National Assessment of Educational Progress

The Nation's Report Card

Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling

NAEP National Assessment of Educational Progress

U.S. Department of Education
NCES 2006-461

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

What Is The Nation's Report Card™?

The Nation's Report Card™, 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, 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.

The National Assessment Governing Board

Darvin M. Winick, Chair
President
Winick & Associates
Dickinson, Texas

Sheila M. Ford, Vice Chair
Former Principal
Horace Mann Elementary School
Washington, D.C.

Francie Alexander
Chief Academic Officer, Scholastic, Inc.
Senior Vice President, Scholastic Education
New York, New York

David J. Alukonis
Chairman
Hudson School Board
Hudson, New Hampshire

Amanda P. Avallone
Assistant Principal & Eighth-Grade Teacher
Summit Middle School
Boulder, Colorado

Honorable Jeb Bush
Governor of Florida
Tallahassee, Florida

Barbara Byrd-Bennett
Executive Superintendent-in-Residence
Cleveland State University
Cleveland, Ohio

Shirley V. Dickson
Educational Consultant
Laguna Niguel, California

David P. Driscoll
Commissioner of Education
Massachusetts Department of Education
Malden, Massachusetts

John Q. Easton
Executive Director
Consortium on Chicago School Research
Chicago, Illinois

David W. Gordon
Sacramento County Superintendent of Schools
Sacramento County Office of Education
Sacramento, California

Kathi M. King
Twelfth-Grade Teacher
Messalonskee High School
Oakland, Maine

Honorable Keith King
Member
Colorado House of Representatives
Colorado Springs, Colorado

Kim Kozbial-Hess
Fourth-Grade Teacher
Fall-Meyer Elementary School
Toledo, Ohio

Andrew C. Porter
Director
Learning Sciences Institute
Peabody College
Vanderbilt University
Nashville, Tennessee

Luis A. Ramos
Community Relations Manager
PPL Susquehanna
Berwick, Pennsylvania

John H. Stevens
Executive Director
Texas Business and Education Coalition
Austin, Texas

Mary Frances Taymans, SND
Executive Director
National Catholic Educational Association
Washington, D.C.

Oscar A. Troncoso
Principal
Socorro High School
Socorro Independent School District
El Paso, Texas

Honorable Thomas J. Vilsack
Governor of Iowa
Des Moines, Iowa

Michael E. Ward
Former State Superintendent of Public Instruction
North Carolina Public Schools
Jackson, Mississippi

Eileen L. Weiser
Member, State Board of Education
Michigan Department of Education
Lansing, Michigan

Grover J. Whitehurst (Ex officio)
Director
Institute of Education Sciences
U.S. Department of Education
Washington, D.C.

Charles E. Smith
Executive Director,
NAGB
Washington, D.C.

IES NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NAEP National Assessment of Educational Progress TM

Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling

U.S. Department of Education
NCES 2006-461

July 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

July 2006

The NCES World Wide Web Home Page 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). *Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling* (NCES 2006-461). 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, the NAEP Education Statistics Services Institute, Pearson Educational Measurement, and Westat.

Executive Summary

The goal of the study was to examine differences in mean National Assessment of Educational Progress (NAEP) reading and mathematics scores between public and private schools when selected characteristics of students and/or schools were taken into account. Among the student characteristics considered were gender, race/ethnicity, disability status, and identification as an English language learner. Among the school characteristics considered were school size and location, and composition of the student body and of the teaching staff. In particular, if the student populations enrolled in the two types of schools differed systematically with respect to background characteristics related to achievement, then those differences would be confounded with straightforward comparisons between school types.

The present report examined results from the 2003 NAEP assessments in reading and mathematics for grades 4 and 8. NAEP draws nationally representative samples of schools and students. In 2003, over 6,900 public schools and over 530 private schools participated in the grade 4 assessments. Over 5,500 public schools and over 550 private schools participated in the grade 8 assessments.

Hierarchical linear models (HLMs) were employed to carry out the desired adjustments. HLMs were a natural choice because they accommodate the nested structure of the data (i.e., students clustered within schools) and facilitate the inclusion of variables derived from student and school characteristics. In this study, the focal parameter was the mean difference between mean NAEP scores for two populations of schools. (This difference was not identical to the difference in mean scores between the two student populations, though the discrepancy was typically small.) HLMs were used to compare all private schools to all public schools, as well as to compare, separately, certain categories of private schools (i.e., those for which sample sizes were sufficient to report reliable estimates) to all public schools. Statistical significance was determined at the .05 level using *t* tests on model results.

Results From Grade 4

Reading

In the first set of analyses, all private schools were compared to all public schools. The average private school mean reading score was 14.7 points higher than the average public school mean reading score, corresponding to an effect size of .41 (the ratio of the absolute value of the estimated difference to the standard deviation of the NAEP fourth-grade reading score distribution). After adjusting for selected student characteristics, the difference in means was near zero and not significant. In the second set of analyses, Catholic schools and Lutheran schools were each compared to all public schools. The results, both with and without adjustments, were similar to the corresponding results for all private schools.

Mathematics

In the first set of analyses, all private schools were again compared to all public schools. The average private school mean mathematics score was 7.8 points higher than the average public school mean mathematics score, corresponding to an effect size of .29. After adjusting for selected student characteristics, the difference in means was -4.5 and significantly different from zero. (Note that a negative difference implies that the average school mean was higher for public schools.) In the second set, Catholic schools and Lutheran schools were each compared to all public schools. The results, both with and without adjustments, were similar to the corresponding results for all private schools.

Results From Grade 8

Reading

In the first set of analyses, all private schools were compared to all public schools. The average private school mean reading score was 18.1 points higher than the average public school mean reading score, corresponding to an effect size of .58. After adjusting for selected student characteristics, the difference in means was 7.3 points and significantly different from zero. In

the second set, Catholic, Lutheran, and Conservative Christian schools were each compared to all public schools. The results, both with and without adjustments, were generally similar to the corresponding results for all private schools. The only exception was that the average difference in adjusted school mean scores between Conservative Christian schools and all public schools was not significantly different from zero.

Mathematics

In the first set of analyses, all private schools were again compared to all public schools. The average private school mean mathematics score was 12.3 points higher than the average public school mean mathematics score, corresponding to an effect size of .38. After adjusting for selected student characteristics, the difference in means was nearly zero and not significant. In the second set, Catholic, Lutheran, and Conservative Christian schools were each compared to all public schools. While the results for Catholic schools, both with and without adjustments, were very similar to the corresponding results for all private schools, the results for the other two types differed.

The initial difference between Lutheran schools and all public schools was substantially larger (19.5 points) than was the case for all private schools. The average difference in adjusted mean mathematics scores between the two types of schools was 4.9 points and significantly different from zero. On the other hand, the initial difference between Conservative Christian schools and all public schools was substantially smaller (5.1 points) and not significant. The average difference in adjusted school means between Conservative Christian schools and all public schools was -7.6 points (i.e., a higher average school mean for public schools) and was significantly different from zero.

Comparison of Results for Grade 4 and Grade 8

Overall, there were many similarities in the results for the two grades. In both reading and mathematics, analyses employing unadjusted NAEP scores indicated that the average private school mean score was higher than the average public school mean score, and the difference was statistically significant. Including selected student characteristics in the model, however, resulted in a substantial reduction in the difference in all four analyses. The reduction varied from 11 to 15 score points. For grade 4 reading and grade 8 mathematics, the average difference in adjusted school mean scores was no longer significant. For grade 4 mathematics, the difference was significant, and the adjusted school mean was higher for public schools. Only for grade 8 reading was the difference still significant with a higher school mean for private schools. For all four analyses, with student characteristics such as gender and race/ethnicity incorporated in the model, the inclusion of school characteristics (e.g., teacher experience, type of school location, school size) had little impact on the estimate of the average difference between the two types of schools.

Variance decompositions yielded similar results for the four grade-subject combinations. Most of the total variance was due to heterogeneity among students within schools rather than heterogeneity among school mean scores. The combination of selected student and school characteristics accounted for about one-third of the total variance for grade 4 and about two-fifths of the total variance for grade 8.

Cautions in Interpretation

When interpreting the results from any of these analyses, it should be borne in mind that private schools constitute a heterogeneous category and may differ from one another as much as they differ from public schools. Public schools also constitute a heterogeneous category. Consequently, an overall comparison of the two types of schools is of modest utility. The more focused comparisons conducted as part of this study may be of greater value. However, interpretations of the results should take into account the variability due to the relatively small sizes of the samples drawn from each category of private school, as well as the possible bias introduced by the differential participation rates across private school categories.

There are a number of other caveats. First, the conclusions pertain to national estimates. Results based on a survey of 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, private schools are "schools of choice." 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 mean scores are confounded with average differences in the student populations, which are not fully captured by the selected student characteristics employed in this analysis.

Summary

In grades 4 and 8 for both reading and mathematics, students in private schools achieved at higher levels than students in public schools. The average difference in school means ranged from almost 8 points for grade 4 mathematics, to about 18 points for grade 8 reading. The average differences were all statistically significant. Adjusting the comparisons for student characteristics resulted in reductions in all four average differences of approximately 11 to 14 points. Based on adjusted school means, the average for public schools was significantly higher than the average for private schools for grade 4 mathematics, while the average for private schools was significantly higher than the average for public schools for grade 8 reading. The average differences in adjusted school means for both grade 4 reading and grade 8 mathematics were not significantly different from zero.

Comparisons were also carried out with subsets of private schools categorized by sectarian affiliation. After adjusting for student characteristics, raw score average differences were reduced by about 11 to 15 points. In grade 4, Catholic and Lutheran schools were each compared to public schools. For both reading and mathematics, the results were generally similar to those based on all private schools. In grade 8, Catholic, Lutheran, and Conservative Christian schools were each compared to public schools. For Catholic and Lutheran schools for both reading and mathematics, the results were again similar to those based on all private schools. For Conservative Christian schools, the average adjusted school mean in reading was not significantly different from that of public schools. In mathematics, the average adjusted school mean for Conservative Christian schools was significantly lower than that of public schools.

Acknowledgments

The 2003 National Assessment of Educational Progress (NAEP) was conducted under the direction of the National Center for Education Statistics (NCES) and overseen by the National Assessment Governing Board (NAGB). NAEP activities are carried out by Educational Testing Service (ETS), Pearson Educational Measurement, NAEP Education Statistics Services 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 Xueli Xu, 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, Linda Myers, and Arlene Weiner. 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: Lisa Bridges, Chris Chapman, Young Chun, Aaron Douglas, Hilary Forster, Arnold Goldstein, Steve Gorman, Andrew Kolstad, Roslyn Korb, Marilyn Seastrom, Alexandra Sedlacek, Alan Vanneman, and Andrew White. William Tirre, the project manager, was instrumental in ensuring the project's completion.

Finally, NAEP would not be possible without the voluntary cooperation of hundreds of school staff and thousands of students. We are grateful to them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There is an extensive research literature that treats questions related to comparisons of public and private schools. An excellent review is provided by McEwan (2000), who argues that, with a few exceptions, there is generally insufficient evidence to reach strong conclusions with regard to such comparisons. Methodological difficulties found in this literature include the size and nature of the available samples of schools and students (e.g., small sample sizes, self-selection into public or private schools), as well as key student, family, school, and community variables that remain unmeasured but may be associated with both public versus private school attendance and student achievement.

A previous National Assessment of Educational Progress (NAEP) report on the achievement of students in private schools (Perie, Vanneman, and Goldstein 2005) compared the NAEP reading and mathematics performance of fourth-, eighth-, and (for some findings) twelfth-grade students attending public and private schools. Results were also presented disaggregated by type of private school. In general, the average scores in reading and mathematics of students in private schools were found to be higher than those of students in public schools.

A natural question is whether these differences can be accounted for by differences in the populations of students attending the various kinds of schools. The previous NAEP report also presented results disaggregated both by school type and by a single student characteristic such as race/ethnicity, gender, or student-reported parents' highest level of education. Generally, the differences between public and private school student performance were diminished somewhat in these disaggregated analyses, but the average scores of private school students remained higher than those of comparable public school students; for example, students of the same race/ethnicity. The further question remains, however, as to whether these observed differences would persist if the comparisons were made between subgroups of private school and public school students who were similar with respect to several characteristics at once, for example, race/ethnicity, gender, and parents' education.

More complex analysis techniques that allow multiple covariates (e.g., race/ethnicity, eligibility for free/reduced-price school lunch) to be statistically controlled within the same analysis can be used to address this question. The present report adds to the previously reported findings by employing a particular type of analysis, called hierarchical linear modeling, the essentials of which are described below. The report examines reading and mathematics data for both grades 4 and 8 from the 2003 NAEP administration. Use of hierarchical linear modeling not only facilitates the examination of multiple covariates simultaneously, but also takes account of the clustering of students within schools.

