DOCUMENT RESUME

ED 441 039 TM 031 205

AUTHOR	McLaughlin, Donald; Drori, Gili
TITLE	School-Level Correlates of Academic Achievement: Student Assessment Scores in SASS Public Schools. Research and Development Report.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-2000-303
PUB DATE	2000-05-00
NOTE	75p.
AVAILABLE FROM	ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll free); Web site: http://nces.ed.gov.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Achievement; Class Size; Elementary Secondary Education; *Institutional Characteristics; Mathematics Tests; National Surveys; *Public Schools; Reading Tests; School Size; *State Programs; Structural Equation Models; Teacher Attitudes; *Testing Programs
IDENTIFIERS	*Schools and Staffing Survey (NCES)

ABSTRACT

The Schools and Staffing Survey (SASS), conducted by the National Center for Education Statistics (NCES), offers the most comprehensive picture available of the education system in the United States. The aim of this study is to show the potential value of a linkage between SASS and data on student academic achievement. To achieve this aim, the study matched 1993-1994 SASS data with state reading and mathematics assessment scores for public schools in 20 states, adjusting for between-state differences in achievement scales by using the state-by-state component of the National Assessment of Educational Progress (NAEP). By combining data sources, school-level correlates of student achievement in a broad sample of U.S. schools were identified. The relationship between a variety of SASS school-level responses and average student assessment scores at the school level was modeled. This model, which was applied separately to public elementary (n=1,123), middle (n=496), and secondary (n=595) schools in the 20 states, related a variety of factors to average achievement. The general correlation of the organization of a school with academic performance can be partitioned into factors of school size, average class size, normative cohesion of staff, and teachers' sense of control in school affairs, each of which, other than school size, has multiple measures in the SASS. Three appendixes contain results from structural equation modeling, factors associated with school climate and achievement, and correlations between state reading assessment and NAEP reading school means. (Contains 66 references and 17 tables.) (SLD)

NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report   May 2000

School-level Correlates of Academic Achievement

Student Assessment Scores in SASS Public Schools

SASS

U.S. Department of Education
Office of Educational Research and Improvement   NCES 2000-303

Customer Survey of OERI Publication Users

To help us improve future editions of this publication and give you better customer service, we would appreciate your comments on this survey form. Please check the appropriate box(es) below for each question. Responses will be kept completely confidential. You may return the survey by mail or FAX. It can be folded and taped closed to allow mailing to the address listed on the reverse side of this form, or it can be returned by FAX to 202–219–1321. **Many thanks for your customer feedback—it is very important to us!**

1a. Name of publication

1b. Publication number

1c. Author name

2. How did you receive a copy of this publication?

- ❑ Bought it
- ❑ Borrowed it
- ❑ Mailing list membership
- ❑ Telephone request
- ❑ Internet request
- ❑ Other (please describe)

3. Was this publication easy to get?

❑ Very ❑ Somewhat ❑ Not at all

4. How did you find out about this and other OERI publications? (Check all that apply.)

- ❑ Conferences
- ❑ Journal articles
- ❑ Teacher/educator
- ❑ Professional associations
- ❑ Internet (WWW)
- ❑ Publication announcement
- ❑ Received in mail
- ❑ OERI staff contact

5. For what purposes did you use this OERI publication? (Check all that apply.)

- ❑ Planning
- ❑ Policy or legislation
- ❑ Administrative decisions
- ❑ Teaching, class material
- ❑ Research/analysis
- ❑ General information
- ❑ Writing news articles, TV or radio material
- ❑ Marketing, sales, or promotion
- ❑ Other (please describe)

6. Did the publication help you accomplish whatever you needed it for?

❑ Yes ❑ No ❑ Partially

7. What is your occupation?

❑ Parent ❑ Teacher ❑ Administrator
❑ Librarian ❑ Researcher ❑ Statistician
❑ Journalist/writer ❑ Policy Analyst ❑ Student
❑ Program Planner
❑ Other (please specify)

8. How could this OERI publication (or other OERI publications) better meet your needs? (Check all that apply.)

- ❑ More important topics in education
- ❑ More timely release of data
- ❑ More text introductions to each section
- ❑ More research statistics
- ❑ Shorter reports (less than 10 pages)
- ❑ Other (please describe)

9. Overall, how satisfied are you with this product?	**Very Satisfied** ❑	**Satisfied** ❑	**Dissatisfied** ❑
a. Comprehensiveness of information	❑	❑	❑
b. Clarity of writing (readability, interpretability)	❑	❑	❑
c. Clarity of presentation (e.g., tables, charts)	❑	❑	❑
d. Timeliness of information	❑	❑	❑
e. Accuracy of information	❑	❑	❑
f. Clarity of technical notes	❑	❑	❑
g. Usefulness of resources and bibliography	❑	❑	❑
h. Organization	❑	❑	❑
i. Length	❑	❑	❑
j. Format	❑	❑	❑

PAPERWORK BURDEN STATEMENT
Office of Educational Research and Improvement (OERI)
Publication Customer Survey

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 1880–0529. The time required to complete this information collection is estimated to average 10 minutes per response, [illegible] the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have any comments [illegible]g the accuracy of the time estimate(s), suggestions for improving this form, or concerns regarding the status of your individual submission of this form, write directly to: Room 204, Media and Information Services, OERI, U.S. Department of Education, 555 New Jersey Avenue NW, Washington, DC 20208–5570.

OERI Publication Customer Survey

Media and Information Services
555 New Jersey Avenue NW—Rm. 202
Washington DC 20208–5570

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 012935 WASHINGTON DC

POSTAGE WILL BE PAID BY US DEPARTMENT OF EDUCATION

V. Allen
Room 200
Media and Information Services
Office of Educational Research and Improvement
US Department of Education
555 New Jersey Avenue NW
Washington DC 20208–5570

Fold on line—TAPE CLOSED—DO NOT STAPLE

10. Do you have any suggestions regarding the content or format of future editions of this publication or other comments?

MIS 1999–6532

NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report May 2000

School-level Correlates of Academic Achievement

Student Assessment Scores in SASS Public Schools

Donald McLaughlin
Gili Drori
American Institutes for Research

Michael Ross, Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 2000-303

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Acting 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.

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
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

May 2000

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/index.asp***

Suggested Citation
U.S. Department of Education. National Center for Education Statistics. *School-level Correlates of Academic Achievement: Student Assessment Scores in SASS Public Schools,* NCES 2000–303, by Donald McLaughlin and Gili Drori. Project Officer, Michael Ross. Washington, DC: 2000.

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

or call toll free 1–877–4ED–Pubs.

Content Contact:
Michael Ross
(202) 219–1565

Acknowledgments

This is the final report of a project carried out by the authors for the National Center for Education Statistics, through the Education Statistics Services Institute. The authors wish to thank the project officer, Dr. Michael Ross, for his thoughtful suggestions and encouragement during the course of the work. We also wish to thank Steven Kaufman, Adam Gamoran, Robert Hauser, Valerie Lee, and Stephen Raudenbush for their thoughtful reviews of an earlier draft. Nevertheless, the conclusions expressed in this paper are solely those of the authors, and no endorsement of these conclusions by reviewers or by the Center should be inferred.

The foundation of this paper is a combination of survey data, including the Schools and Staffing Survey, the Common Core of Data, and the National Assessment of Education Statistics, all of which are conducted by the National Center for Education Statistics, and 20 state assessments, conducted by State Education Agencies. We appreciate the help provided by these agencies in providing the data for this project, and we appreciate the work of staff of the American Institutes for Research, including Mary Anne Arcilla, Grace Wu, Elizabeth Hartka, and Inna Shapotina, in putting the database together. Finally, we recognize the thousands of hours of time spent by the respondents to these surveys and assessments, who provided information that can be used to improve education policies and practices.

Don McLaughlin
Gili Drori

Contents

8

List of Tables

List of Figures

10

Introduction

The Schools and Staffing Survey (SASS), conducted by the National Center for Education Statistics (NCES), offers the most comprehensive picture available of the education system in the United States. Initiated in 1987–88 and repeated in 1990–91 and 1993–94, SASS consists of surveys of districts, schools, principals, and teachers associated with a national sample of schools. It offers information on such issues as policies, programs, services, staffing, and enrollment at both the district and school levels, as well as the principals' and teachers' background, training, experience, perceptions, and attitudes. Given the broad reach of SASS, it can speak to a variety of important educational research and policy questions. The value of SASS would be even greater however, if the relationship between these measures and the level of achievement in schools were known. As noted by others (Boruch and Terhanian 1996, Kaufman 1996), combining this survey information with data from other sources, SASS could more meaningfully inform debates over which factors relate to school effectiveness and could contribute to a broad-based evaluation of school improvement strategies.

The aim of this paper is to show the potential value of a linkage between SASS and data on student academic achievement. To achieve this aim, our approach is two-staged. First, we match the 1993–94 SASS data with state reading and mathematics assessment scores for public schools in 20 states, adjusting for between-state differences in achievement scales by using State NAEP (the state-by-state component of the National Assessment of Educational Progress). Second, by combining these data sources, we identify school-level correlates of student achievement in a broad sample of American public schools. We model the relationship between a variety of SASS school-level responses and average student assessment scores at the school level. Although such a model does not account for individual variation in achievement between students, it is useful for summarizing the relations between the general level of achievement in a school and characteristics of the school.

The model, which is applied separately to public elementary, middle, and secondary schools in 20 states, relates a variety of factors—school attributes, social background of the students, and teachers' characteristics—to the average achievement of students in a school. Of particular interest in this study are the relations among student achievement, average class size, and the school's behavioral climate. Although the data are cross-sectional with respect to time, the large sample size and broad set of measures included in the database create a stable foundation for exploring the relations among these measures. Associations among the measures, and how these associations are affected by the contexts in which schools operate, are examined through partial correlations, ordinary least squares linear regression, and structural equation modeling, both with and without the use of State NAEP to capture between-state covariation. Variations of results by type of school and by analytical method suggest possible explanations for conflicting results in the effective schools research literature (e.g., Purkey and Smith 1983).

Overall, we investigate relationships among these measures in three types of schools:

1,123 public elementary schools, 496 public middle schools, and 595 public high schools, in 20 states. Separate analyses by level are essential for this study, not only because different factors are related to achievement at different levels, but also because different achievement measures are used in each state at the different school levels.[1] The schools in the database developed are a representative sample of schools that enroll about 40 percent of the nation's public school students.

In our model, average student achievement in a school is related to student background factors, school organizational features, teachers' professional characteristics, and school climate. We regard these categories as related to each other, as well as to student achievement, reflecting interdependencies among social factors. Furthermore, the category of organizational features is further refined to focus on particular factors that correlate with the level of student achievement. The general correlation of the organization of a school with academic performance can be partitioned into factors of school size, average class size, normative cohesion of the staff, and teachers' sense of control in school affairs, each of which, other than school size, has multiple measures in SASS.

Combining SASS and State Assessment Data

The construction of this "linked" database, while technical in nature, is critical for the evaluation of the results of this report's analyses. The coverage of the range of educational contexts in the United States by the schools in the sample determines the extent to which inferences based on analyses of the database can be applied to education in the country. Most states currently collect state assessment data[2] on their public schools and thus offer state-specific information on school performance in terms of student test scores. Although many different assessment instruments are used across the states, they all aim to provide an indication of the reading and mathematics achievement levels of their schools, in comparison with other schools in the state. By transforming each school's score to a *z score* relative to other schools tested at the same grade in a state, there is potential for pooling analytical results across states to increase both power and generalizability.

While pooling information from individual states into a single database can add substantial power to analyses to identify school-level correlates of achievement in SASS schools,

[1] An additional sample of 347 combined-grade schools were matched, but are not included in this report because the relationships among measures in these schools proved more difficult to model. Also, an additional 64 SASS schools, for which no teacher data were obtained, were matched, but excluded from these analyses because many of the school measures included in the model are based on average responses of a sample of teachers at the school.

[2] In 1994–95, 45 states had a statewide assessment system; the remaining 5 states either did not have a statewide system at all or had temporarily suspended their programs (National Education Goals Panel 1996). In 1995–96, 46 states administered statewide assessments (Roeber, Bond, and Braskamp 1997). In 1996–97, 45 states administered statewide assessments (Roeber, Bond, and Connealy 1998). Some educational assessment is done in every state, and in most of the few states without statewide testing programs, most districts use nationally standardized tests for assessment.

it does not capture between-state sources of covariation with achievement. State policies, which are intended to provide equal opportunity for all students, and state demographics, which are in most cases less diverse than the nation as a whole, frequently limit the variation of education practices and outcomes in a state, so that within-state associations with achievement are attenuated. To make use of the full range of achievement variation between schools in the country, we also employ State NAEP data. To evaluate whether adjusting state assessment scores based on State NAEP data would add power to the analyses, we focused on states with statewide assessments which also participated in the 1994 State NAEP 4th-grade reading assessment. In each of those states, approximately 2,500 students in 100 schools took tests that have been carefully developed to assess a consensus framework of reading skills.

Although SASS includes both public and private schools, states generally collect assessment data only for public schools, and hence, the SASS-Student Achievement database, created for this report, is limited to information on public schools. In most cases, school-level state assessment scores are public information; and of the 39 states that participated in the 1994 State NAEP 4th-grade reading assessment, 23 provided the American Institutes for Research (AIR) with 1993–94 school-level assessment scores in conjunction with a study of the State NAEP sampling procedures. Scores from three of these states were not usable for this study, either because they were minimum competency tests with no correlation with NAEP, they were not on an easily readable medium, or test usage varied between districts. The 20 states whose state assessment scores are included are Alabama, California, Delaware, Florida, Georgia, Hawaii, Kentucky, Louisiana, Maine, Massachusetts, Michigan, Montana, New Hampshire, New York, Pennsylvania, Rhode Island, Tennessee, Texas, Utah, and Washington.

The linkage of these data to the SASS file required access to restricted information concerning the identification of SASS schools. NCES has established clear criteria for acceptable procedures for storing and using confidential information, and AIR has complied with these criteria. Although the linkage might be possible with information about schools' names and addresses, it was greatly facilitated by the use of an intermediate linkage of both SASS and state assessments through the Common Core of Data (CCD). The 1991–92 CCD file, which served as the sampling frame for SASS, identifies most of the 86,287 public schools in the country by both their federal and state identification codes.[3] In most cases, school records on state assessment files included the state's identification code, which enabled an automated matching procedure.[4]

These 20 states contained 3,785 of the 8,767 SASS public schools. Of these 3,785 SASS public schools, 2,916 had students enrolled in grades corresponding to the state assessment, and

[3] Information about CCD can be found on the NCES Web page: **http://nces.ed.gov/ccd.**

[4] A few of the schools were identified manually by matching their state, city, or zip code, either because the federal identification code was missing from the restricted SASS file or because the state identification code was not included on the assessment file. Details of the file development process can be found in Wu, Royal, and McLaughlin (1997).

