand means, and indexed by i and j as above;[11]

β_{0j} is the mean for school j, adjusted for the predictors $X_1, \ldots, X_p$;

$\beta_{1j} \ldots, \beta_{pj}$ are the regression coefficients for school j, associated with the predictors $X_1, \ldots, X_p$;

e_{ij} is the random error (i.e., residual term) in the level 1 equation, assumed to be independently and normally distributed with mean zero and a common variance σ^2 for all students;

W_{1j} is an indicator of the school type for school j, taking the value 1 for charter schools and 0 for public noncharter schools;

γ_{00} is the intercept for the regression of the adjusted school mean on school type;

γ_{01} is the regression coefficient associated with school type and represents the average difference in adjusted school means between charter and public noncharter schools;

u_{0j} is the random error in the level 2 equation, assumed to be independently and normally distributed across schools with mean zero and variance τ^2; and

$\gamma_{10, \ldots,}\ \gamma_{p0}$ are constants denoting the common values of the p regression coefficients across schools. For example, γ_{10} is the common regression coefficient associated with the first covariate in the level 1 model for each school.

In the level 1 equation, HLM estimates an adjusted mean for each school. In the level 2 equation, these adjusted means are in turn regressed on the school-type indicator. The regression coefficient of primary interest is γ_{01}, and it is referred to as the school-type contrast. (Note that γ_{01} describes a characteristic of the distributions of school-mean scores rather than of the distributions of individual student scores.)

[10] The difference of 5 score points in reading was not statistically significant. The difference of 6 score points in mathematics was statistically significant.

[11] That is, for example, $X_{1ij} = x_{1ij} - x_{1..}$ where x_{1ij} = value of characteristic 1 for student i in school j, and $x_{1..}$ = mean value on characteristic 1 over all students in the sample.

While adjusted school means are allowed to vary from school to school, the other regression coefficients in the level 1 model are all constrained to be constant across schools. This constraint is explicit in the structure of the level 2 equation above,[12] but could be relaxed if desired.

A slightly more general model is given below:

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \ldots + \beta_{pj} X_{pij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{02} W_{2j} + \ldots + \gamma_{0q} W_{qj} + u_{0j}$

$$\beta_{1j} = \gamma_{10}$$
$$\vdots$$
$$\beta_{pj} = \gamma_{p0}$$

In this model, the adjusted mean for school j is regressed on q school characteristics, including school type W_{1j}. In this case, γ_{01} indicates how much of the variation in adjusted school means can be accounted for by the school-type distinction, after taking into account school differences on the other $q - 1$ school characteristics. Thus, not only will the magnitude and statistical significance of the school-type contrast vary from model to model but also the interpretation and relevance to various research questions.[13]

Phase 2 employs models similar to the ones displayed above, with the difference that they now include two indicator variables to distinguish both charter schools affiliated with a PSD and charter schools not affiliated with a PSD from all public noncharter schools. In phase 3, the models from phases 1 and 2 are fit to the subset of schools in a central city location serving a high-minority population.

Charter-school-only analyses

For this set of analyses, models of the same form as above are fit, namely:

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \ldots + \beta_{pj} X_{pij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{02} W_{2j} + \ldots + \gamma_{0q} W_{qj} + u_{0j}$

$$\beta_{1j} = \gamma_{10}$$
$$\vdots$$
$$\beta_{pj} = \gamma_{p0}$$

For all the models in the charter-only analysis, the level 1 regression includes the same collection of student characteristics. (These are the characteristics employed in the level 1 regression of the combined analyses.) The level 2 regression is used to identify those school characteristics that can account for variation among school means, which have been adjusted for differences in student characteristics.

Obviously, the school-type contrast cannot appear here, since only charter schools enter the analysis. On the other hand, there is some interest in estimating the average difference in achievement between those charter schools that are affiliated with a PSD and those that are not. This can be accomplished by defining an indicator variable that distinguishes the PSD-affiliated charter schools from the non-PSD-affiliated charter schools. The regression coefficient associated with this indicator is denoted as the charter-type contrast.

[12] In general, these coefficients could also be modeled to have regressions on school type or other school characteristics. That direction was not explored in these analyses.

[13] A slight complication in interpretation of the school-type contrast arises because some school characteristics (e.g., student absentee rates and student mobility rates) may be partially influenced by school policies.

Estimating standard errors

In fitting models to NAEP data, estimated standard errors of parameter estimates should take account of the stratified, clustered sample design employed by NAEP. The two-level HLM employed for this study reflects the clustering of students within schools. In fact, the intraclass correlation ranges from about 0.20 (reading) to about 0.25 (mathematics). As a result, the estimated standard error for the school-type contrast is appropriately larger than it would be if an analysis were carried out based on the (erroneous) assumption that the NAEP student sample was a simple random sample.

There is, however, an additional complexity because in the NAEP combined sample each state serves as a stratum, and a probability sample of schools is selected from each stratum. While the two-level HLM does not directly reflect these aspects of the design, both design features are taken into account in the results reported below. Unequal school selection probabilities are addressed by incorporating school weights in the analysis. Because of the relatively small number of charter schools and their unbalanced distribution across states, estimating the school-type contrast within each state would be somewhat problematic. Therefore, as indicated above, a parallel analysis is carried out in which average differences in student achievement among states are eliminated. Adjusting for these between-state differences has negligible impact on the estimated standard errors of the school-type contrasts.

Combined analyses: Description of model sequence

In order to examine the differences between charter and public noncharter schools, the series of analyses summarized in figure 1-2 are carried out. One series is conducted for reading and another for mathematics. It should be noted that estimated regression coefficients and their corresponding estimated standard errors are produced for each fitted model. The latter are generated by HLM6 and are intended to capture variability due both to sampling and to measurement errors.[14]

The rationale and a verbal description for each model follow. (As previously mentioned, the coefficient of the charter/public-noncharter indicator is denoted as the school-type contrast.)

Model *a*: This model yields a decomposition of the total variance into within- and between-school components.

Model *b*: In this model, the school-type contrast estimates the average difference in unadjusted school means between charter and public noncharter schools. This estimate should be similar to the estimate obtained in the descriptive analysis (NCES 2004).

Model *c*: This model adjusts school means for differences in students' race/ethnicity. Therefore, the school-type contrast estimates what the average difference between charter and public noncharter schools would be if the NAEP student samples in each of the schools had the same race/ethnicity breakdown.

Model *d*: This model adjusts school means for differences in students' race/ethnicity, as well as other students' characteristics (see figure 1-1) that appear to have a statistically significant relationship to the outcome. The final set of predictor variables is determined by a sequence of exploratory analyses in which different combinations of variables are examined, much as in an ordinary regression analysis.[15] The retained set of variables is not guaranteed to be optimal, and there may be variables that are not included but are correlated with the outcome. The school-type contrast estimates what the average difference in school means between charter and public noncharter schools would be, if all schools' NAEP samples had the same breakdown on all included student variables. This is the focal model in the sequence.

[14] Measurement error is estimated from the variation in results across the five sets of plausible values.

[15] That is, there may be potential predictors having simple correlations with the outcome, but they are not retained in the final model because their partial correlation with the outcome, given the other predictors in the model, may be near zero.

Model *e*: This model builds on model *d* by including school-level variables in addition to the school-type contrast, which now estimates what the average difference in school means between charter and public noncharter schools would be, if all schools' NAEP samples had the same breakdown on included student variables and the same profile on included school variables. As for the student-level variables, the included school-level variables are determined by a sequence of exploratory analyses.

Model *d* is considered the "focal" model in the sequence inasmuch as it provides an estimate of the difference in average achievement between students in the two types of schools after accounting for differences in their student populations with respect to measured student characteristics. Model *e* contributes further insight by providing an estimate of the average difference in achievement after also accounting for differences between school types with respect to measured school characteristics. In this case, the school-type contrast is akin to a partial regression coefficient. As in the case of conventional regression, substantial differences between an ordinary and a partial regression coefficient signal the need for more caution in interpretation of the former.

The pattern in the estimated school-type contrasts obtained from models *b* through *e* can aid in understanding the structure of observed differences in achievement between charter schools and public noncharter schools. The interpretation of these estimates is guided by the *p* values[16] associated with the estimates. Each reported *p* value is calculated with respect to a particular model and not with respect to a sequence of models. Accordingly, if the *p* values are used to conduct significance tests, some of the differences declared significant (at the .05 level, say) may be significant by chance.

The interpretation of the estimated school-type contrasts is also complicated by the historical evolution of the charter school movement. Enabling legislation was passed at different times in the various states, and the level of interest and activity has varied a great deal among states. Consequently, the current distribution of charter school students is concentrated in a relatively small number of jurisdictions. Moreover, this distribution is quite different from the distribution of students in public noncharter schools.

This mismatch in distributions can cause some difficulty because there are differences in average NAEP scores among states. To the extent that, in each state, achievement in all public schools is influenced by general policies and practices (e.g., funding, curriculum, standards, and accountability regulations), national comparisons between charter schools and public noncharter schools will be partially confounded with between-state differences in achievement.

For example, suppose that charter schools are more concentrated in lower-performing states than are public schools overall. The national estimate of the charter school mean will then be more influenced by scores from those states than will the national estimate of the public noncharter school mean. Consequently, the estimate of the difference between the national means (i.e., the school-type contrast) will be larger (i.e., greater disadvantage for charter school students) than would be the case were the distribution of charter schools across states proportional to the distribution of all public noncharter schools.[17] This confounding can possibly be mitigated by the inclusion in the model of other covariates.

[16] The *p* value (two-sided) is the probability that, under the null hypothesis of no average difference between school types, a difference as large in absolute magnitude as the observed difference, or larger than it, would occur.

[17] This can lead to an instance of Simpson's Paradox. See appendix A for an illustration.

The simplest and most direct approach to addressing this confounding is to repeat the analyses described in figure 1-2, but use a different outcome measure. Accordingly, in a parallel series of analyses, the estimated state mean test score was subtracted from each student's outcome before fitting the model. The estimated state means are obtained from the NAEP database. In principle, the uncertainty in the estimated state means should be incorporated into our analyses. However, the corresponding standard errors are much smaller than the standard errors associated with the charter school means and are therefore ignored.

Figure 1-2. Description of the model sequence for the combined analyses

Model	Covariates included in level 1 regression	Covariates included in level 2 regression
a	None	None
b	None	School type
c	Race	School type
d	Race + other student characteristics	School type
e	Race + other student characteristics	School type + other school characteristics

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Thus, if there is a state-specific achievement component common to all schools in the state, but differing among states, then those components will now not contribute to the estimated school-type contrasts in the models described above. This parallel series is carried out in both phase 1 and phase 2. An alternative formulation would be to introduce indicators for each state in the regression model. The disadvantage is that it would greatly increase the number of coefficients to be estimated. In view of the charter school sample sizes, as well as the limitations of the HLM program, this model was not pursued. In the absence of other covariates, the results of the two analyses should be virtually the same. With the introduction of other covariates, results might differ slightly, but there is no reason to expect substantial changes.

Thus, 12 full sets of comparisons are reported: For both reading and mathematics, comparisons of all charter schools to public noncharter schools, using two outcome measures, and comparisons of each of two classes of charter schools to public noncharter schools, using two outcome measures. For each outcome measure, the results obtained from the different models listed in figure 1-2 can be contrasted. Moreover, for each model, the results obtained with standard test scores can be contrasted with those obtained with state mean-deviated test scores.

Charter-school-only analyses: Description of model sequence

The initial results from the charter school pilot study summarized in a descriptive report released earlier show that charter schools constitute a heterogeneous set of institutions (NCES 2004). They vary considerably in philosophy, governance, organization, and the regulatory environment in which they operate. In this section, the sequence of models employed to identify the characteristics of charter schools that account for some of the observed variability in achievement is described. There is particular interest in learning whether those characteristics include some features specific to charter schools, such as type of governance and waivers issued.

The charter school questionnaire was organized into a number of sections, each dealing with a different aspect of the charter school and the environment in which it operates. Inasmuch as the number of school characteristics to be examined is large in relation to the number of charter schools in the sample, there are concerns about multicollinearity. Consequently, a two-phase exploratory strategy was adopted. The responses to the questions in each section of the questionnaire were organized into distinct blocks of variables. Each block was entered separately, and those variables within the block that were statistically significant and/or large in magnitude were retained for inclusion in the second phase, in which variables from different blocks were included in a single analysis.[18] Figure 1-3 shows the model sequence. As is the case for the combined analysis, separate sequences are fitted for reading and for mathematics.

[18] The use of less stringent criteria for variable selection in the early, exploratory stages of a multistage model-fitting process is often recommended. See for example Tukey (1982).

Figure 1-3. Description of the model sequence for the charter-school-only analyses

Model	Student covariates included in level 1 regression	School covariates included in level 2 regression
1	None	None
2	All (student covariates)	None
3	All	All school covariates from combined analyses
4	All	Charter type (PSD[1]/non-PSD)
5.1	All	Waiver block
5.2	All	Monitoring block
5.3	All	Reporting block
5.4	All	Chartering agency block
5.5	All	Population served block
5.6	All	Program content block
5.7	All	Miscellaneous
6	All	Selected (from 3, 4, and 5.1 to 5.7)
7	All	Final set of covariates

[1] Public school district.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

The rationale and a verbal description of each model follow.

Model 1: This model yields a decomposition of the total variance into within- and between-school components.

Model 2: This model adjusts school means for the full set of student characteristics employed in model *d* of the combined analysis. The student characteristics included in model *d* are the ones that were found to have regression coefficients that were statistically significant. Accordingly, they were a natural starting point for selecting student characteristics to be included in the level 1 model for the charter-school-only analysis. As it happens, the estimated regression coefficients were again all statistically significant. The remaining models in the sequence all have this same level 1 structure. The principal purpose of incorporating student characteristics in the level 1 model is to obtain adjusted school means that serve as the criterion in a level 2 model. Since all the student characteristics included in model 2 had statistically significant regression coefficients, it was appropriate to employ the full set of characteristics in order to obtain adjusted school means when fitting models 3 through 7.

Model 3: This model includes the full set of school characteristics used in the exploratory phase of the combined analysis.

Model 4: This model enables an estimation of the average difference among adjusted school means between the PSD-affiliated and non-PSD-affiliated charter schools.

Models 5.1–5.7:
In each of these seven models, a different block of related charter-school-specific characteristics is included at level 2. The blocks are described in appendix A.

Model 6: This model includes those characteristics from models 3, 4, and 5.1–5.7 that were either significant or large in magnitude, as well as all variables describing the racial composition of the school.

Model 7: This model includes those characteristics from model 6 that were statistically significant, along with the charter-type indicator variable.

Chapter 2

Estimating the Average Difference in Achievement Between Charter and Public Noncharter Schools

The comparison of all charter schools to all public noncharter schools (phase 1) begins with the sequence of analyses described in chapter 1 (figure 1-2). This sequence is then repeated in phase 2, but with all charter schools now categorized by whether or not they are affiliated with a public school district (PSD). For both phase 1 and phase 2, the sequence of analyses is run twice—once with standard test scores and once with state-mean-deviated test scores.

In addition to overall comparisons between students in the two types of schools, estimates of a number of more focused comparisons are presented. These comparisons are intended to address the question of whether there are differences in achievement between students in the two types of schools when the schools are limited to those in particular locations and/or serving student populations with specific characteristics. Accordingly, in the third set of analyses, attention is turned to those public schools in central city locations with a population of at least 50 percent Black or Hispanic students. Within this subset of schools, public noncharter schools are compared both to all charter schools and to charter schools categorized by their PSD affiliation. These analyses are referred to as "reduced sample comparisons" (phase 3). Finally, results from the variance decompositions associated with the phase 1 analyses are also presented.

Fitting different hierarchical linear models (HLMs) to NAEP data helps to show how the inclusion of different sets of student covariates changes the estimate of the focal parameter of interest (i.e., the school-type contrast). Thus, in reporting the results of a series of analyses, there is interest not only in the estimate for a specific model, but also in the pattern of estimates through the series. Accordingly, instances may be noted when the estimate is not significant at the usual .05 level, but there will be limited discussion of its magnitude or sign.

Reading

Table 2-1 displays the estimated mean reading scores for students attending schools cross-classified by school type and population served. Recall that the population-served category has been divided into two strata. The stratum of principal interest is defined by a central city location and a high (50 percent or more) proportion of

Table 2-1. Estimated mean reading scores and number of schools, by public school type and location/population served, grade 4: 2003

	Charter schools		PSD[1]-affiliated charter schools		Non-PSD[1]-affiliated charter schools		Public noncharter schools	
Location/population served	NAEP reading mean	Number of schools	NAEP reading mean	Number of schools	NAEP reading mean	Number of schools	NAEP reading mean	Number of schools
All schools	212 (2.1)	148	218 (3.3)	70	208 (3.0)	78	217 (0.3)	6,754
Central city location/ serving high-minority population	197 (2.7)	61	200 (4.4)	25	195 (3.4)	36	197 (0.6)	1,039
All others	220 (2.4)	87	223 (3.3)	45	216 (3.6)	42	220 (0.3)	5,715

[1] Public school district.

NOTE: Standard errors of the estimates appear in parentheses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

disadvantaged minorities in the student body. For all four school types, achievement in this stratum was well below that in the other. About 40 percent of all charter schools fall in this stratum, while about 15 percent of public noncharter schools do. Charter schools in this stratum were slightly more likely to be non-PSD-affiliated than were charter schools overall (59 percent versus 53 percent).

Comparisons of all charter schools to all public noncharter schools

Table 2-2 contains results for models *b*–*e*, for both standard test scores and state-mean-deviated test scores. It displays estimates of the school-type contrast, comparing all charter schools to all public noncharter schools, along with the corresponding *p* values.