One recent study (Lubienski and Lubienski 2006) also employed hierarchical linear models to examine public versus private students' performance on the 2003 NAEP assessment. The general conclusion from that study was "… that demographic differences between students in public and private schools more than account for the relatively high raw scores of private schools. Indeed, after controlling for these differences, the presumably advantageous 'private school effect' disappears, and even reverses in most cases" (Lubienski and Lubienski 2006, p. 3). The analyses conducted for the current report differ from those reported by Lubienski and Lubienski in two main respects. First, they focused exclusively on mathematics in grades 4 and 8, while the current report has a dual focus on reading and mathematics. Second, they divided public school students into those attending charter schools and those attending public noncharter schools—an analysis that is not included in the current report.

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).

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 (www.nagb.org/pubs/pubs.html). Additional information about the design of the 2003 assessments is provided in appendix A.

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 students in the nation, while minimizing the time burden on any individual student or school. However, NAEP is not designed to provide information that directly addresses the issue of how public and private schools contributed to a student's education.

Private schools have a long history in American education. About 10 percent of the entire school population, almost 5.3 million American students, attended private schools during the 2001–2002 school year (Broughman and Puch 2004). NAEP has reported on the performance of private school students since 1977. A recent report drew on results from NAEP assessments administered from 2000 through 2005 (Perie, Vanneman, and Goldstein 2005). In view of the interest in the question of whether different types of schools are equally effective in helping all students to learn, NAGB, which sets policy for NAEP, asked NCES to conduct a special study to supplement the standard report that documents student achievement in private and public schools. 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 how attending public versus private schools relates to students' performance on NAEP.

School and Student Samples

In 2003, over 6,900 public schools and over 530 private schools participated in the NAEP assessments in reading and mathematics at grade 4. Over 5,500 public schools and over 550 private schools participated at grade 8. Within each participating school, 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 in public schools, and a higher proportion participated in reading than in mathematics in private schools. Table 1-1 displays the number of students sampled for the reading and mathematics assessments at each grade by type of school. The number of students sampled is likely to differ slightly from the number of students who were actually assessed as a result of exclusion and absentees. (See appendix A for additional information on school and student participation rates by type of school.) Table 1-2 shows the number of participating schools and students' average scores in reading and mathematics by type of school.

Every effort is made to ensure that all sampled students who are capable of participating in the assessment are assessed. Sampled students who are identified by the school as students with disabilities or as English language learners are included in the assessment unless they do not meet criteria for inclusion established by NAEP.[1] Such students may be assessed with accommodations allowed by NAEP. Five percent of the fourth- and eighth-grade public school students sampled in 2003 were assessed with accommodations in reading, and 7 to 8 percent were assessed with accommodations in mathematics. Of the fourth- and eighth-grade private school students sampled, 2 percent were assessed with accommodations in each subject. At grade 4, public schools in the national sample excluded 4 percent of students in mathematics and 6 percent in reading in 2003. At grade 8, the exclusion rates in public schools were 4 percent in mathematics and 5 percent in reading in 2003. The comparable exclusion rates in private schools were 1 percent or less in both subjects at both grades.

Table 1-1. Student sample size for NAEP reading and mathematics assessments, by type of school, grades 4 and 8: 2003

Type of school	Student sample size by subject	
	Reading	Mathematics
Grade 4		
Public	191,400	191,400
Private	7,500	4,700
Catholic	3,700	2,300
Lutheran	900	600
Conservative Christian	1,000	700
Grade 8		
Public	155,000	153,500
Private	8,300	5,100
Catholic	4,000	2,500
Lutheran	1,000	600
Conservative Christian	1,100	700

NOTE: The number of students sampled for the combined private total includes students in the "other private school" category, which is not listed separately in the table. Sample sizes are rounded to the nearest hundred.
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 Private School Study.

Table 1-2. Number of schools and students' average NAEP reading and mathematics scores, by type of school, grades 4 and 8: 2003

Type of school	Reading		Mathematics	
	Number of schools	Average score	Number of schools	Average score
Grade 4				
Public	6,908	216 (0.3)	6,914	234 (0.2)
Private	542	235 (0.8)	539	244 (0.7)
Catholic	215	235 (1.0)	216	244 (0.8)
Lutheran	90	232 (1.9)	88	245 (1.5)
Conservative Christian	79	‡	78	‡
Grade 8				
Public	5,531	261 (0.2)	5,527	276 (0.3)
Private	568	282 (0.7)	558	292 (1.2)
Catholic	224	281 (0.9)	224	289 (1.4)
Lutheran	101	281 (1.6)	96	296 (1.6)
Conservative Christian	92	276 (1.5)	90	286 (2.6)

‡ Reporting standards not met.
NOTE: Schools participating in the combined private school category include those in the "other private school" category, which is not listed separately in the table. Standard errors of the average scores 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 and Mathematics Private School Study.

[1] The percentages of students identified as students with disabilities or English language learners were 19 to 22 percent in public schools and 3 to 4 percent in private schools (see tables A-7 and A-8 in appendix A).

Cautions in Interpretation

NAEP data are collected as part of an observational study rather than as a randomized experiment. In particular, families choose to enroll their children in private schools, and it is possible that there are systematic differences between those families and the general population of families 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 public school students and private 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." Although this study employs a powerful statistical tool, hierarchical linear models, it cannot fully compensate for the lack of relevant data. The implication is that the estimated effects obtained should not be interpreted in terms of causal relationships.

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

1. A prior measure of achievement, which is often employed as a covariate in such studies. Apparent differences in average achievement between public school students and private school students may simply reflect average differences in achievement between their respective student populations (at entry into the fourth or eighth grade) that are not adequately captured by observed student characteristics.

2. The possible attraction of parents to private schools because they felt that their children were not well served by public schools. The interpretation of the results is further clouded by the fact that, typically, students in the private school sample will have spent different amounts of time in the private school system.

3. The extent to which parents provide 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 private schools in the sample is an order of magnitude smaller than the number of public schools. Consequently, the (estimated) standard errors of the difference estimates will tend to be higher than one might expect, given the total number of schools in the sample, because they are strongly influenced by the size of the smaller sample.

It should be borne in mind that in both grades 4 and 8 the school-level response rates for each type of private school are lower than those for public schools. (See tables A-5 and A-6 in appendix A.) In both grades, Conservative Christian schools have the lowest response rates. This self-selection may introduce bias into the reported comparisons, and this bias would persist even after adjusting for observed student characteristics. Note also that the estimated standard errors that accompany each statistic do not reflect this bias.

Finally, public and private schools differ substantially in the weighted percentages of students sampled who are classified as having disabilities or being English language learners. (See tables A-7 and A-8 in appendix A.) Since classified students tend to perform more poorly on average than other students (Donahue, Daane, and Grigg 2003; Braswell, Daane, and Grigg 2003), such differences in the composition of their student populations contribute to the observed differences in achievement between school types and should be taken into account in their interpretation.

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 and Type II error rates. 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 explanatory 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 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 the public/private designation. 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 in achievement between public and private schools.

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 public or private schools. With a sufficiently large sample, such a procedure would guarantee that, on average, there are no initial differences between students attending public or private schools, and would facilitate a more controlled comparison of the two types of schools. However, students are not randomly assigned to schools; families choose to seek admission for their children to private 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.

[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.

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 covariates at the student level affects the estimated average difference in school mean scores between public and private schools. Secondary interest focuses on the impact of the inclusion of school covariates in level 2 of the model on the estimated mean score difference, as well as on the proportion of variation in the measures at each level that the covariates might account for.[3]

Note that the average difference in school means is, in general, not the same as the average difference in student outcomes. 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, Specifics of HLM Analyses, presented later in this chapter.

For both reading and mathematics at grades 4 and 8, this report presents two sets of analyses. In the first set, HLM is used to estimate the size of the average difference in school mean scores between public and private schools. All private schools are compared to public 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 mean scores between school types for each of the models.[4]

In the second set of analyses, the comparison is between public schools and different categories of private schools: Catholic and Lutheran at grade 4; Catholic, Lutheran, and Conservative Christian at grade 8. Nonreligious private schools at grades 4 and 8 and Conservative Christian schools at grade 4 are left out of this set because the number of such schools participating fell below the minimum standard (i.e., at least 70 percent participation before substitution of replacement schools). In this analysis, contrasts for the different categories of private schools are included in the HLM models so that each category of private school is compared with public schools. There are no comparisons between categories of private schools.

HLM Analysis Software

For this study, the software program HLM6, which carries out the complex calculations associated with fitting HLMs, is used.[5] This program is designed to handle the NAEP data structure, which incorporates five plausible values for each assessed student.[6] 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 dependent variable in the regression. The final estimates are the averages of the results from the five analyses (Mislevy, Johnson, and Muraki 1992). The derivation of the final standard errors follows standard NAEP procedures and combines an estimate of sampling variability based on the first set of plausible values and an estimate of measurement error obtained from the variation in results across the five sets of plausible values. 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 a student factor and a school factor. The student factor is the product of the design weight components related to students, and the school factor is the product of the design weight components related to schools. In the HLM analysis, the student factor is applied at the student level, and the school factor is applied at the school level.

[3] For some purposes, there may also be some interest in the magnitude and sign of the regression coefficients associated with the explanatory variables.

[4] The interpretation of the estimated average difference depends on what characteristics are included in the model.

[5] For information regarding this program, consult Raudenbush et al. (2004).

[6] Plausible values are random draws from the posterior distribution of scale scores for each student. The use of plausible values facilitates unbiased estimation of group statistics and their associated standard errors. The need to employ a set of plausible values for each student, rather than a single estimate of that student's score, arises from the design of the NAEP assessment in which each student is administered a small fraction of the item pool. One consequence is that there is a large uncertainty associated with a summary measure of a student's performance. Moreover, the uncertainty varies across students depending on the number and nature of the items presented. Unless these complexities are appropriately represented in the analysis, estimates of group performance, as well as comparisons between groups, can be seriously biased.

Student and School Variables

The student- and school-level variables used in the analyses are listed in figure 1-1 and were selected from among those available based on information collected from student and school questionnaires. Student-level variables included demographic characteristics and socioeconomic-related information. School-level variables included aggregated information about the teachers and students in the school, as well as school location. More detailed descriptions of the variables that were used are provided in appendix A.

Figure 1-1. Selected student- and school-level NAEP variables: 2003

Student-level variables	School-level variables
Gender	Teacher experience
Race/ethnicity	Teacher certification
Students with disabilities	Student absenteeism
English language learners	Percentage of students excluded
Computer in the home	Percentage of students by race/ethnicity
Eligibility for free/reduced-price school lunch	Student mobility
Participation in Title I	School location
Number of books in the home	Region of the country
Number of absences	Percentage of students eligible for free/reduced-price lunch
	Percentage of students with a disability
	Percentage of English language learners
	Percentage of students in the Title I program
	School size

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 Private School Study.

Specifics of HLM Analyses

Centering

When a covariate 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).

Public versus all private schools analyses

NAEP results typically show a higher average score for private school students than for public school students (Perie, Vanneman, and Goldstein 2005). A question naturally follows: 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., public or private). 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 (i.e., the means over all students), and indexed by i and j as above;

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

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

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 private schools and 0 for public 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 public and private 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.)

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,[7] 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 associated with 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 its interpretation and relevance to various policy questions. In fact, 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.

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

Public versus private school categories analyses

In the second round of analyses, models similar to those displayed above are employed, but include indicator variables distinguishing different categories of private schools—Catholic, Lutheran, and Conservative Christian (grade 8 only)—from public schools.

Description of HLM model sequence

In order to examine the differences between public and private schools, the sequence of analyses summarized in figure 1-2 was carried out. This sequence was conducted for grades 4 and 8 reading and mathematics, for public versus all private schools, and for public versus each of the two or three categories of private schools, for a total of eight analyses. It should be noted that estimated regression coefficients and their corresponding estimated standard errors are produced for each fitted model. The standard errors generated by the HLM program are intended to capture variability due to both sampling and measurement error.

Figure 1-2. Structure of fitted models

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

The rationale and a verbal description for each model follow. (As previously mentioned, the coefficient of the public/private 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 public and private schools. This estimate should be similar to the estimate in the descriptive report based on standard NAEP analysis procedures (Perie, Vanneman, and Goldstein 2005).

Model *c*: This model adjusts school means for differences in students' race/ethnicity, as well as other students' characteristics (see appendix A) that have a statistically significant relationship to the outcome. The final set of explanatory variables is determined by a sequence of exploratory analyses in which different combinations of variables are examined, much as in an ordinary least squares regression analysis. 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.[8] The school-type contrast estimates what the average difference in school means between public and private schools would be, if every school's NAEP student sample had the same breakdown on all included student variables. This is the focal model in the sequence.