2,627 were identified as having both SASS and state assessment information.[5] Of these, 66 had no teacher data, so the final file used for analysis contained 2,561 school records: 1,123 elementary schools, 496 middle schools, 595 secondary schools, and 347 combined-grade schools. The database includes at least 50 schools in each state and constitutes a broad sample of large and small, urban and rural, affluent and impoverished public schools.[6]

The coverage of the range of educational contexts in the United States by the schools in the sample determines the extent to which inferences based on analyses of the database can be expected to generalize to other schools in the country. Although SASS includes both public and private schools, state assessment data are collected for public schools only, and hence, the SASS student achievement subfile created for this report, is limited to information on public schools. The 2,561 public schools included in the study are only slightly different from the general population of American public schools. The grade level distribution of the schools included in our study differs from that of the entire range of SASS public schools, because many states only conduct assessments at the elementary and intermediate levels. On the other hand, SASS selected a smaller percentage of elementary than secondary schools, compensating for this by giving greater weight to each SASS elementary school in estimating national averages. Table 1 provides the comparison between the (unweighted) sample of public schools in this report and the general population of public schools, by two characteristics—school level and locale. It shows that the largest difference involves the inclusion of more combined-grade schools in the SASS-Student Achievement database (13.6 percent, as compared with 2.8 percent). Also, the percentage of schools in the database that are elementary schools is smaller than the percentage of elementary schools in the general population of public schools (43.8 percent, as compared with 60.9 percent). These large differences mean that to be meaningful, all analytical comparisons must be between schools serving the same overall grade levels. In fact, separate analyses are performed for elementary, middle, and secondary schools,[7] while results for combined-grade schools are not included in this report. Concerning locale, differences between the unweighted sample and the population of public schools in America are small. The largest differences between the sample and population are that more of the urban fringe schools in the sample are near mid-size cities,

[5] Of the other 289 SASS schools, 254 did not match with state assessment, 34 merged with state assessment files, but did not have both reading and mathematics scores, and one was found subsequently to have achievement scores that were outliers. The 34 SASS schools missing one of the scores included 24 in California, 2 in Florida, 3 in Maine, 1 in Montana, and 4 in Texas. In addition, 112 of the 254 non-matching schools were special, alternative, or vocational education schools, or schools that had an enrollment of fewer then 10 in the grade assessed; and 86 were not included in one state's (Pennsylvania's) assessment sample in 1993–94.

[6] As a measure of the success of the matching process, 2,662 of the 2,718 SASS public schools that were expected to have matching assessment scores were matched, for a match rate of 97.9 percent.

[7] State math and reading assessment data are used for typical grades in elementary, middle, and high schools. However, the NAEP adjustment for between-state variation at all three levels is based on grade 4 state means for reading and grades 4 and 8 state means for math. Between-state variation in achievement scores tends to be very similar across grade levels and subject matter, so such extrapolations are plausible. Whether these extrapolations reduce the potential to identify correlates of achievement in SASS is a question to be addressed in this study.

rather than large cities. Overall, the differences between the unweighted sample and the population were fairly small. Detailed comparisons on 50 SASS measures are presented by Wu, Royal, and McLaughlin (1997).

Table 1.—Unweighted counts and percentages of SASS-student achievement subfile schools and all American public schools, by grade level and locale

	SASS Achievement Database Schools		All Public Schools	
Elementary	1,123	43.8%	49,154	60.9%
Middle	496	19.4%	12,891	16.0%
Secondary	595	23.2%	15,711	19.5%
Combined grade	347	13.6%	2,235	2.8%
Ungraded	0	0	750	0.9%
Total:	**2,561**	**100%**	**80,741**	**100%**
Large central city	225	8.8%	7,800	9.7%
Mid-size central city	388	15.2%	11,384	14.1%
Urban fringe of large city	263	10.3%	11,733	14.5%
Urban fringe of mid-size city	356	13.9%	7,911	9.8%
Large town	62	2.4%	2,269	2.8%
Small town	577	22.5%	18,176	22.5%
Rural	690	26.9%	21,467	26.6%
Total:	**2,561**	**100%**	**80,740**	**100%**

NOTES: Locale, a SASS measure derived from CCD, is based on 1980 zip code methodology. Percentages of all public schools are based on the 1993–94 SASS data. One outlier school, not included in analyses of achievement, was included in this table.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and state reading and mathematics assessment scores for publics schools in 20 states.

A School-Level Measure of Student Achievement

Average student achievement scores constitute an important descriptive characteristic of public schools in America. Although they do not directly measure school performance, due to their dependence on factors exogenous to schools and out of the schools' control (such as student, family, and community characteristics), and although they have been criticized for a variety of technical considerations (see for example, Meyer 1997), they are still the most widely used indicator of the effectiveness of schools.

Each state selects particular grades for student achievement assessment and particular instruments to use in that assessment. Although there are exceptions, typical state assessments include one elementary grade, one middle school grade, and one high school grade. For example, elementary school assessments are typically administered in third or fourth grade. They are intended to give parents and educators a comparative picture of how well students in each school are acquiring the skills that society expects of them. Considered from a systems perspective, they measure both outcomes of processes in the school, as shaped by background characteristics, and inputs to the school choices made by teachers and parents and subsequent student learning. The important criterion for their use in educational research is that they be comparable across the schools included in a study, so that correlations between differences in processes and differences in scores can be interpreted.

The first step in rendering state assessment scores comparable is to compute each school's score as it relates to other schools' scores in the state at the same grade. That is, the (unweighted) mean score of the schools in the database, for the particular grade and state, is subtracted from each of the scores, to create a score with a mean of zero at each state and grade; and these scores are divided by the standard deviation of the school scores to create a score with a standard deviation of one. Using this measure, third-grade reading scores in one state, fourth-grade reading scores in another, and fifth-grade reading scores in a third are taken to be comparable achievement measures for the purpose of computing within-state correlations across elementary schools with factors such as average class size and school behavioral climate. All achievement differences between states are removed in analyses using this measure, so comparisons with school characteristics would need also to remove between-state variation in school characteristics.

The second step therefore, is to reintroduce between-state variation using a common standard, State NAEP. In a separate study, the State NAEP schools were linked to state assessment scores, and the means, standard deviations, and correlations of State NAEP school means with state assessment school means were computed. Those results were used to create an achievement measure that includes (1) between-state variation in means, (2) between-school variation proportional to between-school variation among State NAEP schools, and (3) a factor that attenuates within-state variation for states in which the assessment is only moderately correlated with State NAEP. The effect of the third factor, multiplying by the correlation between NAEP and the state assessment, "projects" the state assessment variation onto the NAEP scale, capturing that part of the state assessment score that is like NAEP. For example, if for a particular state the correlation of the state assessment with NAEP were zero, all schools in that state would be given the same achievement measure (actually, such a state would be excluded from the analysis). On the other hand, this "spreads schools apart" in states (a) in which school NAEP scores are themselves more varied and (b) in which the state assessment appears to be measuring skills highly related to NAEP. The effect of this spreading is to give greater weight to variations within these states in the computation of correlations of achievement with SASS measures.

Specifically, the school-level achievement score for school j at grade g, in state i is calculated as:

$$X_{gij} = \frac{(\text{school assessment score})_{gij} - (\text{mean school score})_{gi}}{(\text{standard dev of school scores})_{gi}}$$

$$Y_{gij} = ((X)_{gij} \cdot (\text{NAEP standard deviation})_i \cdot (\text{NAEP, state assessment correlation})_i) + (\text{NAEP mean})_i.$$

Finally, to emphasize that these are *not* NAEP scores, they are rescaled to a mean and standard deviation of approximately 50 and 10:

$$\text{ACHIEVE}_{gij} = 50 + \frac{10 \times (Y_{gij} - \text{Mean}(Y))}{\text{Standard Deviation}(Y)}.$$

Important assumptions are needed to apply the NAEP adjustment to the scores at grades other than the fourth and to mathematics scores. The between-state NAEP adjustment was based on the1994 State NAEP fourth-grade assessment for reading and on the 1992 and 1996 State NAEP fourth- and eighth-grade assessments for mathematics. Application of these adjustments to state assessment scores in middle school (for reading) and in high school is based on the assumption that variation in achievement *between states* is stable across grades. The 1992 and 1996 State NAEP mathematics assessment results support this assumption, in that the correlations between the fourth- and eighth-grade state means are .95 and .92, respectively (Mullis, Dossey, Owen, Phillips 1993; Reese, Miller, Mazzeo, & Dossey 1997). However, no information is available regarding State NAEP means at the high school level.

The use of 1992 and 1996 State NAEP state means and standard deviations to construct the 1994 mathematics adjustment assumes that state means varied smoothly, if at all, from 1992 to 1994 to 1996. In fact, the correlations of this study's 20-state NAEP mathematics means between 1992 and 1996 were .91 for fourth grade, and .93 for eighth grade, suggesting that interpolating 1994 figures from 1992 and 1996 figures is reasonable.

Finally, the correlations used in the adjustments were based on fourth-grade reading assessments. Use of these correlations in the adjustment of within-state variation in math scores assumes that between-state variations in reading and math are highly correlated. Because state assessments usually combine reading and math tests from the same publisher and in the same testing session, it is plausible to assume that factors which would affect the reading correlations in different states (e.g., the reliability of the state assessment instrument and distribution of state assessment scores) would also affect the math correlations. The 1992 State NAEP assessments in reading and mathematics in fourth grade indirectly support this assumption, in that the correlation between reading and mathematics state means is .94 (Mullis, Campbell, and Farstrup 1993). Nevertheless, the question remains whether the results of substantive analyses will be diminished by the extrapolation of between-state variation in average achievement from fourth grade to middle and high school variation. Comparative analyses of NAEP-adjusted vs. pooled within-state findings across school levels, carried out in this study, address this question.

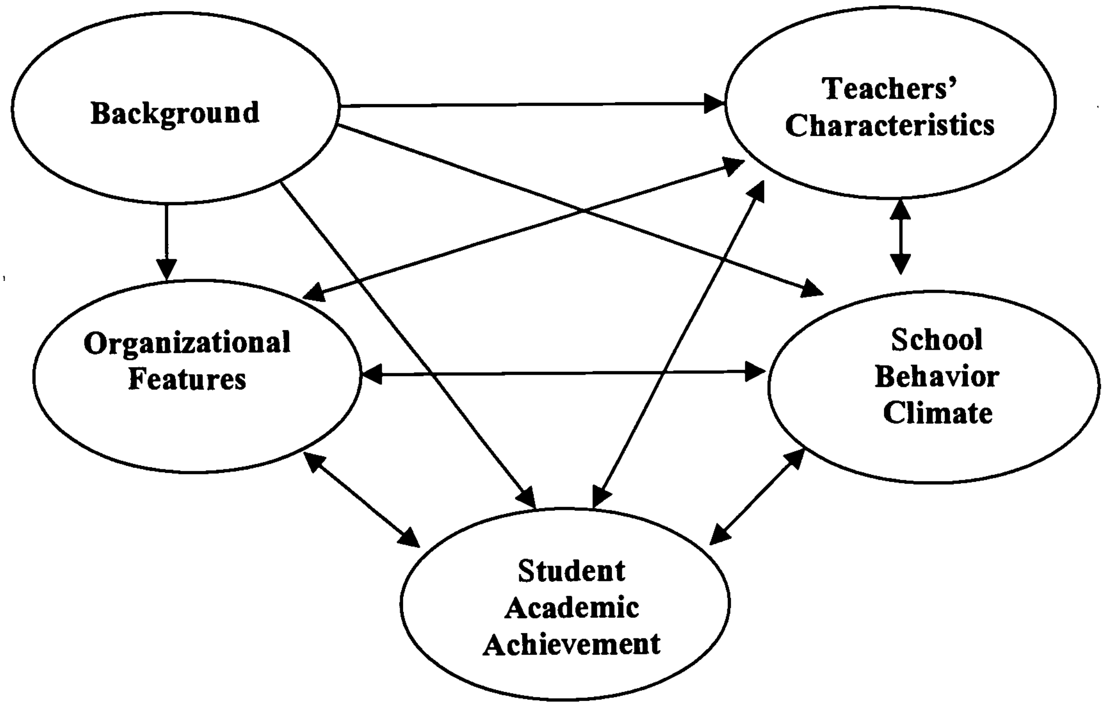
Although State NAEP data were used to capture between-state variation in achievement, it would be highly misleading to interpret the SASS student achievement measures as a surrogate of the school's average NAEP proficiency. First, State NAEP differs from individual state assessments in student sampling (each student takes only a fraction of the NAEP test), administration (a federal government contractor trains test administrators and monitors many testing sessions), motivation (NAEP is a low-stakes assessment with no individual student or school reporting), and item formats (NAEP has a substantial portion of extended open-ended items). The achievement measure developed for this study yields an unbiased estimate of school NAEP means, but with different standard errors in each state. The measure is not dependent on evidence that NAEP is *equated* to the various state assessments, and evidence that they might be "equatable" (i.e., parallel) was not sought. In fact, it is highly unlikely that the states' individual assessments are parallel to NAEP, due to differences in administration, item format, and content frameworks. In other research, we have shown that it is feasible to project state assessment results onto the NAEP scale without assuming that the tests are parallel (McLaughlin 1998). Second, the correlations between NAEP and state assessments differ substantially between states. Although the median correlation in these states in 1994 was .70, the smallest three correlations were between .30 and .50 (Wu, Royal, and McLaughlin 1997). Within-state variation of these synthetic NAEP school means will be smaller than variation of actual NAEP school means, especially in states whose assessments are not highly correlated with NAEP. Correlations are shown in appendix C for 18 states.

School-Level Correlates of Achievement

Student academic performance is shaped by multiple factors, relating to the school, the teaching process, the students' social and family background, and the community; and a school's reputation for academic performance can affect parents' decisions, students' behavior, and teachers' attitudes and decisions. We model student achievement in American public schools as related to four types of factors: (a) students' background, (b) organizational features of the school, (c) professional characteristics of the teachers, and (d) school behavior climate. While all these factors affect student academic success, they also interact with each other, and organizational characteristics and teacher choices can be affected by achievement at the school. We therefore, conceptualize the interrelationships among the five categories in this model as a web of interactions. Figure 1 graphically describes the general model.

The model shown in figure 1 refers to the school as a unit. Of course, achievement is an individual student characteristic, and the majority of variation in achievement is among students in the same classroom and between classrooms in the same school. Nevertheless, there is substantial reliable variation in achievement between schools, and from a policy perspective, there may be actions that can improve the overall achievement level in a school.

Figure 1.—School-level Correlates of Student Achievement: General Model



While analyses at the school level may not shed light on individual variation in learning, they can provide evidence on the correlation between school reform policies and achievement. Hierarchically structured data, with both individual student data and schools and staffing data, such as collected in NAEP and NELS:88, facilitate understanding of the correlates of individual student achievement. These data are much more costly to collect on a school sample the size of SASS than is the construction of a synthetic achievement measure from existing NAEP and state assessment data. In any case, the existence of within-school variability does not threaten the validity of analyses at the school level.

Using the SASS student achievement subfile, the model in figure 1 can be further specified in a variety of ways, one of which is shown in figure 2. The background category in the model is represented by three factors: percentages of students in poverty, with language barriers, and in racial/ethnic minorities. The organizational category is represented by four factors: school size, average class size, teachers' influence, and normative cohesion. The aim of the analyses to be carried out is to test hypotheses about the relations among these factors, either in terms of partial correlations or in terms of fits of linear models, such as that represented graphically in figure 2. By testing these models, inferences can be made about correlates of achievement across a wide range of public schools, although the "arrows" cannot, in most cases, be taken to indicate a direction of causality because of the alternative explanations of many of the correlations.

The arrows in figure 2 represent a "structural" model for the data; that is, a model for the relations among factors one would expect if there were no measurement error. Each of the factors in the model, except school size, is represented by multiple measures in SASS, as indicated in figure 3. In structural equation terminology, figure 3 presents the *measurement* model corresponding to the structural model in figure 2. The arrows in figure 3 indicate assumptions about the sources of variance in the *observed* measures. In addition, not shown, variation in each observed measure (shown in a rectangle in figure 3) is assumed to include an error component. Except for two correlations: between (1) the free lunch eligible and minority percentages and (2) teachers' perception of a racial/ethnic problem and the second climate composite; these errors are assumed to be independent. Each of the factors in figure 2, except school size, is represented by at least two indicators, providing the capacity for estimating the contribution of measurement error to variance in the indicators. Also indicated in figure 3 is a factor representing a common response pattern among five of the measures. These measures may be more positively intercorrelated than other measures, because they all represent teachers' subjective opinions about their school.