Consider first the estimates obtained when the outcome measure is the student's standard plausible values. For model *b*, the estimate is –5.2; that is, the average of the mean NAEP reading scores among charter schools is estimated to be about 5 points lower than the average of the mean NAEP reading scores among public noncharter schools. This difference is significant at the .05 level.

When student race is introduced at level 1 (model *c*), the estimated school-type contrast is no longer significant. However, with model *d*, when a larger set of student-level covariates is included (see figure 1-1 in chapter 1), the estimated school-type contrast is again significant. Such changes in magnitude and significance are not entirely unexpected when data are disaggregated with respect to different combinations of correlated characteristics.[1] Typically, results obtained from the more comprehensive set of characteristics are considered more credible, although limitations of sample size usually constrain the extent of disaggregation that is feasible.[2]

Accordingly, comparing models *b* and *d*, adjusting for differences in the full set of measured student characteristics reduces the gap between the two types of schools by about 1 score point (the differences are –4.2 for model *d* and –5.2 for model *b*). While the estimate of the gap is somewhat smaller, the estimated standard error of that estimate is reduced by 35 percent. The corresponding *t* statistic is significant. When school covariates at level 2 are also included (model *e*), the estimated school-type contrast is –3.3 and significant.

While a *p* value conveys the level of statistical significance of an estimate, it does not necessarily capture how interesting or meaningful the result is from a substantive point of view. For the latter purpose, it is common to express the estimate as an effect size (Cohen 1988). The purpose of computing an effect size for a statistic is to provide an indication of its practical

Table 2-2. Estimated average difference between mean reading scores in charter schools and public noncharter schools, by model, grade 4: 2003

Model			NAEP reading score		NAEP reading score (state-mean-deviated)	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value[2]	Estimate[3]	*p* value[2]
b	None	School type	– 5.2 (2.62)	.05	– 3.7 (2.58)	.15
c	Race	School type	– 2.4 (2.14)	.26	– 1.1 (2.14)	.61
d	Race and other student characteristics	School type	– 4.2 (1.70)	.01	– 3.0 (1.76)	.09
e	Race and other student characteristics	School type and other school characteristics	– 3.3 (1.53)	.03	– 4.2 (1.63)	.01

[1] Estimate of average difference in school means between charter schools and public noncharter schools, adjusted for other variables in the model.
[2] The *p* value is not adjusted for multiple significance tests.
[3] Estimate of average difference in school means between charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[1] This is analogous to Simpson's Paradox (see appendix A). For a numerical illustration, see table 7 of Carnoy, et al. (2005).
[2] See Cohen (1986) for a technical analysis of the problem.

import. By scaling the magnitude of the statistic through division by a measure of the spread of the distribution of test scores, a dimensionless quantity is obtained, which can be compared across different settings.

In this context, one measure of the effect size corresponding to an estimate of the school-type contrast is the ratio of the absolute value of the estimate to the standard deviation of the NAEP fourth-grade reading score distribution. Since the standard deviation is 37, the effect size of the estimate for model *d* is 4.2/37 = 0.11. An alternative approach to calculating an effect size is detailed in appendix A.

Turning to the analysis of the plausible values that have been state mean deviated for models *b*, *c*, and *d*, the estimates of the school-type contrast are smaller (i.e., less negative) than those obtained with standard plausible values, and none are statistically significant. For model *e*, the estimated effect of –4.2 is statistically significant and about 30 percent larger than the corresponding estimate for model *e* using standard plausible values.

Comparisons of two classes of charter schools to all public noncharter schools

Table 2-3 contains the results of fitting models in which a distinction is made between charter schools that are affiliated with a PSD and those that are not. The average difference in achievement between each class of charter schools and all public noncharter schools is estimated. As in table 2-2, the estimates for models *b*–*e* for both outcome measures are presented.

Consider first the estimates when the outcome measure is the student's standard plausible values. For the PSD-affiliated schools, none of the estimated effects for the four fitted models are significantly different from zero. The effect size for the estimate for model *d* is 2.0/37 = 0.05. On the other hand, the estimates for the non-PSD-affiliated schools are all negative and statistically significant for models *b*, *d*, and *e*. The adjustment for student covariates (model *d*) yields an estimated effect, –6.3, which is smaller than the estimated effect of –9.3 when there is no adjustment (model *b*). Note that for each fitted model, the effect size of the estimate is larger than the effect size of the corresponding estimate for the PSD-affiliated schools. The effect size for model *d* is 6.3/37 = 0.17.

Table 2-3. Estimated average difference between mean reading scores in two types of charter schools and public noncharter schools, by model, grade 4: 2003

Model			NAEP reading score				NAEP reading score (state-mean-deviated)			
	Level 1 covariates	Level 2 covariates	Estimate[1] (PSD[5]-affiliated)	*p* value[6]	Estimate[2] (non-PSD[5]-affiliated)	*p* value[6]	Estimate[3] (PSD[5]-affiliated)	*p* value[6]	Estimate[4] (non-PSD[5] affiliated)	*p* value[6]
b	None	School type	– 0.6 (3.70)	.88	– 9.3 (3.44)	.01	1.6 (3.60)	.66	– 8.6 (3.43)	.01
c	Race	School type	0.2 (3.12)	.96	– 4.8 (2.79)	.09	2.2 (3.15)	.48	– 4.1 (2.77)	.14
d	Race and other student characteristics	School type	– 2.0 (2.43)	.41	– 6.3 (2.25)	.01	0.0 (2.59)	1.00	– 5.7 (2.23)	.01
e	Race and other student characteristics	School type and other school characteristics	– 2.4 (2.23)	.29	– 4.2 (2.03)	.04	– 3.4 (2.47)	.17	– 4.9 (2.05)	.02

[1] Estimate of average difference in school means between PSD-affiliated charter schools and public noncharter schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between non-PSD-affiliated charter schools and public noncharter schools, adjusted for other variables in the model.
[3] Estimate of average difference in school means between PSD-affiliated charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
[4] Estimate of average difference in school means between non-PSD-affiliated charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
[5] Public school district.
[6] The *p* value is not adjusted for multiple significance tests.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Turning to the analysis of plausible values that have been state mean deviated, none of the estimated effects for PSD-affiliated schools are significantly different from zero. For non-PSD-affiliated schools, the estimated effects for all models are negative and statistically significant for models *b, d,* and *e.* The estimated effect in model *d* (–5.7) is smaller than the estimated effect for model *b* (–8.6). Note that for each fitted model, the effect size of the estimate is larger than the effect size of the corresponding estimate for the PSD-affiliated schools. The effect size for model *d* is 5.7/37 = 0.15.

Reduced sample comparisons of charter schools to all public noncharter schools

Table 2-4 presents results of fitting models *b*, *d*, and *e* for all charter schools and all public noncharter schools serving a high-minority population in a central city location. It also presents the results of fitting models *b*, *d*, and *e* for those same charter schools, now distinguished by whether or not they were affiliated with a PSD. The same sets of student and school characteristics (level 1 and level 2) were employed as in the analyses reported in the section comparing all charter schools to all public noncharter schools. The data are the standard plausible values, and the estimated regression coefficients at level 1 are similar to those in the earlier analysis.

When comparing all charter schools to all public noncharter schools in this stratum, the estimated school-type contrasts for all three models are not statistically significant. Similarly, when comparing PSD-affiliated and non-PSD-affiliated charter schools to all public noncharter schools, the estimated average differences in achievement for all models are not significantly different from zero.

Variance decompositions for comparisons of all charter schools to all public noncharter schools

As indicated earlier, an analysis based on HLM decomposes the total variance[3] of NAEP reading scores into the fraction attributable to differences among students within schools and the fraction attributable to differences among schools. Table 2-5 presents the variance decompositions corresponding to models *a*–*e* for the standard NAEP score outcome, comparing all charter schools to all public noncharter schools. The numbers in the second and fourth columns represent the percentage reduction in the residual variance achieved by that level of the model, treating the residual variance in model *a* as the baseline.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 1403 = 1101 + 302. That is, nearly 80 percent of

Table 2-4. Estimated average difference between mean reading scores in charter schools and public noncharter schools in a central city location and serving a high-minority population, grade 4: 2003

Model			All charter schools		PSD[1]-affiliated charter schools		Non-PSD[1]-affiliated charter schools	
	Level 1 covariates	Level 2 covariates	Estimate	*p* value[2]	Estimate	*p* value[2]	Estimate	*p* value[2]
b	None	School type	– 0.2 (3.85)	.95	3.1 (4.20)	.47	– 2.0 (5.34)	.71
d	Race and other student characteristics	School type	– 0.6 (2.76)	.82	1.8 (3.39)	.59	– 1.9 (3.66)	.59
e	Race and other student characteristics	School type and other school characteristics	0.9 (2.50)	.71	1.3 (3.05)	.68	0.8 (3.36)	.82

[1] Public school district.
[2] The *p* value is not adjusted for multiple significance tests.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[3] The total variance is the variance in scores in the full NAEP sample of students.

Table 2-5. Variance decompositions for reading scale scores, grade 4: 2003

Model	Level 1 covariates	Level 2 covariates	Between students, within schools: Variance	Between students, within schools: Percent of variance in model *a* accounted for	Between schools: Variance	Between schools: Percent of variance in model *a* accounted for
a[1]	None	None	1101	†	302	†
b	None	School type	1101	#	301	#
c	Race	School type	1067	3	169	44
d	Race and other student characteristics	School type	888	19	100	67
e	Race and other student characteristics	School type and other school characteristics	887	19	71	76

† Not applicable.
# Rounds to zero.
[1] Model *a* is unstructured.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

the total variance (1101/1403) is attributable to within-school heterogeneity, and slightly more than 20 percent of the total variance (302/1403) is attributable to between-school heterogeneity.[4] These figures are typical of two-level analyses of school achievement data (Raudenbush and Bryk 2002).

The introduction of the charter school contrast at level 2 (model *b*) accounts for a negligible proportion of between-school variance (i.e., zero to two decimal places), despite the fact that the corresponding regression coefficient is statistically significant. This seems counterintuitive, but the explanation is that there are relatively few charter schools in the overall school sample, and that most of the variance among school means is due to differences in means among schools within each school type. Consequently, eliminating the difference in means between school types would have almost no impact on the total variance among school means.

Turning to model *c,* including student race at level 1 reduces the within-school variance component by about 3 percent ((1101 – 1067)/1101), but the between-school component by about 44 percent ((302 – 169)/302). That is, the variance among school means adjusted for students' races is 56 percent as large as the variance among unadjusted school means. This result is consistent with the proposition that within-school populations are relatively homogeneous with respect to race and ethnicity, but that there is some variation among schools.[5] Since average differences in achievement between racial/ethnic groups are substantial, school means adjusted for those differences would be much less variable than unadjusted means. On the other hand, such adjustments would have little effect within a school in which the averages for the groups most strongly represented were similar.

In model *d,* all student-level covariates are included and together account for 19 percent of the within-school variance. However, as with model *c,* the impact on the variation at level 2 is much greater. In fact, the variance among school means adjusted for the full set of student characteristics (100) is now one-third as large as the variance among unadjusted school means (302). Finally, when school-level covariates are added (model *e*), the residual variance among adjusted school means is reduced to 71, representing an additional 9 percent (76 percent minus 67 percent) of the initial between-school variance accounted for. This incremental contribution of 9 percent seems rather small. However,

[4] Between-school heterogeneity refers to the variance among (unadjusted) school means.
[5] More exactly, in most schools the student population is largely drawn from racial/ethnic groups with similar average scores. For example, in one school most students might be Black or Hispanic, two groups with lower average achievement scores. In another school, most students might be White or Asian, two groups with higher average achievement scores. Relatively few schools have roughly equal numbers of students drawn from both higher- and lower-scoring groups.

the reduction in variation among adjusted school means from model *d* to model *e* is from 100 to 71 or 29 percent of the model *d* variance.

With the student and school variables available, more between-school than within-school heterogeneity is accounted for. The within-school variance component represents about 80 percent of the total variance, and about 20 percent of that component can be accounted for. The between-school variance component represents about 20 percent of the total variance, and about 80 percent of that component can be accounted for. All together, about one-third of the total variance is accounted for.

The variance decompositions are also carried out for the analyses using the student plausible values that have been state mean deviated. This affects only the between-school variance component, which is reduced by about 10 percent. The percentages of between-school variance explained for models *b*–*e* are very similar to those for the standard student outcomes (see appendix C for the full set of results). Variance decompositions for the analyses employing two classes of charter schools are identical to those for one category and so do not require separate discussion.

The variance decompositions reported in table 2-5 are an important by-product of these HLM analyses, the principal purpose of which is to yield estimates of parameters of interest and unbiased estimates of the corresponding standard errors. The contribution of the school-type contrast to the reduction in between-school variance is strongly determined by the relative proportions of the two types of schools. Consequently, the variance decomposition results do not directly help in interpreting the school-type contrast. On the other hand, these results do enhance understanding of the context in which these analyses take place. Specifically, comparing the results for models *b* and *d* indicates that schools (in general) differ widely in measured characteristics of students which are associated with achievement. Consequently, when those characteristics are adjusted for, the heterogeneity in school means is reduced by two-thirds. This leaves one-third of the between-school variance to be explained by other unmeasured student characteristics and by other unmeasured school characteristics. In a sense, that establishes a limit on the relative importance of differences in school characteristics in comparison to differences in student characteristics in explaining the variation in achievement between schools.

Mathematics

Table 2-6 displays the estimated mean mathematics scores for students attending schools cross-classified by school type and population served. Recall that population served has been divided into two strata. As was the case for reading, achievement in central city schools with a high percentage of minority students is well

Table 2-6. Estimated mean mathematics scores and number of schools, by public school type and location/population served, grade 4: 2003

	Charter schools		PSD[1]-affiliated charter schools		Non-PSD[1]-affiliated charter schools		Public noncharter schools	
Location/population served	NAEP mathematics mean	Number of schools	NAEP mathematics mean	Number of schools	NAEP mathematics mean	Number of schools	NAEP mathematics mean	Number of schools
All schools	228 (2.0)	150	234 (3.1)	70	225 (2.3)	80	234 (0.2)	6,761
Central city location/ serving high-minority population	216 (1.7)	61	218 (3.0)	25	214 (2.1)	36	218 (0.5)	1,040
All others	236 (2.0)	89	240 (3.0)	45	231 (2.7)	44	237 (0.2)	5,721

[1] Public school district.

NOTE: Standard errors of the estimates appear in parentheses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

below that in the other schools, for all school types. The cell counts are almost identical to those in table 2-1.

Comparisons of all charter schools to all public noncharter schools

Table 2-7 contains mathematics results for models *b*–*e*, for both standard test scores and state-mean-deviated test scores. It displays estimates of the school-type contrast, comparing all charter schools to all public noncharter schools, along with the corresponding *p* values.

Consider first the estimates obtained when the measure is the student's standard plausible values. For model *b,* the estimate is –5.8; that is, the average of the mean NAEP mathematics scores among charter schools is estimated to be nearly 6 points lower than the average of the mean NAEP mathematics scores among public noncharter schools. The difference is significant at the .05 level.

When student race at level 1 is introduced (model *c*), the estimated school-type contrast is no longer significant. However, when all student-level covariates are included (model *d*), the estimated school-type contrast is again significant. That is, adjusting for differences in the full set of measured student characteristics reduces the gap between the two types of schools by about 1 score point [–4.7 – (–5.8) = 1.1]. When school covariates are also included at level 2 (model *e*), the estimated school-type contrast is –3.5 and significant.

As in the case of the reading analysis, the effect size of an estimate was employed as a means of conveying the substantive meaning of the result. Since the standard deviation of the NAEP fourth-grade mathematics score distribution is 28, the effect size of the estimate for model *d* is 4.7/28 = 0.17. Note that this is about 50 percent larger than the effect size for the comparable model for reading.

Turning to the analysis of the outcomes that have been state mean deviated, for models *b, c,* and *d,* the estimated school-type contrasts are similar to, but about 10 to 20 percent smaller than, those obtained with standard plausible values. As before, the estimated school-type contrasts for models *b* and *d* are significant, but not significant for model *c*. The effect size for the estimate for model *d* is 4.1/28 = 0.15. For model *e,* the estimated school-type contrast is –4.6, which means that after accounting for student and school characteristics, as well as differences in state means, mean achievement in charter schools is 4.6 points lower than mean achievement in public noncharter schools. The estimated effect of –4.6 is about 30 percent larger than the corresponding estimate for model *e* (–3.5) with the standard plausible values.

Table 2-7. Estimated average difference between mean mathematics scores in charter schools and public noncharter schools, by model, grade 4: 2003

Model			NAEP mathematics score		NAEP mathematics score (state-mean-deviated)	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value[2]	Estimate[3]	*p* value[2]
b	None	School type	– 5.8 (1.99)	.00	– 5.2 (2.00)	.01
c	Race	School type	– 2.9 (1.64)	.07	– 2.3 (1.64)	.15
d	Race and other student characteristics	School type	– 4.7 (1.46)	.00	– 4.1 (1.47)	.01
e	Race and other student characteristics	School type and other school characteristics	– 3.5 (1.31)	.01	– 4.6 (1.39)	.00

[1] Estimate of average difference in school means between charter schools and public noncharter schools, adjusted for other variables in the model.
[2] The *p* value is not adjusted for multiple significance tests.
[3] Estimate of average difference in school means between charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Comparisons of two classes of charter schools to all public noncharter schools

Table 2-8 contains the results of fitting models in which a distinction is made between those charter schools that are part of a PSD and those not part of a PSD. The average difference in achievement between each class of charter schools and all public noncharter schools is estimated. Estimates for models *b*–*e* for both outcome measures are presented.

Consider first the estimates when the outcome is the student's standard plausible values. For the PSD-affiliated schools, none of the estimated effects for the four fitted models are significantly different from zero.