Model *d*: This model builds on model *c* by including school-level variables in addition to the school-type contrast, which now estimates what the average difference in school means between public and private schools would be, if every school's NAEP student sample had the same breakdown on included student variables and the same profile on included school variables. The included school-level variables are determined by a sequence of exploratory analyses.

[8] In this setting, the term "optimal" refers to the best possible selection of variables from among those available. In addition, the problem of model misspecification (i.e., variables that are correlated with the outcome but not available for analysis) must be taken into account. An example would be measures of prior academic achievement.

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 2

Estimating the Mean Difference in Fourth-Grade Achievement Between Private and Public Schools

For each subject, the comparison of all private schools to all public schools (set 1) follows the sequence of analyses described in chapter 1. This sequence is then repeated with two categories of private schools compared to all public schools (set 2). Finally, results from the variance decompositions associated with the set 1 analyses are also presented.

Fitting different models to NAEP data shows how the inclusion of different combinations of variables changes the estimate of the focal parameter of interest (i.e., the school-type contrast). Thus, in reporting the results of a sequence of analyses, there is interest not only in the estimate for a specific model but also in the pattern of estimates through the sequence. Accordingly, instances when the estimate is not significant at the .05 level may be noted, but its magnitude and sign are not discussed. Results of the analyses for grade 4 reading and mathematics are presented in the following sections.

Reading

Comparisons of all private schools to all public schools

Table 2-1 contains results for models *b–d*. It displays estimates of the school-type contrast, comparing all private schools to all public schools, along with the corresponding standard errors and *p* values.[1] For model *b*, the estimate is 14.7; that is, the average of the mean NAEP reading scores among private schools is estimated to be almost 15 points higher than the average of the mean NAEP reading scores among public schools. This difference is significant at the .05 level. However, when all student-level covariates are included (model *c*), the estimated school-type contrast is not significantly different from zero. Finally, when school-level covariates are included at level 2 (model *d*), the estimate remains nonsignificant.

While a *p* value conveys the level of statistical significance of an estimate, it does not necessarily convey how useful 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). Effect size is a statistical term that refers to a difference between two estimates that has been scaled by some indicator of the variability of the underlying phenomenon. It complements a test of statistical significance but is not intended to imply a cause-and-effect relationship.

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 36, the effect size of the estimate for model *b* is 14.7/36 = .41; that is, the model estimate corresponds to a difference in mean fourth-grade reading scores equal to 41 percent of one standard deviation in the fourth-grade

Table 2-1. Estimated school-type contrasts, comparing private schools and public schools on mean NAEP reading scores, grade 4: 2003

Model			Difference in average school means	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value
b	None	School type	14.7 (1.10)	.00
c	Race + other student characteristics	School type	0.1 (0.83)	.94
d	Race + other student characteristics	School type + other school characteristics	– 1.1 (0.87)	.20

[1] Estimate of average difference in school means between private schools and public schools, adjusted 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 Private School Study.

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

reading score distribution. An alternative approach to calculating an effect size is detailed in appendix A.[2] It involves computing the ratio of the school-type contrast to the standard deviation of school means, based on the rationale that both numerator and denominator refer to characteristics of the distribution of school means.

Comparisons of two categories of private schools to all public schools

Table 2-2 contains the results of estimating models in which only Catholic or Lutheran schools are compared to public schools. At level 2, a distinction is made between private schools designated as Catholic or Lutheran. Consider first the results for Catholic schools. For model *b*, the estimate is 16.1; that is, the average of the mean NAEP reading scores among Catholic schools is estimated to be 16 points higher than the average of the mean NAEP reading scores among all public schools. The difference is significant at the .05 level. This corresponds to an effect size of 16.1/36 = .45. However, when all student-level covariates are included (model *c*), the estimated school-type contrast is not significant. Finally, when school-level covariates are included at level 2 (model *d*), the estimate remains nonsignificant.

This pattern is similar to the pattern obtained for all private schools and for Lutheran schools. Note that the standard errors associated with the estimates for Catholic schools are generally much less than those associated with the estimates for Lutheran schools. This is due to the larger number of Catholic schools and the larger number of students enrolled in those schools.

Variance decompositions for comparisons of all private schools to all public schools

An analysis based on HLM yields a decomposition of the total variance[3] of the NAEP reading scores into two components: One component can be associated with the variation among students within schools and one with the variation among schools. Each component is said to be "accounted for" by the corresponding units (i.e., students or schools). Table 2-3 presents the variance decompositions corresponding to models *a–d,* comparing all private schools to all public schools. It also presents the percentage reduction in the variances achieved by each level of the model, treating the variances in model *a* as the baseline.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 1,301 = 976 + 325; that is, about 75 percent of the total variance (976/1,301) is attributable to within-school heterogeneity, and about 25 percent of the total variance (325/1,301) is attributable to between-school heterogeneity.[4]

The introduction of the school-type contrast at level 2 (model *b*) accounts for 10 percent of the between-school variance, despite the fact that the corresponding regression coefficient is large and statistically significant. This seems counterintuitive, but the explanation is

Table 2-2. Estimated school-type contrasts, comparing two types of private schools and public schools on mean NAEP reading scores, grade 4: 2003

Model			Catholic		Lutheran	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value	Estimate[2]	*p* value
b	None	School type	16.1 (1.21)	.00	13.3 (2.39)	.00
c	Race + other student characteristics	School type	1.5 (0.93)	.11	– 2.7 (1.73)	.12
d	Race + other student characteristics	School type + other school characteristics	– 0.5 (1.00)	.59	– 1.6 (1.58)	.32

[1] Estimate of average difference in school means between Catholic schools and public schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between Lutheran schools and public schools, adjusted 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 Private School Study.

[2] Using the method explained in appendix A, the effect size for model *b* is 14.7/18.0 = .82, since the standard deviation of school means is 18.
[3] The total variance is the variance in scores of the full NAEP sample of students.
[4] Between-school heterogeneity refers to the variance among (unadjusted) school means.

Table 2-3. Variance decompositions for NAEP reading scores, grade 4: 2003

Model	Level 1 covariates	Level 2 covariates	Between students, within schools: Residual variance	Between students, within schools: Percentage of variance in model *a* accounted for	Between schools: Residual variance	Between schools: Percentage of variance in model *a* accounted for
a	None	None	976	†	325	†
b	None	School type	976	#	291	10
c	Race + other student characteristics	School type	782	20	100	69
d	Race + other student characteristics	School type + other school characteristics	782	20	85	74

† 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 Private School Study.

that the number of private schools is small in comparison to the number of public schools and that most of the variance at the school level is due to differences among schools within each school type. Consequently, despite the large regression coefficient associated with the school-type indicator, eliminating the difference in means between school types would have only a modest impact on the total variance among school means.

Turning to model *c*, including all student-level covariates accounts for 20 percent of the within-school variance. However, the impact on the variation at level 2 is greater. In fact, the variance among school means adjusted for student characteristics and school type (100) is now less than one-third as large as the variance among unadjusted school means (325). Finally, when school-level covariates are added (model *d*), the residual variance among adjusted school means is reduced to 85, representing an additional 5 percent (74 – 69) of the initial between-school heterogeneity accounted for.

This incremental contribution of 5 percent seems rather small. However, the reduction in between-school variance from model *c* to model *d* is from 100 to 85, or 15 percent of the model *c* variance. Thus, a substantively meaningful proportion of the heterogeneity among adjusted school means can be accounted for by measured school characteristics.

When both student and school variables are 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 (or 14 percent of the total variance) can be accounted for by the covariates in the model. Additionally, the between-school variance component represents about 25 percent of the total variance, and about 74 percent of that component (or 19 percent of the total variance) can be accounted for by the covariates in the model. In summary, the covariates included in the model account for about one-third (33 percent) of the total variance. The results for the variance decompositions based on the analyses that distinguish between two categories of private schools are very similar to those discussed above and are not presented.

The results of the variance decomposition enhance the understanding of the likely sources of heterogeneity in student achievement and, consequently, of the context in which the HLM analyses take place. A comparison of the variance between schools for models *b* and *c* indicates that schools (in general) differ widely in measured characteristics of students that are associated with reading achievement. Specifically, when an adjustment is made for those characteristics, the heterogeneity among school means is reduced by almost two-thirds. The variance decomposition associated with model *d* indicates that about one-quarter of the total variance remains to be accounted for 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, compared to differences in student characteristics, in accounting for the variation in reading achievement among students.

Mathematics

Comparisons of all private schools to all public schools

Table 2-4 contains results for models *b–d*. It displays estimates of the school-type contrast, comparing all private schools to all public schools, along with the corresponding standard errors and *p* values. For model *b*, the estimate is 7.8; that is, the average of the mean NAEP mathematics scores among private schools is estimated to be about 8 points higher than the average of the mean NAEP mathematics scores among public schools. This difference is significant at the .05 level. Since the standard deviation of grade 4 mathematics scores is 27, the corresponding effect size is 7.8/27 = .29.[5] When all student-level covariates are included (model *c*), the estimated school-type contrast is -4.5 and significant at the .05 level. That is, when school means are adjusted for differences in measured student characteristics, the average of private school means is significantly lower than the average of public school means. Finally, when school-level covariates are included at level 2 (model *d*), the estimated school-type contrast is -4.1 and is significant at the .05 level. The effect sizes corresponding to the estimates from models *c* and *d* are .17 and .15, respectively.

Table 2-4. Estimated school-type contrasts, comparing private schools and public schools on mean NAEP mathematics scores, grade 4: 2003

Model			Difference in average school means	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value
b	None	School type	7.8 (0.86)	.00
c	Race + other student characteristics	School type	– 4.5 (0.67)	.00
d	Race + other student characteristics	School type + other school characteristics	– 4.1 (0.72)	.00

[1] Estimate of average difference in school means between private schools and public schools, adjusted 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 Private School Study.

Comparisons of two categories of private schools to all public schools

Table 2-5 contains the results of estimating models in which only Catholic or Lutheran schools are compared to public schools. At level 2, a distinction is made between private schools designated as Catholic or Lutheran. Consider first the results for Catholic schools. For model *b*, the estimate is 8.2; that is, the average of the mean NAEP mathematics scores among Catholic schools is estimated to be about 8 points higher than the average of the mean NAEP mathematics scores among all public schools. The difference is significant at the .05 level. When all student-level covariates are included (model *c*), the estimated school-type contrast is -4.3 and significant at the .05 level. Finally, when school-level covariates are included at level 2 (model *d*), the estimate is -4.5 and significant at the .05 level. Note that the negative estimates in models *c* and *d* mean that the average of school means is higher for public schools.

This pattern is similar to the pattern obtained for all private schools, but not for Lutheran schools. For the latter, the estimate for model *b* is 10.0 and significant at the .05 level. When all student-level covariates are included, the estimated school-type contrast is -2.9 and significant at the .05 level. However, when school-level covariates are included at level 2, the estimate is not significant.

Note that the standard errors associated with the estimates for Catholic schools are generally less than those associated with the estimates for Lutheran schools. This is due to the larger number of Catholic schools and the larger number of students enrolled in those schools.

Variance decompositions for comparisons of all private schools to all public schools

As indicated earlier, an analysis based on HLM decomposes 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 2-6 presents the variance decompositions corresponding to models *a–d*,

[5] Using the method explained in appendix A, the effect size for model *b* is 7.8/14.7 = .53, since the standard deviation of school means is 14.7.

Table 2-5. Estimated school-type contrasts, comparing two types of private schools and public schools on mean NAEP mathematics scores, grade 4: 2003

Model			Catholic		Lutheran	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value	Estimate[2]	*p* value
b	None	School type	8.2 (1.04)	.00	10.0 (1.71)	.00
c	Race + other student characteristics	School type	– 4.3 (0.79)	.00	– 2.9 (1.41)	.04
d	Race + other student characteristics	School type + other school characteristics	– 4.5 (0.85)	.00	– 1.3 (1.37)	.35

[1] Estimate of average difference in school means between Catholic schools and public schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between Lutheran schools and public schools, adjusted 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 Private School Study.

comparing all private schools to all public schools. It also presents the percentage reduction in the residual variances achieved by each level of the model, treating the residual variances in model *a* as the baseline.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 746 = 531 + 215. That is, 71 percent of the total variance (531/746) is attributable to within-school heterogeneity, and 29 percent of the total variance (215/746) is attributable to between-school heterogeneity.

The introduction of the school-type contrast at level 2 (model *b*) accounts for 4 percent of the between-school variance, despite the fact that the corresponding regression coefficient is large and statistically significant. This seems counterintuitive, but the explanation, again, is that the number of private schools is small in comparison to the number of public schools and that most of the variance at the school level is due to differences among schools within each school type. Consequently, eliminating the difference in means between school types would have only a modest impact on the total variance among school means.