The two indicators of school behavior climate are based on 20 items concerning teachers' perceptions of problems in the school. The two parallel measures were constructed by averaging balanced halves of these items. For example, drug abuse is in one set, alcohol abuse in the other; student absenteeism in one and dropping out in the other, vandalism in one and robbery in the other. Included in both sets were two topics on which there were multiple items: tardiness and attacks on teachers. Table 2 describes the particular indicators used to estimate these various types of factors, and the SASS data fields on which they are based. The field numbers shown in table 2 refer to the standard coding of SASS measures. Teacher variables in the model are represented by the average of the up-to-five sampled teachers who participated in the survey in each participating school.

Figure 2.—One Possible Path Model Relating Achievement and School and Background Factors

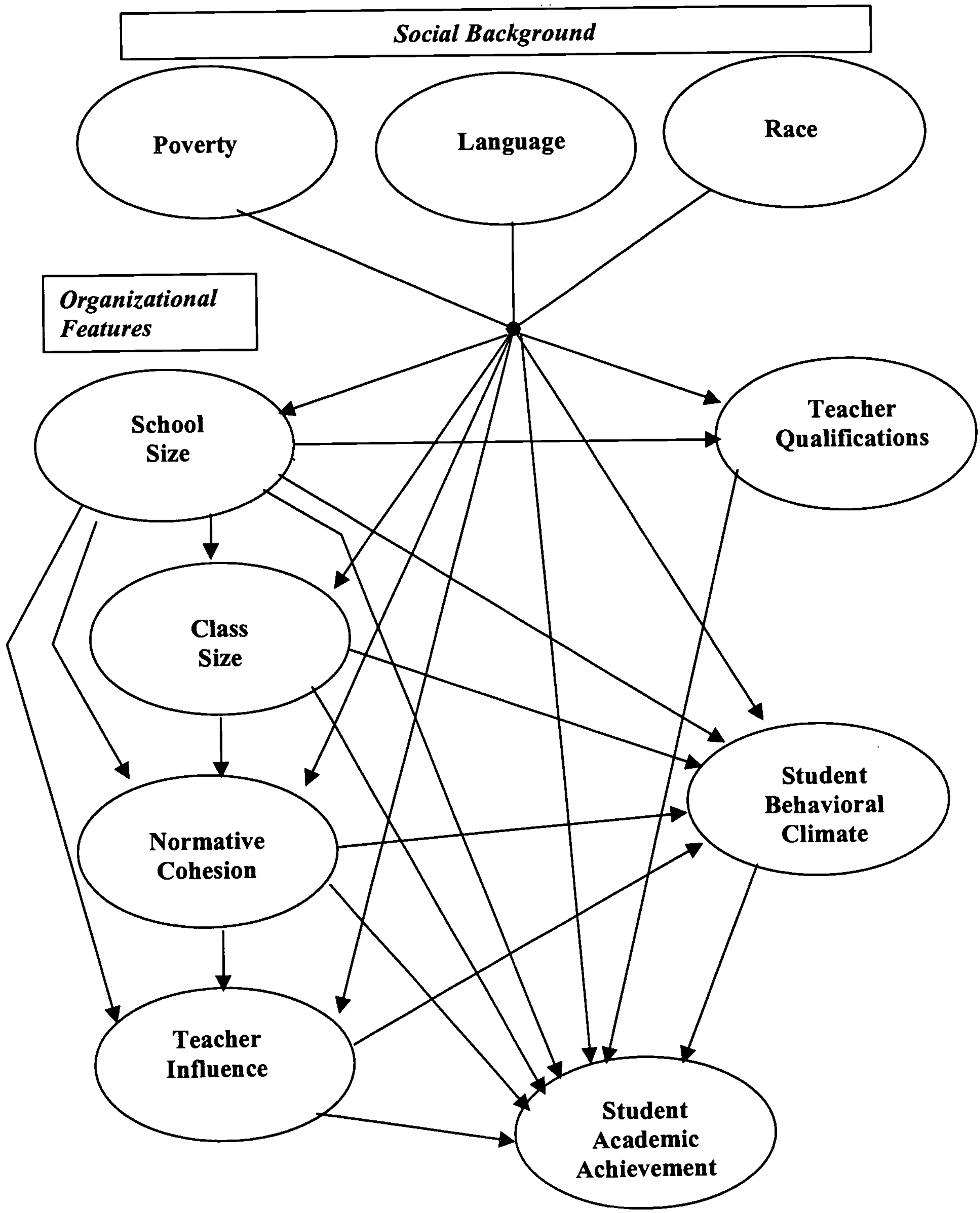
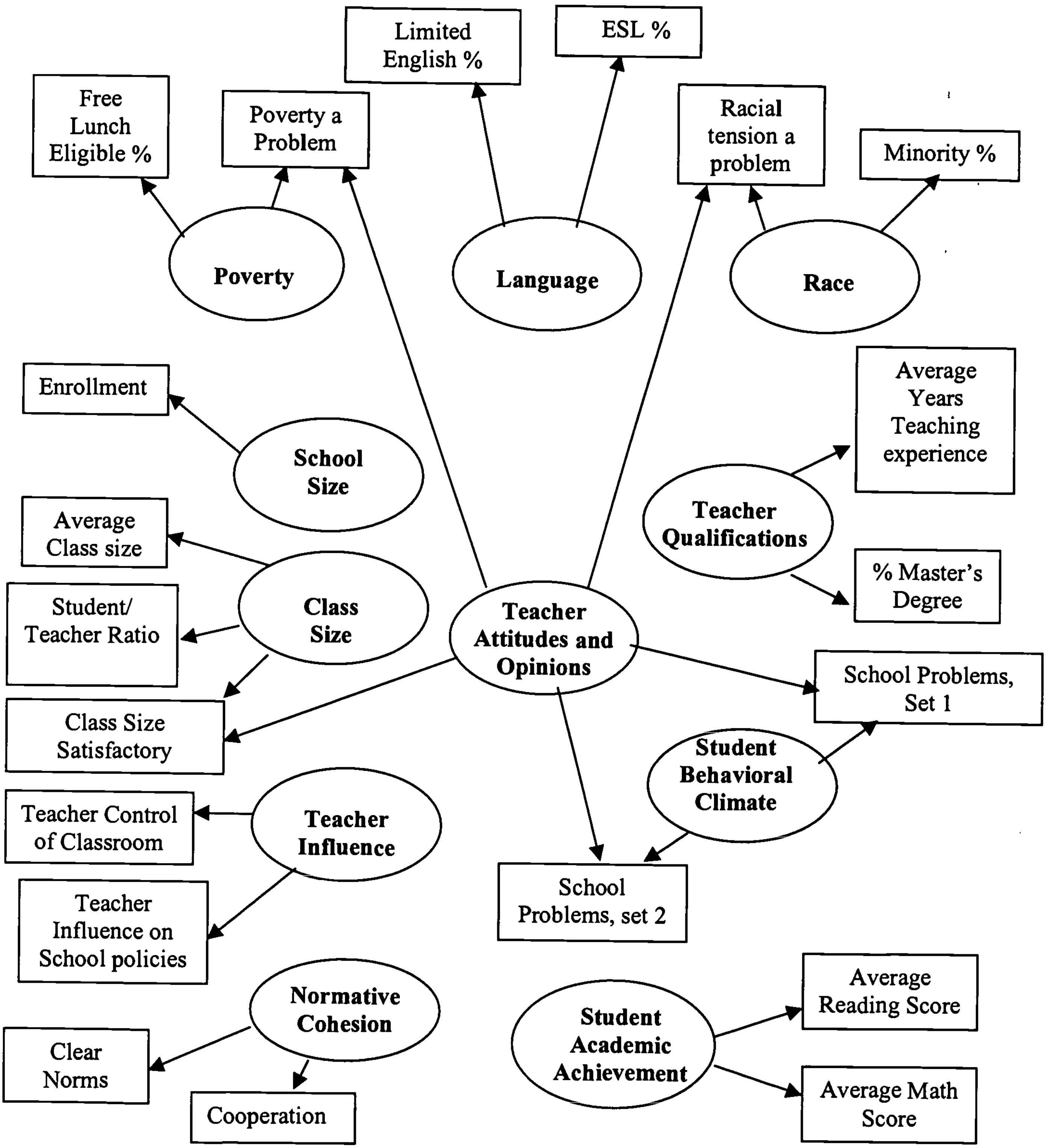


Figure 3.—Measurement Model for School-Level Correlates of Achievement



The SASS measures vary between elementary and secondary schools, as shown in table 3.[8] For example, fewer secondary students were reported to be eligible for the free school lunch program and fewer were limited in English proficiency, similar to results found by NAEP. Secondary schools also tended to be larger, to have more student behavior problems, and to have more teachers with master's degrees who felt they had more control over their classrooms, but less cooperation among teachers. As a result, since the achievement measure is computed in such a way as to have essentially the same mean and standard deviation at each grade level, it is essential that the correlates of achievement be determined separately for each grade level. Combining the data across grade levels would yield artifactual relations among the SASS measures, as well as masking relations with the achievement measure.

Furthermore, determining the correlates separately for elementary, middle, and high schools provides an opportunity to explore the patterns of change in the correlations over the school years. Much like Herriot and Firestone's (1984) arguments that the images of schools vary by level[9] and that each school level operates differently,[10] we anticipate that the relative importance of the various factors to student academic achievement will vary by school level.

A great deal of research has been done on these possible school-level correlates of achievement, but none have had the combined benefit of such a large and varied sample of schools and set of school measures. To set the context for discussion of analyses of the SASS student achievement subfile, we briefly review the literature on student background measures, school organization measures, teacher qualifications, and school climate, as they relate to student achievement.

---

[8] Because changes were made in the file after the standardization of the achievement scores, the mean and standard deviation are not exactly 50 and 10.

[9] Elementary schools are imagined to be more rational and bureaucratic, while high schools are seen as more anarchic and envisioned to be a loosely coupled system.

[10] In elementary schools, curriculum is more limited; while in high schools, curriculum is broad. Also, the operations of elementary schools are more centralized and consensus-drive, while in high schools there are high levels of complexity and differentiation.

Table 2.—Specification of the components of the SASS achievement correlates model

Category	Factor	What it Means	How We Measured It? Our Indicators	SASS Item
Student Background	Poverty	Low SES	Percent of lunch program eligibility, and teachers' identification of a poverty problem	S1655 S1660 ()by S0255) T1165
	Race	Racial tension and problems	Percent of white students, and teachers' identification of a problem of racial tension	S0425 ()by S0255) T1170
	English Language proficiency	Low proficiency of English	Percents identified as having language problems (participation in English enhancing programs, LEP) and teachers' identification of limited English as a problem for students	S1410 S1295 S0415 ()by S0255) T1190
Teaching Quality	Teacher qualification	Teachers' degrees and teaching experience	Professional education (i.e., having a master's degree) and years of teaching experience	T0235 T0095 T0100 T0105 T0110
School Climate	School atmosphere and problems	Student behavior	Teachers' identification of problems with (set #1) tardiness, dropping out, lack academic challenge, vandalism, drug abuse, physical conflicts, verbal abuse of teachers, physical attacks on teachers, and teacher absenteeism, and (set #2) student absenteeism, cutting class, apathy, robbery or theft, disrespect of teachers, alcohol abuse; and weapons in school	T1005 T1075 T1080 T1085 T1090 T1095 T1100 T1105 T1115(5)T1150 T1325 T1330 T1340 T1345
Organization Features	School size	Organization size	Number of students enrolled	S0255
	Class size	Resource per student; crowding	Average class size, ratio of teachers to students, and teachers' sense of satisfaction with class sizes	S0255 T0830-T0970 S0255 S0910 S0850 T1285
	Teacher Influence	Teachers' control over school policies and classroom arrangements	Teachers' sense of influence over such school matters as setting discipline policy, determining content of in-service programs, hiring, school budget, teacher evaluation, and establishing curriculum; And, teachers' sense of influence over such classroom matters as selecting textbooks and other instructional materials, selecting content, topic, and skills to be taught, selecting teaching technique, evaluating and grading, disciplining students, and determining amount of homework assigned	T1015 T1020 T1025 T1030 T1035 T1040 T1045 T1050 T1055 T1060 T1065 T1070
	Normative cohesion	Clarity of norms	Colleagues share beliefs and values; principal enforces rules and back-up staff; receive support from parents; principal lets staff know expectations; goals and priorities are clear; principal knows what wants and communicates to staff; behavior rules are consistently enforced	T1200 T1225 T1245 T1255 T1260 T1265 T1295
		Cooperation among school staff	Coop among staff; whether teachers coordinate course with other teachers; sense that administration-staff relations are supportive and encouraging; teachers integrate library/media sources into curriculum	T1205 T1270 T1290 T1310

NOTE: SASS items beginning with "T" came from the teacher questionnaire, while those beginning with "S" came from the school questionnaire.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and state reading and mathematics assessment scores for publics schools in 20 states.

Table 3. Means and standard deviations of indicators used in the analyses, by grade level

Indicator Name	Definition	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
Math.Achieve	School average mathematics score	49.8 (8.2)	49.7 (8.0)	49.3 (8.4)
Read.Achieve	School average reading score	49.1 (7.8)	48.7 (7.7)	48.2 (8.7)
Pct_Lunch	Percent students eligible: lunch program	44 (31)	40 (27)	28 (25)
Item T1165	Teachers say poverty is a problem (1 = serious, 4 = no)	2.36 (0.78)	2.31 (0.63)	2.33 (0.62)
Pct_Minority	Percent of minority students	31 (33)	34 (32)	30 (31)
Item T1170	Teachers say there is racial tension (1 = serious, 4 = no)	3.32 (0.61)	2.86 (0.66)	2.96 (0.64)
Pct_LimEng	Percent of students with limited English proficiency	5 (13)	4 (11)	3 (9)
Pct_ESL	Percent of students in ESL program	4 (10)	3 (9)	2 (5)
Masters degree	Percent of teachers with a master's degree	40 (32)	40 (28)	46 (26)
Yrs_Texp	Average years of teaching experience	14.6 (5.6)	13.9 (4.5)	15.0 (4.3)
School Climate Set #1	Teachers note student behavior problems (higher = fewer problems)	36.1 (3.6)	31.1 (4.0)	29.9 (3.6)
School Climate Set #2	Teachers note student behavior problems (higher = fewer problems)	35.6 (3.5)	29.7 (4.1)	28.2 (3.8)
School Size	Log (enrollment)	6.1 (.59)	6.5 (.75)	6.6 (.89)
Avg_Clsz	Average class size	23.5 (8.5)	25.4 (7.0)	23.1 (6.1)
Ratio_ST	Student/teacher ratio	18.2 (4.4)	18.0 (4.9)	16.9 (4.8)
Item T1285	Teachers satisfied with classroom size (1, 4 = strongly agree, disagree)	2.25 (0.75)	2.39 (0.62)	2.07 (0.55)
T_I_Class	Teacher sense of influence over classroom affairs (higher = more)	23.4 (3.1)	24.1 (2.4)	24.5 (2.4)
T_I_S_Pol	Teacher sense of influence over school policies (higher = more)	11.0 (3.8)	10.0 (3.4)	10.2 (3.1)
Clear norms	Teacher perception of clarity of norms (higher = more)	3.15 (0.38)	2.95 (0.35)	2.91 (0.32)
Cooperation	Teacher perception of cooperation among staff (higher = more)	3.12 (0.36)	3.01 (0.31)	2.97 (0.28)

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and state reading and mathematics assessment scores for publics schools in 20 states.