On the other hand, the estimates for the non-PSD-affiliated schools are all negative and statistically significant. The adjustment for student covariates (model *d*) yields an estimated school-type contrast, –6.4, which is smaller than the estimated school-type contrast of –9.8 when there is no adjustment (model *b*). The effect size for model *d* is 6.4/28 = 0.23.

Turning to the analysis of outcomes that have been state mean deviated, none of the estimated effects for PSD-affiliated schools are significantly different from zero.

For non-PSD-affiliated schools, the estimated effects for all models are negative and statistically significant. The estimated effect in model *d* (–7.1) is smaller than the estimated effect in model *b* (–10.4). The effect size for model *d* is 7.1/28 = 0.25. Augmenting model *d* by the inclusion of other school characteristics (i.e., model *e*) does not result in changes in the statistical significance of the school-type contrast.

Reduced-sample comparisons of charter schools to all public noncharter schools

In table 2-9, results of fitting models *b, d,* and *e* are presented both for all charter schools serving a high-minority population in a central city location and those same charter schools classified by whether or not they were affiliated with a PSD. In each case, the charter schools are compared to all public noncharter

Table 2-8. Estimated average difference between mean mathematics scores in two types of charter schools and public noncharter schools, by model, grade 4: 2003

Model			NAEP mathematics score				NAEP mathematics score (state-mean-deviated)			
	Level 1 covariates	Level 2 covariates	Estimate[1] (PSD[5]-affiliated)	*p* value[6]	Estimate[2] (non-PSD[5]-affiliated)	*p* value[6]	Estimate[3] (PSD[5]-affiliated)	*p* value[6]	Estimate[4] (non-PSD[5]-affiliated)	*p* value[6]
b	None	School type	– 1.2 (3.13)	.71	– 9.8 (2.18)	.00	0.9 (3.06)	.77	– 10.4 (2.15)	.00
c	Race	School type	– 0.7 (2.65)	.80	– 4.9 (1.87)	.01	1.4 (2.63)	.60	– 5.5 (1.79)	.00
d	Race and other student characteristics	School type	– 2.6 (2.40)	.28	– 6.4 (1.62)	.00	– 0.5 (2.41)	.82	– 7.1 (1.54)	.00
e	Race and other student characteristics	School type and other school characteristics	– 3.1 (2.16)	.16	– 3.9 (1.49)	.01	– 3.3 (2.25)	.14	– 5.7 (1.60)	.00

[1] Estimate of average difference in school means between PSD-affiliated charter schools and public noncharter schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between non-PSD-affiliated charter schools and public noncharter schools, adjusted for other variables in the model.
[3] Estimate of average difference in school means between PSD-affiliated charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
[4] Estimate of average difference in school means between non-PSD-affiliated charter schools and public noncharter schools, adjusted for differences in state means and for other variables in the model.
[5] Public school district.
[6] The *p* value is not adjusted for multiple significance tests.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Table 2-9. Estimated average difference between mean mathematics scores in charter schools and public noncharter schools in a central city location and serving a high-minority population, grade 4: 2003

Model			All charter schools		PSD[1]-affiliated charter schools		Non-PSD[1]-affiliated charter schools	
	Level 1 covariates	Level 2 covariates	Estimate	*p* value[2]	Estimate	*p* value[2]	Estimate	*p* value[2]
b	None	School type	– 3.5 (1.91)	.07	– 2.8 (3.19)	.38	–3.8 (2.39)	.11
d	Race and other student characteristics	School type	– 4.8 (1.48)	.00	– 4.7 (2.71)	.09	–4.8 (1.79)	.01
e	Race and other student characteristics	School type and other school characteristics	– 2.0 (1.61)	.22	– 4.3 (3.09)	.16	–0.7 (1.61)	.68

[1] Public school district.
[2] The *p* value is not adjusted for multiple significance tests.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

schools serving a high-minority population in a central city location. The same sets of student and school characteristics (level 1 and level 2) are employed as in the analysis reported in the section comparing charter schools to all public noncharter schools. The data are the standard plausible values, and the estimated regression coefficients at level 1 are similar to those in the earlier analysis.

Comparing all charter schools to all public noncharter schools in this stratum, the estimated school-type contrast is negative for all three models, but is only significantly different in model *d*. That is, when school means have been adjusted for differences in student characteristics, the average difference in achievement between the two school types is –4.8 (model *d*). The corresponding effect size is 4.8/28 = 0.17. (Note that this is nearly the same as the effect size of 0.17 for the full school sample.)

Similarly, comparing PSD-affiliated and non-PSD-affiliated charter schools to all public noncharter schools, the estimated average difference in mean achievement is negative for all three models but, with one exception, not statistically significant. That exception occurs for non-PSD-affiliated schools, after adjusting for student covariates (model *d*). The corresponding effect size is 4.8/28 = 0.17.

Variance decompositions for comparisons of all charter schools to all public noncharter schools

In parallel to the analysis of the reading data, HLM is used to decompose the total variance of NAEP mathematics scores into the fraction attributable to differences among students within schools and the fraction attributable to differences among schools. Table 2-10 presents the variance decompositions corresponding to models *a*–*e* for the standard NAEP scores, comparing all charter schools to all public noncharter schools.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 818 = 608 + 210. Consequently, nearly 75 percent of the total variance (608/818) is attributable to within-school heterogeneity, and slightly more than 25 percent of the total variance (210/818) is attributable to between-school heterogeneity.

As before, the introduction of the charter school contrast at level 2 (model *b*) accounts for a negligible proportion of between-school variance (i.e., zero to two decimal places), despite the fact that the corresponding regression coefficient is statistically significant. Again, the explanation is that there are relatively few charter schools in the overall school sample, and that most of the variance at the school level is due to differences among schools within each type of school.

Turning to model *c,* including student race at level 1 reduces the within-school variance component by about 5 percent, but the between-school component by about 44 percent. That is, the variance among school means adjusted for students' race is 56 percent as large as the variance among unadjusted school means. Again, this is consistent with the relative homogeneity within schools, and the relative heterogeneity across schools, with respect to race and ethnicity. (See the comparable section for reading presented earlier in this chapter.)

In model *d,* all student-level covariates are included and together account for about 21 percent of the within-school variance. The variance among adjusted school means (76) is slightly more than one-third as large as the variance among unadjusted school means in model *a* (210). Finally, when school-level covariates are added (model *e*), an additional 10 percent (74 percent minus 64 percent) of the initial between-school variance can be accounted for. As in the case of reading, the incremental contribution of 10 percent is rather small. However, the reduction in variation among adjusted school means from model *d* to model *e,* that is, from 76 to 55 or 28 percent, is substantively meaningful.

Thus, with the student and school variables available, more between-school than within-school heterogeneity is accounted for. The within-school variance component represents about 75 percent of the total variance, and about 20 percent of that component can be accounted for. On the other hand, about 75 percent of the between-school variance component, which represents about 25 percent of the total variance, is accounted for. All together, as was the case for reading, about one-third of the total variance is accounted for.

The variance decompositions are also carried out for the analyses using the student plausible values that have been state mean deviated. This affects only the between-school variance component, which is reduced by about 8 percent. The percentages of between-school variance explained for models *b* through *e* are very similar to, but smaller than, those for the standard student outcomes (see appendix C for the full set of results). As before, variance decompositions for the analyses employing two classes of charter schools are identical to those for one category and so do not require separate discussion.

Table 2-10. Variance decompositions for mathematics scale scores, grade 4: 2003

Model			Between students, within schools		Between schools	
	Level 1 covariates	Level 2 covariates	Variance	Percent of variance in model *a* accounted for	Variance	Percent of variance in model *a* accounted for
a[1]	None	None	608	†	210	†
b	None	School type	608	#	210	#
c	Race	School type	577	5	117	44
d	Race and other student characteristics	School type	482	21	76	64
e	Race and other student characteristics	School type and other school characteristics	482	21	55	74

† Not applicable.
# Rounds to zero.
[1] Model *a* is unstructured.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

As was the case for reading, the results of the variance decomposition enhance understanding of the context in which the HLM analyses take place. A comparison of the variance between schools for models *b* and *d* indicates that schools (in general) differ widely in measured student characteristics associated with achievement. Specifically, when those characteristics are adjusted, the heterogeneity among school means is reduced by almost two-thirds. This leaves about one-third of the variance to be explained by other unmeasured student characteristics and by other unmeasured school characteristics. In a sense, this establishes a limit on the relative importance of differences in school characteristics, in comparison to differences in student characteristics, in explaining the variation in achievement among students.

Summary

The comparisons of all charter schools to all public noncharter schools for reading are qualitatively similar for the two outcome measures examined. In particular, after inclusion of student covariates (model *d*), mean NAEP scores in charter schools are lower on average than those in public noncharter schools, with differences of 4.2 (for standard NAEP scores) and 3.0 points (for state-mean-deviated NAEP scores). However, only the former is significant at the .05 level. Next, both PSD-affiliated and non-PSD-affiliated charter schools are compared to public noncharter schools. Means in PSD-affiliated schools do not differ significantly from means in all public noncharter schools. However, there is a statistically significant difference between means in non-PSD-affiliated charter schools and those in public noncharter schools.

In the last phase of the analysis, attention focused on those schools serving urban, high-minority-student populations. For this reduced school sample, there are no significant differences in mean reading scores between all charter schools and public noncharter schools or between either of the two types of charter schools and public noncharter schools.

For the full school sample, results for mathematics generally resemble those for reading, with respect to the estimated size of the disadvantage attached to charter schools in each model. The main difference between the two subjects is that, typically, for each model the effect size in mathematics is greater than that for reading. This is likely because both the between-student within-school and the between-school variance components are smaller in mathematics. Furthermore, the school-type contrast for mathematics is statistically significant at the .05 level for both outcome measures.

Another difference arises in the analysis of data from the reduced school sample (i.e., those schools serving central city, high-minority populations). In contrast to reading, estimates of the school-type contrast for model *d* (i.e., the model that includes student covariates at level 1) indicate statistically significant differences in mean mathematics achievement between charter schools and public noncharter schools, as well as between non-PSD-affiliated charter schools and public noncharter schools.

For the full school data, the patterns in variances accounted for by the different models (relative to the baselines established by fitting the unstructured model *a*) are remarkably similar for the two subjects. In both cases, about three-quarters of the total variance is due to differences between students within schools. However, the total variance is smaller in mathematics than in reading. For both reading and mathematics, adjusting school means for all measured student covariates results in a two-thirds reduction in the variance between school means.

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 3

Examining Characteristics of Charter Schools Associated With Student Achievement

This chapter examines the relationships between various characteristics of charter schools and the achievement of the students enrolled in those schools. As in the previous chapter, hierarchical linear models (HLMs) are employed to represent the relationships among the variables. In the previous chapter, attention focused on a single regression coefficient, the school-type contrast. In this chapter, however, attention focuses on the contribution of each school-level variable in accounting for the variance in school means. There is also some interest in the magnitudes and signs of all the regression coefficients. Parallel analyses for reading and mathematics are conducted, employing standard plausible values as the criterion.

Reading

Results are presented for the model sequence for the charter-only analyses (figure 1-3 in chapter 1). Model 1 (with no predictors) provides a baseline for the decomposition of variance and is discussed later. Using model 2 (with student covariates only), the estimated regression coefficients for the student covariates can be compared with those in model *d* for the combined analysis, which includes the same student covariates (see table E-1 in appendix E). The comparison is of interest, since the model *d* estimates are largely determined by data from the public noncharter schools. Indeed, the estimated coefficients for each covariate are strikingly similar, supporting the credibility of the findings of the combined analysis.

The school-level regression in model 3 (with both student and school covariates) includes all the school variables available for the combined analysis. The first column of table 3-1 lists those school characteristics with regression coefficients that are statistically significant. Nonsignificant coefficients are also included if they had an associated *p* value of less than .10 or a value that was larger than one-third of the range of the variable. (Because of the collinearity among some of the variables, as well as the exploratory nature of the process, a rather liberal criterion for retaining variables in the analysis sequence was employed. See table E-2 in appendix E for results for all the variables included in model 3.) These characteristics are then included in the set of predictors employed in the school-level regression in model 6. The last three columns of table 3-1 display the regression coefficients, their standard errors, and their associated *p* values. Note that the differences in magnitude of the coefficients for region in comparison to the others are the result of differences in scaling: The region variables are coded as 0 or 1, while the variables measured in percentages assume values from 0 to 100. The years of teaching experience range from 0 to 56 with an average of 14.

Table 3-1. Regression coefficients for selected charter school characteristics, model 3, grade 4 reading: 2003

School characteristic	Regression coefficient	*p* value
Percentage eligible for free/reduced-price school lunch	– 0.1 (0.06)	.10
Percentage of students with a disability	– 0.9 (0.17)	.00
Average years of teaching experience	0.4 (0.22)	.11
Region[1]		
Midwest	– 14.2 (5.97)	.02
South	– 11.1 (5.89)	.06
West	– 21.3 (5.73)	.00
Percentage of Black students	– 0.1 (0.06)	.03
Reporting 6 percent or more students absent on an average day	– 4.8 (3.39)	.16

[1] The comparison region was Northeast.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Of the 19 school characteristics included in the regression equation, 8 were selected for inclusion in model 6. The negative coefficients for percentage of students with a disability and percentage of Black students indicate that the higher the percentages of those groups in the school population, the lower the average score for the school.[1] On average, charter school means in the Midwest, South, and West regions of the country were lower than school means in the Northeast[2] (see figure A-3 in appendix A for a list of states included in each region).

In model 4, the school-level regression includes a single characteristic: Whether or not the school is affiliated with a public school district (PSD). The estimated coefficient is 4.6, with an estimated standard error of 3.46, and is not significant (see table E-3 in appendix E). At first, this may seem inconsistent with the results from the combined analysis. Recall, however, that in the combined analysis the two types of charter schools were each compared to public noncharter schools and not to each other. Referring to table 2-3 for model *d,* the estimated disadvantage for PSD-affiliated charter schools compared to public noncharter schools was –2.0, while for non-PSD-affiliated charter schools, it was –6.3. Only the latter was significant. The estimated difference in the disadvantage between the two types of charter schools was 4.3, which is very near the estimate of 4.6 obtained in the present analysis.[3]

Models 5.1–5.7 were each fitted with a different block of characteristics specifically related to charter schools. That is, the data for these characteristics were obtained from the charter school questionnaires. Table 3-2 displays, for each block, the variables with regression coefficients that either achieved significance at the .05 level or were large in comparison to the range of the variable (see table E-4 in appendix E for results for all the variables included in models 5.1–5.7). These variables are all then included in the school-level regression in model 6, the results of which determine the variables used in model 7.

Restricting attention to those variables with significant regression coefficients, the results in table 3-2 indicate that charter schools that had a curriculum requirement waiver or that were monitored for student attendance had a higher average achievement than those that did not have those characteristics. On the other hand, charter schools that were monitored for school finances or compliance with state/federal regulations performed lower on average than those that were not. Further, charter schools serving at-risk students or located in states with strong chartering laws also performed lower on average than schools that did not have those characteristics.

Table 3-2. Regression coefficients for selected charter school characteristics in models 5.1–5.7, grade 4 reading: 2003

Charter school characteristic	Regression coefficient	*p* value
Waiver of curriculum requirements (model 5.1)	8.8 (4.38)	.05
Areas monitored (model 5.2)		
Student achievement	– 8.5 (4.44)	.06
Student attendance	10.8 (5.21)	.04
School finances	– 12.0 (4.72)	.01
Compliance with state/federal regulations	– 22.2 (5.90)	.00
Report to chartering agency (model 5.3)	– 5.9 (3.83)	.12
Serve at-risk students (model 5.5)	– 12.1 (6.17)	.05
Focus of program content (model 5.6)		
No specialized area	8.0 (3.72)	.03
Particular education philosophy	9.6 (5.24)	.10
Strong state chartering law (model 5.7)	– 7.7 (3.43)	.03

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[1] The stated interpretation of the regression coefficients always assumes that the values of the other variables in the model are held constant. That is, these are all partial regression coefficients. It is possible that both the magnitude and sign of the coefficient of a given variable will change when another set of variables is included in the model.

[2] Note that the coefficient for the South was not significantly different from zero at the .05 level.

[3] This difference in the estimates can be accounted for by the differences in the fitted student-level regression model, which lead to slightly different adjusted school means.

Schools that had a comprehensive curriculum with no specialized area of focus tended to have higher means than schools with other types of curricula, while schools located in states with strong chartering laws had lower means on average in comparison to schools located in states with weak chartering laws. Interestingly, none of the variables that code for the different entities that grant charters to schools approached significance.[4]

It is problematic to relate these findings directly to the efficacy of a particular policy. For example, it cannot be determined whether the higher achievement in schools for which attendance was monitored, in comparison to schools for which attendance was not monitored, was the result of the monitoring. Similarly, schools for which student achievement was monitored had lower mean achievement than schools that were not monitored. Was that due to the monitoring or, rather, was it that schools that come to the notice of the chartering authority for some reason and were more closely monitored as a result, were also doing more poorly on average? Other interpretations are also possible, and the present data do not allow for distinguishing among them.

In an attempt to work with models that required fewer degrees of freedom, three composite variables at the school level were constructed: the total number of waivers, the total number of areas monitored, and the total number of groups schools are required to report to. As indicated in appendix A, there are a maximum of seven policies from which a school could obtain waivers, seven areas in which a school could be monitored, and eight groups to which a school could report. A model was fitted in which these three composite variables were the sole school covariates. None proved significant.