Turning to model *c*, including all student-level covariates accounts for 22 percent of the within-school variance. However, the impact on the variation at level 2 is much greater. In fact, the variance among school means adjusted for student characteristics and school type (77) is about one-third as large as the variance among unadjusted school means (215). Finally, when school-level covariates are added (model *d*), the residual variance among adjusted school means is reduced to 59, representing an additional 9 percent (73 – 64) of the initial between-school heterogeneity accounted for.

This incremental contribution of 9 percent seems rather small. However, the reduction in between-school

Table 2-6. Variance decompositions for NAEP mathematics scores, grade 4: 2003

Model			Between students, within schools		Between schools	
	Level 1 covariates	Level 2 covariates	Residual variance	Percentage of variance in model *a* accounted for	Residual variance	Percentage of variance in model *a* accounted for
a	None	None	531	†	215	†
b	None	School type	531	#	206	4
c	Race + other student characteristics	School type	415	22	77	64
d	Race + other student characteristics	School type + other school characteristics	415	22	59	73

† 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 Private School Study.

variance from model *c* to model *d* is from 77 to 59, or 23 percent of the model *c* variance. Thus, a considerable proportion of the remaining heterogeneity among adjusted school means can be accounted for by school characteristics other than school type.

When both student and school variables are available, more between-school than within-school heterogeneity is accounted for. The within-school variance component represents about 71 percent of the total variance, and about 22 percent of that component (or 15 percent of the total variance) can be accounted for by the covariates in the model. Additionally, the between-school variance component represents about 29 percent of the total variance, and about 73 percent of that component (or 21 percent of the total variance) can be accounted for by the covariates in the model. In summary, the covariates included in the model account for about one-third (36 percent) of the total variance. The results for the variance decompositions based on the analyses that distinguish between two categories of private schools are very similar to those discussed above and are not presented.

As was the case for reading, the results of the variance decomposition enhance the understanding of the likely sources of heterogeneity in student achievement and, consequently, of the context in which the HLM analyses take place. A comparison of the variance between schools for models *b* and *c* indicates that schools (in general) differ widely in measured characteristics of students that are associated with mathematics achievement. Specifically, when adjusted for those characteristics, the heterogeneity among school means is reduced by almost two-thirds. The variance decomposition associated with model *d* indicates that about one-quarter of the total variance remains to be accounted for 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, compared to differences in student characteristics, in accounting for the variation in mathematics achievement among students.

Summary

The sequence of analyses comparing all private schools to all public schools in grade 4 reading yields informative results. With no adjustment for differences among students, the average of the mean NAEP reading scores is higher among private schools than among public schools. However, when student covariates are included in the model, the averages of the adjusted school means for the two types of schools are not significantly different. The pattern is repeated for the separate comparisons of all public schools to Catholic schools and to Lutheran schools.

The variance decompositions associated with the analyses comparing all private schools to all public schools indicate that there is greater variance among students within schools than among schools. However, a smaller fraction of the within-school variance component than of the between-school component is accounted for. About one-third of the total variance can be accounted for by a combination of student and school covariates.

The sequence of analyses comparing all private schools to all public schools in grade 4 mathematics yields somewhat different results. With no adjustment for differences among students, the average of the mean NAEP mathematics scores is higher among private schools than among public schools. However, when student-level covariates are included in the model, the averages of private school means is lower than the averages of public school means. The same pattern holds for the separate comparisons of all public schools to Catholic schools and to Lutheran schools.

Although the total variance in mathematics is slightly more than one-half the total variance in reading, the pattern in variance decompositions is similar to that of reading. The between-student, within-school variance component is greater than the between-school variance component. However, a smaller fraction of the between-student, within-school variance than of the between-school variance is accounted for. About one-third of the total variance can be accounted for by a combination of student and school covariates available for use in these analyses.

Chapter 3

Estimating the Mean Difference in Eighth-Grade Achievement Between Private and Public Schools

For each subject, the comparison of all private schools to all public schools (set 1) follows the sequence of analyses described in chapter 1. This sequence is then repeated with three categories of private schools compared to all public schools (set 2). Results from the variance decompositions associated with the set 1 analyses are also presented.

Fitting different models to NAEP data shows how the inclusion of different combinations of variables changes the estimate of the focal parameter of interest (i.e., the school-type contrast). Thus, in reporting the results of a sequence of analyses, there is interest not only in the estimate for a specific model but also in the pattern of estimates through the sequence. Accordingly, instances when the estimate is not significant at the .05 level may be noted, but its magnitude and sign are not discussed. Results of the analyses for grade 8 reading and mathematics are presented in the following sections.

Reading

Comparisons of all private schools to all public schools

Table 3-1 contains results for models *b*–*d*. It displays estimates of the school-type contrast, comparing all private schools to all public schools, along with the corresponding standard errors and *p* values.[1] For model *b*, the estimate is 18.1; that is, the average of the mean NAEP reading scores among private schools is estimated to be about 18 points higher than the average of the mean NAEP reading scores among public schools. This difference is significant at the .05 level. Since the standard deviation of reading scores is 31, the corresponding effect size is 18.1/31 = .58.[2] When all student-level covariates are included (model *c*), the estimated school-type contrast is 7.3 and significant at the .05 level. Finally, when school-level covariates are included at level 2 (model *d*), the estimate is 5.7 and significant at the .05 level. The effect sizes corresponding to models *c* and *d* are 7.3/31 = .24 and 5.7/31 = .18, respectively.

Table 3-1. Estimated school-type contrasts, comparing private schools and public schools on mean NAEP reading scores, grade 8: 2003

Model			Difference in average school means	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value
b	None	School type	18.1 (1.26)	.00
c	Race + other student characteristics	School type	7.3 (0.91)	.00
d	Race + other student characteristics	School type + other school characteristics	5.7 (1.00)	.00

[1] Estimate of average difference in school means between private schools and public schools, adjusted 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 Private School Study.

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

[2] If the alternative approach detailed in appendix A is adopted, the effect size is 18.1/18 = 1.0, since the standard deviation of school means is 18. See discussion of effect size at the beginning of chapter 2.

Table 3-2. Estimated school-type contrasts, comparing three types of private schools and public schools on mean NAEP reading scores, grade 8: 2003

Model			Catholic		Lutheran		Conservative Christian	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value	Estimate[2]	*p* value	Estimate[3]	*p* value
b	None	School type	19.3 (1.08)	.00	20.5 (1.76)	.00	12.3 (2.75)	.00
c	Race + other student characteristics	School type	8.0 (0.77)	.00	7.1 (1.46)	.00	1.1 (2.03)	.59
d	Race + other student characteristics	School type + other school characteristics	5.4 (0.81)	.00	4.9 (1.38)	.00	0.8 (1.92)	.67

[1] Estimate of average difference in school means between Catholic schools and public schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between Lutheran schools and public schools, adjusted for other variables in the model.
[3] Estimate of average difference in school means between Conservative Christian schools and public schools, adjusted 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 Private School Study.

Comparisons of three categories of private schools to all public schools

Table 3-2 contains the results of estimating models in which only Catholic, Lutheran, or Conservative Christian schools are compared to public schools. At level 2, a distinction is made among private schools designated as Catholic, Lutheran, and Conservative Christian. Consider first the results for Catholic schools. For model *b*, the estimate is 19.3; that is, the average of the mean NAEP reading scores among Catholic schools is estimated to be about 19 points higher than the average of the mean NAEP reading scores among all public schools. The difference is significant at the .05 level. When all student-level covariates are included (model *c*), the estimated school-type contrast is 8.0 and significant at the .05 level. Finally, when school-level covariates are included at level 2 (model *d*), the estimate is 5.4 and significant at the .05 level. The effect sizes corresponding to models *c* and *d* are 8.0/31 = .26 and 5.4/31 = .17, respectively.

This pattern is similar to the pattern obtained for all private schools. It is repeated for Lutheran schools but not for Conservative Christian schools. In particular, for the latter, the estimates for models *c* and *d* are not significantly different from zero. Note that the standard errors associated with the estimates for Catholic schools are generally less than those associated with the estimates for Lutheran and Conservative Christian schools. This is due to the larger number of Catholic schools and the larger number of students enrolled in those schools.

Variance decompositions for comparisons of all private schools to all public 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 3-3 presents the variance decompositions corresponding to models *a–d,* comparing all private schools to all public schools. It also presents the percentage reduction in the variances achieved by each level of the model, treating the variances in model *a* as the baseline.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 977 = 649 + 328; that is, about 66 percent of the total variance (649/977) is attributable to within-school heterogeneity, and about 34 percent of the total variance (328/977) is attributable to between-school heterogeneity. The introduction of the school-type contrast at level 2 (model *b*) accounts for 25 percent of the between-school variance.

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

Table 3-3. Variance decompositions for NAEP reading scores, grade 8: 2003

Model	Level 1 covariates	Level 2 covariates	Between students, within schools: Residual variance	Between students, within schools: Percentage of variance in model *a* accounted for	Between schools: Residual variance	Between schools: Percentage of variance in model *a* accounted for
a	None	None	649	†	328	†
b	None	School type	650	#	245	25
c	Race + other student characteristics	School type	502	23	97	70
d	Race + other student characteristics	School type + other school characteristics	501	23	82	75

† 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 Private School Study.

Turning to model *c*, including all student-level covariates accounts for 23 percent of the within-school variance. However, the impact on the variation at level 2 is greater. In fact, the variance among school means adjusted for student characteristics and school type (97) is now less than one-third as large as the variance among unadjusted school means (328). Finally, when school-level covariates are added (model *d*), the residual variance among adjusted school means is reduced to 82, representing an additional 5 percent (75 – 70) of the initial between-school heterogeneity accounted for.

This incremental contribution of 5 percent seems rather small. However, the reduction in between-school variance from model *c* to model *d* is from 97 to 82, or 15 percent of the model *c* variance. Thus, a considerable proportion of the remaining heterogeneity among adjusted school means can be accounted for by school characteristics other than school type.

When both student and school variables are available, more between-school than within-school heterogeneity is accounted for. The within-school variance component represents about 66 percent of the total variance, and about 23 percent of that component can be accounted for. On the other hand, 75 percent of the smaller between-school variance component is accounted for. In summary, the covariates included in the model account for about two-fifths (41 percent) of the total variance. The results for the variance decompositions based on the analyses that distinguish among three categories of private schools are very similar to those discussed above and are not presented.

As was the case for grade 4 reading, the results of the variance decomposition enhance the understanding of the context in which the HLM analyses take place. A comparison of the variance between schools for models *b* and *c* indicates that schools (in general) differ widely in measured characteristics of students that are associated with reading achievement. Specifically, when those characteristics are adjusted for, the heterogeneity among school means is reduced by almost two-thirds. The variance decomposition associated with model *d* indicates that about one-quarter of the total variance remains to be accounted for 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, compared to differences in student characteristics, in accounting for the variation in reading achievement among students.

Mathematics

Comparisons of all private schools to all public schools

Table 3-4 contains results for models *b–d*. It displays estimates of the school-type contrast, comparing all private schools to all public schools, along with the corresponding standard errors and *p* values. For model *b*, the estimate is 12.3; that is, the average of the mean NAEP mathematics scores among private schools is estimated to be about 12 points higher than the average of the mean NAEP mathematics scores among public schools. This difference is significant at the .05 level. Since the standard deviation of mathematics scores is 32, the corresponding effect size is 12.3/32 = .38.[4] When all student-level covariates are included (model *c*), the estimated school-type contrast is not significantly different from zero. Finally, when school-level covariates are included at level 2 (model *d*), the estimated school-type contrast is again not significantly different from zero.

Table 3-4. Estimated school-type contrasts, comparing private schools and public schools on mean NAEP mathematics scores, grade 8: 2003

Model			Difference in average school means	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value
b	None	School type	12.3 (1.21)	.00
c	Race + other student characteristics	School type	– 0.2 (0.93)	.83
d	Race + other student characteristics	School type + other school characteristics	– 0.6 (0.97)	.56

[1] Estimate of average difference in school means between private schools and public schools, adjusted 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 Private School Study.

Comparisons of three categories of private schools to all public schools

Table 3-5 contains the results of estimating models in which only Catholic, Lutheran, or Conservative Christian schools are compared to public schools. At level 2, a distinction is made among private schools designated as Catholic, Lutheran, and Conservative Christian. Consider first the results for Catholic schools. For model *b*, the estimate is 13.2; that is, the average of the mean NAEP mathematics scores among Catholic schools is estimated to be about 13 points higher than the average of the mean NAEP mathematics scores among all public schools. The difference is significant at the .05 level. The corresponding effect size is 13.2/32 = .41. When all student-level covariates are included (model *c*), the estimated school-type contrast is not significantly different from zero. Finally, when school-level covariates are included at level 2 (model *d*), the estimate is not significantly different from zero.