Student Background

Family background and socioeconomic status have consistently been shown to be related to student achievement (Coleman et al. 1966, Hanushek 1986). Previous studies called attention to students' health and nutrition, the physical environment of their homes, their family structure (e.g., single-parent homes), the parenting styles, beliefs, and expectations of the students' parents, and finally, their inherited intelligence.[11] These studies lead us to expect that adverse background conditions (such as poverty and lack of parental involvement) lower student academic performance. While the major purpose of research on schools is to identify characteristics of *schools* that contribute to student achievement, it is essential to take background characteristics into account, because they affect the intercorrelations of school measures. For example, suppose that poor children were found to be attending schools with chipping paint. We would expect to find a correlation between chipping paint and test scores, but one should not infer from that correlation that painting schools will improve test scores. It is the pattern of correlates among schools with students with similar backgrounds that is of interest.

It is particularly important to include background factors in this study of school-level factors, because the database does not contain longitudinal achievement data on individual students or student cohorts. School characteristics and policies are more likely to be correlated with school-level variation in gains in achievement than with achievement measured at one point in a student's career. Including student background measures in the model serves to control for much of the between-school variation in achievement potential that students bring to school. Nevertheless, the lack of "pre" measures of achievement underlines the need to interpret the results of analyses of SASS data relevant to the model in figures 1 and 2 as causally indeterminate.

To maintain a distinction among diverse background issues, we focus on three types of background factors: (1) poverty, (2) English language proficiency, and (3) racial/ethnic minority status. *Poverty* is a major component of the effect of a student's background on his or her academic achievement. Students are at a dramatically higher risk of academic failure when they come from a poor background, or from low SES families. Similarly, *language* is currently the focus of a heated debate in education circles, because there is evidence that lack of English proficiency puts students at a disadvantage when they are trying to learn not only how to read and write, but also how to solve mathematics and science problems. Last, *racial/ethnic affiliation* is repeatedly argued to be an important factor relating to academic performance. Non-white or Hispanic students (i.e., minority students) are assumed to be disadvantaged in schools, while white students reap greater benefits in education. One need not address issues about the sources of the impact of these background variations on achievement in school to realize the need to control for these factors in modeling school processes.

[11] For literature reviews, see Scott-Jones 1984, Henderson and Berla 1994.

We measure each background factor by multiple responses in SASS surveys.[12] The unobserved factor "poverty" is composed of two indicators, originating from the SASS questionnaire: (1) the percentage of students who qualify for the national school lunch program and (2) teachers' identification of poverty as a problem.[13] The factor of English language proficiency is composed of two similar indicators: (1) the percentage of students identified as limited English proficient (LEP) and (2) the percentage of students participating in the school's English as a Second Language (ESL) program. Last, the factor of "race" is composed of by two SASS-based indicators: (1) the percentage of white students in the school and (2) teachers' identification of racial tension as a problem.

While all three factors represent various dimensions of the students' social background, they are still quite distinct from each other. So while the correlation scores among the various indicators across latent variable lines are generally high and significant,[14] the eigenvalue for each factor is significant. The eigenvalues are 2.76 for "language," 1.49 for "poverty," and 1.45 for "race."

This multiple measurement potential is a unique strength of the SASS database. Two of the three background factors are each measured by both an objective indicator, the percentage of students with the corresponding characteristic, and a subjective indicator, the perception of a sample of teachers in the school that that factor is a source of problems in the school. For example, the underlying factor of poverty is only imperfectly measured by the percentage of children who are eligible for the federal free lunch program due to differences in cost of living and in eligibility counting procedures between school districts, as well as in the impact of poverty on readiness to learn in different settings. Incorporating the perceptions of a sample of teachers in the school that poverty is a problem in the school can eliminate some of the error of measurement of poverty. Of course, there are different kinds of error in teacher perceptions, ranging from different interpretations of survey items to sampling error, but the combination of the two indicators can be expected to have greater validity for the impact of poverty on learning in the school than either separately. Because SASS has many related objective and subjective measurements, one can control for measurement error associated with variation in teachers' use of the response scale (e.g., some teachers mean something more serious by "problem" than others do) by estimating the extent to which each teacher tends to be a positive or negative responder to attitude and opinion items.

[12] As mentioned in the methodological section, each such conceptual factor, or latent variable, is composed based on both empirical and substantive grounds: we combine multiple questionnaire responses based on confirmatory factor analysis, yet we also maintain a judgment as to the substantive coherence of each factor.

[13] The correlation between these two measures of poverty is .49 ($p<.01$; $n = 2561$). See Appendix A for further details.

[14] For example, in the general sample of public schools ($n = 2561$) the SASS response to teachers identification of a poverty problem (which is an indicator of the latent variable "poverty") correlates at .38 with the proportion of minority students (an indicator of "race") and at .24 with the ratio of students having limited English (LEP; an indicator of "language").

Alternative indicators of a particular factor, such as subjective and objective assessments of poverty must be correlated, but they need not be highly intercorrelated; however, a low intercorrelation is likely to limit the power of the data to measure correlations with achievement. The intercorrelations of the indicators included in each composite factor are shown in appendix A. For example, for poverty, the intercorrelations between the objective and subjective indicators are .55, .51, and .47 for elementary, middle, and secondary schools, respectively.[15] These intercorrelations are themselves limited by the "reliability" of the indicators that are based on averages of teachers' responses. If different teachers see the same school very differently, the average of their responses cannot tell a great deal about the school, per se. The percentage of variance in school means that is attributable to between-school variation, which is a measure of this reliability, is also shown in appendix A. For the average rating of whether poverty is a moderate or serious problem, the reliabilities are .73, .79, and .84 for teachers in elementary, middle, and high schools, respectively.

School Organizational Features

Both objective features of a school's organization, such as its class sizes, and subjective features, such as the level of cooperation among its staff, may be correlated with achievement. However, unlike the social background factors, these features are endogenous and to varying extents under control of the principals and teachers in the school. These characteristics may affect the overall goal attainment of the school, as indicated by the academic success of its students; and to some extent academic achievement can affect these characteristics. For example, a magnet school with a reputation for attracting students with potential for careers in science is likely to have higher test scores than other schools, purely as a function of the students who enroll, and its magnet status may affect class sizes, either enlarging them to respond to demand or lowering them as a result of special funding as a magnet school. Although the specific models presented in this report focus on accounting for variation in achievement in terms of variation on organizational characteristics, the direction of causality is not determined in these data. The contribution of the data is to constrain the set of plausible hypotheses and force a search for alternative explanations when the data do not fit preconceived theories.

[15] The corresponding intercorrelations for the minority and language factors are approximately .45 and .75. The language factor, unlike the minority and poverty factors, is defined by two indicators that are both objective counts.

SASS has a wide range of information about schools, obtained from the principal in the administrator and school questionnaires and from the teachers in the teacher questionnaire. Four organizational features have been selected for inclusion in the model for this report: (1) school size, (2) average class size, (3) the teachers' sense of influence over school affairs, and (4) the normative cohesion of the school's staff. These four features, while sharing an organizational perspective that emphasizes a structural dimension and reflects on its impact on the operations of an organization, are quite distinct from each other. Each such feature represents a distinct dimension of organizational life: school size reflects the organization's potential scope; class size reflects resources; teachers' sense of influence represents the authority structure within the school; and the normative cohesion of the staff refers to the organizational culture. We should point out that this report focuses on a limited array of the measures available in SASS. A more elaborate model might include a variety of other SASS organizational measures, such as organizational complexity, as reflected in the number of different kinds of positions in the school; organizational goals, as expressed by the principal; perceptions of outside influence by state agencies and local school boards on decisions; and staff diversity. The decision to select four organizational factors was made to demonstrate the potential of this database.

School size has been shown to have a significant effect on the school's performance, yet the direction of the effect is inconsistent. As Lee and Smith (1993) note, two conflicting perspectives can be found in the literature on the effects of school size—one perspective highlights the positive consequences of increased size, drawing on the economics of scale to focus on the greater diversity of program offerings and on resource issues, while the other perspective focuses on negative consequences, centering on the more formal and impersonal social interactions as impeding success. Recently, research evidence is accumulating to support the argument that student achievement is lower in bigger schools. Greater school size is shown to be associated with reduced academic engagement among eighth-grade students (Lee and Smith 1993), and at the high school level with lower student participation in school activities (Lee and Smith 1995), less satisfaction with school experiences and lower school attendance (Lindsay 1982), and higher dropout rates (Pittman and Haughwout 1987). Based on these research findings, many reforms call for dividing big schools into smaller subunits (e.g., Oxley 1994, Boyer 1995), in order to provide students with a more stable and supportive educational experience. To assess these arguments we measure school size as simply the SASS indicator of the number of students enrolled in the school.

In regards to class size, common sense suggests that it has an effect on students' academic performance: It is now generally assumed that smaller classes, where greater attention is given to each student, foster academic Meta-analyses of existing studies (Glass et al. 1982), along with large-scale research projects based on experimental designs (Finn and Achilles 1990) have confirmed the link between reduced class size and greater student learning,[16] while

[16] For a review, see Mosteller, Light, and Sachs 1996. For a critical assessment, see Slavin 1989.

Blatchford and Mortimore (1994) offer a number of possible explanations for these relationships between class size and student achievement.

Nevertheless, none of the prior studies has a database of the size and richness of SASS, which can be used to estimate class size effects while controlling for other exogenous and endogenous factors. SASS has three indicators of a school's average class size: (1) average class size as reported by the sampled teachers;[17] (2) the school's teacher/student ratio, calculated as the total number of students in school divided by the sum of the full-time teachers in the school and one-half the number of part-time teachers in the school; and (3) the average of the sampled teachers' ratings of satisfaction with the size of their classes. Although none of these by itself is a highly reliable measure of a school's class sizes, their combination reduces measurement error of this factor.[18]

The third organizational characteristic refers to *teachers' sense of influence*. While teachers' sense of influence over their work environment is sometimes referred to as representing their efficacy, expectancy, and sense of control, we regard it as an indication—albeit from the teachers' perspective—of the allocation of authority within the school. Here we tap onto current debates over the role of school leadership, or of cooperative management, in schools. The direction of effect of this factor is, however, unclear. Moreover, as Ingersoll (1996) points out, the particular domain of teacher influence may determine the magnitude of its effect on student academic performance. This factor is calculated from teachers' responses to SASS questions about their sense of influence over school policies[19] and influence over matters concerning their own class.[20]

The fourth organizational factor whose correlation with achievement we investigated is the *normative cohesion* of the staff. This factor refers to the cultural solidarity among staff members in the school, or the collective norms that govern staff behavior in this organization. Cohesive interpersonal communication serves as an important foundation for consensus formation, social support, effective access to institutional resources, satisfaction, and eventually, enhanced performance. Therefore, while in research on workplace cohesion interpersonal

[17] For a teacher teaching a single self-contained classroom, class size is the total number of students enrolled in the teacher's class at the school. By contrast, for a teacher teaching multiple departmentalized classrooms, class size is calculated as the average number of students in the teachers' classes.

[18] Among elementary schools, the reliability estimate of teachers reported class sizes as an indicator of the school is low, .31, and the intercorrelations of that with the other class size indicators are also low, .28 and .29. This may limit the potential elementary school correlations of class size with achievement.

[19] Included items refer to influence over school policies in discipline, the content of inservice training programs, hiring new full-time teachers, deciding priorities in spending school budget, evaluating teachers, and establishing curriculum. The reliability of the school mean is about .62.

[20] Included items refer to control in one's own classroom over selection of textbooks and other instructional materials, selection of contents, topics and skills to be taught, selection of teaching techniques, evaluating and grading students, disciplining students, and determining the amount of homework to be assigned. The reliability of the school mean ranges from .36 at the elementary level to .50 at the secondary level.

networks are sometimes considered disruptive (e.g., Granovetter 1982), it is plausible to expect that within the education system cohesion is related to effectiveness of teaching and therefore to student performance. Friedkin and Slater (1994) show that the cohesiveness of the collegial network among teachers and the centrality of the principal's advice are highly associated, and that the related position of principals as school leaders leads to improved school performance. The causal direction of a positive relation between normative cohesion and achievement is ambiguous, however. Serving a high-achieving student body can, itself, be a motivator and facilitator for teachers' working together.

To estimate normative cohesion, we constructed two SASS composite measures: (1) a score for the clarity of norms,[21] and (2) a score for staff cooperation.[22] While, in general, the normative structure is assumed to affect such features as organizational coupling, formalization of goals, and legitimization of practices, we estimate its relationship with teacher perceptions (teachers' sense of influence), school climate (behavioral problems), and school performance, or, in this case, student achievement.

Last is the issue of interrelations among the multitude of factors affecting student achievement. The impact of the students' social background on academic achievement, while largely outside of school influence, is not likely to be entirely independent of school arrangements. Rather, schools themselves may be influenced in multiple ways by the background of their students. Schools located in disadvantaged social environments tend, for example, to be bigger; such schools also have less control over their recruitment, and, thus, end up with a teaching staff that is less normatively coherent. Hence, any valid model of relations between school-based organizational factors and achievement must take the effects of background factors such as poverty, race, and language barriers on both a school's organization and its average achievement into account. The direction of causality is somewhat less of an issue in this case, because it is much less plausible, at least in the short term, that a public school's organizational characteristics would affect the demographics of its students than the reverse. Although the organizational characteristics of a school might easily affect the choices of teachers to teach at the school, it is less frequent that these characteristics affect the choices of parental residence for the students.

Teachers' Qualifications

[21] Calculated as the average of teachers' responses to SASS questions about sharing beliefs and values with colleagues in school; receiving support from parents; goals and priorities in school being clear; rules being consistently enforced by all teachers; the principal backing up staff members; the principal letting staff know what is expected of them; and the principal having a clear vision for the school and communicating this vision to the staff. The mean correlation among these seven measures is .42, and the reliability of the school mean is about .57.

[22] Calculated as the average teachers' responses to SASS questions about getting cooperative effort among staff members; making a conscious effort to coordinate course content with other teachers; planning with media specialist or librarian; integration of their specialty into teaching; and viewing the behavior of school administration as being supportive and encouraging. The mean correlation among these four measures is .20, and the reliability of the school mean is about .48.

The third group of factors represents an attempt to take into account quality of teaching by measuring teachers' objective *qualifications*. Teachers' qualifications are calculated from a measure of years of teaching experience and the percentage of teachers who acquired a master's degree. This factor, while focusing on credentials and seniority, conveys a sense of the teachers' professional ability. More qualified teachers are more experienced, can better handle their students, colleagues, and their school system, and are more familiar with teaching practices and with the substantive material. Overall, therefore, it is easier for experienced, qualified teachers to create a more effective learning experience for students. Unfortunately, although teachers' qualifications are included in the analysis, the factor is relatively weak, compared to the other factors.[23] A wide variety of other SASS measures of teacher qualifications could be explored in a study focusing particularly on teaching quality and achievement. These qualifications include out-of-field teaching, selectivity of the colleges the teachers attended, amount of inservice training, and hours spent on school-related activities, training to teach limited English proficient students, as well as other measures.

Teachers' qualifications may play a complex role in a general model of school processes and the relations between these qualifications and achievement may differ between schools and between states. For example, as shown by Rowan, Chiang, and Miller (1997) the effects of teachers' ability and motivation on their students' academic performance varies across different work situations. The correlation between teachers' qualifications and achievement may also reflect the choices of where teachers teach: schools with students from more advantaged backgrounds, or which maintain orderly and disciplined school environments, may find it easier to recruit and retain highly qualified teachers. On the other hand, a negative correlation might indicate that high achieving school districts allocate their resources to other strategies (e.g., smaller classes), rather than paying for higher qualified teachers, or that states with chronically low achievement have systems in place that ensure that large numbers of their teachers have advanced education and are experienced. As with school organizational factors, the direction of causality of the relations between teacher qualifications and achievement cannot be inferred directly from correlations, but these correlations provide valuable evidence of relations that must be explained in some manner.