As indicated above, model 6 incorporated those variables that appeared to be of interest based on the analyses up to that point. All the school characteristics related to the racial/ethnic demographic breakdown of the schools, as well as the charter-type contrast, were also included. Based on the results of fitting model 6 (see table E-5 in appendix E), those variables with regression coefficients that were statistically significant or large in absolute magnitude (i.e., at least one-tenth of the range of the variable) comprised the variables employed in the final model, 7.[5] Table 3-3 displays the results for model 7.

With the exception of years of teaching experience and charter type, all the variables in the school-level regression of model 7 are significant at the .05 level. The magnitudes of the estimated regression coefficients are very nearly the same as they were in the preliminary regressions just described (i.e., models 3, 4, and 5.1–5.7). Note that, in addition to charter type (i.e., PSD or non-PSD), only two of the variables from the charter school questionnaire (monitoring for student achievement and compliance with state/federal regulations) are included in this final model.

Table 3-3. Regression coefficients for selected charter school characteristics, model 7, grade 4 reading: 2003

Charter school characteristic	Regression coefficient	*p* value
Percentage eligible for free/reduced-price school lunch	– 0.1 (0.04)	.03
Percentage of students with a disability	– 0.7 (0.18)	.00
Years of teaching experience	0.3 (0.22)	.13
Region[1]		
Midwest	– 12.9 (4.98)	.01
South	– 10.8 (4.84)	.03
West	– 21.1 (5.08)	.00
Percentage of Black students	– 0.1 (0.05)	.03
Areas monitored		
Student achievement	– 10.7 (4.33)	.02
Compliance with state/federal regulations	– 21.5 (5.06)	.00
Charter type (PSD[2]/non-PSD)	4.2 (2.99)	.17

[1] The comparison region was Northeast.
[2] Public school district.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[4] Such entities could be public school districts, universities, specially constituted authorities, etc. See appendix A.
[5] Although the charter-type contrast was not significant, it was included in model 7.

The variance decomposition results for some of the models considered are found in table 3-4. Model 1 yields the basic decomposition. The total variance is the sum of the two displayed components: 1331 = 861 + 470. Consequently, about 65 percent of the total variance is attributable to within-school heterogeneity, and about 35 percent is attributable to between-school heterogeneity.[6] Recall that in the combined analysis, about 20 percent of the total variance was attributed to between-school heterogeneity.

Introducing the full set of student covariates (model 2) accounts for 13 percent of the within-school variance but 57 percent of the between-school variance. That is, the variance among school means adjusted for student covariates is 43 percent as large as the variance among unadjusted school means. When the general school characteristics are included (model 3), 82 percent of the between-school variance is accounted for. For the exploratory models 4 and 5.1–5.7, the between-school variance is about 150. Turning to the final model, 7, the between-school variance is 74, so the variance accounted for is 84 percent.

A key question in this chapter is whether differences in the characteristics of charter schools can account for differences in the achievement of their students. For this question, model 2 is an appropriate baseline. Comparing model 7 to model 2, the observed reduction in variance is 64 percent [100 x (203 – 74)/203]. That is, school-level characteristics can account for about three-fifths of the heterogeneity in the school means that have been adjusted for differences among students.

Mathematics

As is the case for reading, model 1 provides a baseline for the decomposition of variance and is discussed later. With model 2, the estimated regression coefficients for the student covariates can be compared with those in model *d* for the combined analysis. Again, the estimated coefficients for each covariate are strikingly similar, supporting the credibility of the findings of the combined analysis (see table E-6 in appendix E).

The school-level regression in model 3 includes all the school variables available for the combined analysis. A complete list is provided in table E-7 in appendix E. The first column of table 3-5 lists those school characteristics with regression coefficients that were either statistically significant at the .05 level or large in relation to the range of the variable. These characteristics are then included in the set of predictors employed in the school-level regression in model 6. The last three columns of table 3-5 display the regression coefficients, their estimated standard errors, and their associated *p* values. Note that the differences in magnitude of the coefficients for region in comparison to the others are the result of differences in scaling. The region variables are coded as 0 or 1, while the variables measured in percentages assume values from 0 to 100.

Table 3-4. Variance decompositions for reading scale scores in selected models for charter schools only, grade 4: 2003

Model			Between students, within schools		Between schools	
	Level 1 student covariates	Level 2 school covariates	Variance	Percent of variance in model 1 accounted for	Variance	Percent of variance in model 1 accounted for
1	None	None	861	†	470	†
2	All student covariates	None	753	13	203	57
3	All student covariates	All school covariates from combined analysis	754	12	83	82
7	All student covariates	Final set of covariates	751	13	74	84

† Not applicable.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[6] The total variance is about 5 percent smaller than the total variance in the combined analysis.

Table 3-5. Regression coefficients for selected charter school characteristics, model 3, grade 4 mathematics: 2003

School characteristic	Regression coefficient	*p* value
Percentage of students with a disability	– 0.2 (0.12)	.06
Region[1]		
Midwest	– 6.1 (3.16)	.06
South	– 2.4 (3.91)	.54
West	– 12.7 (3.97)	.00
Percentage of students enrolled the last day of school	– 1.6 (0.86)	.08

[1] The comparison region is Northeast.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Of the 19 characteristics included in the regression equation, 5 were selected for inclusion in model 6. Again, charter school means in the Midwest, South, and West regions appeared lower, on average, than school means in the Northeast. The difference was statistically significant for the West region. In model 4, the school-level regression includes a single characteristic: Whether or not the school was affiliated with a PSD. The estimated coefficient is 4.6, with an estimated standard error of 2.89, and is not significant (see table E-8 in appendix E).

The models 5.1–5.7 were each fitted with a different block of characteristics related to charter schools. Table 3-6 displays, for each block, the variables with regression coefficients that achieved either significance at the .05 level or a value that was larger than one-third of the range of the variable (see table E-9 in appendix E for results for all the variables included in models 5.1–5.7). These variables are all then included in the school-level regression in model 6, the results of which determined the variables used in model 7.

Restricting attention to those characteristics with significant regression coefficients, the results in table 3-6 indicate that charter schools with curriculum requirement waivers had higher average achievement than charter schools without such waivers. Those with a waiver for student assessment had a lower average achievement than those without a waiver. Schools whose charters were granted by a state agency achieved at lower levels than those whose charters were granted by some other agency, and those that had a specialized curriculum achieved at lower levels than those without a specialized curriculum. Of course, the relationships cited here simply summarize patterns of statistical association and should not be interpreted in terms of causal linkages.

As indicated above, model 6 incorporated those variables that appeared to be of interest based on the analyses up to that point. All the school characteristics related to the racial/ethnic demographic breakdown of

Table 3-6. Regression coefficients for selected charter school characteristics in models 5.1–5.7, grade 4 mathematics: 2003

Charter school characteristic	Regression coefficient	*p* value
Waivers (model 5.1)		
Curriculum requirements	6.3 (3.07)	.04
Student attendance requirements	6.8 (4.00)	.09
Student assessment requirements	– 9.7 (3.10)	.00
Areas monitored (model 5.2)		
Student attendance	6.6 (4.19)	.12
School governance	4.8 (3.13)	.13
Charter-granting agency (model 5.4)		
Postsecondary institution	– 4.5 (3.33)	.18
State charter-granting agency	– 11.5 (4.75)	.02
Focus of program content (model 5.6)		
Special curriculum	– 10.2 (3.67)	.01
Particular education philosophy	5.3 (4.06)	.22
Miscellaneous (model 5.7)		
Strong state chartering law	– 5.0 (2.98)	.10
Report to a state entity on progress	– 5.2 (4.03)	.20

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

the schools were included, as well as the charter-type contrast. Those variables with regression coefficients that were statistically significant or large in absolute magnitude (see table E-10 in appendix E) comprised the variables employed in the final model 7. Table 3-7 displays the results for model 7.

First note that the magnitudes of the estimated regression coefficients are very nearly the same as they were in the preliminary regressions just described. Three regression coefficients were statistically significant: Those associated with indicators of waivers from curriculum requirements and student assessment requirements, as well as with the indicator of a charter granted by a state agency. All the variables in the final model are derived from the charter school questionnaire—none of the general school characteristics are present.

Table 3-7. Regression coefficients for selected charter school characteristics, model 7, grade 4 mathematics: 2003

Charter school characteristic	Regression coefficient	*p* value
Waivers		
Curriculum requirements	6.8 (2.64)	.01
Student attendance requirements	7.1 (3.92)	.07
Student assessment requirements	– 8.8 (3.70)	.02
Areas monitored		
Student attendance	6.1 (4.75)	.20
School governance	4.9 (2.58)	.06
Charter granted by state agency	– 10.6 (4.01)	.01
Charter type	3.3 (2.29)	.15

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

The variance decomposition results for some of the models considered are found in table 3-8. Model 1 yields the basic decomposition. The total variance is the sum of the two displayed components: 745 = 497 + 248. Consequently, about 67 percent of the total variance is attributable to within-school heterogeneity, and about 33 percent is attributable to between-school heterogeneity.

Introducing the full set of student covariates (model 2) accounts for 18 percent of the within-school variance and about 55 percent of the between-school variance. That is, the variance among school means adjusted for student covariates is 45 percent as large as the variance among unadjusted school means. When the general school characteristics are included as well (model 3), about 69 percent of the between-school variance is accounted for. For the exploratory models 4 and 5.1–5.7, the between-school variance is about 100 to 115. Turning to the final model, 7, the between-school variance is 85, so the variance accounted for is about 66 percent.

Table 3-8. Variance decompositions for mathematics scale scores in selected models for charter schools only, grade 4: 2003

Model	Level 1 student covariates	Level 2 school covariates	Between students, within schools: Variance	Between students, within schools: Percent of variance in model 1 accounted for	Between schools: Variance	Between schools: Percent of variance in model 1 accounted for
1	None	None	497	†	248	†
2	All student covariates	None	410	18	112	55
3	All student covariates	All school covariates from combined analysis	414	17	77	69
7	All student covariates	Final set of covariates	411	17	85	66

† Not applicable.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

A key question in this chapter is whether differences in the characteristics of charter schools can account for differences in the achievement of their students. For this question, model 2 is an appropriate baseline. Comparing model 7 to model 2, the observed reduction in variance is about 24 percent [100 x (112 – 85)/112]. Thus, school-level characteristics account for less than a quarter of the heterogeneity in the school means that have been adjusted for differences among students.

Summary

The sequence of analyses for reading culminates in a fitted school-level regression (model 7) that includes 10 school covariates, three of which (monitoring student achievement, monitoring compliance with state/federal regulations, and charter type) are derived from the charter school questionnaire. Indicators for region and percentage of the school population who are students with disabilities are significant.

Turning to the decomposition of variance in the reading analyses, the total variance of student outcomes across charter schools is about 5 percent smaller than the total variance across all schools. However, for charter schools, the proportion of the total variance that is accounted for by differences among school means is nearly twice that in the case of all public schools (35 percent versus 20 percent). The between-school variance for adjusted school means, after including student characteristics in the level 1 model, is less than one-half the variance for unadjusted school means. Almost two-fifths $\left(1-\frac{751+74}{861+470}\right)$ of the total variance is accounted for by the final model, 7.

Comparison of the results in tables 2-5 and 3-4 indicates that school-level characteristics of charter schools play a greater role in accounting for differences in student achievement in reading than do school-level characteristics of all public schools. Among the latter, characteristics specific to charter schools account for less variation than do general school characteristics.

The sequence of analyses for mathematics culminates in a fitted school-level regression (model 7) that includes seven school covariates, all of which are derived from the charter school questionnaire. That is, the model does not include any of the general school covariates. Note that the set of school-level variables in model 7 for mathematics is different from the set of school-level variables in model 7 for reading. In the absence of any relevant theory, the models were determined entirely by the analysis of the data and standard variable selection procedures in regression analysis.

Turning to the decomposition of variance, the total variance of student outcomes across charter schools is about 10 percent smaller than the total variance across all schools. For charter schools, the proportion of the total variance that is accounted for by differences among school means is slightly larger than that in the case of all public schools (33 percent versus 25 percent). The variance for adjusted school means, after including student characteristics in the level 1 model, is less than one-half the variance among unadjusted school means. The proportion of the total variance accounted for by the final model 7 is about one-third $\left(1-\frac{411+85}{497+248}\right)$.

Comparison of the results in table 2-10 and table 3-8 indicates that school-level characteristics of charter schools play a greater role in accounting for differences in student achievement in mathematics than do school-level characteristics of all public schools. In contrast to the results for reading, characteristics specific to charter schools account for more variation than do general school characteristics.

Thus, in considering charter schools only, the results for reading and mathematics are not parallel, except for the findings that the charter-type contrast is not significant and that the final models can account for about one-third to two-fifths of the total variance.

References

Bifulco, R., and Ladd, H.F. (2004, August). *The Impacts of Charter Schools on Student Achievement: Evidence From North Carolina.* Working Papers Series No. SAN04-01. Durham, NC: Terry Sanford Institute of Public Policy, Duke University.

Carnoy, M., Jacobsen, R., Mishel, L., and Rothstein, R. (2005). *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement.* Washington, DC: Economic Policy Institute and Teachers College Press.

Center for Education Reform. (2003, January 21). *CER Releases 8th National Charter School Directory*. CER Press Release. Retrieved September 2, 2004, from www.edreform.com/index.cfm?fuseAction=document&documentID=1032.

Center for Education Reform. (2004). *Charter School Laws Across the States: Ranking and Scorecard, 8th Edition*. The Center for Education Reform. Retrieved January 31, 2005, from www.edreform.com/_upload/charter_school_laws.pdf.

Chambers, R.L. (2003). *Which Sample Survey Strategy? A Review of Three Different Approaches*. S^3RI Methodology Working Paper M03/20. Southampton, UK: Southampton Statistical Sciences Research Institute.

The Charter School Achievement Consensus Panel. (2006, May). *Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines.* NCSRP White Paper Series, No. 2. Seattle, WA: Center on Reinventing Public Education, University of Washington.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J.E. (1986). An Uncertainty Principle in Demography and the Unisex Issue. *American Statistician, 40*: 32–39.

Finnigan, K., Adelman, N., Anderson, L., Cotton, L., Donnelly, M.B., and Price, T. (2004). *Evaluation of the Public Charter Schools Program: Final Report.* (Policy and Program Studies Service.) U.S. Department of Education. Jessup, MD: Education Publications Center.

Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (2002). *The Impact of Charter Schools on Academic Achievement*. Retrieved October 15, 2004, from http://edpro.stanford.edu/eah/papers/charters.aea.jan03.pdf.

Hoxby, C.M. (2004, September). *A Straightforward Comparison of Charter Schools and Regular Public Schools in the United States.* Unpublished manuscript, Cambridge, MA: Harvard University.

Little, R. (2003). To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. *The University of Michigan Department of Biostatistics Working Paper Series.* Working Paper 4. Ann Arbor, MI: University of Michigan School of Public Health. Retrieved January 28, 2005, from www.bepress.com/umichbiostat/paper4.

Lubienski, C., and Lubienski, S.T. (2006). *Charter, Private, Public Schools and Academic Achievement: New Evidence From NAEP Mathematics Data.* New York, NY: Columbia University.

Mislevy, R.J., Johnson, E.J., and Muraki, E. (1992). Scaling Procedures in NAEP. *Journal of Educational Statistics*, *17*(2): 131–154.

National Assessment of Educational Progress. (1988). Mathematics Objectives: 1990 Assessment. Princeton, NJ: Author.

National Center for Education Statistics. (2004). *America's Charter Schools: Results From the NAEP 2003 Pilot Study* (NCES 2005-456). U.S. Department of Education, National Center for Education Statistics. Washington, DC: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

Nelson, F.H., Rosenberg, B., and Van Meter, N. (2004, August). *Charter School Achievement on the 2003 National Assessment of Educational Progress.* Washington, DC: American Federation of Teachers, AFL-CIO.

Pfeffermann, D., Moura, F., and Nascimento Silva, P. (2004). *Multi-Level Modelling Under Informative Sampling*. S^3RI Methodology Working Paper M04/09. Southampton, UK: Southampton Statistical Sciences Research Institute.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., and Rasbash, J. (1998). Weighting for Unequal Selection Probabilities in Multilevel Models. *Journal of the Royal Statistical Society*, *60*(1): 23–40.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., and Congdon, R. (2004). *HLM6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Roy, J., and Mishel, L. (2005, April). *Advantage None: Re-Examining Hoxby's Finding of Charter School Benefits.* Briefing Paper 158. Washington, DC: Economic Policy Institute. Retrieved May 23, 2006, from http://www.epinet.org/content.cfm/bp.

Tukey, J.W. (1982). Introduction to Styles of Data Analysis Techniques. In R. Launer and A.F. Siegel (Eds.), *Modern Data Analysis* (pp. 1–11). New York, NY: Academic Press, Inc.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix A

Overview of Procedures

This appendix provides information about the methods and procedures used in NAEP's 2003 pilot study of reading and mathematics achievement among fourth-grade charter school students.

The Assessment Design

The National Assessment Governing Board (NAGB) is responsible for formulating policy for NAEP and is charged with developing assessment objectives and test specifications. These specifications are outlined in subject-area frameworks developed by NAGB with input from a broad spectrum of educators, parents, and members of the general public. An overview of the frameworks and structure of the reading and mathematics assessments are presented in this section. The complete frameworks are available on the NAGB website (http://nagb.org/pubs/pubs.html).

2003 NAEP reading assessment

The reading framework sets forth a broad definition of "reading literacy" that includes developing a general understanding of written text, thinking about it, and using various texts for different purposes. In addition, the framework views reading as an interactive and dynamic process involving the reader, the text, and the context of the reading experience. For example, readers may read stories to enjoy and appreciate the human experience, study science texts to form new hypotheses about knowledge, or follow directions to fill out a form. NAEP reflects current definitions of literacy by differentiating among two contexts for reading and four aspects of reading at grade 4. The contexts for reading and aspects of reading make up the foundation of the NAEP reading assessment.