This pattern is similar to the pattern obtained for all private schools. But in this case, the results for Lutheran and Conservative Christian schools follow a different pattern. For Lutheran schools, the estimate for model *b* is 19.5 and significant at the .05 level. When all student-level covariates are included, the estimate is 4.9 and significant at the .05 level. The effect sizes corresponding to models *b* and *c* are 19.5/32 = .61, and 4.9/32 = .15, respectively. However, when school-level covariates are included at level 2, the estimate is not significantly different from zero. Consider now the Conservative Christian schools. The estimate for model *b* is not significantly different from zero. When all student-level covariates are included, the estimate is -7.6 and significant at the .05 level; that is, when school means are adjusted for differences in student characteristics, Conservative Christian schools have lower adjusted mean scores, on average, than public schools. Finally, when school-level covariates are included at level 2, the estimate is -7.7 and significant at the .05 level. The effect sizes corresponding to models *c* and *d* are 7.6/32 = .24 and 7.7/32 = .24, respectively. (Note

[4] Since the standard deviation of school means is 19, the alternative effect size calculation yields 12.3/19 = .65.

Table 3-5. Estimated school-type contrasts, comparing three types of private schools and public schools on mean NAEP mathematics scores, grade 8: 2003

Model			Catholic		Lutheran		Conservative Christian	
	Level 1 covariates	Level 2 covariates	Estimate[1]	*p* value	Estimate[2]	*p* value	Estimate[3]	*p* value
b	None	School type	13.2 (1.43)	.00	19.5 (2.32)	.00	5.1 (3.38)	.13
c	Race + other student characteristics	School type	0.8 (1.08)	.48	4.9 (1.89)	.01	– 7.6 (2.62)	.00
d	Race + other student characteristics	School type + other school characteristics	– 1.1 (1.16)	.36	3.4 (2.00)	.09	– 7.7 (2.60)	.00

[1] Estimate of average difference in school means between Catholic schools and public schools, adjusted for other variables in the model.
[2] Estimate of average difference in school means between Lutheran schools and public schools, adjusted for other variables in the model.
[3] Estimate of average difference in school means between Conservative Christian schools and public schools, adjusted 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 Private School Study.

that these effect sizes reflect higher performance among public school students.)

The standard errors associated with the estimates for Catholic schools are generally less than those associated with the estimates for the Lutheran and Conservative Christian schools. This is due to the larger number of Catholic schools and the larger number of students enrolled in those schools.

Variance decompositions for comparisons of all private schools to all public schools

As indicated earlier, an analysis based on HLM decomposes 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 3-6 presents the variance decompositions corresponding to models *a*–*d*, comparing all private schools to all public schools. It also presents the percentage reduction in the variances achieved by each level of the model, treating the variances in model *a* as the baseline.

Model *a* yields the basic decomposition. The total variance is simply the sum of the two displayed components: 1,024 = 666 + 358. That is, about 65 percent of the total variance (666/1,024) is attributable to within-school heterogeneity, and about 35 percent of the total variance (358/1,024) is attributable to between-school heterogeneity.

The introduction of the school-type contrast at level 2 (model *b*) accounts for 9 percent of the between-school variance, despite the fact that the corresponding regres-

Table 3-6. Variance decompositions for NAEP mathematics scores, grade 8: 2003

Model			Between students, within schools		Between schools	
	Level 1 covariates	Level 2 covariates	Residual variance	Percentage of variance in model *a* accounted for	Residual variance	Percentage of variance in model *a* accounted for
a	None	None	666	†	358	†
b	None	School type	666	#	325	9
c	Race + other student characteristics	School type	492	26	127	65
d	Race + other student characteristics	School type + other school characteristics	492	26	107	70

† 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 Private School Study.

sion coefficient is large and statistically significant. This seems counterintuitive, but the explanation, again, is that the number of private schools is small in comparison to the number of public schools and that most of the variance at the school level is due to differences among schools within each school type. Consequently, eliminating the difference in means between school types would have only a modest impact on the total variance among school means.

Turning to model *c*, including all student-level covariates accounts for 26 percent of the within-school variance. However, the impact on the variation at level 2 is greater. In fact, the variance among school means adjusted for student characteristics and school type (127) is about one-third as large as the variance among unadjusted school means (358). Finally, when school-level covariates are added (model *d*), the residual variance among adjusted school means is reduced to 107, representing an additional 5 percent (70 - 65) of the initial between-school heterogeneity accounted for.

This incremental contribution of 5 percent seems rather small. However, the reduction in between-school variance from model *c* to model *d* is from 127 to 107, or 16 percent of the model *c* variance. Thus, a considerable proportion of the remaining heterogeneity among adjusted school means can be accounted for by school characteristics other than school type.

When both student and school variables are available, more between-school than within-school heterogeneity is accounted for. The within-school variance component represents about 65 percent of the total variance, and about 26 percent of that component (or 17 percent of the total variance) can be accounted for by the covariates in the model. Additionally, the between-school variance component represents about 35 percent of the total variance, and about 70 percent of that component (or 25 percent of the total variance) can be accounted for by the covariates in the model. In summary, the covariates included in the model account for about two-fifths (42 percent) of the total variance.

As was the case for grade 4 mathematics, the results of the variance decomposition enhance the understanding of the context in which the HLM analyses take place. A comparison of the variance between schools for models *b* and *c* indicates that schools (in general) differ widely in measured characteristics of students that are associated with mathematics achievement. Specifically, when those characteristics are adjusted for, the heterogeneity among school means is reduced by almost two-thirds. The variance decomposition associated with model *d* indicates that about 30 percent of the total variance remains to be accounted for 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, compared to differences in student characteristics, in accounting for the variation in mathematics achievement among students.

Summary

The sequence of analyses comparing all private schools to all public schools in grade 8 reading yields results generally similar to those obtained for the separate comparisons of all public schools to Catholic schools and Lutheran schools, but not to Conservative Christian schools. With no adjustment for differences among students, the average of the mean NAEP reading scores is higher among private schools (as well as Catholic and Lutheran schools separately) than among public schools. When student covariates are included in the model, all private schools, as well as Catholic and Lutheran schools, maintain a higher average school mean than public schools.

The variance decompositions associated with the analyses comparing all private schools to all public schools indicate that there is greater variance among students within schools than among schools. However, a smaller fraction of the within-school variance component than of the between-school component is accounted for. About two-fifths of the total variance can be accounted for by a combination of student and school covariates.

The analysis comparing all private schools to all public schools in grade 8 mathematics, with no adjustment for differences among students, shows a higher average school mean for private schools than for public schools. This initial difference between private and public schools in mathematics is smaller than the one in reading. When student-level covariates are included in the model, the difference in adjusted averages between private and public schools is not significantly different from zero. The same pattern holds for the separate comparisons of all public schools to Catholic schools, but not to Lutheran or Conservative Christian schools. The initial difference showing a higher average school mean for Lutheran schools than for public schools is significantly different from zero, as is the difference in school means adjusted for student characteristics. With respect to the Conservative Christian schools, the average difference in unadjusted means with public schools is not significantly different from zero. (That is, there is no initial difference between Conservative Christian schools and public schools.) When the average difference in school means is adjusted for student characteristics, the average school mean is higher for public schools than for Conservative Christian schools.

In this grade, the total variance in mathematics is almost equal to the total variance in reading. Moreover, the patterns in variance decompositions are very similar for the two subjects. The between-student, within-school variance component is greater than the between-school variance component. However, a smaller fraction of the larger component than of the smaller one is accounted for. Slightly more than two-fifths of the total variance can be accounted for by a combination of student and school covariates.

References

Allen, N.L., Donoghue, J.R., and Schoeps, T.L. (2001). The NAEP 1998 Technical Report (NCES 2001-509). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved January 31, 2006, from http//:nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001509.

Braswell, J., Daane, M., and Grigg, W. (2003). *The Nation's Report Card: Mathematics Highlights 2003* (NCES 2004-451). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Broughman, P.S., and Puch, K.W. (2004). *Characteristics of Private Schools in the United States: Results From the 2001–2002 Private School Universe Survey* (NCES 2005-305). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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.

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

Donahue, P., Daane, M., and Grigg, W. (2003). *The Nation's Report Card: Reading Highlights 2003* (NCES 2004-452). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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.

McEwan, P.J. (2000). The Potential Impact of Large-Scale Voucher Programs. *Review of Educational Research*, *70*(2): 103–149.

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 Assessment Governing Board (1996). *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Washington, DC: Author.

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

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

Perie, M., Vanneman, A., and Goldstein, A. (2005). *Student Achievement in Private Schools: Results From NAEP 2000–2005* (NCES 2006-459). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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.

THIS PAGE INTENTIONALLY LEFT BLANK.

Appendix A

Overview of Procedures

This appendix provides an overview of the 2003 NAEP reading and mathematics assessments, samples, and analysis procedures.

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 is presented in this section. The complete frameworks is available on the NAGB website (www.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 three contexts for reading and four aspects of reading. 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. When *reading to perform a task*, readers search quickly for specific pieces of information.

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 reading assessment. Included in the figure are sample questions that illustrate how each aspect of reading is assessed within each reading context. (Note that reading to perform a task is not assessed at grade 4.)

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

Context for reading	Aspect of 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.*
Reading to perform a task	*What time can you get a nonstop flight to X?*	*What must you do before step 3?*	*Describe a situation in which you would omit step 5.*	*Is the information in this brochure easy to use?*

SOURCE: *Reading Framework for the 2003 National Assessment of Educational Progress* (NAEP), National Assessment Governing Board.

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 assessment, are presented in tables A-1 and A-2.

Table A-1. Target and actual percentage distribution of NAEP reading questions, by context for reading, grades 4 and 8: 2003

Grade	Context for reading: Reading for literary experience	Reading for information	Reading to perform a task
Grade 4			
Target	55	45	†
Actual	50	50	†
Grade 8			
Target	40	40	20
Actual	28	41	30

† Not applicable. Reading to perform a task was not assessed at grade 4.
NOTE: Detail may not sum to totals because of rounding.
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 on the NAEP reading assessment, by aspect of reading, grades 4 and 8: 2003

Grade	Aspect of reading: Forming a general understanding/ developing interpretation	Making reader/text connections	Examining content and structure
Grade 4			
Target	60	15	25
Actual	61	17	22
Grade 8			
Target	55	15	30
Actual	56	18	26

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.

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 Mathematics* (1989). This project resulted in a "content-by-ability" matrix design (National Assessment of Educational Progress 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 (NAGB 1996). Figure A-2 describes the five content areas that constitute the NAEP mathematics assessment. The questions designed to test the various content areas at each grade generally reflect the expectations normally associated with instruction at that level.

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, and an understanding of the meaning of addition, subtraction, multiplication, and division. At grade 8, number sense is extended to include positive and negative numbers, as well as properties and operations involving whole numbers, fractions, decimals, integers, and rational numbers.
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. At grade 8, this content area includes these measurement concepts, but the focus shifts to more complex measurement problems that involve volume or surface area or that require students to combine shapes and to translate and apply measures. Eighth-grade students also solve problems involving proportional thinking (such as scale drawing or map reading) and do applications that involve the use of complex measurement formulas.
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. At grade 8, students are asked to expand their understanding to include properties of angles and polygons. They are also asked to apply reasoning skills to make and validate conjectures about transformations and combinations of shapes.
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. At grade 8, students are asked to analyze statistical claims and to design experiments, and they are asked to use simulations to model real-world situations. This content area focuses on eighth graders' basic understanding of sampling, their ability to make predictions based on experiments or data, and their ability to use some formal terminology related to probability, data analysis, and statistics.
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. At grade 8, the assessment includes more algebraic notation, stressing the meaning of variables and an informal understanding of the use of symbolic representations in problem-solving contexts. Students are asked to use variables to represent a rule underlying a pattern. Eighth graders are asked to demonstrate a beginning understanding of equations and functions and the ability to solve simple equations and inequalities.

SOURCE: *Mathematics Framework for the 2003 National Assessment of Educational Progress* (NAEP), National Assessment Governing Board.

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 distributions for content areas as specified in the framework for grades 4 and 8 are 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 NAEP mathematics items, by content area, grades 4 and 8: 2003

	Percentage of items	
Content area	Grade 4	Grade 8
Number sense, properties, and operations	40	25
Measurement	20	15
Geometry and spatial sense	15	20
Data analysis, statistics, and probability	10	15
Algebra and functions	15	25

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

Common design features of the reading and mathematics assessments

Each student who participated in the 2003 NAEP assessment received a booklet containing three or four sections: a set of general background questions, a set of subject-specific background questions, and either two sets of cognitive questions in mathematics or two 25-minute sections or one 50-minute section in reading (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. The grade 8 mathematics assessment consisted of 10 cognitive blocks, and the reading assessment consisted of twelve 25-minute blocks and one 50-minute block. 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

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.