School Climate

The various background, teacher, and organizational factors shape the climate of the school. School climate refers to the atmosphere of student behavior in the school and serves as a metaphor for the nature of school experience for students, teachers, and parents.[24] Whereas some

[23] There is substantial within-school variation in years of teaching experience (reliabilities of .29, .30, and .37 at the three grade levels), and correlations with percent of responding teachers with a masters degree range from .14 to .27.

[24] See Anderson (1982) for a review of literature on school climate.

studies conceptualize such an atmosphere as it relates to learning, discipline, or expectations,[25] and whereas previous studies of school climate called attention to a variety of attributes, such as teacher commitment and morale or the emphasis on academic goals and criteria,[26] we focus on behavior patterns. Specifically, we examine the prevalence of various student behavioral problems. Such behavioral problems, refer to students' respect for teachers and staff, peers, and property, disorderly and disruptive behavior, and substance abuse. By focusing on student behavior problems, we call attention to the tone set by students themselves, and not by teachers, principals, parents, nor the policy guidelines. Thus, our conceptualization of school climate is not merely a synonym for "better school;" it is not a personality trait of the school; and it is not a relationship-based assessment (Hoy and Hannum 1997[27]). Rather, our measure identifies behavioral patterns, while still regarding school climate as an organizational feature at the school level (see, Sirotkin 1980). We also identify specific behavioral tendencies in schools (e.g., violence, substance abuse), which reflect both the physical and psychological health of the school as an environment for learning.

The general environment set in school touches upon the cultural resources, or support, available for students, and upon the impediments or disruptions to academic concentration. The school climate influences, therefore, how schools function as an educational environment for their students and as a workplace for their teachers. Arguably, providing a safe, orderly, peaceful, and "bully proofed" school environment contributes to student academic success, while having multiple behavioral problems disrupts proper education. For the measurement of this factor, we rely on teachers' identification of such problems as tardiness, absence, cuts, dropouts, apathy, verbal and physical conflicts, theft, vandalism, weapons, alcohol, and drug abuse. The perceived prevalence of these behavioral problems sets a discouraging atmosphere in schools, while in schools where such problems are minor or nonexistent, the culture sets a positive value and expectation to support learning and academic achievement. These characteristics are difficult to measure in a uniform manner across a large sample of schools, but SASS has attempted such a measurement by asking teachers to indicate which of two dozen different types of potential problems are serious, moderate, minor, or nonexistent in their school. To assess the extent of measurement error in these perceptions, two composite measures can be constructed by arbitrarily dividing the problem ratings into halves.[28]

School climate factors are likely to be related to school organizational features, such as

[25] For example, Pallas (1988) and following him, Raudenbush, Rowan, and Kang (1991) expand the concept of school climate to include principal leadership, staff cooperation, teacher control over school and class policy, teacher efficacy, and teacher satisfaction.

[26] For a short summary of previous studies, see Taylor and Tashakkori 1995: 218–219.

[27] For another example of a relational approach to school climate, while investigating its effect on the alternative output measure effective teaching, see Tarter, Sabo, and Hoy 1995.

[28] The reliabilities of the two climate composites range from .70 to .86 for the three school levels, and their intercorrelation ranges from .50 to .72.

size (e.g., Raudenbush, Rowan, and Kang 1991), and the general model reflects these relationships. Although a positive relationship between school climate and achievement can be expected, the direction of causality is particularly ambiguous. There is likely to be a positive feedback between students' focus on achievement and teachers' perceptions of their behavior and demeanor. Nevertheless, evidence about the significance of this correlation in schools serving different grade levels is valuable.

Finally, in designing structural equations to represent the model, we added one additional factor that describes a method effect. This factor, labeled *teacher attitudes and opinions* in figure 3, may account for some of the variance in those indicators that call for teachers' subjective opinions (namely, their perceptions that poverty or racial tension is a problem, satisfaction with class size, perceptions of influence and cooperation, and the components of the school climate). These items may be more highly correlated with each other than with other indicators, because of a halo effect in teachers' tendencies to make ratings of their schools. Thus, for example, teachers who identified one behavioral problem as serious tended also to estimate other behavioral problems as serious. Similarly, teachers who seemed to be satisfied with one of their work conditions (e.g., level of normative clarity) tended also to be satisfied with other terms. While our method partially addresses this matter of personal variations by relying on *average* teacher responses per school, we still found a consistent tendency for teacher responses to be highly correlated among themselves and to be less correlated with the SASS principals' responses on similar questions.

In summary, our model is consistent with views that average student achievement in a school is affected by multiple factors that represent various facets of the education system and of the social environment within which education occurs. Our model is consistent with views that student achievement may affect other characteristics of a school, such as school climate or the choices of teachers to teach at the school. Using structural equation modeling, each factor other than school size is a latent variable that is estimated by multiple indicators. Finally and most importantly, the modeling permits the investigation of interrelations among these facets of public school education.

Analytical Method

The SASS student achievement subfile contains school-level statistics on hundreds of measures for over 2,000 public schools in 20 states. The potential for analyses of this database is enormous. In this report, we have constructed a set of 18 composites of SASS items. For a baseline demonstration of analytical feasibility, the database has essentially been reduced to the intercorrelation matrix of these 18 composites, along with two assessment scores for schools at each of three grade levels, plus, for the purpose of standard error estimation, an indicator of the state in which each school resides.[29] The raw correlations between the 18 composites and the two

[29] Although exploration of other functional forms may yield additional insights, estimation purely in terms of linear (continued)

achievement measures do not provide meaningfully interpretable information, however, because many of the apparent correlations are mere reflections of correlations among other measures, and other "real" correlations are masked by confounding measures and can only be uncovered by controlling for those confounders. Thus, the first meaningful stage of analysis is to examine partial correlations between achievement measures and SASS composite measures of school organization and climate and teacher qualifications, controlling for social background factors.

Partial correlation analysis, as a method of testing for significant relations between school composites and achievement measures, has the advantage that it does not presuppose a causal ordering among the school composites and achievement measures. However, the picture that partial correlations provide of the correlates of achievement is limited in that the multivariate structure is not apparent. For example, the partial correlations of both normative cohesion and school climate with achievement may be positive, but one cannot discern from them whether this is due to a common factor in climate and cohesion or to independent factors. A form of multivariate regression is needed to identify the more complex structure.

The simplest form of multivariate regression is ordinary least squares (OLS) linear regression. This methodology can reveal the multivariate structure if its assumptions are satisfied, but an important assumption in OLS modeling, which limits its value for educational research, is that the "predictor" measures are measured without error. With many databases, this assumption is untestable, because only a single measure of each construct is available. However, the SASS database, with its multiple sources of information about school-related constructs, offers the opportunity to take measurement error into account when modeling the structural relations among factors. Structural Equation Modeling (SEM) jointly models the structural relations among latent factors (as in figure 2), while simultaneously modeling measurement error (as in figure 3). Computer programs such as LISREL, EQS, and SAS PROC CALIS can be used to estimate the variance components in SEM (see Bollen 1989). The primary analytical results presented in this report are SEM analyses based on SAS PROC CALIS estimation.

Structural equation modeling is particularly helpful in specifying, estimating, and testing hypothesized relationships among meaningful concepts or factors,[30] by allowing such concepts to be estimated from several indicators or measures. Employing SEM, we are able to base our estimation of each concept on multiple indicators, to evaluate the interrelations among the indicators (covariance on the measurement level), and to evaluate the relationships among the concepts in the model. As mentioned briefly earlier, the composition of each factor or latent variable is determined using both empirical and substantive criteria. We combine multiple questionnaire responses based on confirmatory factor analysis, yet we also maintain a judgment as to the substantive coherence of each factor.

---

models is an efficient initial step, because software packages are readily available and most important relations among educational factors are monotonic and therefore "visible" to linear analyses.

[30] Technically referred to as latent variables, yet also known as unobserved or unmeasured variables.

The model specification for our estimation of student achievement is:

$$0_{Achievement} = \ni_{11} >_{Background} + \ni_{12}\, 0_{Organization} + \ni_{13}\, 0_{Teachers} + \ni_{14}\, 0_{Climate} + *_1$$

where:

0_{factor} is a latent factor that may be affected by other factors in the system;
$>_{factor}$ is a latent factor that is not affected by the other factors in the system;
$\ni_{11}$ is the vector of coefficients for the vector of three independent social background factors;
$\ni_{12}$ is the vector of coefficients for the vector of four school organization factors;
$\ni_{13}$ is the coefficient for the teachers' qualification factor;
$\ni_{14}$ is the coefficient for the school climate factor; and
$*_1$ represents other unmeasured factors that may affect a school's achievement level.[31]

More specifically, the model for correlates of average academic achievement relates variation on achievement to variation on eight latent variables. A complementary measurement model relates each of the latent predictors to multiple SASS measures, with the exception that school size is based on a single measure. The SEM specification is:

$$0_{Achievement} = \ni_{11}\, 0_{School\ climate} + \ni_{12}\, 0_{Tchr\ qualifications} + \ni_{13}\, 0_{Teacher\ control} + \ni_{14}\, 0_{School\ size}$$
$$+ \ni_{15}\, 0_{Class\ size} + \ni_{16}\, 0_{Norm\ cohesion} + \ni_{17} >_{Language} + \ni_{18} >_{Poverty} + \ni_{19} >_{Race} + *_1$$

$$0_{School\ size} = \ni_{21} >_{Language} + \ni_{22} >_{Poverty} + \ni_{23} >_{Race} + *_2$$

$$0_{Class\ size} = \ni_{31}\, 0_{School\ size} + \ni_{32} >_{Language} + \ni_{33} >_{Poverty} + \ni_{34} >_{Race} + *_3$$

$$0_{Tchr\ qualifications} = \ni_{41}\, 0_{School\ size} + \ni_{42} >_{Language} + \ni_{43} >_{Poverty} + \ni_{44} >_{Race} + *_4$$

---

[31] $0_{organization}$ and $>_{background}$ consist of multiple individual factors and therefore, numerous measurement-level specifications are included in this model, as shown in figure 3. Here are the measurement specifications, using the > factor "poverty" as an example:

$$X_{R.Lunch} = 8_1 >_{Poverty} + ,_{x1}$$

$$X_{Poverty.Problem} = 8_2 >_{Poverty} + ,_{x2}$$

Where, $X_{R.Lunch}$ is the observed indicator of $>_{Poverty}$ that identifies the ratio of students that are eligible for the school's lunch program, 8_1 is the coefficient relating to $>_{Poverty}$, and $,_{x1}$ is the related measurement error term; and $X_{Poverty.Problem}$ is the observed indicator of $>_{Poverty}$ that identifies the average teachers' response to whether poverty is a problem at the school. Note that SASS teacher responses that identify prevalent problems, such as this observed indicator of poverty, are also assumed to be affected by a general "teacher attitude and opinions" latent variable, as indicated in figure 3.

Similar to the identification of exogenous latent variables, the measurement-level specification for the achievement factor is:

$$Y_{Read.Achieve} = 8_{Y1}\, 0_{Achievement} + ,_{Y1}$$

$$Y_{Math.Achieve} = 8_{Y2}\, 0_{Achievement} + ,_{Y2}$$

Where, $Y_{Read.Achieve}$ is the observed indicator of the latent endogenous factor student achievement ($0_{Achievement}$), 8_{Y1} is the coefficient relating to $0_{Achievement}$, and $,_{Y1}$ is the measurement error term. Similarly, $Y_{Math.Achieve}$ is the second observed indicator of the latent endogenous factor student achievement ($0_{Achievement}$), 8_{Y2} is its coefficient relating to $0_{Achievement}$, and $,_{Y2}$ is its measurement error term.

$$0_{\text{Norm cohesion}} = \ni_{51} 0_{\text{School size}} + \ni_{52} >_{\text{Language}} + \ni_{53} >_{\text{Poverty}} + \ni_{54} >_{\text{Race}} + *_{5}$$

$$0_{\text{Tchr control}} = \ni_{61} 0_{\text{School size}} + \ni_{62} 0_{\text{Norm cohesion}} + \ni_{63} >_{\text{Language}} + \ni_{64} >_{\text{Poverty}} + ,_{65} \varsigma_{\text{Race}} + *_{6}$$

$$0_{\text{School climate}} = \ni_{71} 0_{\text{Tchr control}} + \ni_{72} 0_{\text{School size}} + \ni_{73} 0_{\text{Class size}} + \ni_{74} 0_{\text{Norm cohesion}}$$

$$+ \ni_{75} >_{\text{Language}} + \ni_{76} >_{\text{Poverty}} + \ni_{77} >_{\text{Race}} + *_{7}$$

Because the data are correlational, it should be pointed out that the analysis is also consistent with views that characteristics such as school climate, cohesive norms, cooperation, and satisfaction are affected by the kinds of skills and attitudes that children bring to the school, which are best reflected in achievement scores. For example, if the student peer norm is to focus on class work, there is likely to be less of a problem with absenteeism, tardiness, and class cuts. Likewise, a negative correlation between average class size and average achievement may indicate that smaller classes facilitate achievement, but it may also be due to a socioeconomic variation between school communities, which affects both class size and achievement. Nevertheless, the methodology for the analyses is a variant of linear modeling, with asymmetric "independent" and "dependent" variables. Therefore, although the correlational results may appear to be couched in terms of "effects" of some factors on others, these "effects" merely indicate the partitioning of variance in the "dependent" factors into covariances with "independent" factors. This use of the term "effect" should not be confused with its use in the description of causal relations. Without controlled experimental studies, the direction of causality cannot be inferred, only conjectured.

In addition to SEM analyses and in order to verify the robustness of the findings, we also carried out ordinary least squares (OLS) linear regression analyses.[32] For these analyses, we created composites for poverty, English language proficiency, race/ethnic minority, teacher qualifications, school climate, school size, class size, teacher control, and normative cohesion, using the same measures as in the SEM analyses. The linear regression equations are exactly the same equations as the structural equations given above, with factors replaced by the explicit composites indicated in figure 3. Each composite was created by adding together the corresponding measures after scaling each in standard deviation units. Separate regressions were performed for elementary, middle, and secondary schools. Finally, because this report was not focusing on differences in achievement correlates between states, the powerful analytical technique of hierarchical linear modeling (HLM) was not used. Hierarchical modeling would be a powerful tool in the use of these data to study the effects of state educational policies on school

[32] The major difference between the two methodologies is in how they treat measurement error in predictive measures. The basic assumption of OLS linear regression is that predictors are measured "without error." That is, each estimated coefficient represents the extent to which a measure, *per se*, not an underlying construct that it measures, accounts for variation in the dependent variable. SEM analysis, on the other hand, accounts for measurement error and allows for the specification of correlations among both the constructs and the measurement errors. SEM's greater flexibility must be weighed not only against the additional computational complexity, but also against the additional complexity of interpretation. Relations among latent variables do not have the same simplicity as relations among observable variables. Therefore, our approach is to examine and compare the results of both methodologies to identify robust patterns in the correlates of school-level achievement.

factors and achievement.

Weights and Standard Errors

In order to provide the basis for publishing state-by-state figures, the SASS sample of public schools is selected as a stratified random sample, with similar numbers of schools in each state. One result of this sampling method is that to support national estimates, SASS schools in states with a population of many schools have substantially higher weights than SASS schools in states with fewer schools. Although this variation in weights removes bias in national estimates, it also reduces the precision of (weighted) estimates. For research goals of understanding relations among school factors, rather than of generating unbiased national estimates of counts and averages, the use of weights is not only unnecessary, but also limits the power of the SASS data to address the goals.