The "contexts for reading" dimension of the NAEP reading framework provides guidance for the types of texts to be included in the assessment. Although many commonalities exist among the different types of reading contexts, different contexts do lead to real differences in what readers do. For example, when *reading for literary experience*, readers make plot summaries and abstract major themes. They describe the interactions of various literary elements (e.g., setting, plot, characters, and theme). When *reading for information*, readers critically judge the organization and content of the text and explain their judgments. They also look for specific pieces of information. A third context defined in the framework, *reading to perform a task*, is not assessed at grade 4.

The "aspects of reading" dimension of the NAEP reading framework provides guidance for the types of comprehension questions to be included in the assessment. The four aspects are 1) *forming a general understanding*, 2) *developing interpretation*, 3) *making reader/text connections*, and 4) *examining content and structure*. These four aspects represent different ways in which readers develop understanding of a text. In *forming a general understanding*, readers must consider the text as a whole and provide a global understanding of it. As readers engage in *developing interpretation*, they must extend initial impressions in order to develop a more complete understanding of what was read. This involves linking information across parts of a text or focusing on specific information. When *making reader/text connections*, the reader must connect information in the text with knowledge and experience. This might include applying ideas in the text to the real world. Finally, *examining content and structure* requires critically evaluating, comparing and contrasting, and understanding the effect of different text features and authorial devices.

Figure A-1 shows the relationship between these reading contexts and aspects of reading in the NAEP fourth-grade reading assessment. Included in the figure are sample questions that illustrate how each aspect of reading is assessed within each reading context.

The assessment framework specifies not only the particular dimensions of reading literacy to be measured, but also the percentage of assessment questions that should be devoted to each. The target percentage distribution for contexts of reading and aspects of reading as specified in the framework, along with the actual percentage distribution in the fourth-grade assessment, are presented in tables A-1 and A-2.

2003 NAEP mathematics assessment

The mathematics framework used for the 2003 assessment had its origins in a framework developed for the 1990 mathematics assessment under contract with the Council of Chief State School Officers (CCSSO). The CCSSO project considered objectives and frameworks for mathematics instruction at the state, district, and school levels. The project also examined curricular frameworks on which previous NAEP assessments were based, consulted with leaders in mathematics education, and considered a draft version of the National Council of Teachers of Mathematics (NCTM) *Curriculum and Evaluation Standards for School*

Figure A-1. Sample NAEP questions, by aspects of reading and contexts for reading specified in the reading framework for grade 4: 2003

	Aspect of reading			
Context for reading	Forming a general understanding	Developing interpretation	Making reader/text connections	Examining content and structure
Reading for literary experience	What is the story/plot about?	How did this character change from the beginning to the end of the story?	What other character that you have read about had a similar problem?	What is the mood of this story and how does the author use language to achieve it?
Reading for information	What point is the author making about this topic?	What caused this change?	What other event in history or recent news is similar to this one?	Is this author biased? Support your answer with information about this article.

SOURCE: National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

Table A-1. Target and actual percentage distribution of questions, by context for reading, grade 4: 2003

Context for reading	Target	Actual
Reading for literary experience	55	50
Reading for information	45	50

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Table A-2. Target and actual percentage distribution of student time, by aspect of reading, grade 4: 2003

Aspect of reading	Target	Actual
Forming a general understanding/developing interpretation	60	61
Making reader/text connections	15	17
Examining content and structure	25	22

NOTE: Actual percentages are based on the classifications agreed upon by NAEP's Instrument Development Panel. It is recognized that making discrete classifications for these categories is difficult and that independent efforts to classify NAEP questions have led to different results.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Mathematics (1989). This project resulted in a "content-by-ability" matrix design (NAEP 1988) that was later updated for the 1996 assessment to allow questions to be classified in more than one content area and to include categories for mathematics ability and process goals. Figure A-2 describes the five content areas that constitute the NAEP mathematics assessment. The questions designed to test the various content areas at grade 4 generally reflect the expectations normally associated with instruction at that level.

The assessment framework specifies not only the particular areas that should be assessed, but also the percentage of the assessment questions that should be devoted to each of the content areas. The target percentage distribution for content areas as specified in the framework for grade 4 is presented in table A-3. The distribution of items among the content areas is a critical feature of the assessment design, since it reflects the relative importance and value given to each.

Table A-3. Target percentage distribution of items, by mathematics content area, grade 4: 2003

Content area	Percentage of items
Number sense, properties, and operations	40
Measurement	20
Geometry and spatial sense	15
Data analysis, statistics, and probability	10
Algebra and functions	15

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment.

Figure A-2. Descriptions of the five NAEP mathematics content areas

Number sense, properties, and operations	This content area focuses on students' understanding of numbers (whole numbers, fractions, decimals, integers, real numbers, and complex numbers), operations, and estimation, and their application to real-world situations. At grade 4, the emphasis is on the development of number sense through connecting various models to their numerical representations, as well as an understanding of the meaning of addition, subtraction, multiplication, and division.
Measurement	This content area focuses on an understanding of the process of measurement and the use of numbers and measures to describe and compare mathematical and real-world objects. Students are asked to identify attributes, select appropriate units and tools, apply measurement concepts, and communicate measurement-related ideas. At grade 4, the focus is on time, money, temperature, length, perimeter, area, capacity, weight/mass, and angle measure.
Geometry and spatial sense	This content area is designed to extend beyond low-level identification of geometric shapes to include transformations and combinations of those shapes. Informal constructions and demonstrations (including drawing representations) along with their justifications take precedence over more traditional types of compass-and-straightedge constructions and proofs. At grade 4, students are asked to model properties of shapes under simple combinations and transformations, and to use mathematical communication skills to draw figures from verbal descriptions.
Data analysis, statistics, and probability	This content area emphasizes the appropriate methods for gathering data, the visual exploration of data, various ways of representing data, and the development and evaluation of arguments based on data analysis. At grade 4, students are asked to apply their understanding of numbers and quantities by solving problems that involve data. Fourth graders are asked to interact with a variety of graphs, to make predictions from data and explain their reasoning, to deal informally with measures of central tendency, and to use the basic concepts of chance in meaningful contexts.
Algebra and functions	This content area extends from work with simple patterns at grade 4 to basic algebra concepts at grade 8. The grade 4 assessment involves informal demonstration of students' abilities to generalize from patterns, including the justification of their generalizations. Students are expected to translate between mathematical representations, to use simple equations, and to do basic graphing.

SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

Common design features of the reading and mathematics assessments

Each student who participated in the 2003 NAEP assessment received a booklet containing four sections: a set of general background questions, a set of subject-specific background questions, and two sets of cognitive questions in either reading or mathematics (there were no booklets that contained both reading and mathematics questions). The sets of cognitive questions are referred to as "blocks." The 2003 grade 4 reading and mathematics assessments each consisted of 10 blocks of cognitive questions. Each block contained a combination of multiple-choice, short constructed-response, and extended constructed-response questions.

The design of the NAEP reading and mathematics assessments allows maximum coverage of a range of content while minimizing the time burden for any one student participating in the assessment. This was accomplished through the use of matrix sampling, in which representative samples of students took various portions of the entire pool of assessment questions. Individual students were required to take only a small portion of the total pool of assessment questions, but the aggregate results across the entire assessment allow for broad reporting of reading and mathematics abilities for the targeted population.

In addition to matrix sampling of questions, the NAEP assessment designs utilized a procedure for distributing blocks across booklets that controlled for position and context effects. Students received different blocks of questions in their booklets according to a procedure that assigned blocks of questions in a way that balanced the positioning of blocks across booklets (i.e., a given block did not appear in the same position in every booklet), balanced the pairing of blocks within booklets (i.e., pairs of blocks occurred the same number of times), and ensured that every block of questions was paired with every other block. The procedure also cycles the booklets for administration so that, typically, only a few students in any assessment session receive the same booklet.

Teacher, school, and students with disabilities/ limited-English-proficient student questionnaires[1]

In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a teacher questionnaire, a school questionnaire, and a questionnaire for students with disabilities (SD) and limited-English-proficient (LEP) students. The teacher questionnaire was administered to the reading or mathematics teachers of students participating in the corresponding assessment. The questionnaire took approximately 20 minutes to complete and focused on the teacher's general background and experience, the teacher's background related to reading or mathematics, and information about classroom instruction.

The school questionnaire was given to the principal or other administrator in each participating school and took about 20 minutes to complete. The questions asked about school policies, programs, facilities, and the demographic composition and background of the students and teachers at the school.

The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students selected to participate in the assessment who were identified as having an Individualized Education Program (IEP) or equivalent plan (for reasons other than being gifted or talented) or being limited English proficient. An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire took approximately three minutes to complete and asked about the student and the special-education programs in which he or she participated.

[1] In 2003, NAEP questionnaires referred to English language learners (ELL) as limited-English-proficient (LEP). Elsewhere in this report, the current description (ELL) is used.

Charter school survey

In addition to the standard NAEP questionnaires, the charter schools identified for the pilot study were administered a separate charter school survey. The purpose of collecting additional information about the charter schools was to provide a context for interpreting the charter school students' performance on NAEP. The survey asked questions regarding the establishment of the charter school, its management, student population served, curriculum focus, governance, and autonomy. The questions from the charter school study are available on the NAEP website (http://nces.ed.gov/nationsreportcard/studies/charter).

Staff from Westat, Inc., a NAEP contractor, administered the survey via telephone by calling each school in the charter school sample and speaking with the school principal or school coordinator. Survey data were collected for a total of 150 schools, 138 initially identified through the school questionnaire, plus 12 additional schools verified by state coordinators.

Sample Design

The results presented in this report are based on nationally representative probability samples of fourth-grade public school students. The samples were drawn as part of the 2003 NAEP state assessments in reading and mathematics. Note that the 2003 reading and mathematics assessments were administered to samples of both fourth- and eighth-grade students, but only the fourth-grade sample was targeted for the pilot study of charter schools. An oversampling of fourth-grade charter schools was incorporated into the sample design in order to supplement the information about charter schools identified in the state samples. Table A-4 contains the target populations and actual sample sizes for the 2003 grade 4 reading and mathematics assessments.

The sampling frame consisted of public schools having the relevant grade in each state, according to information from the 2000–01 NCES Common Core of Data (CCD)[2] Public Elementary and Secondary School Universe file. The samples were selected based on a two-stage sample design. In the first stage, schools were selected from stratified frames within participating states. Charter school status was one of three stratification variables used in the 2003 NAEP state sample design (the others were type of location classification and classification by percentage of Black and Hispanic students enrolled). In the second stage, students were selected from within schools.

Table A-4. Student sample size and target population, by type of public school and subject assessed, grade 4: 2003

Subject	Sample size: Charter schools	Sample size: Public noncharter schools	Target population: Charter schools	Target population: Public noncharter schools
Reading	3,296	188,148	46,476	3,562,077
Mathematics	3,238	188,201	46,187	3,557,115

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Stratification of schools according to charter school status was determined with information from the 2000–01 NCES CCD. Because the status of schools as charter or regular public schools can change, NAEP state coordinators were asked to update the charter school status of the selected sample during the process of recruiting schools to participate in the 2003 assessments. The state coordinators were asked to confirm which of the schools designated by CCD as being chartered were indeed charter schools, and to identify any other schools in the NAEP sample that might have become charter schools since the CCD list of schools had last been updated. This yielded a sample of 142 schools. In addition to the charter schools identified using this procedure, eight schools were subsequently identified as charter schools using a question on the NAEP school questionnaire.

[2] The Common Core of Data (CCD) is a program of NCES that annually compiles information about the nation's public schools and school districts, and makes this information available through a public database. For more information, see http://nces.ed.gov/ccd/.

In order to obtain a representative sample of charter school students, several states were targeted for an oversampling of charter schools. The 2000–01 NCES CCD indicated that 12 states in the country had more than 2 percent of their student population attending charter schools that included grade 4. Charter schools were oversampled in three states that together contained almost half (49 percent) of all charter school students. These states were California (26 percent of all charter schools), Michigan (16 percent), and Texas (6 percent). The sampling rates for charter schools in these three states were adjusted to meet the increased student sample size targets.

Participation of schools and students in the NAEP samples

Table A-5 provides a summary of the school and student participation rates for the 2003 grade 4 reading and mathematics assessment samples. Participation rates are presented for charter schools and public noncharter schools. Four different rates are presented.

The first rate is a student-centered, weighted percentage of schools participating in the assessment. This rate is based only on the schools that were selected for the assessment. The numerator of this rate is the estimated number of students represented by the selected schools that participated in the assessment. The denominator is the estimated number of students represented by the selected schools that had eligible students enrolled.

The second school participation rate is a school-centered, weighted percentage of schools participating in the assessment. This rate is based only on the schools that were selected for the assessment. The numerator of this rate is the estimated number of schools represented by the selected schools that participated in the assessment. The denominator is the estimated number of schools represented by the selected schools that had eligible students enrolled.

The student-centered and school-centered school participation rates differ if school participation is associated with the size of the school. If the student-centered rate is higher than the school-centered rate, this indicates that larger schools participated at a higher rate than smaller schools. If the student-centered rate is lower, smaller schools participated at a higher rate than larger schools.

Also presented in table A-5 are weighted student participation rates. Some students sampled for NAEP are not assessed because they cannot meaningfully participate. The numerator of this rate is the estimated number of students who are represented by the students assessed (in either an initial session or a makeup session). The denominator of this rate is the estimated number of students represented by the eligible sampled students in participating schools.

Table A-5. School and student participation rates, by type of public school and subject assessed, grade 4: 2003

	School participation			Student participation	
Type of public school	Student-centered weighted percentage	School-centered weighted percentage	Number of schools participating	Student-weighted percentage	Number of students assessed
Reading					
Charter schools	100	100	150	92	3,115
Public noncharter schools	100	100	6,764	94	175,898
Mathematics					
Charter schools	100	100	150	92	3,154
Public noncharter schools	100	100	6,764	94	181,171

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Participation of students with disabilities and/or English language learners in the NAEP samples

Testing all sampled students is the best way for NAEP to ensure that the statistics generated by the assessment are as representative as possible of the performance of the populations of participating jurisdictions. Therefore, every effort is made to ensure that all selected students who are capable of participating in the assessment are assessed. However, all groups of students include certain proportions that cannot be tested in large-scale assessments (such as students who have profound mental disabilities) or who can only be tested through the use of testing accommodations such as extra time, one-on-one administration, or use of magnifying equipment. Some students with disabilities (SD) and some English language learners (ELL) cannot show on a test what they know and can do unless they are provided with accommodations.

In 2003, NAEP inclusion rules were applied, and accommodations were offered when a student had an Individualized Education Program (IEP) because of a disability, was protected under Section 504 of the Rehabilitation Act of 1973[3] because of a disability, and/or was identified as being an ELL student; all other students were asked to participate in the assessment under standard conditions.

The number and percentages of SD and/or ELL students in the 2003 charter and public noncharter school samples are presented in table A-6. The data in this table include the number and percentage of students *identified* as SD and/or ELL, the number and percentage of students *excluded*, the number and percentage of SD and/or ELL students *assessed*, the number and percentage *assessed without accommodations*, and the number and percentage *assessed with accommodations*.

Table A-7 displays the percentages of SD/ELL students assessed with the variety of available accommodations. It should be noted that students assessed with accommodations typically received some combination of accommodations. The percentages presented in the table reflect only the primary accommodation provided. For example, students assessed in small groups (as compared with standard NAEP sessions of about 30 students) usually received extended time. In one-on-one administrations, students often received assistance in recording answers (e.g., use of a scribe or computer) and were afforded extra time. Extended time was considered the primary accommodation only when it was the sole accommodation provided.

Data Collection and Scoring

The 2003 NAEP reading and mathematics assessments were conducted from January to March 2003 by contractors to the U.S. Department of Education. Trained field staff from Westat conducted the data collection. Materials from the 2003 assessment were shipped to Pearson Educational Measurement, where trained staff evaluated the responses to the constructed-response questions using scoring guides prepared by Educational Testing Service. Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four- and five-level guides, and many of the short constructed-response questions were rated according to three-level guides that permitted partial credit. Other short constructed-response questions were scored as either correct or incorrect.

[3] Section 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 794 (2002), is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal assistance.

Table A-6. Students with disabilities and/or English language learners identified, excluded, and assessed, by type of public school and subject assessed, grade 4: 2003

Students' status	Charter schools		Public noncharter schools	
	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled
Reading				
SD and/or ELL				
Identified	652	17	39,282	22
Excluded	181	4	12,250	6
Assessed	471	13	27,032	16
Without accommodations	340	9	16,118	10
With accommodations	131	4	10,914	5
SD				
Identified	329	10	26,961	14
Excluded	125	3	9,342	5
Assessed	204	7	17,619	9
Without accommodations	120	3	8,075	4
With accommodations	84	3	9,544	5
ELL				
Identified	394	9	15,877	10
Excluded	84	2	4,394	3
Assessed	310	7	11,483	8
Without accommodations	250	6	9,227	7
With accommodations	60	2	2,256	1
Mathematics				
SD and/or ELL				
Identified	662	18	39,391	22
Excluded	84	2	7,030	4
Assessed	578	16	32,361	18
Without accommodations	346	8	15,881	10
With accommodations	232	8	16,480	8
SD				
Identified	327	10	26,992	14
Excluded	75	2	5,530	3
Assessed	252	8	21,462	11
Without accommodations	119	3	7,810	4
With accommodations	133	6	13,652	7
ELL				
Identified	403	9	15,845	11
Excluded	29	1	2,438	2
Assessed	374	8	13,407	9
Without accommodations	257	6	9,222	7
With accommodations	117	3	4,185	2

NOTE: SD = Students with disabilities. ELL = English language learners. Detail may not sum to totals because of rounding. Within each grade level the combined SD/ELL portion of the table is not a sum of the separate SD and ELL portions because some students were identified as both SD and ELL. Such students would be counted separately in the bottom portions but counted only once in the top portion.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Table A-7. Students with disabilities and/or English language learners assessed with accommodations in reading and mathematics, by type of primary accommodation and type of public school, grade 4: 2003

	Weighted percentage of assessed students	
Type of accommodation	Charter schools	Public noncharter schools
Reading		
Large-print book	0.04	0.05
Extended time	1.64	1.27
Small group	2.62	4.09
One-on-one	#	0.16
Scribe/computer	0.07	0.13
Other	0.04	0.08
Mathematics		
Bilingual book	1.43	0.84
Large-print book	#	0.06
Extended time	0.64	0.95
Read aloud	1.00	0.58
Small group	4.41	5.61
One-on-one	0.46	0.33
Scribe/computer	#	0.19
Other	0.01	0.08

Rounds to zero.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading and Mathematics Charter School Pilot Study.