Sample Design

The results presented in this report are based on nationally representative probability samples of fourth- and eighth-grade students. The samples consisted of public school students assessed in each state, and an additional sample of private school students. Table A-4 contains the target populations and sample sizes for the 2003 fourth- and eighth-grade reading and mathematics assessments.

The samples were selected based on a two-stage sample design. In the first stage, schools were selected from stratified frames within participating states. In the second stage, students were selected from within schools.

The rounded sample sizes presented in the table are likely to differ slightly from the number of students assessed presented in the following section as a result of exclusion and absentees.

Table A-4. Student sample size and target population for NAEP reading and mathematics assessments, by type of school, grades 4 and 8: 2003

	Reading		Mathematics	
Type of school	Student sample size	Target population	Student sample size	Target population
Grade 4				
Public	191,400	3,609,000	191,400	3,603,000
Private	7,500	373,000	4,700	378,000
Catholic	3,700	194,000	2,300	195,000
Lutheran	900	20,000	600	19,000
Conservative Christian	1,000	58,000	700	63,000
Grade 8				
Public	155,000	3,579,000	153,500	3,575,000
Private	8,300	354,000	5,100	360,000
Catholic	4,000	189,000	2,500	193,000
Lutheran	1,000	16,000	600	16,000
Conservative Christian	1,100	45,000	700	47,000

NOTE: The number of students sampled for the combined private total includes students in the "other private school" category, which is not listed as a separate private category in the table. Student sample sizes are rounded to the nearest hundred, and target populations are rounded to the nearest thousand.

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 Private School Study.

Participation of schools and students in the NAEP samples

Tables A-5 and A-6 provide a summary of the school and student participation rates for the 2003 reading and mathematics assessment samples at grades 4 and 8. Participation rates are presented for public and private schools. One student and four different school participation rates are presented.

The first rate is a student-centered, weighted percentage of schools participating in the assessment, before substitution of demographically similar schools.[1] This rate is based only on the schools that were initially selected for the assessment. The numerator of this rate is the estimated number of students represented by the initially selected schools that participated in the assessment. The denominator is the estimated number of students represented by the initially selected schools that had eligible students enrolled.

The second school participation rate is a student-centered, weighted participation rate after substitution. The numerator of this rate is the estimated number of students represented by the participating schools, whether originally selected or selected as a substitute for a school that chose not to participate. The denominator is the estimated number of students represented by the initially selected schools that had eligible students enrolled. (This is the same as that for the weighted participation rate for the sample of schools before substitution.) Because of the common denominators, the weighted participation rate after substitution is at least as great as the weighted participation rate before substitution.

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

The fourth school participation rate is a school-centered, weighted participation rate after substitution. The numerator is the estimated number of schools represented by the participating schools, whether originally selected or selected as a substitute for a school that did not participate. The denominator is the estimated number of schools represented by the initially 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 tables A-5 and A-6 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.

[1] The initial base sampling weights were used in weighting the percentages of participating schools and students. An attempt was made to preselect one substitute school for each sampled public school and one for each sampled private school. To minimize bias, a substitute school resembled the original selection as much as possible in affiliation, type of location, estimated number of grade-eligible students, and minority composition.

Table A-5. Percentage of schools and students participating in NAEP reading and mathematics assessments, by type of school, grade 4: 2003

Type of school	School participation: Student-centered weighted percentage, Before substitution	School participation: Student-centered weighted percentage, After substitution	School participation: School-centered weighted percentage, Before substitution	School participation: School-centered weighted percentage, After substitution	School participation: Number of schools participating after substitution	Student participation: Student weighted percentage	Student participation: Number of students assessed
Reading							
Public	100	100	100	100	6,908	94	179,013
Private	79	80	74	76	542	95	7,488
Catholic	91	91	–	–	215	95	3,659
Lutheran	89	90	–	–	90	96	920
Conservative Christian	68	69	–	–	79	94	957
Mathematics							
Public	100	100	100	100	6,914	94	184,325
Private	79	80	74	76	539	95	4,718
Catholic	91	91	–	–	216	94	2,285
Lutheran	89	90	–	–	88	95	555
Conservative Christian	68	69	–	–	78	97	651

– Not available.
NOTE: Schools and students participating in the combined private school category include those in the "other private school" category, which is not listed separately in the table.
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 Private School Study.

Table A-6. Percentage of schools and students participating in NAEP reading and mathematics assessments, by type of school, grade 8: 2003

Type of school	School participation: Student-centered weighted percentage, Before substitution	School participation: Student-centered weighted percentage, After substitution	School participation: School-centered weighted percentage, Before substitution	School participation: School-centered weighted percentage, After substitution	School participation: Number of schools participating after substitution	Student participation: Student weighted percentage	Student participation: Number of students assessed
Reading							
Public	100	100	100	100	5,531	91	146,351
Private	74	76	75	78	568	94	8,324
Catholic	85	88	–	–	224	94	4,018
Lutheran	94	94	–	–	101	95	1,005
Conservative Christian	75	76	–	–	92	93	1,071
Mathematics							
Public	100	100	100	100	5,527	91	147,600
Private	74	76	75	78	558	95	5,073
Catholic	85	88	–	–	224	95	2,463
Lutheran	94	94	–	–	96	96	605
Conservative Christian	75	76	–	–	90	95	659

– Not available.
NOTE: Schools and students participating in the combined private school category include those in the "other private school" category, which is not listed separately in the table.
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 Private School 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 who 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[2] because of a disability, and/or was identified as ELL; all other students were asked to participate in the assessment under standard conditions.

The percentages of students identified as SD and/or ELL in the 2003 public and private school samples are presented in table A-7 for grade 4 and A-8 for grade 8. The tables also include the percentage of students *excluded*, the percentage of SD and/or ELL students *assessed*, the percentage *assessed without accommodations*, and the percentage *assessed with accommodations*.

[2] 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-7. Students with disabilities and/or English language learners identified, excluded, and assessed in NAEP reading and mathematics assessments, by type of school, grade 4: 2003

Students' status	Weighted percentage of students sampled	
	Public	Private
Reading		
SD and/or ELL		
Identified	22	4
Excluded	6	1
Assessed	16	3
Without accommodations	10	1
With accommodations	5	2
SD		
Identified	14	3
Excluded	5	1
Assessed	9	3
Without accommodations	4	1
With accommodations	5	2
ELL		
Identified	10	1
Excluded	2	#
Assessed	8	#
Without accommodations	7	#
With accommodations	1	#
Mathematics		
SD and/or ELL		
Identified	22	4
Excluded	4	#
Assessed	18	4
Without accommodations	10	1
With accommodations	8	2
SD		
Identified	14	4
Excluded	3	#
Assessed	11	3
Without accommodations	4	1
With accommodations	7	2
ELL		
Identified	11	1
Excluded	1	#
Assessed	9	#
Without accommodations	7	#
With accommodations	2	#

Rounds to zero.
NOTE: SD = Students with disabilities. ELL = English language learners. Detail may not sum to totals because of rounding. 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 Private School Study.

Table A-8. Students with disabilities and/or English language learners identified, excluded, and assessed in NAEP reading and mathematics assessments, by type of school, grade 8: 2003

Students' status	Weighted percentage of students sampled	
	Public	Private
Reading		
SD and/or ELL		
Identified	19	3
Excluded	5	#
Assessed	13	3
Without accommodations	8	1
With accommodations	5	2
SD		
Identified	14	3
Excluded	4	#
Assessed	10	3
Without accommodations	5	1
With accommodations	5	2
ELL		
Identified	6	#
Excluded	2	#
Assessed	5	#
Without accommodations	4	#
With accommodations	1	#
Mathematics		
SD and/or ELL		
Identified	19	3
Excluded	4	#
Assessed	15	3
Without accommodations	8	1
With accommodations	7	2
SD		
Identified	14	3
Excluded	3	#
Assessed	11	3
Without accommodations	5	1
With accommodations	6	2
ELL		
Identified	6	#
Excluded	1	#
Assessed	5	#
Without accommodations	4	#
With accommodations	1	#

Rounds to zero.
NOTE: SD = Students with disabilities. ELL = English language learners. Detail may not sum to totals because of rounding. 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 Private School Study.

Tables A-9 and A-10 display the percentages of SD and/or 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.

Table A-9. Percentage of students with disabilities and/or English language learners assessed with accommodations in NAEP reading and mathematics assessments, by type of primary accommodation and type of school, grade 4: 2003

Type of accommodation	Weighted percentage of assessed students	
	Public	Private
Reading		
Large-print book	0.05	0.01
Extended time	1.28	0.72
Small group	4.07	0.90
One-on-one	0.16	0.12
Scribe/computer	0.13	0.01
Other	0.08	#
Mathematics		
Bilingual book	0.85	#
Large-print book	0.06	#
Extended time	0.95	0.86
Read aloud	0.59	0.16
Small group	5.60	1.13
One-on-one	0.33	0.18
Scribe/computer	0.19	0.02
Other	0.08	#

Rounds to zero.
NOTE: The percentages in the table reflect only the primary accommodation provided.
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 Private School Study.

Table A-10. Percentage of students with disabilities and/or English language learners assessed with accommodations in NAEP reading and mathematics assessments, by type of primary accommodation and type of school, grade 8: 2003

Type of accommodation	Weighted percentage of assessed students	
	Public	Private
Reading		
Large-print book	0.02	#
Extended time	1.74	1.20
Small group	3.64	0.37
One-on-one	0.06	0.06
Scribe/computer	0.07	0.03
Other	0.06	#
Mathematics		
Bilingual book	0.29	#
Large-print book	0.03	#
Extended time	1.55	1.37
Read aloud	0.32	0.03
Small group	4.55	0.40
One-on-one	0.15	0.11
Scribe/computer	0.07	0.02
Other	0.07	0.05

Rounds to zero.
NOTE: The percentages in the table reflect only the primary accommodation provided.
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 Private School Study.

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 (ETS). 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.

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 both grades 4 and 8. 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 both grades 4 and 8.

Summary Data

The hierarchical linear modeling (HLM) program requires that the input data be organized in a summary data file. Separately for the reading and mathematics assessments at each grade, appropriate summary data files were generated for the following samples:

- public and all private schools at grades 4 and 8; and
- public, Catholic, and Lutheran schools at grade 4, and public, Catholic, Lutheran, and Conservative Christian schools at grade 8.

This resulted in eight 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 are 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.

Use of Sampling Weights

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.

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 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 analysis comparing all private schools to all public schools, 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.

This analysis employed HLM6 (Raudenbush et al. 2004), 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 et al. 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.[3]

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 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.

Variance Estimation and Nonsampling Errors

To account for measurement error, NAEP produces plausible values rather than single scores for each assessed student. Variance estimates for statistics incorporate an estimate of measurement error that is based on the differences among the plausible values for each student. For further details, see Allen, Donoghue, and Schoeps (2001) and Raudenbush et al. (2004).

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 instrumenta-

[3] 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.

tion 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 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 analysis have uncertainty associated with them because of measurement and sampling error. The uncertainty of an estimate is reported as a standard error. The results of the HLM analyses are reported in the main text 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.

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 grade 4 reading, the standard deviation among school means is $\sqrt{325}$ = 18.0 (see table 2-3). The corresponding effect size is 14.7/18.0 = 0.82, a value twice as large as the 0.41 presented in the main text. 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 for student groups defined by variables on the student, teacher, and school administrator questionnaires. There were 9 student-level and 13 school-level variables used in the various HLM analyses. A number of these variables were 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. Descriptions of the variables used in the HLM analyses are presented in the following sections.

Student-level variables

Nine student-level variables were used in the HLM analysis: gender, race/ethnicity, whether students had a disability or were English language learners, computer in the home, eligibility for free/reduced-price school lunch, participation in the Title I program, number of books in the home, and number of absences.

Gender: Results are available for male and female students.

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 groups: 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 (SD): Students who had an IEP or were protected under Section 504 of the Rehabilitation Act of 1973.

English language learners (ELL): All students identified based on school records as receiving 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: Students were asked whether 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 U.S. 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 2002–03 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.

Student participates in the Title I program: Title I is a U.S. Department of Education program that provides financial assistance to local education agencies for eligible public and private schools with high percentages of poor students. Although private schools are less likely than public schools to participate in the program, local education agencies are required to provide comparable Title I services for both types of schools. Based on available school records, students are classified as either currently participating in a Title I program, receiving Title I services, or not receiving such services. The classification applies only to the school year when the assessment is administered (i.e., the 2002–03 school year) and is not based on participation in previous years. If the school does not offer any Title I programs or services, all students in that school would be classified as not participating.

Number of books in the home: Students 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 together along 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. Teaching experience was the mean of the experience of teachers in that school. 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 or private) was substituted.