Therefore, to identify generalizable correlates of school average student achievement, analyses were carried out without use of SASS weights. This has the effect of weighting less populous states equally with more populous states; but this is more likely to affect the estimates of the sizes of effects, not the outcomes of tests of the statistical significance of overall relations. In fact, the unweighted distribution of the grade levels and locales of the schools on the SASS student achievement subfile more closely matched the total population of public schools than did the weighted distribution of those schools. Evidently, the categories of schools with the largest SASS weights also tended to be the categories most likely to have state assessment data.

The value of the SASS student achievement subfile depends on whether meaningful patterns of statistically significant results emerge from analyses of relations between SASS measures and achievement. Thus, assessment of statistical significance is central to the study. Unfortunately, usual tests of statistical significance available in common statistical packages are based on the assumptions of simple random sampling, and the SASS student achievement subfile is far from a simple random sample. For estimation of standard errors, SASS provides replicate weights. However, because we did not use weights, the SASS replicate weights could not be used to estimate standard errors. Since the SASS student achievement subfile is stratified by state and there are definite between-state components of variance in the measures used,[33] estimates of standard errors based on simple random sampling formulas tend to underestimate true standard errors by factors of as much as two. To compute better estimates of standard errors, we carried out analyses on repeated random half-samples of the schools on the subfile. Specifically, 100 randomly selected half-samples of states were selected; the analytical results were obtained for each half-sample; and the standard deviation of these results served as the estimate of the standard errors.[34] It should be noted that this method of estimation of standard errors does not

[33] In particular, the achievement measures were constructed specifically to include a component of between-state variance based on State NAEP results.

[34] The loss of precision from using half-samples instead of the full sample essentially exactly compensated for the fact that the repeated samples were each half of the "population" from which they were sampled. Both the sample (continued)

treat the 20 states as consisting of a universe of "strata," but rather as a set of "clusters" of schools, supporting generalization to a superpopulation of states that are like those included. If one were to use this database merely to generalize to schools in the selected 20 states, smaller estimates of standard errors would be obtained, because between-state variation would be treated as fixed, rather than random. Those standard errors could be estimated by replicating half-samples of schools within each state, rather than half-samples of states.

---

size effect and the finite sampling correction involved a factor of two.

Results

Two major questions are addressed by these analyses of the SASS student achievement subfile:

> Are organizational factors and school behavioral climate correlates of school mean assessment scores?[35]
>
> How is empirical evidence on the correlates of achievement affected by the removal of between-state variation from measures?

The value of the SASS student achievement subfile is tested by the answers it gives to the first question. Although the analyses reported here merely scratch the surface of the potential for analyses of these data, they should provide evidence of a meaningful pattern of relations between school-level factors and assessment scores. The second question addresses the issue of the importance of using the State NAEP results to represent the between-state variation in achievement which state assessments cannot capture.

Relations among factors characterizing schools are likely to be attenuated by measurement error in the measures of those factors, as well as by less than perfect test score reliability; however, because the SASS database contains multiple measures of each factor and state assessment scores are available for two different areas, math and reading/language arts, structural equation modeling (SEM) can be used to reduce the analytical "noise" caused by measurement error. The SEM estimation program in SAS (PROC CALIS) was used to find the best fitting model for these data, but because the data were clustered in states, standard error estimates produced by that program were only useful as general guidelines for exploratory analyses. Estimates of standard errors used in tests of statistical significance were obtained by a series of 100 random half-sample replications of the estimation program—half samples were drawn at the state level to incorporate between state variation in the estimation of standard errors.

To corroborate the results of SEM analyses, as well as to bridge the analyses at the individual state level, for which the database was not sufficiently large for stable SEM estimation, ordinary least squares (OLS) linear regression model estimation was carried out for the same models. In the OLS analyses, the SEM latent factors were replaced by balanced linear combinations of the same measurements used in the SEM analyses to define the factors.

Because a substantial percentage of variation in school characteristics is between elementary and middle and middle and high schools, meaningful analyses must compare only schools at the same level. Therefore, all analyses were carried out separately for elementary, middle, and secondary schools.[36] For the three separate levels, the same qualitative SEM

[35] Originally, teacher qualifications were also to be included in this question. However, the data did not support the use of the proposed composite of average sampled teacher's education and experience.

[36] Files were also created for combined-grade (K–12) schools. However, the analyses for combined-grade schools are not reported, due to problems in obtaining a good model fit.

specifications were used, although different parameter estimates were obtained.

Two categories of school-based factors are included in these analyses: school behavior climate, based on teachers' perceptions of problems in the school, and four organizational characteristics, school and class sizes, normative cohesion, and teachers' sense of control and influence. As a first step in exploring the relations between these factors and achievement scores, partial correlations of the school-based SASS factors with achievement and with each other are shown in table 4. These partial correlations represent the bivariate relations among the factors, partialing out the three background factors of poverty, language barriers, and race/ethnicity.

Although the partial correlations are neutral with respect to direction of causality, they do not capture multivariate relations among the school-based factors. For example, are the negative partial correlations of school and class size with achievement in middle schools independent effects, or is one of the correlations a byproduct of the high intercorrelation between school and class size? A more interpretable picture of the interrelations among the factors can be obtained from multiple regression and structural equation modeling.

Ordinary least squares (OLS) multiple regression is a second step in the analysis of school-based correlates of achievement. Separate equations can be modeled for a composite of reading and mathematics achievement and for composites of measures representing school climate, normative cohesion, teachers' influence, class size, and school size. (In this analysis, the observed measures were weighted equally in defining each composite, by transforming each to a standard score before averaging.) Unlike partial correlations, multiple regression differentiates between "predictor" and "dependent" variables, although the interpretation of equations remains merely that a combination of predictors is correlated with the dependent variable. Estimates of standardized regression coefficients for the equations indicated in figure 2 are shown in table 5.[37]

An important limitation of OLS regression is that while variance in each (intermediate) endogenous factor is partially accounted for by other factors, each factor is assumed to be measured without error in predicting other factors. For survey measures, that assumption is not supportable, and a preferable method of analysis is structural equation modeling, which takes measurement error into account in estimating the proportions of variance in factors accounted for by other factors. Estimates of coefficients for the SEM corresponding to figure 2 are shown in table 6.[38]

If all of the school-based predictors in the OLS regression equations were uncorrelated with each other, then the standardized regression weights displayed in table 5 would, to a first

[37] The same factor definitions are used in tables 4 and 5. Estimates of coefficients for background factors are not shown in table 3, because the coefficients are not germane to the exploration of school-based correlates of achievement.

[38] The corresponding structural equation coefficients for the background factors are not shown in table 6, because they are not germane to the exploration of school-based correlates of achievement. As expected, SEM analyses confirm that all measurement variables are indeed significantly related with the corresponding factors, as indicated in figure 3. See appendix A for results of SEM on the measurement level.

approximation, be the same as the partial correlations displayed in table 4. Furthermore, *if* all of the indicators of each factor in the SEM analysis were perfectly intercorrelated, then the results in table 6 should be approximately the same as those in table 5. (The approximation would not be exact, because the SEM equations are estimated jointly, while the OLS equations are estimated separately.) Thus, substantial differences in corresponding coefficients between these tables demonstrate the impact of measurement error and multivariate interactions.

Although SEM is the preferred method of analysis for such analyses, because it takes into account the measurement error in predictors, estimation of SEM coefficients requires a large number of degrees of freedom, and the resulting estimates can be sensitive to intercorrelations between the indicators of the various factors. Thus, to assess the robustness of the SEM results, it is useful to compare the results from the three approaches. Where they are in conflict, further study of the fit of the data to the assumptions of the method is warranted.

For example, compare the SEM and OLS regression results for the relations between achievement, school size, and class size in middle schools. In table 5, we see that according to OLS regression analysis, neither is a significant relation; but in table 6, the structural relation between the class-size factor and achievement is significant. This apparent conflict is due to the fact that the school-size and class-size factors are so closely correlated that their individual coefficients cannot be stably estimated by OLS regression, given the size of the sample.

SEM results must be interpreted carefully however, because the apparent implied causal direction in the regressions is not valid. Simple interpretation of the results in table 6 requires that one assume that the arrows in the causal model shown in figure 2 imply causality.[39] Any significantly positive or negative coefficient might also reflect either a causal relation in the opposite direction or the existence of a third factor that has a causal relation to both factors.[40, 41]

Simultaneously interpretation of the results presented in tables 4, 5, and 6 provides information about the school-based correlates of achievement and about the extent to which measurement error and correlated predictors can distort analytical results. Through these analyses, we address the following question:

> Are school climate, class size, school size, and teachers' perceptions of normative cohesion and sense of control statistically significantly related to reading and mathematics achievement, based on the SASS student achievement subfile?

For the purposes of inferring statistical significance, standard errors of the tabulated estimates were estimated by replicating the analyses on 100 random half samples of the database. To

[39] The selection of factors for exclusion from particular structural equations was based on trial estimation runs in which the goodness of fit index was improved by their deletion.

[40] Table 6 does not include the coefficients for the measurement equations portrayed in figure 3. These coefficients are presented in appendix A.

[41] Parameter estimates for a variant of the model in figure 2, in which the positions of school climate and achievement are reversed, are given in appendix B.

incorporate the design structure, each half sample included either all or none of the SASS schools in a particular state. Because this is an exploratory study based on 20 states, some results are described as "approaching significance" when corresponding Student's t-values are near ∀1.96. These are results which might well be significant at the .05-level if a few more states were added to the database.

Although a teacher qualification factor is included in the model shown in figure 2 as noted previously, the reliability and coherence of the SASS school-level composite of teacher experience and education, as a characteristic of a school (but based on a small sample of teachers) is weak. Unlike the factors based on teachers' perceptions (school climate, normative cohesion, and sense of control), which can have a communality based on common perceptions of the school as a workplace environment, teachers' experience and education are naturally highly variable within a school as new teachers are continually added to replace highly experienced teachers who retire. Investigation of these teacher qualification indicators as correlates of achievement appears to require achievement data at the individual classroom level.

In fact, the SEM analyses frequently failed to converge when the teacher qualification factor was included, and when they did, the results just as frequently indicated a negative relation as a positive one. Therefore, the teacher qualification factor has been omitted from the anlayses. In the next sections, we evaluate the evidence that school climate and organizational characteristics are correlates of achievement at the school level.

School Climate

The partial correlations in table 4 indicate that a positive school behavioral climate is a correlate of higher average achievement at the middle and secondary levels, although there is not a significant relationship at the elementary level. The increasing correlation over grade levels is expected, because the frequency of behavior problems that are included in the SASS questionnaire increases with students' ages. It may be that in elementary schools, serious problems are sufficiently rare that no pattern of relations with achievement can emerge.

Although the OLS regression results in table 5 confirm this finding, the SEM results in table 6 do not. The reason for this lies in the fact that the school climate factor is, itself, highly predictable from the other factors. The inclusion of the structural equation predicting the school climate factor causes the other factors to be estimated as slightly more similar to school climate than they would otherwise be. Thus, in the structural equation for achievement, they jointly take the place of school climate. In the OLS regression equation for achievement, on the other hand, the school climate composite is correlated with achievement over and beyond its relations with other composites.

Table 4.— Partial correlations of organizational and climate factors associated with student achievement in public elementary, middle, and secondary schools, controlling for background factors

Factor 1	Factor 2	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
School size	**Student achievement**	–0.10	–0.19*	+0.03
Class size		–0.06	–0.19*	–0.13*
Normative cohesion		+0.00	+0.01	–0.08*
Teachers' influence		+0.04	+0.16*	+0.17
School climate		+0.03	+0.09*	+0.11*
	School climate			
School size		–0.14*	–0.26*	–0.36*
Class size		–0.15	–0.24*	–0.25*
Normative cohesion		+0.34*	+0.41*	+0.44*
Teachers' influence		+0.18*	+0.20*	+0.26*
School size	**Teachers' self-perceptions of influence**	–0.13*	–0.14*	–0.22*
Class size		–0.09	–0.03	–0.13*
Normative cohesion		+0.35*	+0.27*	+0.26*
School size	**Normative cohesion**	+0.02	–0.01	–0.09*
Class size		–0.07	–0.03	–0.07
	Class size			
School size		+0.33*	+0.48*	+0.54*

NOTES: Table entries are partial correlations, partialing out poverty, language, and race/ethnic factors.
(*) $p<.05$ based on repeated half-sample standard deviations.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and state reading and mathematics assessment scores for publics schools in 20 states.

Table 5.—OLS standardized regression weights for organizational and climate factors associated with student achievement in public elementary, middle, and secondary schools

Independent effects	Dependent factor	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
	Student achievement			
School size		–0.07	–0.07	+0.18*
Class size		–0.02	–0.10	–0.15*
Normative cohesion		–0.01	–0.05	–0.15*
Teachers' influence		+0.02	+0.11*	+0.15
School climate		+0.00	+0.03	+0.16*
r^2		0.46	0.55	0.51
	School climate			
School size		–0.09*	–0.15*	–0.23*
Class size		–0.06	–0.11	–0.06
Normative cohesion		+0.25*	+0.29*	+0.31*
Teachers' influence		+0.04	+0.05	+0.08
r^2		0.51	0.58	0.57
School size	**Teachers' self-perceptions of influence**	–0.15*	–0.15*	–0.21*
Normative cohesion		+0.36*	+0.26*	+0.22*
r^2		0.18	0.15	0.20
	Normative cohesion			
School size		+0.02	–0.01	–0.10*
r^2		0.07	0.02	0.03
	Class size			
School size		+0.35*	+0.51*	+0.55*
r^2		0.14	0.28	0.38

NOTES: (*) p<.05 based on repeated half-sample standard deviations.
r^2 values include effects of three background factors in addition to factors shown.
SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and state reading and mathematics assessment scores for publics schools in 20 states.

Table 6.—SEM associations of organizational and climate factors with student achievement in public elementary, middle, and secondary schools

Independent effects	Dependent factor	Elementary (n = 1123)	Middle (n = 496)	Secondary (n =5 95)
	Student achievement			
School size		+0.04	+0.06	+0.32*
Class size		–0.25	–0.38*	–0.36*
Normative cohesion		–0.06	–0.01	–0.06
Teachers' influence		+0.03	+0.07	–0.01
School climate		–0.11	–0.05	–0.08
r^2		0.72	0.87	0.82
	School climate			
School size		–0.09	–0.14	–0.28*
Class size		–0.12	–0.18	–0.00
Normative cohesion		+0.26*	+0.33*	+0.29*
Teachers' influence		+0.01	+0.09	+0.13
r^2		0.58	0.62	0.68
	Teachers' self-perceptions of influence			
School size		–0.11	–0.06	–0.20*
Normative cohesion		+0.39*	+0.35*	+0.31*
r^2		0.22	0.26	0.45
	Normative cohesion			
School size		–0.02	–0.11	–0.18*
r^2		0.09	0.03	0.09
	Class size			
School size		+0.47*	+0.57*	+0.63*
r^2		0.32	0.42	0.51
	Statistical Summary Measures			
	GFI (AGFI)	0.96 (0.93)	0.92 (0.87)	0.93 (0.88)
	Π2 (d.f.)	426 (100)	395 (100)	432 (100)

NOTES: Entries in table standardized gamma coefficients.