Approximately 3.9 million constructed responses were scored for the 2003 reading assessment. The within-year average percentage of agreement for the 2003 national reliability sample for reading was 90 percent at grade 4. Approximately 4.7 million constructed responses were scored for the 2003 mathematics assessment. The within-year average percentage of agreement for the 2003 national reliability sample for mathematics was 95 percent at grade 4.

Weighting and Variance Estimation

As described in previous sections of this appendix, a multistage, stratified, clustered sampling design was used to select the students to be assessed. The properties of a sample obtained through such a complex design are very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which observations from different students can be considered statistically independent of one another.

Typically, sampling weights are used in the estimation process to account for the fact that the probabilities of selection were not identical for all students. For the HLM analysis, weights were handled in a special way, following a suggestion of Pfeffermann et al. (1998). For further details, consult appendix B.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources: inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate and, because of their nature, the impact of such errors cannot be reflected in the databased estimates of uncertainty provided in NAEP reports.

Drawing Inferences From the Results

Regression estimates in the HLM analyses have uncertainty associated with them due to measurement and sampling error. The uncertainty of an estimate is reported as a standard error. In both the main text and in appendix B, the results of the HLM analyses are reported as regression estimates together with the corresponding *p* values obtained from simple *t* tests of significance. When the regression coefficient is associated with an indicator of a school's membership in one of two groups, the estimate of the coefficient represents the difference in average scale scores between the groups. The *p* value associated with the estimate is the probability that a difference of this magnitude would occur if the null hypothesis of no difference between groups were true.

Simpson's paradox

The consequences of the confounding between school-type differences and heterogeneity among states in mean achievement can be illustrated by a simple example, framed in terms of populations (rather than samples). Suppose there are two states, denoted A and B, with charter schools in A having an average score 15 points higher than charter schools in B, and public noncharter schools in A also having an average score 15 points higher than public noncharter schools in B. Further suppose that in each state charter school means are 5 points higher than public noncharter school means. Finally, it is assumed that the number of charter schools is small relative to the number of public noncharter schools. The situation is represented in table A-8.

Clearly, there are proportionately more charter schools in B, the lower-performing state, while there are proportionately more public noncharter schools in A, the higher-performing state. Direct computation shows that the population mean for charter schools is 190, which is also the population mean for public noncharter schools. Thus, one would conclude that there is no difference between the two types of schools, despite the 5-point advantage of charter schools in each state. However, if school means are adjusted for the difference in state means, the 5-point advantage of charter schools is essentially recovered.

Table A-8. Illustrative example of Simpson's Paradox

	Charter schools		Public noncharter schools	
State	Average scale score	Number of schools	Average scale score	Number of schools
A	200	10	195	300
B	185	20	180	150

Alternative effect size calculation

In the context of fitting HLMs to data, there is an alternative effect size calculation. Since the parameter of interest refers to the difference in average school means between the two types of schools, it is reasonable to compute the effect size as the ratio of the magnitude of the statistic to the standard deviation of the distribution of school means. The latter can be obtained from the variance decomposition provided by an unstructured HLM (model *a*). For reading, the standard deviation among school means is $\sqrt{302} = 17.4$ (see table 2-5). The corresponding effect size is 4.2/17.4 = 0.24, a value somewhat larger than 0.11 presented in the main text. For mathematics, this quantity is $\sqrt{210} = 14.5$. The corresponding effect size is 4.7/14.5 = 0.32. Expert opinion is divided on which form of the effect size is more appropriate (S. Raudenbush, personal communication, April 27, 2005). Accordingly, the more conservative calculation, which uses the standard deviation of student test score distribution, was employed in the main text.

Variable Descriptions

NAEP reports average scores and percentages of students for groups of students defined by data from the student, teacher, and school administrator questionnaires. In addition to the standard NAEP questionnaires, information was collected from a survey specifically designed to address issues relevant to charter schools. Descriptions of the variables used in the charter school HLM analyses are presented in the following sections.

Student-level variables

Eight student-level variables were used in the HLM analysis: gender, race/ethnicity, whether a student had an Individualized Education Plan (IEP) or was an English language learner, whether there was a computer in the home, eligibility for free/reduced-price school lunch, number of books in the home, and number of absences.

Gender: Results are available for male and female students as reported by the school.

Race/ethnicity: Based on information obtained from school records, students who participated in the 2003 NAEP assessments were identified as belonging to one of six mutually exclusive racial/ethnic subgroups: White, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, or unclassifiable. When school-reported information was missing, student-reported data were used to determine race/ethnicity. Students whose race based on school records was unclassifiable or, if school data were missing, who self-reported their race as "multicultural" but not Hispanic, or who did not self-report racial/ethnic information, were categorized as unclassifiable.

Students with disabilities: Students who had an IEP or were protected under Section 504 of the Rehabilitation Act of 1973 were included in the NAEP assessment, except in the following cases:

> The school's IEP team determined that the student could not participate.
>
> The student's cognitive functioning was so severely impaired that he or she could not participate.
>
> The student's IEP required that the student had to be tested with an accommodation or adaptation that NAEP does not allow, and the student could not demonstrate his or her knowledge without that accommodation.

English language learners (ELL): All students who received academic instruction in English for three years or more were included in the assessment. Those ELL students who received instruction in English for less than three years were included unless school staff judged them to be incapable of participating in the assessment in English.

Computer in the home: Fourth-grade students were asked if there was a computer at home that they could use. Students could respond either "yes" or "no" to the question.

Eligibility for free/reduced-price school lunch: NAEP collects data on students' eligibility for free or reduced-price school lunch as an indicator of family economic status. As part of the Department of Agriculture's National School Lunch Program, schools can receive cash subsidies and donated commodities in return for offering free or reduced-price lunches to eligible children. Based on available school records, students were classified as either currently eligible for free/reduced-price school lunch or not eligible. Eligibility for the program is determined by a student's family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level, and reduced-price lunch qualification is set at between 130 and 185 percent of the poverty level. The classification applies only to the school year when the assessment was administered (i.e., the 2002–2003 school year) and is not based on eligibility in previous years. If school records were not available, or if the school did not participate in the program, the student was classified as not eligible.

Number of books in the home: Fourth-graders who participated in the assessment were asked about how many books there were in their homes. Response options included "a few (0–10)," "enough to fill one shelf (11–25)," "enough to fill one bookcase (26–100)," or "enough to fill several bookcases (more than 100)." For the purpose of this analysis, the first two response categories were combined, along with any missing responses, and the last two categories were combined.

Number of absences: As part of the student questionnaire, students were asked how many days they had been absent from school in the last month. Response options included "none," "1 or 2 days," "3 or 4 days," "5 to 10 days," or "more than 10 days." Students who indicated "none" made up one category in the analysis, and those who indicated "1 or more days" were combined with students who had missing responses.

School-level variables

Most of the school-level variables used in the HLM analyses were based on teachers' and school administrators' responses to selected questions from the standard NAEP questionnaires. A few variables were created using information collected about sampled students as part of the administration process.

Years of teaching experience: Teachers whose students participated in the fourth-grade NAEP assessment were asked to indicate the number of years they had worked as an elementary or secondary teacher (including full-time teaching assignments, part-time teaching assignments, and long-term substitute assignments, but not student teaching). The variable was the aggregated value for all students matched with the teacher questionnaire. If the number of years reported was 60 or more, it was set to "missing." If the value was missing for the entire school, the mean for the school type (public noncharter or charter) was substituted.

Teacher certification: Teachers of participating students were asked to indicate the type of teaching certificate they held (choosing from five possible options) or if they held no certificate. Results for students whose teachers indicated having a regular or provisional certificate were categorized as having a "certified" teacher. Students whose teachers indicated having a probationary, temporary, or emergency certificate (or if the response was missing) were categorized as having a teacher who was not certified. The variable was the aggregated value for a school of all students matched with a teacher questionnaire. The categories for the analysis were "all teachers in the school were certified," "some teachers in the school were certified," and "no teachers in the school were certified."

Student absenteeism: School-level information related to student absenteeism was obtained in several different ways. In the first of two variables from the school questionnaire, administrators were asked to indicate the degree to which student absenteeism was a problem in their school. Three categories were used in the analysis: one indicating it was not a problem, one indicating it was a minor problem, and one indicating it was either a moderate or serious problem. Missing values were coded as part of the third category.

In the second variable, administrators were asked to indicate the percentage of students absent on an average day. Response options included "0–2%," "3–5%," "6–10%," and "more than 10%." In the case of missing responses, the results were combined together with the "0–2%" category. The "6–10%" and "more than 10%" categories were also combined for the analysis.

A third variable was created to reflect the percentage of students absent on the day of the assessment. The number of students who were reported absent on the administration schedule was divided by the total number of sampled students in the school (i.e., the number of students assessed plus the number of students absent).

Percentage of students excluded: The percentage of students excluded from the assessment was calculated by dividing the number of sampled students who were excluded by the total number of sampled students in the school (i.e., the number of students assessed plus the number of students absent or excluded).

Percentage of students in racial/ethnic subgroups: The percentage of students by racial/ethnic categories was based on information provided by the schools and maintained by Westat, the contractor responsible for NAEP data collection.

Student mobility: Student mobility was measured based on school administrators' responses to a question that asked about the percentage of students who were enrolled at the beginning of the school year and who were still enrolled at the end of the school year. Response categories included "98–100%," "95–97%," "90–94%," "80–89%," "70–79%," "60–69%," "50–59%," and "less than 50%." Responses indicating "less than 50%," "50–59%," and "60–69%" were combined for the analysis. Missing values were imputed to the median value for the "80–89%" category.

Type of location: Results from the 2003 assessment are reported for students attending schools in three mutually exclusive location types: central city, urban fringe/large town, and rural/small town.

Following standard definitions established by the Federal Office of Management and Budget, the U.S. Census Bureau (see http://www.census.gov/) defines "central city" as the largest city of a metropolitan statistical area (MSA) or a consolidated metropolitan statistical area (CMSA). Typically, an MSA contains a city with a population of at least 50,000 and includes its adjacent areas. An MSA becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has a population of 1,000,000 or more, its component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation. In the NCES Common Core of Data (CCD), locale codes are assigned to schools. For the definition of "central city" used in this report, two locale codes of the survey are combined. The definition of each school's type of location is determined by the size of the place where the school is located and whether or not it is in an MSA or CMSA. School locale codes are assigned by the U.S. Census Bureau. For the definition of central city, NAEP reporting uses data from two CCD locale codes: large city (a central city of an MSA or CMSA with the city having a population greater than or equal to 25,000) and midsize city (a central city of an MSA or CMSA having a population less than 25,000). Central city is a geographical term and is not synonymous with "inner city."

The "urban fringe" category includes any incorporated place, census-designated place, or nonplace territory within a CMSA or MSA of a large or midsized city and defined as urban by the U.S. Census Bureau, but which does not qualify as a central city. A large town is defined as a place outside a CMSA or MSA with a population greater than or equal to 25,000.

"Rural" includes all places and areas with populations of less than 2,500 that are classified as rural by the U.S. Census Bureau. A small town is defined as a place outside a CMSA or MSA with a population of less than 25,000, but greater than or equal to 2,500.

Region of the country: As of 2003, to align NAEP with other federal data collections, NAEP analysis and reports have used the U.S. Census Bureau's definition of "region." The four regions defined by the U.S. Census Bureau are Northeast, South, Midwest, and West. Figure A-3 shows how states are subdivided into these census regions. All 50 states and the District of Columbia are listed.

Figure A-3. States within regions of the country defined by the U.S. Census Bureau

Northeast	South	Midwest	West
Connecticut	Alabama	Illinois	Alaska
Maine	Arkansas	Indiana	Arizona
Massachusetts	Delaware	Iowa	California
New Hampshire	District of Columbia	Kansas	Colorado
New Jersey	Florida	Michigan	Hawaii
New York	Georgia	Minnesota	Idaho
Pennsylvania	Kentucky	Missouri	Montana
Rhode Island	Louisiana	Nebraska	Nevada
Vermont	Maryland	North Dakota	New Mexico
	Mississippi	Ohio	Oregon
	North Carolina	South Dakota	Utah
	Oklahoma	Wisconsin	Washington
	South Carolina		Wyoming
	Tennessee		
	Texas		
	Virginia		
	West Virginia		

SOURCE: U.S. Census Bureau. Retrieved January 20, 2005, from http://www.census.gov/geo/www/us_regdiv.pdf.

Percentage of students eligible for free/reduced-price school lunch: The percentage of students eligible for free/reduced-price school lunch was based on aggregated data from among the students assessed.

Percentage of students with an IEP: The percentage of students with an IEP was based on aggregated data from among the students assessed.

Percentage of students identified as ELL: The percentage of students identified as ELL was based on aggregated data from among the students assessed.

Charter school variables

The complete survey administered to administrative staff in the sampled charter schools is available on the NAEP website (http://nces.ed.gov/nationsreportcard/pdf/studies/CharterSchoolSurvey.pdf). The questions from the survey that were used in the HLM analysis are described below.

Waivers: School personnel responded “yes,” “no,” or “don’t know” for each category listed as part of the following question: Does your school’s charter include waivers or exemptions from the following state or district policies?

Teacher certification requirements

Teacher/staff hiring/firing policies

Curriculum requirements

Student attendance/seat time requirements

Student assessment requirements

Control of finances/budget

Incentives, rewards, or sanctions due to school performance

Monitoring: School personnel responded “yes,” “no,” or “don’t know” for each category listed as part of the following question: In which of the following areas is your school monitored by the state or your school’s charter-granting agency?

Instructional practices

Student achievement

Student behavior

Student attendance

School governance

School finances

Compliance with state or federal regulations

Progress reporting: School personnel responded “yes,” “no,” or “don’t know” when asked to which of the following groups they were required to make a report on their school’s progress:

Chartering agency

Private funders

Parents

Community/general public

School governing board

State board of education

State department of education (if this is not the chartering agency)

Legislature

A separate variable (reporting to a state entity) was created based on administrators’ responses to just the categories for reporting to the state board or department of education.

Charter-granting agency: Administrators were asked to indicate who granted their school’s charter. Response options included “school district,” “state board of education,” “postsecondary institution,” “state charter-granting agency,” “other,” or “don’t know.”

Student population served: Administrators indicated the primary type of student population served by the charter school from among the following options:

All students

At-risk students

Students with disabilities

Gifted/talented students

Some other population

Don't know

A separate variable was created based on whether or not administrators indicated serving at-risk students as part of their response.

New or pre-existing school: Administrators were asked to indicate if their charter school was a newly created school or a pre-existing school.

Program content: School personnel were asked to indicate which of the following statements best described their charter school's primary focus in terms of program content:

We have a comprehensive curriculum with no specialized area of focus.

We have a special curricular focus, for example, the arts, math/science, foreign language immersion.

Our curriculum is based on a particular educational philosophy, for example, Montessori, open school.

Our curriculum is based on a particular philosophy or set of values, for example, Eastern philosophy, religion.

Other

Don't know

School management: Charter school administrators indicated whether or not their school was operated by an organization or company that also managed other schools.

Part of another school district: The last question from the charter school survey used in the analysis asked if the school was part of another public school district or local education agency, or if it was a charter school district by itself.

Strong laws: Schools were categorized according to whether or not they were in states with strong charter school laws based on information provided in a special report from the Center for Education Reform (2004). The report lists the following 10 criteria for a strong charter school law:

number of schools (unlimited or substantial number of autonomous charter schools);

multiple chartering authorities/binding appeals process;

variety of applicants;

new starts;

formal evidence of local support;

automatic waiver from laws and regulations;

legal/operational autonomy;

guaranteed full funding;

fiscal autonomy; and

exemption from collective bargaining agreements/district work rules.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix B

Considerations in the Use of Sampling Weights in Multilevel Models

The approach to estimation in this study differs from that taken in standard NAEP reports. The latter is referred to as "design-based" (Chambers 2003) since it does not employ parametrized stochastic models to motivate the estimation procedure. Rather, it computes weighted averages of students' plausible values, where the weights are derived from the survey design, to estimate the target quantities (estimands). Such estimates are approximately unbiased with respect to the distribution generated by repeated sampling. Estimates of the variance of these estimates are obtained by applying a specific jackknife procedure, which is structured to account for some aspects of the survey design, rather than a particular model for the data.

On the other hand, an analysis based on a hierarchical linear model (HLM) is "model-based" (Chambers 2003) since the estimates are obtained by solving a set of likelihood equations derived from the postulated model. Accordingly, estimates obtained through this procedure will be influenced by the form of the model, as well as by the degree of congruence between the model and the data. Such estimates may or may not be unbiased in finite samples, but can be more efficient than the design-based estimates if the model is approximately correct.