Teachers whose students participated in the NAEP assessment at grade 8 were asked to indicate the number of years they taught reading, writing, language arts, or mathematics in grades 6 through 12.

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 two different ways. In the first variable from the school questionnaire, 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 second 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 groups: The percentage of students by racial/ethnic categories was based on information provided by the schools and maintained by Westat, the statistical agency 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, it 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 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 analyses 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.

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.

Percentage of students in the Title I program: The percentage of students in the Title I program was based on aggregated data from among the students assessed.

School size: School size was based on the number of students currently enrolled as reported in the school questionnaire.

Figure A-3. States and the District of Columbia 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. Department of Commerce Economics and Statistics Administration.

Appendix B

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. Of course, with real data, homogeneity of variance can only hold approximately. If the variances vary randomly over schools, then the impact on the estimation of the level 2 coefficients is slight. However, if variance heterogeneity is associated with either level 1 or level 2 covariates, then the impact may be more serious (Raudenbush and Bryk 2002, p. 263).

The tests for homogeneity of variance in the reading data for grades 4 and 8 are presented and discussed. Tests for heterogeneity of variance were performed on the mathematics data from the same grades with similar results. As a result, conclusions drawn from the discussion of the reading results will be assumed to be valid for mathematics, and separate results for the mathematics data will not be presented. 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 covariates $X_1, \ldots, X_p$;

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

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 a common variance σ^2;

W_{1j} is an indicator of the school type (public or private) 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 private and public 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 in grades 4 and 8. The results for grade 4 are displayed in table B-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 B-1. Test for homogeneity of level 1 variance in mean NAEP reading scores, grade 4: 2003

Chi-square statistic	= 40104
Number of degrees of freedom	= 4947
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 Private School Study.

For the total group of schools (public and private), the empirical distribution of level 1 residual variances is given in figure B-1. The 18 highest variances were considered outliers in these analyses. Figure B-2 displays the same variances plotted against school size. The smaller schools have residual variances that are very variable, indicated by the broad scatter of points on the left side of the figure. 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 fewer. Based on standard statistical theory, smaller schools would be expected to be more variable than larger schools.

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

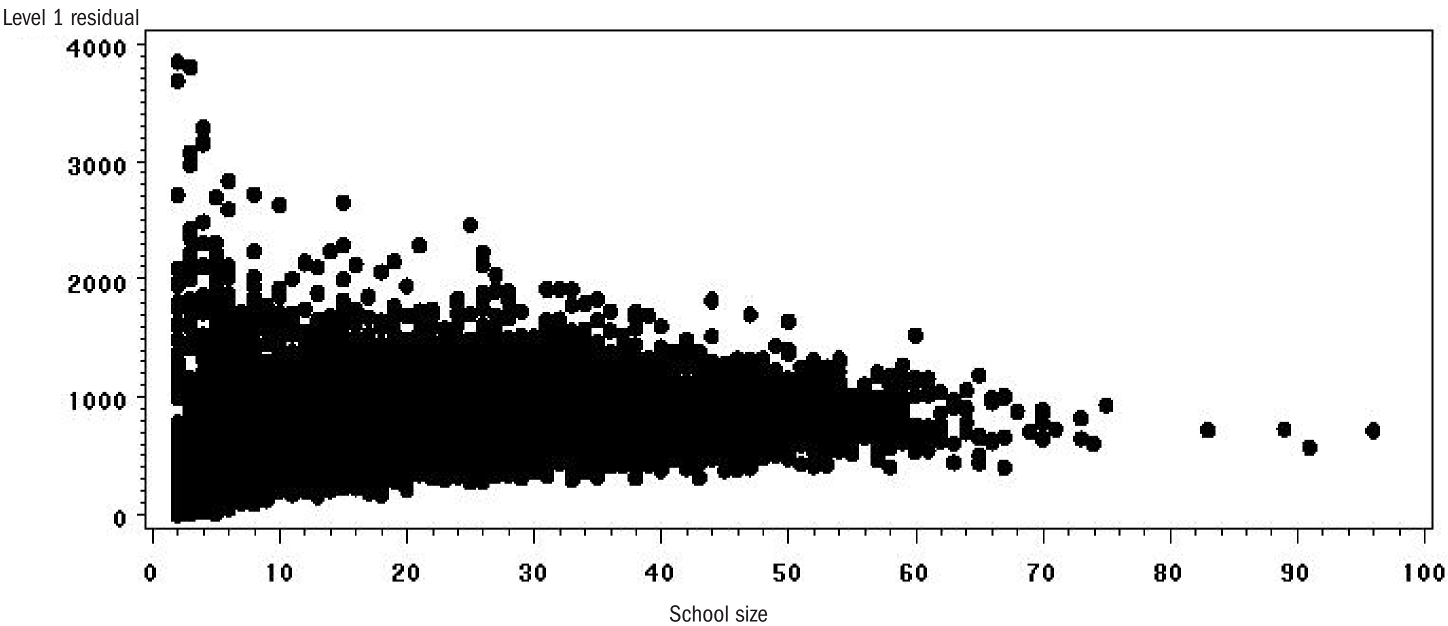
```
Level 1 residual                                                                                   Frequency
   75 |***                                                                                                68
  225 |*******                                                                                           143
  375 |************************                                                                          474
  525 |***************************************************                                              1028
  675 |******************************************************************************                   1558
  825 |***************************************************************************                      1508
  975 |****************************************************                                             1035
 1125 |*********************************                                                                 659
 1275 |******************                                                                                350
 1425 |**********                                                                                        193
 1575 |*****                                                                                              93
 1725 |***                                                                                                66
 1875 |**                                                                                                 30
 2025 |*                                                                                                  15
 2175 |*                                                                                                  18
 2325 |                                                                                                    5
 2475 |                                                                                                    3
 2625 |                                                                                                    4
 2775 |                                                                                                    3
 2925 |                                                                                                    1
 3075 |                                                                                                    2
 3225 |                                                                                                    2
 3375 |                                                                                                    0
 3525 |                                                                                                    0
 3675 |                                                                                                    1
 3825 |                                                                                                    2
      S-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-----
             100     200     300     400     500     600     700     800     900    1000    1100    1200    1300    1400    1500

                                                Frequency
```

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

The scatterplot of level 1 residual variances for public schools only is not displayed. Since these schools constitute the vast majority of all schools in the study, the plot is almost identical to that in figure B-2, so presenting that plot would not further illuminate the issue.

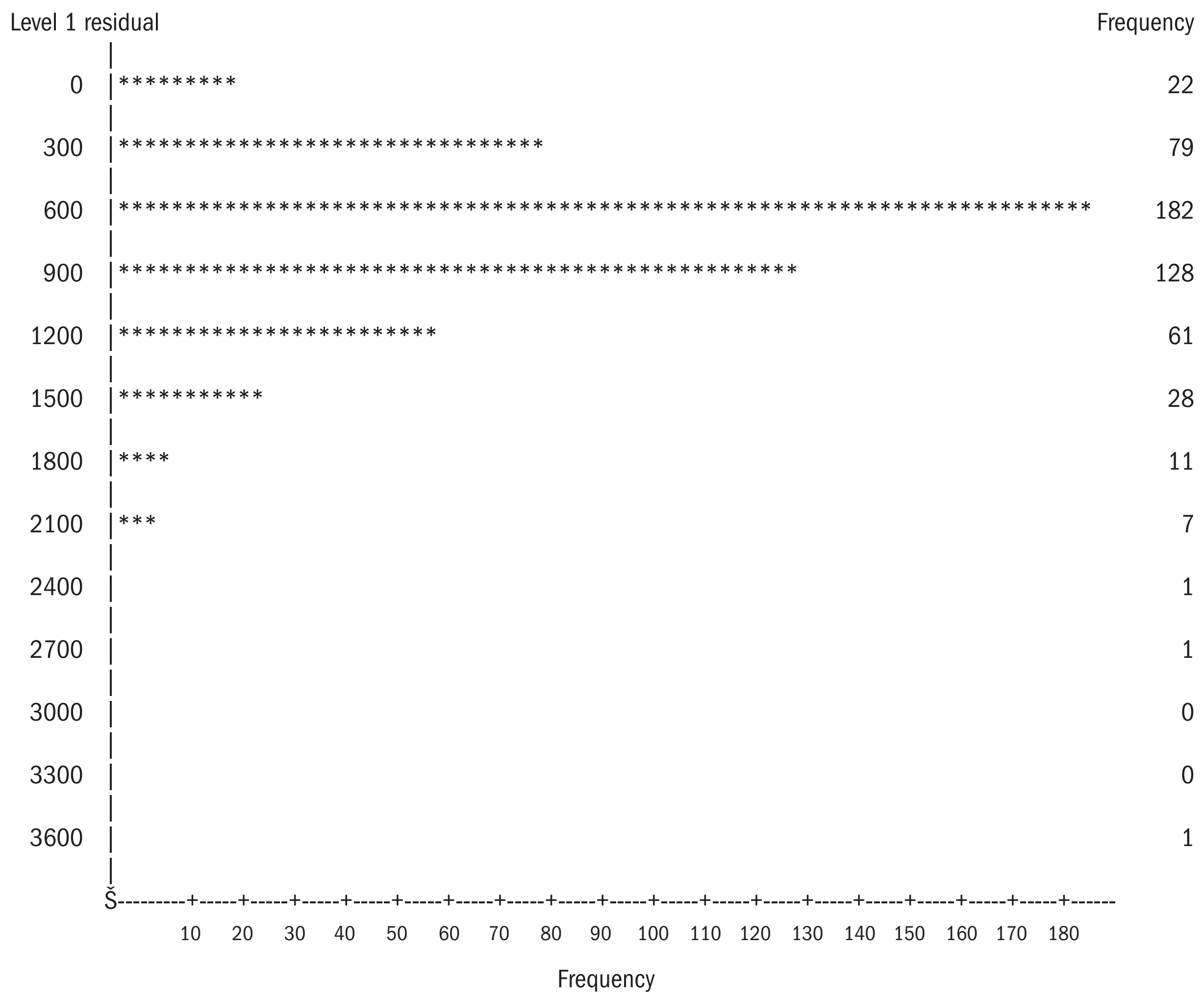
Figure B-2. Plot of school residual variance from HLM analysis against school size for all schools, NAEP reading scores, grade 4: 2003



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

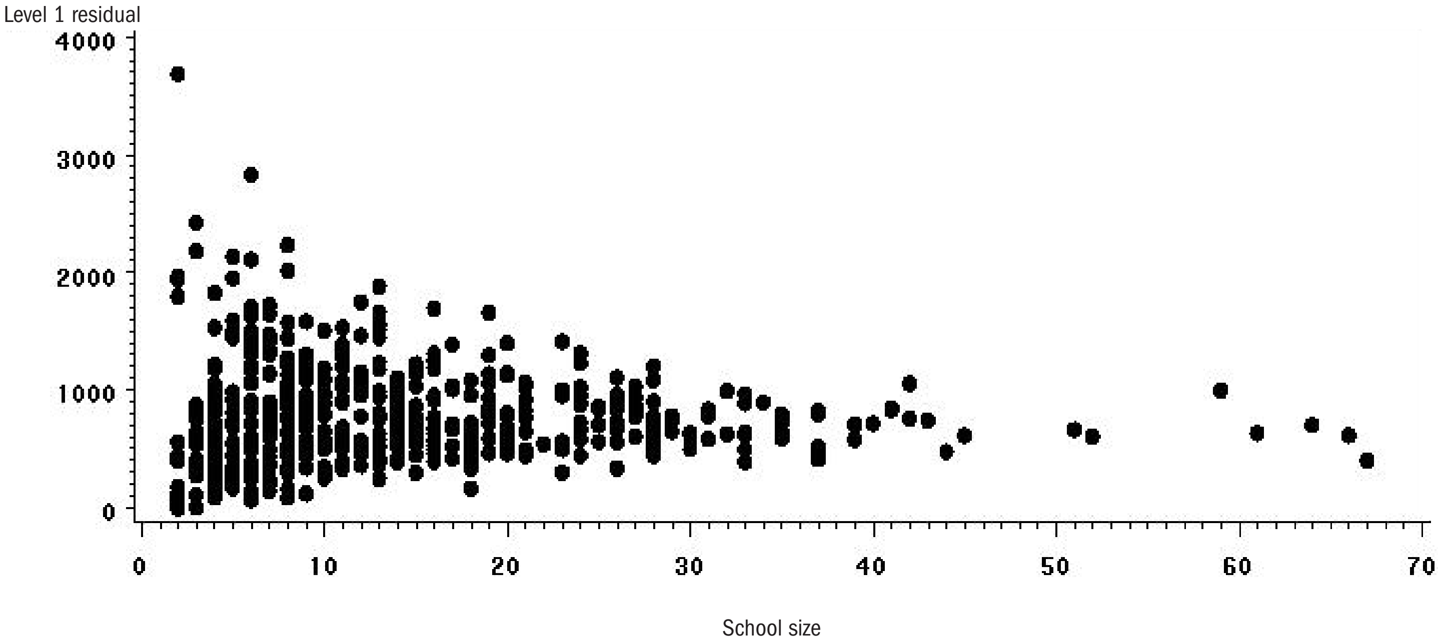
Although there is a strong association between heterogeneity of variance and school size, it is important to examine whether there is also a relationship between heterogeneity of variance and school type. Figure B-3 displays the empirical distribution of level 1 residual variances for private schools only. Figure B-4 displays the residual variances for private schools plotted against school size. As in figure B-2, 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 0 and 10. There is a suggestion that, conditional on school sample size, variance heterogeneity is slightly greater in public schools than in private schools.