(*) $p<.05$ based on repeated half-sample standard deviations.

r^2 values include effects of three background factors in addition to factors shown.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

So, is school climate a correlate of achievement in middle and secondary schools? Should educators expect that improving school climate might contribute to improved test scores? Examination of the correlates of school climate in tables 4, 5, and 6 sheds some light on this issue. The results from all three sets of analyses suggest that school behavior climates are better in schools with high normative cohesion (i.e., where teachers feel that they have common goals and cooperate) and in smaller schools, especially among students in higher grade levels. Although cohesion and climate are correlated with each other, the partial correlation of achievement with cohesion is smaller than its correlation with climate, so in the OLS regression, cohesion becomes a moderator, with a negative coefficient, magnifying the positive coefficient for climate. Two factors in the SEM model eliminate the effect, suggesting that it is an artifact of OLS regression. First, the contribution of normative cohesion to school climate is explicitly taken into account, and second, the "method" factor representing teachers' positive or negative response tendencies is included. The ambiguity of this relation is reinforced by the reversal of the equations for school climate and achievement. The fit of the data to the reversed model, shown in appendix B, is essentially the same as the fit of the data shown in table 6.

The message from these analyses, based on the assumptions embodied in figure 2, is that steps to increase normative cohesion among staff may go hand-in-hand with improving school climate, but that these factors are not strong correlates of assessment scores when measurement error is taken into account.

Class Size

As predicted by most current research, reduced *class size* is related to higher academic performance. Blatchford and Mortimore (1994) offer a number of possible explanations for the relationship between class size and student achievement. They point out that small class size may facilitate a more individualized and more effective instruction, more complete curriculum coverage, and greater student involvement in classroom activities. Such an environment of personal attention also shapes a more settled and more orderly school environment.

In view of this, it is not surprising that the partial correlations in table 4 are significant at the middle and secondary school levels, the regression coefficient in table 5 at the secondary level is significant, and the SEM coefficients in table 6 are large and significant or nearly significant at all three grade levels.

It seems clear that average class size in a school is a correlate of achievement among secondary schools, and the inconsistency of the evidence at the middle school level is minor, since the nonsignificant OLS regression weight is in the consistent direction, with $t = -1.33$. At the elementary level, the evidence is somewhat weaker, in that, although the SEM coefficient approaches significance ($t=-1.93$), the partial correlation coefficient ($t=-0.57$) and the OLS regression coefficient ($t=-0.09$) are not different from random variation. This appears to be an instance where the flexibility of the SEM method may enable it to uncover a relation that is hidden by the imperfect relations between observed indicators and underlying factors. Partial correlations and regression equations are based on predefined unweighted averages of the indicators of achievement, reading and mathematics scores, and of the class size measures; but SEM can give greater weight to the more predictable measure of the achievement factor, which in elementary schools is reading, as opposed to mathematics, and to the more predictable class-size measures, which are student/teacher ratios and teachers' satisfaction with class size, rather than teachers' reports of average class size. (See Appendix A for the relative weights given to the measures in the SEM analysis.) Subsequent SEM analyses for elementary schools corroborate this by indicating that the class size factor accounts for a statistically significant percentage of the variance in reading scores (McLaughlin and Drori, 1998).

School Size

The SEM and OLS results in tables 5 and 6 both indicate that there is higher achievement in larger secondary schools, but that there is no significant relation between school size and achievement among elementary and middle schools. The bivariate partial correlations shown in table 4 generally indicate a more negative relation between school size and achievement than do the SEM and OLS multivariate analyses. This is because the correlations fail to take into account a confounding measure: school size and class size are highly correlated factors, and the negative partial correlation between school size and achievement among middle schools is due to the significant correlation of both of these factors with class size.

School size also appears to be a correlate of school climate—smaller schools tend to have better behavioral climates, as perceived by teachers. However, according to the SEM analysis, this relationship is only statistically significant at the secondary school level ($t_{elementary}=-1.64$, $t_{middle}=-0.35$, $t_{secondary}=-3.44$). A similar finding holds for teachers' perceptions of control over their classrooms and influence on school policies: according to the bivariate partial correlations and OLS regressions, perception of influence is greater at small schools at all levels, and the SEM results are in the same direction, but not statistically significant. This may be a case in which the database is not sufficiently large for the SEM methodology to identify a relation as significant: due to the flexibility of the method, fits to different half-samples of states yield more varying parameter estimates than ordinary least squares or bivariate correlations would indicate.

To summarize the correlates of school size, we find that although teachers in small schools perceive fewer problems, test scores are not higher, other things equal. On the contrary, in secondary schools, higher average test scores are found in larger schools. These findings are not fully consistent with the latest wave of empirical research (for example, Lindsay 1982, Pittman and Haughwout 1987, Lee and Smith 1993, 1995, albeit in schools other then at secondary level) and the reform agenda raised on their basis (e.g., Oxley 1994, Boyer 1995). While these studies refer to school "smallness" as an advantage and a desired quality, in spite of its merely indirect effect on achievement, our findings suggest that (1) larger secondary schools may have resources that contribute to higher achievement and (2) any relation between small enrollments and high achievement in elementary and middle schools may be explainable in terms of the smaller class sizes in these schools.

Normative Cohesion

Normative cohesion among the staff members is centered on organizational culture, rather then a structural feature like size. In these analyses it is identified with a combination of perceived clarity of school norms and perceived cooperation among staff. While the expectation is that the cultural cohesion establishes a stable foundation for performance, the SASS student-achievement subfile reveals no such relationship between school staff cohesion and student academic achievement. In fact, there is a suggestion of a negative relationship at the secondary level. Yet, school staff cohesion is a strong correlate of school behavioral climate and teachers' sense of influence over school affairs. One explanation of this paradoxical situation may be that administrators in some lower achieving schools feel a greater need to convey clear norms to teachers and to foster cooperation.

Sense of Control and Influence

According to bivariate partial correlations and OLS regression results, teachers' sense of control over classroom practices and influence in shaping policy in the school appear to form a significantly positive correlate of achievement in middle schools, and in secondary schools, the relation approaches significance ($t_{correlations}$=1.88, $t_{regression}$=1.87). However, SEM analyses do not find this relationship to be significant, suggesting that the relation may be due to interactions with other measures, such as normative cohesion and school size.

The finding that normative cohesion is a positive correlate of sense of control and influence is consistent with Ingersoll's (1996) conclusion that teachers' autonomy and influence "make an important difference for the amount of cooperation or conflict in schools" (p. 171). Relying on similar items in the SASS 1987–88 survey, he adds that this relationship varies by the locus of teachers' control: when locus is fundamentally social (i.e., "selection, maintenance, and transmission of behaviors and norms;" p. 171), rather than concerned with curriculum and instruction (i.e., "selection of textbooks, topics, materials, and teaching techniques;" p. 171), then the association between teachers' lack of power and conflict in school is strongest.

Overall, there is evidence in the SASS student-achievement subfile that organizational characteristics of schools are correlates of student achievement. SASS offers an abundance of opportunity to assess organizational characteristics in American schools (see Baker 1996). Hence, while we examined merely four organizational characteristics in this report, future work may concentrate on such organizational features as organizational inertia (for example, in regards to personnel tenure); organizational change (for example, in regards to reform issues); or autonomy (both an intra-organizational feature—room for initiatives—or an inter-organizational features—dominance of the school district over major decision-making issues).

The interpretations of the results in tables 4, 5, and 6 as they relate to the model in figure

2 are limited by the lack of a valid and reliable school-level teacher qualification factor. It is likely that some of the variance in school-level achievement is explained by average differences in teachers' qualifications and that some of the covariance with other factors is mediated by the missing teacher qualification factor. Addition of data from other sources, such as information on the selectivity of the colleges from which the teachers were graduated, or district per-pupil expenditures for instruction, adjusted for cost-of-education variations, may fill this gap. Otherwise, study designs such as NELS:88 and NAEP (when specific teacher questionnaires were matched to specific students' performance) are more appropriate for assessing the correlation between teacher qualifications and achievement.

Comparisons of Between- and Within-State Correlates of Achievement

The findings presented in the preceding section are based on achievement measures which combine (a) state assessment information for within-state variation and (b) State NAEP information for between-state variation. These analyses could also be carried out relying purely on the state assessment information, essentially aggregating the results of 20 separate within-state analyses into summary findings. Within-state analyses would be less powerful, to the extent that relations of interest are between-state relations. For example, if class sizes are mandated to be fairly uniform within states, then most of the variation in class sizes may be between states. In that case, while analyses using State NAEP information uncover a relation between class size and achievement, similar within-state analyses would not. Considering variation in achievement correlates to occur at three levels: between states, between schools in the same state, and between students in the same school, pooled within-state analyses would be sensitive only to relations at the second of these levels.

On the other hand, the inclusion of between-state variation based on State NAEP has two costs. First, additional assumptions are required because the only State NAEP information available for the 1993–94 school year is in fourth-grade reading. The generalization of the results to middle and secondary schools, and to an achievement factor that includes mathematics, assumes that between-state variation in achievement is relatively uniform across grades and between reading and mathematics. As noted in the discussion of the development of this measure, comparisons of state means from State NAEP in 1992 and 1996 support this assumption. Furthermore, the results presented in tables 4, 5, and 6 indicate that the results are not attenuated at the middle and secondary level relative to the elementary level, contrary to expectations (if the extrapolations to higher grades were not valid). In fact, there are more statistically significant relations at the secondary level than at the elementary level.

Second, the effective sample sizes may be dramatically reduced when between-state variation is taken into account. Although there are 1,123 elementary schools, 496 middle schools, and 595 secondary schools in the database, they are found in only 20 states (in only 17 states for middle and secondary schools). If the primary variation in achievement or in SASS measures is between states, then the effective sample sizes may be closer to 20 than to 500, 600, or 1,100. The question of how great the reduction in sample size is, depends on the percentages of variance of the estimates that are between- and within-state. These percentages and the corresponding effective sample sizes for means are shown in table 8.[42] When compared with the effective sample sizes for pooled within-state analyses, which are essentially the number of schools, minus the number of states, it becomes apparent that if the design effects on the SEM estimates, OLS regression weights, and partial correlations are as great as the design effects on mean estimates, then much larger effects are needed for statistical significance in the analyses presented in the preceding section than would be needed in pooled within-state analyses.[43]

The largest relative between-state variances in this sample are in achievement, school sizes, and class sizes, as well as percentages of minorities and of limited English proficient (LEP) students. The percentages of minorities and LEP students were included in all of the analyses, and school size was included in all of the SEM and OLS regression analyses. On the other hand, most of the variation in teachers' perceptions of school behavior climate and normative cohesiveness was between schools in the same state, not between states; and the percentage of students eligible for the national free lunch program was also nearly uniform across the 20 states.

To determine whether the information added by using State NAEP to capture between-state variation compensates for the reduction in effective sample sizes in partial correlation, SEM, and OLS regression estimation, parallel pooled within-state analyses were carried out by subtracting state means from all measures used in the analysis. The results are displayed in tables 8, 9, and 10, which parallel tables 4, 5, and 6.

First, the pooled within-state results which involve only SASS measures, not the achievement measures, follow patterns similar to the patterns when between-state variance is included. That is, the correlations between school size and average class size, and among teachers' perception of influence, normative cohesion, and school climate, are positive whether or not between-state variance is considered.

Second, the within-state relations of school-based factors to achievement differ from the overall relations, because the overall relations are sums of between- and within-state components. Therefore, to interpret the differences between tables 8, 9, and 10 and tables 4, 5, and 6, we must simultaneously examine the between-state relations. With few states, regression estimates based on states as units are not stable; however, the standardized coefficients in these regression functions, shown in table 11, help to explain the results in tables 8, 9, and 10.

[42] Some random variation between states would be expected, as a result of between-school variation. However, the proportion of variance that is systematically between states is estimated from a comparison of the variance of estimated state means to the pooled within-state variance, taking into account the effective sample size (number of schools) in each state sample and subtracting the expected random variation between states.

[43] That is, the standard errors computed in testing for statistical significance by repeatedly randomly selecting half of the states for analysis were much larger than they would if there were no systematic variation between states. Generally, standard errors are proportional to the reciprocal square root of the effective sample sizes, other things equal.

Table 7.—Percentage of between-school variance that is systematically between states

	Elementary (n = 1,123)		Middle (n = 496)		Secondary (n = 595)	
	Percent between state variance	Effective sample size	Percent between state variance	Effective sample size	Percent between state variance	Effective sample size
Achievement	32 %	53	30 %	43	39 %	25
Poverty	8 %	199	8 %	132	8 %	104
Racial/ethnic minority	29 %	62	33 %	39	48 %	20
Language barriers	27 %	65	13 %	88	34 %	28
School size	27 %	65	48 %	28	25 %	40
Class size	34 %	53	50 %	26	48 %	20
School climate	6 %	246	9 %	120	11 %	78
Teachers' perception of control	12%	133	19 %	64	28 %	35
Teachers' perception of cohesiveness	2 %	488	8 %	125	6 %	140

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table 8.—Pooled within-state partial correlations of organizational and climate factors associated with student achievement in public elementary, middle, and secondary schools, controlling for background factors

Factor 1	Factor 2	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
	Student achievement			
School size		+0.03	–0.06	+0.12
Class size		+0.02	–0.09*	+0.01
Normative cohesion		+0.08*	+0.09	–0.02
Teachers' influence		+0.00	+0.08	+0.10
School climate		+0.10*	+0.23*	+0.18*
	School climate			
School size		0.18*	–0.21*	–0.35*
Class size		–0.03	–0.19*	–0.26*
Normative cohesion		+0.33*	+0.44*	+0.43*
Teachers' influence		+0.22*	+0.29*	+0.37*
	Teachers' self-perceptions of influence			
School size		–0.15*	–0.13*	–0.23*
Class size		–0.14*	–0.13	–0.25*
Normative cohesion		+0.39*	+0.41*	+0.44*
	Normative cohesion			
School size		–0.06	–0.02	–0.16*
Class size		–0.07*	–0.08*	–0.14*
	Class size			
School size		+0.32*	+0.35*	+0.55*

NOTES: Table entries are partial correlations, partialing out poverty, language, and race/ethnic factors.
(*) $p<.05$ based on repeated half-sample standard deviations.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table 9.—Pooled within-state OLS standardized regression weights for organizational and climate factors associated with student achievement in public elementary, middle, and secondary schools

Independent effects	Dependent factor	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
	Student achievement			
School size		+0.03	–0.00	+0.21*
Class size		+0.01	–0.03	–0.04
Normative cohesion		+0.05	–0.01	–0.13*
Teachers' influence		–0.03	+0.01	+0.08
School climate		+0.10*	+0.22*	+0.28*
r^2		0.41	0.52	0.40
	School climate			
School size		–0.13*	–0.12*	–0.20*
Class size		+0.04	–0.07*	–0.04*
Normative cohesion		+0.23*	+0.29*	+0.26*
Teachers' influence		+0.07	+0.07*	+0.14*
r^2		0.54	0.62	0.55
	Teachers' self-perceptions of influence			
School size		–0.13*	–0.13*	–0.17*
Normative cohesion		+0.40*	+0.42*	+0.42*
r^2		0.19	0.20	0.24
	Normative cohesion			
School size		–0.06	–0.02	–0.16*
r^2		0.11	0.05	0.09
	Class size			
School size		+0.32*	+0.36*	+0.57*
r^2		0.11	0.14	0.34

NOTE: (*) $p<.05$ based on repeated half-sample standard error estimates.
r^2 values include effects of three background factors in addition to factors shown.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table 10.—Pooled within-state SEM associations of organizational and climate factors with student achievement in public elementary, middle, and secondary schools