It is important to keep in mind that the model parameter estimated in a student-level analysis is not generally the same as the model parameter estimated in a multilevel analysis. In the combined analysis, for example, the estimand in the student-level analysis represents the average difference between students attending the two types of schools, while the estimand in the basic school-level analysis represents the average difference in school means between the two types of schools. In simple situations with equal numbers of students per school and random sampling at both levels, the two parameters coincide. In unbalanced situations with a complex sampling design, they can and do differ, although the differences should generally not be large.

For this analysis, HLM6 (Raudenbush et al. 2004) was employed, which is capable of fitting a broad range of hierarchical models. HLM6 uses a modified pseudo-maximum likelihood method, with the modification consisting of weighting the contribution of each unit to the likelihood function by the inverse of its probability of selection. It employs a combination of the EM algorithm and Fisher scoring to obtain parameter estimates.

The problem of whether and how to incorporate weights in fitting HLMs to survey data is an area of active research (Chambers 2003; Little 2003; Pfeffermann et al. 1998; Pfeffermann, Moura, and Nascimento Silva 2004). As the discussion following the earlier Pfeffermann paper indicates, there is no unanimity in the field with respect to this question, even as to whether weights should be used at all. Alternative suggestions are made, but there is no consensus on a preferred approach.[1]

In view of the complexity of the NAEP survey, it is not surprising that the sampling weight associated with each assessed student is the product of a large number of components, each reflecting a different aspect of the survey design and its implementation. Following

[1] A fully Bayesian approach is detailed in Pfeffermann et al. (2004). Little (2003) also argues in favor of a Bayesian approach. In practice, however, non-Bayesian methods are still more popular, partly because of tradition and partly because of computational feasibility.

the recommendation of Pfeffermann et al. (1998), the student weight was factored into a school weight and a student-within-school weight. The school weight is the product of three components:

a school base weight,

a school trimming factor, and

a school nonresponse adjustment factor.

Of the four remaining components, three are related to the conditional probability of selection of the student, given that the school was selected, and one is an adjustment for student nonresponse. The latter incorporates information across schools, and it would be inappropriate to employ such information in an analysis that focuses on comparisons among schools. Therefore, that component was eliminated. The product of the other three components is a constant that, after appropriate normalization within schools, is equal to unity for all students in all schools. Thus, the results presented in the main text are derived from analyses that employed variable school weights at level 2 and constant (equal to one) student weights at level 1. This combination of weights is referred to as "standard weights" in the tables that follow.

Of course, other combinations of weights are possible and, given the lack of consensus in the field, it is informative to examine how much the estimates for a particular model would vary with different sets of weights. In effect, this would constitute a sensitivity analysis to evaluate the robustness of the reported results. If the patterns of results are similar across weighting schemes, then there is more confidence in the results reported in the main text. If they vary considerably, then the question of which weights should be preferred should be reconsidered.

Accordingly, the sequence of analyses was replicated for the NAEP scores presented in table 2-2 (reading) and table 2-7 (mathematics) with two alternative sets of weights, employing different combinations of student and school weights. The first set employs the full student weight[2] at level 1, but no school weights. The second set employs the aggregated student weights as school weights (level 2), but no student weights (level 1). For purposes of comparison, the results of employing HLM5 in conjunction with the standard weights are also included.[3] The estimates of the average difference in school means between charter schools and public noncharter schools, as well as the corresponding *p* values, are displayed in table B-1 for reading and table B-2 for mathematics.

Table B-1. Estimated average difference between mean reading scores in charter schools and public noncharter schools using different sets of weights and different versions of the HLM software, grade 4: 2003

Model			HLM version 6						HLM version 5	
			Standard weights[1]		Alternative 1[2]		Alternative 2[3]		Standard weights[1]	
	Level 1 covariates	Level 2 covariates	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value
b	None	School type	– 5.2 (2.62)	.05	– 4.3 (2.58)	.10	– 4.3 (2.57)	.09	– 7.9 (2.04)	.00
d	Race and other student characteristics	School type	– 4.2 (1.70)	.01	– 4.0 (1.67)	.02	– 3.9 (1.66)	.02	– 5.0 (1.31)	.00
e	Race and other student characteristics	School type and other school characteristics	– 3.3 (1.53)	.03	– 3.0 (1.50)	.05	– 2.9 (1.48)	.05	– 3.7 (1.22)	.00

[1] School weights only.
[2] Student weights only.
[3] School weights equal to aggregate student weights only.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

[2] This weight includes the adjustment for student nonresponse.
[3] HLM5 is the earlier version of the software released by SSI and uses a different methodology for model fitting and for incorporating weights into the analysis. The output from HLM5 is identical to that produced by the routine PROC MIXED in SAS.

Table B-2. Estimated average difference between mean mathematics scores in charter schools and public noncharter schools using different sets of weights and different versions of the HLM software, grade 4: 2003

Model			HLM version 6						HLM version 5	
			Standard weights[1]		Alternative 1[2]		Alternative 2[3]		Standard weights	
	Level 1 covariates	Level 2 covariates	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value
b	None	School type	– 5.8 (1.99)	.00	– 5.4 (1.95)	.01	– 5.3 (1.95)	.01	– 6.7 (1.44)	.00
d	Race and other student characteristics	School type	– 4.7 (1.46)	.00	– 4.6 (1.42)	.00	– 4.5 (1.43)	.00	– 4.0 (0.99)	.00
e	Race and other student characteristics	School type and other school characteristics	– 3.5 (1.31)	.01	– 3.3 (1.24)	.01	– 3.3 (1.24)	.01	– 2.7 (0.98)	.01

[1] School weights only.
[2] Student weights only.
[3] School weights equal to aggregate student weights only.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

The results for reading (table B-1) show that the estimates of the average difference under model *b* are all larger (i.e., more negative) than the estimate (– 4.1) of the average difference between students presented in the previously published report on charter schools (NCES 2004). The estimate obtained from HLM5 is the most discrepant, while the estimates obtained from HLM6 with the alternative weighting schemes are the closest. At the same time, the patterns in the estimates across models *b*, *d,* and *e* are similar for all the methods considered. In each case, the estimates become smaller in absolute value as covariates are added to each level of the model. As it happens, considering the results for HLM6 only, the estimates reported in the main text are the largest in absolute value, although the differences are minimal for models *d* and *e*. This suggests that the inclusion of student and school covariates has accounted for some of the contributions of the weights. For model *d,* all three estimates indicate an estimated average difference of about 4 score points in adjusted (for student covariates) school means between the two types of schools. The corresponding *p* values are all less than .02. When school covariates are added (model *e*), the estimates are reduced by about 1 score point, and the corresponding *p* values are less extreme.

Turning now to the results for mathematics (table B-2), the estimate obtained from HLM5 under model *b* is the most discrepant from the estimate of – 5.5 of the average difference between students that was presented in the previously published report on charter schools (NCES 2004). Under model *b*, the estimates obtained from HLM6 with the alternative weighting schemes are slightly smaller (i.e., less negative) than the estimate obtained with the preferred weighting scheme. The patterns in the estimates across models *b*, *d*, and *e* are very similar for all the methods considered. In each case, the estimates become smaller in absolute value as covariates are added to each level of the model. Considering the results for HLM6 only, the estimates reported in the main text are slightly larger in absolute value than those based on the alternative weighting schemes. For model *d,* all three estimates indicate an estimated average difference of about 4.5 score points in adjusted (for student covariates) school means between the two types of schools. The corresponding *p* values are all rounded to less than .01. When school covariates are added (model *e*), the estimates are reduced by a little more than 1 score point, and the corresponding *p* values are slightly less extreme.

Summary

In summary, for both reading and mathematics, the conclusions in chapter 2 with respect to average differences in adjusted school means between charter and public noncharter schools are very similar to those that would be reached with two different plausible sets of weights.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix C

Variance Decompositions Using State-Mean-Deviated Test Scores

Table C-1. Variance decompositions for state-mean-deviated reading scale scores, grade 4: 2003

Model	Level 1 covariates	Level 2 covariates	Between students, within schools: Variance	Between students, within schools: Percent of variance in model *a* accounted for	Between schools: Variance	Between schools: Percent of variance in model *a* accounted for
a	None	None	1102	†	273	†
b	None	School type	1102	#	272	#
c	Race	School type	1067	3	170	38
d	Race and other student characteristics	School type	887	20	113	59
e	Race and other student characteristics	School type and other school characteristics	887	20	88	68

† Not applicable.
# Rounds to zero.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table C-2. Variance decompositions for state-mean-deviated mathematics scale scores, grade 4: 2003

Model	Level 1 covariates	Level 2 covariates	Between students, within schools: Variance	Between students, within schools: Percent of variance in model *a* accounted for	Between schools: Variance	Between schools: Percent of variance in model *a* accounted for
a	None	None	608	†	194	†
b	None	School type	608	#	194	#
c	Race	School type	577	5	117	40
d	Race and other student characteristics	School type	482	21	82	58
e	Race and other student characteristics	School type and other school characteristics	481	21	63	68

† Not applicable.
# Rounds to zero.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix D

Homogeneity of Variance Assumption in the HLM Analysis

This appendix examines the assumption of homogeneity of level 1 variances that is made in the HLM analyses reported in the main text. This assumption asserts that the residual variance of the outcome, after adjusting for the level 1 covariates, is the same for all schools. A typical two-level model takes the form:

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \ldots + \beta_{pj}X_{pij} + e_{ij}$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\vdots$$
$$\beta_{pj} = \gamma_{p0}$$

where i indexes students within schools, j indexes schools;

y_{ij} is the outcome for student i in school j;

$X_1, \ldots, X_p$ are p student characteristics, centered at their grand means, and indexed by i and j as above;

β_{0j} is the mean for school j, adjusted for the predictors $X_1, \ldots, X_p$;

$\beta_{1j}, \ldots, \beta_{pj}$ are the regression coefficients for school j, associated with the predictors $X_1, \ldots, X_p$;

e_{ij} is the random error (i.e., unexplained deviation) in the level 1 equation, assumed to be independently and normally distributed with a mean zero and common variance σ^2;

W_{1j} is an indicator of the school type (charter or public noncharter) for school j;

γ_{00} is the intercept for the regression of the adjusted school mean on school type;

γ_{01} is the regression coefficient associated with school type and represents the average difference in adjusted school means between charter and public noncharter schools;

u_{0j} is the random error in the level 2 equation, assumed to be independently and normally distributed across schools with a mean zero and a variance of τ^2; and

$\gamma_{10}, \ldots, \gamma_{p0}$ are constants denoting the common values of the p regression coefficients across schools. For example, γ_{10} is the common regression coefficient associated with the first covariate in the level 1 model for each school.

The focal assumption is that, indeed, σ^2 is the same for all schools.

The HLM6 program has a chi-square test for homogeneity of variance. This test was run for the reading data and is displayed in table D-1. The p value of .00 leads to rejection of the null hypothesis that the level 1 variances are homogeneous across schools.

One approach to investigating the departure from homogeneity is to look for outliers that may be associated with some variable that is left out of the model (Raudenbush and Bryk 2002).

Table D-1. Test for homogeneity of level 1 variance, grade 4 reading: 2003

Chi–square statistic	= 49405
Number of degrees of freedom	= 4513
p value	= .00

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

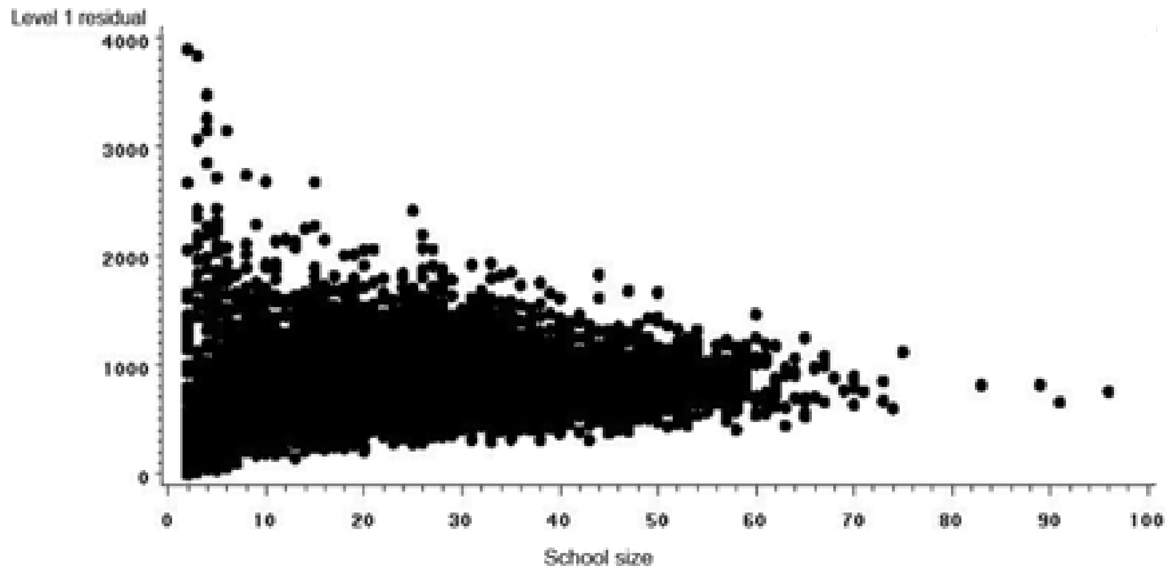
For the total group of schools (charter and public noncharter), the empirical distribution of level 1 residual variances is given in figure D-1. It appears that the 16 highest variances are clear outliers. Figure D-2 displays the same variances plotted against school size. The number of students in the schools where students were assessed in reading ranged from 1 to 96, with an average of 25. The smaller schools have residual variances that are very variable, indicated by the broad scatter of points on the left side of the figure. However, residual variances are less variable with increasing school size, indicated by the narrow scatter of points on the right side of the figure. Essentially, all of the outlying values are associated with schools with sample sizes of 10 or less.

Figure D-1. Histogram of residual variances from HLM analysis for all schools, grade 4 reading: 2003

```
Level 1 residual                                                                                    Frequency
          |
      75  |**                                                                                          41
     225  |*****                                                                                      107
     375  |******************                                                                         353
     525  |********************************************                                               878
     675  |***********************************************************************                   1413
     825  |************************************************************************                  1443
     975  |*****************************************************                                     1060
    1125  |*********************************                                                        664
    1275  |*****************                                                                        344
    1425  |**********                                                                               199
    1575  |*****                                                                                    105
    1725  |***                                                                                       50
    1875  |**                                                                                        31
    2025  |*                                                                                         15
    2175  |*                                                                                         15
    2325  |                                                                                           6
    2475  |                                                                                           3
    2625  |                                                                                           3
    2775  |                                                                                           2
    2925  |                                                                                           1
    3075  |                                                                                           3
    3225  |                                                                                           1
    3375  |                                                                                           0
    3525  |                                                                                           1
    3675  |                                                                                           0
    3825  |                                                                                           2
          |
          Š-------+-------+-------+--------+--------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
                 100     200     300      400      500     600     700     800     900    1000    1100    1200    1300    1400
                                                          Frequency
```

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

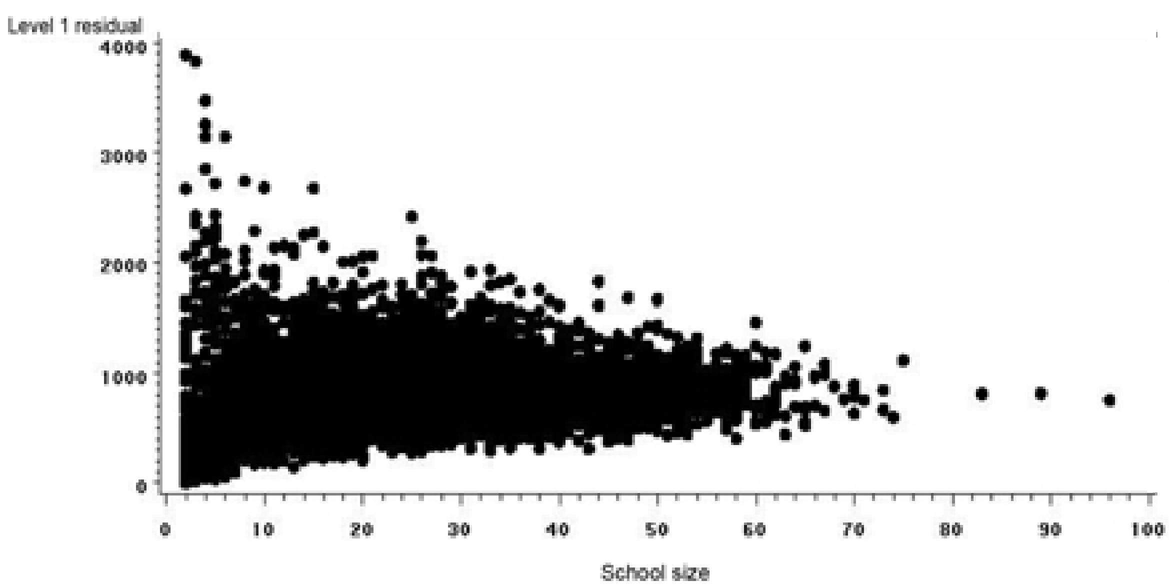
Figure D-2. Plot of school residual variance from HLM analysis against school size for all schools, grade 4 reading: 2003



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Figure D-3 displays the scatterplot of level 1 residual variances for public noncharter schools only. Since these schools constitute the vast majority of all schools in the study, the plot is almost identical to that in figure D-2.