Figure B-3. Histogram of residual variances from HLM analysis for private schools, NAEP reading scores, grade 4: 2003



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

Figure B-4. Plot of school residual variance from HLM analysis against school size for private schools, NAEP reading scores, grade 4: 2003



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

For comparison with the grade 4 analysis, table B-2 gives the results of the HLM heterogeneity test for reading results at grade 8. Once again the *p* value of .00 leads to rejection of the null hypothesis that the level 1 variances are homogeneous across schools.

Table B-2. Test for homogeneity of level 1 variance in mean NAEP reading scores, grade 8: 2003

Chi-square statistic	= 28967
Number of degrees of freedom	= 4534
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 Private School Study.

Figure B-5 gives the distribution of level 1 residual variances for the total group of schools. In this case, the largest 15 variances were considered to be outliers. In figure B-6, the same residual variances are plotted against school size. As with grade 4 schools, variance appears to be a function of sample size, with the outlying values associated with schools with sample sizes of 15 or less.

As before, it is important to determine whether type of school is a confounding factor contributing to school heterogeneity. Figure B-7 shows the empirical distribution of level 1 residual variances for private schools in grade 8 reading. In this case, the range of values is about the same as for the set of all schools, with six outliers. The scatterplot of variances plotted against school size is given in figure B-8, where it is clear that the outlier values occur for schools with sample sizes between 0 and 10.

In reading, for both grades 4 and 8, there is considerable heterogeneity in residual variance. Greater dispersion is strongly associated with school sample size (i.e., a function of sampling variability) and only weakly associated with school type for fixed school sample size. As Raudenbush and Bryk (2002) point out, variance heterogeneity could also be due to misspecifying the model. In this case, there are at least two possible causes for such misspecification: omitted predictors and treating level 1 regression coefficients as fixed.

Figure B-5. Histogram of residual variances from HLM analysis for all schools, NAEP reading scores, grade 8: 2003

```
Level 1 residual                                                                             Frequency
      0 |**                                                                                         46
    200 |*********                                                                                 214
    400 |***************************************                                                   986
    600 |******************************************************************************           1953
    800 |****************************************************************                         1610
   1000 |*****************************                                                             733
   1200 |***********                                                                               274
   1400 |****                                                                                       98
   1600 |*                                                                                          34
   1800 |*                                                                                          14
   2000 |                                                                                            7
   2200 |                                                                                            7
   2400 |                                                                                            2
   2600 |                                                                                            0
   2800 |                                                                                            6
   3000 |                                                                                            2
   3200 |                                                                                            0
   3400 |                                                                                            0
   3600 |                                                                                            2
   3800 |                                                                                            1
   4000 |                                                                                            1
   4200 |                                                                                            0
   4400 |                                                                                            1
   4600 |                                                                                            1
   4800 |                                                                                            1
        Š-- -----------+------------+------------+------------+------------+------------+------------+------------+------------+--------
                      200          400          600          800         1000         1200         1400         1600         1800

                                                  Frequency
```

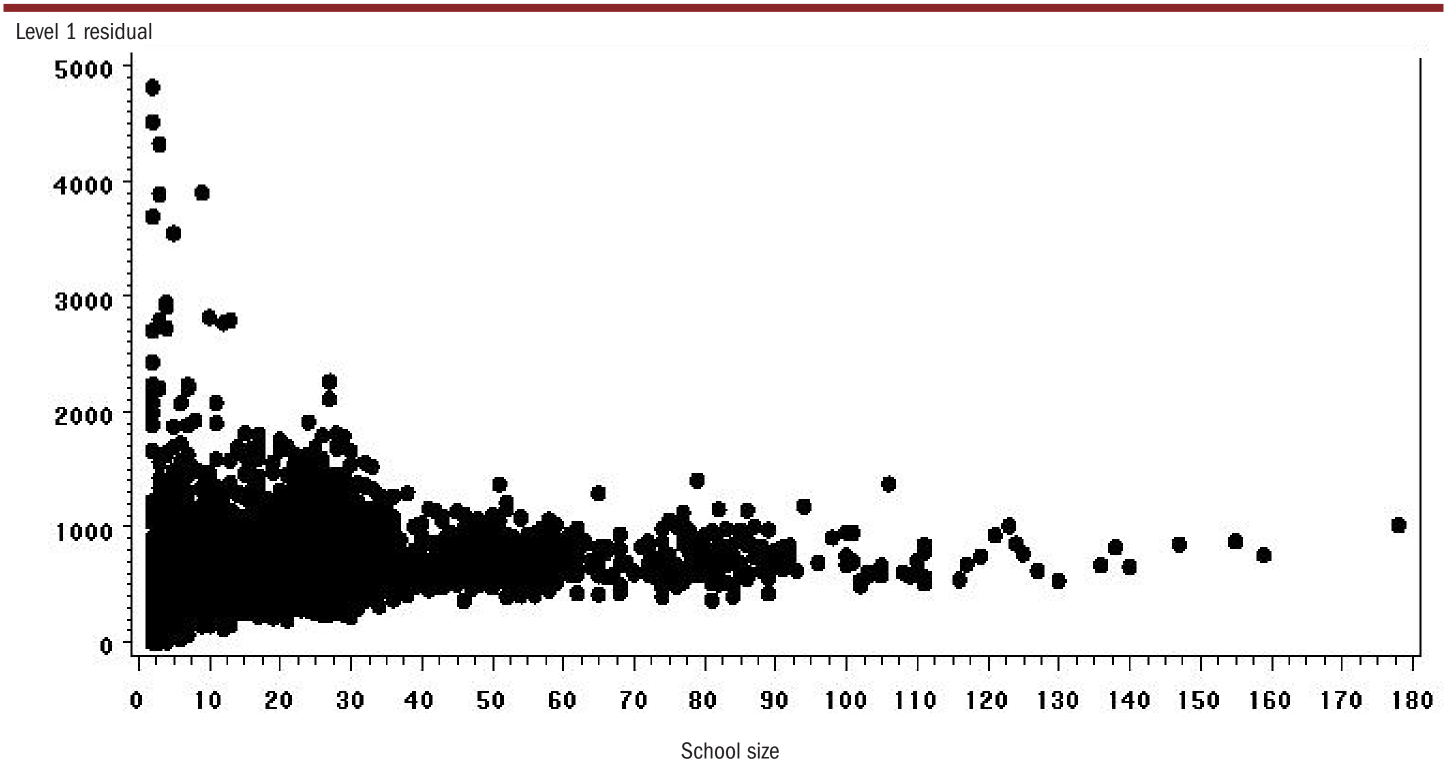
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Private School Study.

With respect to the former, it has already been noted that the absence of an indicator of prior academic achievement can result in biased estimates. It is also possible that the omission of such an indicator can induce apparent heterogeneity in residual variances, with a corresponding impact on the bias of estimates of level 2 parameters, such as school type.

In a set of exploratory analyses, the null hypothesis of fixed regression coefficients was accepted, and, consequently, the simple HLMs described at the beginning of this appendix were adopted. It is still possible, of course, that there is substantial variation in slopes across schools and that this heterogeneity could also contribute to bias in the estimates of level 2 parameters.

In sum, the difficulties entailed in establishing the approximate unbiasedness of the estimate of the school-type effect are inherent in the design of the NAEP survey and the nature of the data collected by NAEP. Ultimately, these issues must be resolved through the analysis of the richer databases increasingly available at the state level.

Figure B-6. Plot of school residual variance from HLM analysis against school size for all schools, NAEP reading scores, grade 8: 2003



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

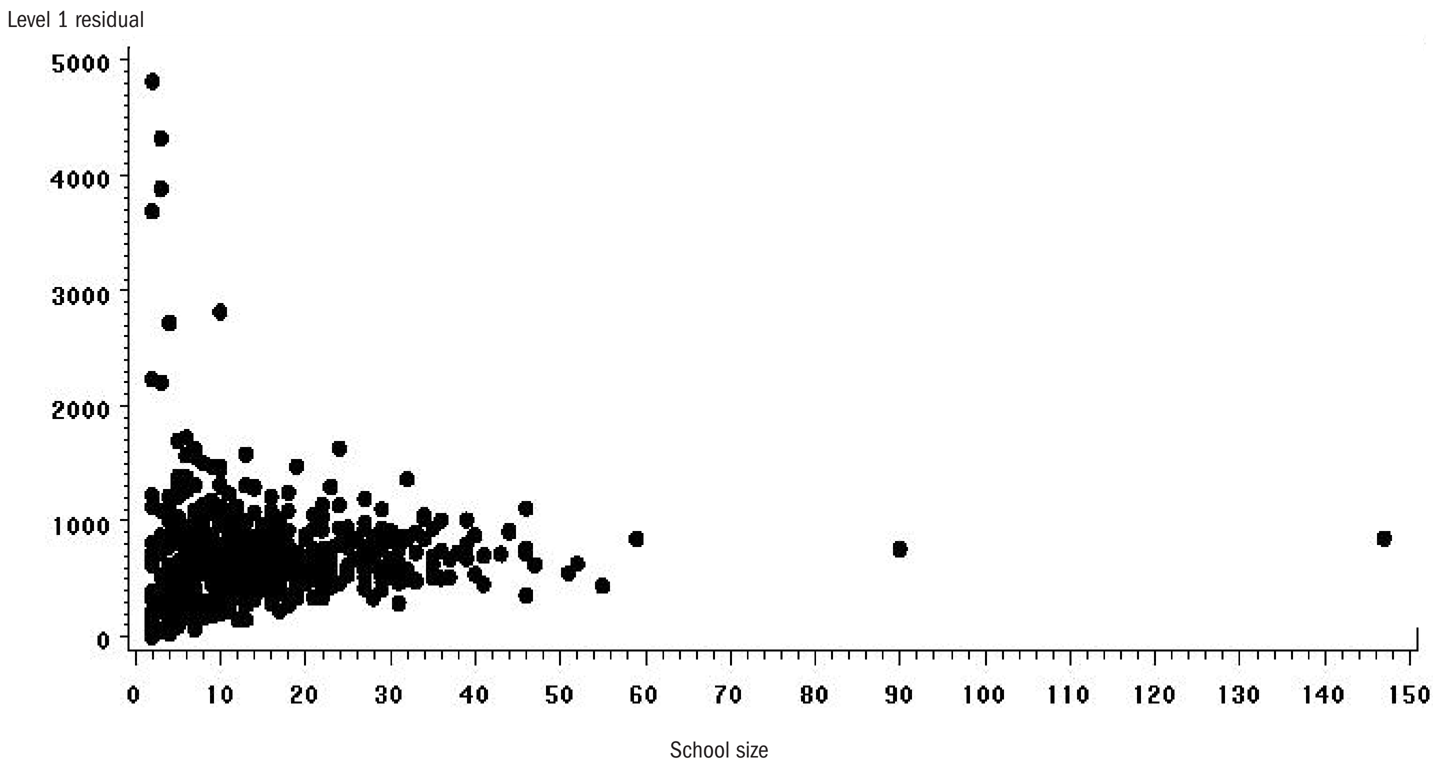
Figure B-7. Histogram of residual variances from HLM analysis for private schools, NAEP reading scores, grade 8: 2003

```
Level 1 residual                                                     Frequency
          |
     0    |*******                                                      37
          |
   400    |*****************************************                   204
          |
   800    |**********************************************              229
          |
  1200    |***********                                                  57
          |
  1600    |**                                                           12
          |
  2000    |                                                              0
          |
  2400    |                                                              2
          |
  2800    |                                                              2
          |
  3200    |                                                              0
          |
  3600    |                                                              1
          |
  4000    |                                                              1
          |
  4400    |                                                              1
          |
  4800    |                                                              1
          |
          Š --------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
                   20    40    60    80   100   120   140   160   180   200   220

                                  Frequency
```

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

Figure B-8. Plot of school residual variance from HLM analysis against school size for private schools, NAEP reading scores, grade 8: 2003



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

THIS PAGE INTENTIONALLY LEFT BLANK.

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