Figure D-3. Plot of school residual variance from HLM analysis against school size for public noncharter schools, grade 4 reading: 2003



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

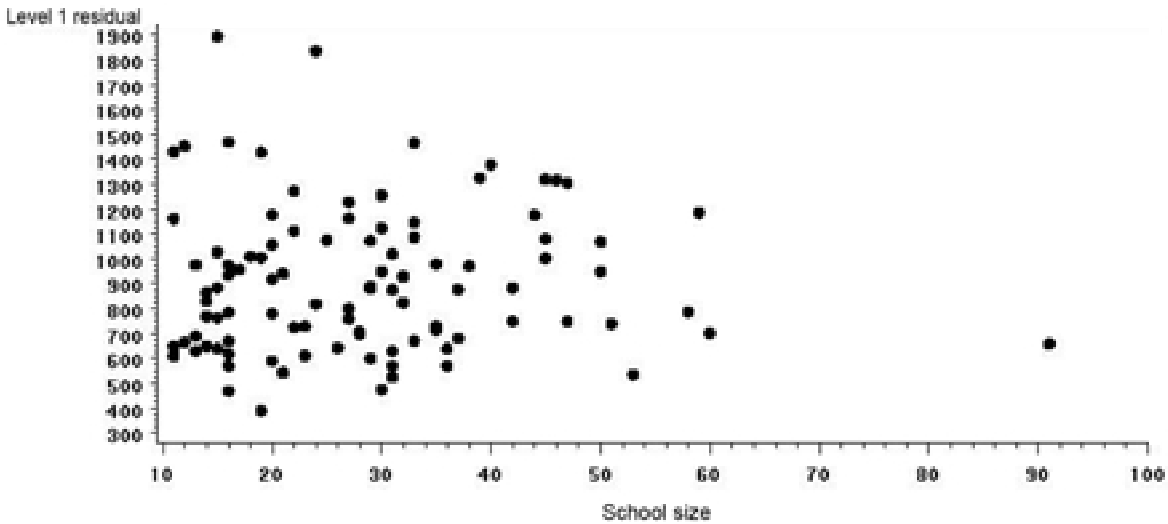
It is of some interest to determine whether charter schools have contributed disproportionately to variance heterogeneity. Figure D-4 displays the empirical distribution of level 1 residual variances for charter schools only. The range of values is smaller than that of the set of all schools, with one outlier at a value of 3000. Figure D-5 displays the residual variances for charter schools plotted against school size. As in figures D-2 and D-3, the heterogeneity among residual variances is a function of school size. More widely scattered values are found to the left of the figure for sample sizes between 10 and 20. Note, however, that the range of values is smaller than that for public noncharter schools of similar size. Thus, the inclusion of charter schools in the analysis does not contribute to the heterogeneity of residual variances.

Figure D-4. Histogram of residual variances from HLM analysis for charter schools, grade 4 reading: 2003

```
Level 1 residual                                                     Frequency
      |
  200 |*******                                                               7
      |
  600 |***********************************************************          59
      |
 1000 |****************************************************                 52
      |
 1400 |**********************                                               22
      |
 1800 |***                                                                   3
      |
 2200 |*                                                                     1
      |
 2600 |                                                                      0
      |
 3000 |*                                                                     1
      |
      Š--------+-------+-------+--------+--------+-------+-------+-------+-------+-------+-------+-------+---
               5      10      15       20       25      30      35      40      45      50      55
                                              Frequency
```

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Figure D-5. Plot of school residual variance from HLM analysis against school size for charter schools, grade 4 reading: 2003



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Although the level 1 residual variances are heterogeneous, there is not a systematic association with school type. Greater dispersion appears to be associated with school sample size. In such a situation, HLM estimates of level 2 fixed effects (such as the school-type contrast) and the corresponding standard errors are approximately unbiased (Raudenbush and Bryk 2002, chapter 9).

It is possible that, in addition to the relation to school sample size, the variance heterogeneity can be due in part to unidentified slope heterogeneity in the level 1 model. That is, if some of the regression coefficients at level 1 varied across schools but were constrained to be constant, then this misspecification could have resulted in apparent variance heterogeneity (Raudenbush and Bryk 2002, chapter 9). Such misspecification can also lead to biased estimates of the level 2 coefficients. It should be recalled that, as part of the analyses carried out, the variance components associated with the level 1 regression coefficients were tested and not found to be significantly different from zero. Consequently, following standard practice, the slopes were treated as constants in subsequent models. Nonetheless, it is certainly possible that there is sufficient heterogeneity among schools to contribute to the observed variance heterogeneity. With HLM6, it is not possible to model a unique residual variance for each school. It would be possible, however, to model one residual variance for charter schools and a different one for public noncharter schools.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix E

Data Appendix for the Charter-School-Only Analyses

Table E-1. Regression coefficients for student characteristics in model 2 of the charter-school-only analyses, grade 4 reading: 2003

Student characteristic	Regression coefficient	*p* value
Gender	5.4 (1.41)	.00
Race/ethnicity		
White	– 12.2 (3.48)	.00
Black	– 4.7 (3.08)	.13
Hispanic	6.2 (4.34)	.15
Asian/Pacific Islander	– 7.3 (9.84)	.46
American Indian/Alaska Native	2.4 (7.64)	.75
Students with disabilities	– 25.1 (4.00)	.00
English language learners	– 20.2 (5.20)	.00
Computer in the home	5.1 (2.77)	.08
Eligible for free/reduced-price school lunch	– 13.8 (2.32)	.00
Number of books in the home	8.5 (2.16)	.00
Number of absences	3.6 (1.87)	.06

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table E-2. Regression coefficients for school characteristics in model 3 of the charter-school-only analyses, grade 4 reading: 2003

School characteristic	Regression coefficient	*p* value
Percentage of students excluded	0.2 (0.12)	.15
Percentage of students absent	0.1 (0.15)	.60
Percentage of students eligible for free/reduced–price school lunch	– 0.1 (0.06)	.10
Percentage of students with a disability	– 0.9 (0.17)	.00
Percentage of English language learners	0.1 (0.08)	.26
Years of teaching experience	0.4 (0.22)	.11
Teacher certification	– 0.2 (3.12)	.95
Type of location		
Urban fringe	0.7 (4.47)	.88
Central city	4.2 (4.64)	.36
Region[1]		
Midwest	– 14.2 (5.97)	.02
South	– 11.1 (5.89)	.06
West	– 21.3 (5.73)	.00
Percentage of students by race/ethnicity		
Black	– 0.1 (0.06)	.03
Hispanic	– 0.0 (0.06)	.67
Asian/Pacific Islander	0.1 (0.21)	.49
American Indian/Alaska Native	– 0.3 (0.16)	.11
Student mobility (percentage of students enrolled the last day of school)	– 0.9 (0.89)	.31
Percentage reporting 3 to 5 percent students absent on an average day	– 0.2 (2.94)	.94
Percentage reporting 6 percent or more students absent on an average day	– 4.8 (3.39)	.16

[1] The comparison region was Northeast.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table E-3. Regression coefficients for school characteristics in model 4 of the charter-school-only analyses, grade 4 reading: 2003

School characteristic	Regression coefficient	*p* value
Affiliation with a PSD[1]	4.6 (3.46)	.18

[1] Public school district.
NOTE: Standard error of the estimate appears in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table E-4. Regression coefficients for school characteristics in models 5.1–5.7 of the charter-school-only analyses, grade 4 reading: 2003

School characteristic	Regression coefficient	*p* value
Waiver block (5.1)		
Teacher certification requirements	– 4.6 (5.30)	.39
Teacher/staff hiring/firing policies	– 0.3 (4.41)	.94
Curriculum requirements	8.8 (4.38)	.05
Student attendance/seat time requirements	6.8 (5.74)	.24
Student assessment requirements	– 5.5 (4.55)	.23
Control of finances/budget	– 1.3 (5.46)	.81
Incentives, rewards, or sanctions due to school performance	1.5 (4.38)	.73
Monitor block (5.2)		
Instructional practices	– 6.0 (4.30)	.17
Student achievement	– 8.5 (4.44)	.06
Student behavior	0.9 (4.21)	.83
Student attendance	10.8 (5.21)	.04
School governance	8.0 (5.26)	.13
School finances	– 12.0 (4.72)	.01
Compliance with state or federal regulations	– 22.2 (5.90)	.00
Reporting block (5.3)		
Chartering agency	– 5.9 (3.83)	.12
Private funders	0.4 (4.59)	.93
Parents	– 0.1 (4.91)	.98
Community/general public	4.0 (5.44)	.46
School governing board	– 2.0 (6.15)	.74
State board of education or state department of education (if this is not the chartering agency)	– 3.4 (6.46)	.60
Legislature	– 4.6 (4.28)	.28
Chartering agency block (5.4)		
State board of education	0.5 (3.82)	.89
Postsecondary institution	– 6.4 (4.38)	.15
State charter-granting agency	– 12.2 (9.04)	.18
Multiple response (other, don't know, or omitted)	– 8.7 (6.01)	.15
Population served block (5.5)		
At-risk students	– 12.1 (6.17)	.05
Students with disabilities	2.7 (12.71)	.83
Gifted/talented students	0.2 (4.34)	.97
Content block (5.6)		
Comprehensive curriculum (no specialized primary focus)	8.0 (3.72)	.03
Special curricular focus (e.g., arts, mathematics/science, foreign language immersion)	– 6.2 (10.29)	.55
Educational philosophy (e.g., Montessori, open school)	9.6 (5.24)	.10
Philosophy or set of values (e.g., Eastern philosophy, religion)	3.0 (4.02)	.46
Miscellaneous (5.7)		
Strong state chartering law	– 7.7 (3.43)	.03
New or pre-existing school	– 2.2 (4.16)	.60
School management	4.0 (3.47)	.25

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table E-5. Regression coefficients for school characteristics in model 6 of the charter-school-only analyses, grade 4 reading: 2003

School characteristic	Regression coefficient	*p* value
Percentage eligible for free/reduced-price school lunch	– 0.1 (0.05)	.14
Percentage of students with a disability	– 0.6 (0.17)	.00
Years of teaching experience	0.3 (0.20)	.10
Region[1]		
Midwest	– 8.0 (5.73)	.17
South	– 6.8 (5.12)	.19
West	– 16.8 (6.05)	.01
Percentage of Black students	– 0.1 (0.06)	.09
Percentage of Hispanic students	0.0 (0.06)	.75
Percentage of Asian/Pacific Islander students	0.2 (0.24)	.34
Percentage of American Indian/Alaska Native students	– 0.1 (0.16)	.49
Percentage reporting 6 percent or more students absent on an average day	– 3.2 (3.02)	.30
Waiver of curriculum requirements	3.4 (2.60)	.20
Areas monitored		
Student achievement	– 11.1 (4.57)	.02
Student attendance	4.9 (4.19)	.25
School finances	– 3.5 (3.90)	.38
Compliance with state/federal regulations	– 19.1 (6.66)	.01
Report to chartering agency	– 3.4 (2.92)	.25
Affiliation with a PSD[2]	3.2 (2.80)	.26
Serve at-risk students	– 2.7 (4.26)	.53
Focus of program content		
No specialized area	3.2 (2.66)	.23
Particular education philosophy	– 3.7 (8.04)	.66
Strong state chartering law	– 0.1 (2.92)	.98

[1] The comparison region was Northeast.
[2] Public school district.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Charter School Pilot Study.

Table E-6. Regression coefficients for student characteristics in model 2 of the charter-school-only analyses, grade 4 mathematics: 2003

Student characteristic	Regression coefficient	*p* value
Gender	– 3.3 (1.38)	.02
Race/ethnicity		
White	– 13.2 (2.24)	.00
Black	– 10.1 (2.11)	.00
Hispanic	9.8 (5.91)	.10
Asian/Pacific Islander	– 6.2 (8.82)	.48
American Indian/Alaska Native	– 11.1 (5.28)	.04
Students with disabilities	– 20.2 (2.35)	.00
English language learners	– 10.4 (2.16)	.00
Computer in the home	3.8 (1.64)	.02
Eligible for free/reduced-price school lunch	– 8.1 (1.78)	.00
Number of books in the home	10.0 (1.26)	.00
Number of absences	4.5 (1.09)	.00

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Table E-7. Regression coefficients for school characteristics in model 3 of the charter-school-only analyses, grade 4 mathematics: 2003

School characteristic	Regression coefficient	*p* value
Percentage of students excluded	– 0.0 (0.18)	.93
Percentage of students absent	– 0.0 (0.11)	.85
Percentage of students eligible for free/reduced–price school lunch	– 0.0 (0.06)	.75
Percentage of students with a disability	– 0.2 (0.12)	.06
Percentage of English language learners	0.1 (0.08)	.18
Years of teaching experience	0.1 (0.15)	.68
Teacher certification	2.2 (2.71)	.41
Type of location		
Urban fringe	3.0 (3.65)	.41
Central city	– 2.1 (3.88)	.59
Region[1]		
Midwest	– 6.1 (3.16)	.06
South	– 2.4 (3.91)	.54
West	– 12.7 (3.97)	.00
Percentage of students by race/ethnicity		
Black	– 0.1 (0.05)	.16
Hispanic	0.0 (0.06)	.52
Asian/Pacific Islander	0.2 (0.19)	.30
American Indian/Alaska Native	0.2 (0.18)	.23
Student mobility (percentage of students enrolled the last day of school)	– 1.6 (0.86)	.08
Percentage reporting 3 to 5 percent students absent on an average day	2.4 (2.62)	.36
Percentage reporting 6 percent or more students absent on an average day	– 3.1 (3.23)	.34

[1] The comparison region was Northeast.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Table E-8. Regression coefficients for school characterisitcs in model 4 of the charter-school-only analyses, grade 4 mathematics: 2003

School characteristic	Regression coefficient	*p* value
Affiliation with a PSD[1]	4.6 (2.89)	.12

[1] Public school district.
NOTE: Standard error of the estimate appears in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Table E-9. Regression coefficients for school characteristics in models 5.1–5.7 of the charter–school-only analyses, grade 4 mathematics: 2003

School characteristic	Regression coefficient	*p* value
Waiver block (5.1)		
Teacher certification requirements	– 2.0 (2.71)	.46
Teacher/staff hiring/firing policies	– 2.3 (3.13)	.47
Curriculum requirements	6.3 (3.07)	.04
Student attendance/seat time requirements	6.8 (4.00)	.09
Student assessment requirements	– 9.7 (3.10)	.00
Control of finances/budget	1.2 (3.43)	.72
Incentives, rewards, or sanctions due to school performance	2.8 (3.20)	.38
Monitor block (5.2)		
Instructional practices	– 2.1 (3.22)	.52
Student achievement	– 1.7 (6.19)	.79
Student behavior	0.8 (2.46)	.76
Student attendance	6.6 (4.19)	.12
School governance	4.8 (3.13)	.13
School finances	– 5.4 (4.58)	.25
Compliance with state or federal regulations	20.2 (22.90)	.38
Reporting block (5.3)		
Chartering agency	– 0.9 (3.06)	.78
Private funders	– 0.4 (3.44)	.92
Parents	– 0.4 (3.52)	.92
Community/general public	– 0.6 (3.29)	.85
School governing board	– 2.5 (4.33)	.57
Legislature	– 4.6 (3.62)	.21
Chartering agency block (5.4)		
State board of education	– 3.3 (3.30)	.31
Postsecondary institution	– 4.5 (3.33)	.18
State charter–granting agency	– 11.5 (4.75)	.02
Multiple response (other, don't know, or omitted)	1.0 (4.64)	.82
Population-served block (5.5)		
At–risk students	0.5 (4.08)	.90
Students with disabilities	6.2 (11.75)	.60
Gifted/talented students	0.8 (5.60)	.88
Content block (5.6)		
Comprehensive curriculum (no specialized primary focus)	1.4 (3.39)	.68
Special curricular focus (e.g., arts, mathematics/science, foreign language immersion)	– 10.2 (3.67)	.01
Educational philosophy (e.g., Montessori, open school)	5.3 (4.06)	.22
Philosophy or set of values (e.g., Eastern philosophy, religion)	– 3.2 (4.47)	.48
Miscellaneous (5.7)		
Report to state board of education or state department of education (if this is not the chartering agency)	– 5.2 (4.03)	.20
Strong state chartering law	– 5.0 (2.98)	.10
New or pre–existing school	– 0.1 (2.88)	.97
School management	3.6 (3.32)	.28

NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

Table E-10. Regression coefficients for school characteristics in model 6 of the charter-school-only analyses, grade 4 mathematics: 2003

School characteristic	Regression coefficient	*p* value
Percentage of students with a disability	– 0.1 (0.11)	.20
Region[1]		
Midwest	– 8.6 (5.81)	.14
South	– 0.3 (5.06)	.95
West	– 3.6 (6.26)	.57
Student mobility (percentage of students enrolled the last day of school)	– 1.1 (0.77)	.16
Waivers		
Curriculum requirements	6.2 (2.41)	.01
Student attendance/seat time requirements	6.2 (4.26)	.15
Student assessment requirements	– 7.7 (3.47)	.03
Areas monitored		
Student attendance	3.5 (3.57)	.33
School governance	3.9 (2.67)	.14
Strong state chartering law	– 3.1 (3.83)	.43
Report to state board of education or state department of education (if this is not the chartering agency)	– 3.0 (4.08)	.47
Charter granted by		
Postsecondary institution	5.7 (7.06)	.42
State charter–granting agency	– 7.0 (4.51)	.12
Focus of program content		
Special curricular focus	– 6.3 (5.18)	.23
Particular education philosophy	6.7 (5.56)	.24
Affiliation with a PSD[2]	2.6 (2.62)	.33

[1] The comparison region was Northeast.
[2] Public school district.
NOTE: Standard errors of the estimates appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Charter School Pilot Study.

THIS PAGE INTENTIONALLY LEFT BLANK.

U.S. Department of Education
ED Pubs
8242-B Sandy Court
Jessup, MD 20794-1398

Official Business
Penalty for Private Use, $300

U.S. POSTAGE PAID
U.S. DEPARTMENT
OF EDUCATION
PERMIT NO. G-17