Independent effects	Dependent factor	Elementary (n = 1123)	Middle (n = 496)	Secondary (n = 595)
	Student			
School size	**achievement**	+0.05	–0.04	+0.28*
Class size		–0.13	–0.01	–0.17*
Normative cohesion		–0.04	–0.11*	–0.17
Teachers' influence		+0.00	+0.03	+0.12
School climate		–0.04	+0.21*	+0.21
r^2		0.63	0.78	0.69
	School climate			
School size		–0.20*	–0.00	–0.27*
Class size		+0.17	–0.20	+0.00
Normative cohesion		+0.24*	+0.36*	+0.22*
Teachers' influence		+0.08	+0.07	+0.25*
r^2		0.63	0.66	0.66
	Teachers' self-perceptions of			
School size	**influence**	–0.10	–0.07	–0.19*
Normative cohesion		+0.46*	+0.49*	+0.57*
r^2		0.25	0.28	0.44
	Normative			
School size	**cohesion**	–0.10	–0.11*	–0.25*
r^2		0.13	0.07	0.13
	Class size			
School size		+0.49*	+0.57*	+0.72*
r^2		0.37	0.34	0.56
	Statistical Summary Measures			
	GFI (AGFI)	0.97 (0.94)	0.95 (0.91)	0.94 (0.90)
	Π2 (d.f.)	365 (100)	270 (100)	329 (100)

NOTES: Entries in table are standardized gamma coefficients
(*) $p<.05$ based on repeated half-sample standard deviations.
r^2 values include effects of three background factors in addition to factors shown.
SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table 11.—Standardized regression weights for OLS school-based predictors of achievement, with state as the unit

	Elementary schools (n=20)	Middle schools (n=17)	Secondary schools (n=17)
School size	-.05	-.20	+.24
Class size	-.37	-.32	-.47
Normative cohesion	-.04	+.10	+.52
Teachers' influence	+.11	+.37	+.15
School climate	-.31	-.31	-.36
r^2	.76	.78	.68

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

The most noticeable differences between the overall and within-state correlates of achievement are (a) the weaker relation between smaller classes and higher achievement in the pooled within-state results than it is overall and (b) the more positive relation between school climate and achievement in the pooled within-state results. In both cases, these correspond to the pattern of negative between-state coefficients for class size in table 11. The correlation between class size and achievement appears primarily to be a between-state phenomenon. As can be seen in table 7, a large percentage of the variation in the class size factor is between states, suggesting that most of these states have systems that minimize variation on class size between schools at the same level. Based on the logic of the path diagram in figure 2, if one were to propose a national initiative to lower class sizes to a particular level, its effect on achievement would be felt primarily in the states with largest average class sizes.[44]

The relation between school climate and achievement follows a different pattern. Overall, the relation is not statistically significant (tables 4, 5, and 6); but within states there appears to be a positive relation. Balancing the positive within-state relation is a pattern of negative school climate coefficients in table 11. If we consider each state to provide a common context for its schools, then within that context, a better behavioral climate and higher scores occur at the same schools. However, as contexts change between states, there is no relation (or even a slightly

[44] Since these data were collected in 1994, the state with largest elementary school class sizes among those included in this study (California) has in fact, allocated a substantial increase in state funding for education specifically for the purpose of reducing class sizes in all elementary schools in the state.

negative relation) between states with higher scores and states whose teachers perceive better school behavioral climates.

Finally, the increased effective sample sizes for estimating means when between-state variance is removed, suggested by the results in table 7, did not generalize to smaller estimated standard errors for the coefficients in tables 8, 9, and 10. That is, although the effects were on the same order of magnitude in tables 8, 9, and 10, compared to tables 4, 5, and 6, there were not substantially more statistically significant results in tables 8, 9, and 10. Eliminating the between-state variance in the *means* of the measures included in the analysis did not substantially reduce the between-state variance in the *relations* among these measures. Stated in different terms, information about the design effects on estimates of means for this sample did not generalize to information about design effects on correlations, regression coefficients, and structural equation model parameters.

The interpretation of results involving both between- and within-state relations would be facilitated by the use of direct hierarchical structural equation modeling methodology. With that methodology, between state components of relations could be measured directly. Nevertheless, the results presented in this section indicate the value of including the NAEP between-state achievement adjustments in the analysis. Meaningful results relating class size to achievement measures were found in the overall analyses that would not have been found in pooled within-state analyses.

The importance of aggregating the database over many states is evident when these analyses are carried out for each state separately. Because there are only a few dozen SASS public schools at each grade level in each state, attempts to carry out these analyses within a single state yield much less reliable information than is provided by aggregating the results across states. The results of individual state analyses, shown in table 12, indicate that the coefficient estimates can range from positive to negative on all of these correlates. The correlates of average school achievement are complex, and the SASS sample in an individual state is not sufficiently large to separate systematic relations from the variety of unmeasured correlates of achievement.

One explanation for variation in results between states is the differences among the state assessments. State assessments are designed to serve a variety of purposes, and these purposes are sometimes in conflict with the goal of providing a valid and reliable measure of between-school average student achievement variation. In fact, one might take the correlation with NAEP as an indicator of a state assessment's relevance to this purpose of identifying relations between achievement and school-level factors that are generalizable across states. In fact, the correlation between (a) a state assessment's correlation with NAEP at fourth grade shown in appendix C and (b) that state assessment's correlation with SASS school-level measures should be positive. At the elementary and middle school levels, the correlations are significant (.48 at the elementary level and .63 at the middle school level). Apparently, the state assessments that were more highly correlated with NAEP were also more closely related to SASS measures. It is important to point out, in this regard, that a higher correlation with NAEP is an indicator of the reliability of a test as much as of its similarity in content and format to NAEP.

Table 12.—Average and range of standardized linear regression coefficients for within-state variation in achievement, by grade level

	Elementary (n = 20)			Middle (n = 16)			Secondary (n = 14)		
Independent	**mean**	**min**	**max**	**Mean**	**min**	**max**	**mean**	**min**	**max**
School size	–.01	–.35	+.46	–.05	–.44	+.36	+.21	–.41	+.99
Class size	+.04	–.23	+.61	+.07	–.34	+1.49	–.16	–.90	+.39
Teacher control	+.04	–.12	+.24	–.06	–.63	+.30	+.07	–.73	+1.2
Cohesion	+.04	–.58	+.67	+.04	–.53	+1.30	–.17	–1.83	+.26
School climate	+.12	–.14	+.35	+.36	+.03	+1.70	+.31	–.42	+1.2

NOTE: Only states with 10 or more schools at the grade level on the SASS student achievement subfile are included.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

At the secondary level, where the numbers of states with at least 10 schools was smaller (14), the correlation was not significant. One possible reason for the nonsignificant pattern at the secondary level is that the state assessment tests at the secondary level tend to be quite different from the fourth-grade assessments whose correlations with NAEP were measured.

Finally, differences between states exhibit a high level of stability across grade levels. Not only are NAEP State means highly correlated between fourth and eighth grades, but the other factors, which are based on different SASS schools and different teachers at each grade level, are also highly correlated. For example, as shown in table 13, states in which teachers in SASS elementary schools perceived that they had more control and influence were also the states in which secondary teachers also had that perception (r = .58). The strongest relation was in language barriers, representing the large between-state variation in percentages of Hispanic students. Although some of the results are easily attributable to population and wealth differences, other correlations suggest that there may be state policies that foster, for example, greater normative cohesion in schools.

Table 13.—State-level cross-grade correlations of school correlates of achievement

	Elementary-Middle	Elementary-Secondary	Middle-Secondary
Poverty percentage	.73	.48	.60
Language barriers	.96	.88	.88
Minority percentage	.93	.84	.87
School size	.85	.62	.63
Class size	.89	.63	.74
Teacher control	.77	.58	.82
Normative cohesion	.49	.63	.91
School climate	.51	.28	.74

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Conclusions

The objective of this report has been to demonstrate and evaluate the strategy of combining a large scale national survey of schools (SASS), which lacks measures of student achievement, with school-level assessment data from a large number of individual states. If application of this strategy yields new insights about schools or identifies questions that lead to new avenues of research, then its value is demonstrated. If the substantive findings are empty, the strategy is less attractive.

To demonstrate the strategy, a set of 18 composites of SASS data, including student background information, organizational information, teachers' qualifications, and school climate perceptions were constructed and merged with school reading and mathematics mean scores. The resulting data were analyzed using correlational, multiple regression, and structural equation model analyses. These analyses only begin to tap the richness of the SASS database: selection of other subsets of the SASS data or other analytical methods could add to the evaluation of the strategy.

Substantive Findings

The clearest result with respect to correlates of achievement is that average achievement scores are higher in schools with smaller class sizes. This result, obtained from structural equation modeling using both state assessment data and NAEP adjustments for between-state variance in achievement, is consistent across grade levels (see table 6), although it is only significant at the .05 level in middle and secondary schools. While there are alternative causal explanations for this finding, such a finding in a large sample of public schools in 20 states is an important corroboration of the controlled research results that indicate that class size makes a difference.

The positive relation between small classes and achievement was stronger for secondary schools than for elementary schools. In secondary schools, the positive association with achievement included both large schools and small classes. An important aspect of the relation between class size and achievement, shown by the comparison of results with and without between-state variance components, is that it is primarily a between-state phenomenon. Restricting the study to within-state comparisons and then aggregating the results across 20 states yields much less evidence of a class size relation to achievement scores. This may be due to state policy-related limitations on variation in class sizes.

Substantive findings were not limited to class size. There was limited evidence of a positive relation between teachers' perceptions of the school's behavior climate and achievement scores. In particular, this relation was only statistically significant when between-state variation was omitted from the data; and although all three analytical methods found it to be significantly positive in middle schools, it was not statistically significant in the structural equation analyses in elementary and high schools.

Based on these findings, one cannot avoid the conclusion that combining the SASS data with a school-level student achievement measure has the potential for addressing important policy questions about school-based strategies for improving student performance. Because the data are not longitudinal, causal inferences must be treated much more tentatively than conclusions based on data on the achievement gains of a specified set of students over time. Also, because the data are school means, they cannot address the factors that differentially affect the achievement of different students in the same school. Nevertheless, findings from analyses of the SASS student-achievement subfile, based on over 2,000 schools in 20 states, can contribute to the overall educational policy database.

Methodological Findings

The primary conclusion reached in this study is that the strategy of matching school-level assessment scores to a national survey (a) is feasible and not costly, because the data are readily available, and (b) leads to valid and reliable conclusions about correlates of public school achievement across much of the United States. The additional step of linking the database to State NAEP to capture between-state achievement variation is also feasible and not costly and provides additional informational value.

It is clear from these analyses that between-state variation in achievement and in its correlates is an important component of the national database on education, because the contexts within states reduce the variance on key factors to the point that important relations disappear. In a sense, that is the goal of many state policies—to provide resources to schools in such a way that students in all schools in a state have equal opportunities to achieve at high levels. However, in this database of 20 states, nearly half of the variance in school sizes and class sizes is between states (see table 7), and a third or more of the variation in percentage of minority enrollment is between states. Studies that focus purely on variation in schools within states will miss the effects of these factors on educational achievement.

The methods used in this paper focused on overall correlates of achievement, including between-state variation, but comparison with analyses of within-state relations indicates the potential value of applying a multilevel analysis to these data. No state-level variables were included in this analysis, but combining this database with information on the educational policies of these 20 states, in a hierarchical linear structural equation model, would provide the basis for addressing many educational policy issues.

A positive methodological finding was the generalizability of the between-state achievement measures across grade levels. Although state assessment scores were available for grades from 3 to 11, NAEP reading scores for individual states were only available for grade 4 in 1994. If the ordering of states in reading achievement changed substantially from grade 4 to grades 8 and 11, then the results of overall analyses of middle school and high school data would be diluted by linkage error. This dilution should not affect the within-state analyses, however.

The extension of the NAEP adjustment proved valid, in that the findings for secondary schools, using the between-state data (compare tables 5 and 9, or 6 and 10) are as meaningful as the findings for elementary schools. This conclusion is not surprising, given the very high correlation of State NAEP means in different grades and subjects, but its support in this study may suggest new uses of State NAEP data in conjunction with state assessment data.

A limitation on the validity of aggregating teacher data for school level analyses became apparent in the findings concerning teacher qualifications (average years of teaching experience and percent having a masters degree). These measures, unlike the teachers' responses to questions about school policies and school behavioral climate, had very low reliability *as measures of the school,* because there was relatively little systematic between-school variation: most of the variation was between teachers at the same school. This problem was manifest in the low intercorrelation between these measures; and as a result preliminary analytical findings concerning the relations of this teacher qualifications factor were uninterpretable. Ultimately, the decision was made to omit this factor from the analyses reported here.

Finally, athough the data are purely correlational, there are logical constraints, such as that school factors probably do not cause differences in student background characteristics in the short term ("white flight" notwithstanding). Interpretation of the results of structural equation modeling in terms of hypothetical path models, such as shown in figure 2, can lead to fruitful suggestions for avenues of research and policy development.

Future Research

Three broad areas of research stemming from this study appear to be fruitful: hierarchical analyses to examine the relations between state education policies and the SASS correlates of achievement; development of a measure of a school's achievement gains over time which can be associated with SASS measures; and further refinement of the linkage functions between state assessments and NAEP.

The findings of this study clearly indicate different patterns of correlates of achievement within states and between states. Schools in the same state tend to operate within common frameworks of funding, staff accreditation, curriculum, testing, and school reform programs. With uniformity in these aspects of education, variations in other factors are more likely to manifest correlations with achievement. On the other hand, to the extent that state frameworks affect achievement outcomes, between-state correlates of achievement can emerge. An analysis methodology which simultaneously models the within- and between-state relations among variables and takes measurement error into account is needed for this. With such a methodology, and with the addition of a database of quantitative indicators of relevant state policies, the SASS student achievement subfile would increase in value.

Every school addresses the needs of a different student population with different resources, and it is therefore unfair to hold all schools accountable to the same achievement standard. However, a number of states are turning to reform criteria that base decision making on measures of *gains* in achievement over years. Although SASS cannot easily add longitudinal student growth data, it is certainly feasible to add other years' school-level achievement data to the subfile. Specifically, the addition of 1997–98 reading scores, linked to the 1998 fourth and eighth grades State NAEP reading assessment and CCD data on changing enrollment patterns and resources over the intervening years, would provide the basis for identifying SASS factors (measured in 1994) that are predictive of gains in achievement. For example, one wonders whether staff turnover rates would portend gains, other things equal. Of course, states continue to develop and refine assessment systems, and the state assessment scores for a school in 1998 may not be equivalent to scores obtained in 1994, so linkage of measures of achievement gains over time to repetitions of State NAEP is an essential requirement for the development of a longitudinal database.

The power of the database for longitudinal analyses can be greatly enhanced with the addition of the next cohort of SASS. If a subsample of schools included in SASS in 1994 are also included in 1999–2000, then using the 2000 State NAEP assessment for adjustment of mathematics scores would enable matching of longitudinal changes in SASS school-based factors with longitudinal changes in achievement, controlling for longitudinal changes in student background factors.

A third line of research would focus on improving the achievement measures included in the SASS student achievement subfile. The linkages used for the analyses presented in this report were based entirely on the means, standard deviations, and correlations between State NAEP and state assessment school means. The errors in these linkages can be diminished significantly by more detailed analysis of the relations among the scores. In particular, current research NCES has found that linkages to NAEP can be improved by considering nonlinear terms and by including demographic indicators. For example, all state reading assessments are sensitive to race/ethnic differences, but some are more sensitive than others. Their sensitivities could be matched to NAEP's measurement of the distribution of race/ethnic achievement differences by explicitly including that matching factor in the NAEP adjustment step in constructing the SASS school-level achievement score. The result would be increased comparability of within-state variation in the achievement measure across states.

Bibliography

Alexander, Karl L. & Griffin, Larry J. (1976). School District Effects on Academic Achievement: A Reconsideration; Comments on Bidwell and Kasarda. *American Sociological Review*, 41(1):144–152.

Anderson, C.S. (1982). The search for School Climate: A Review of Research. *Review of Educational Research*, 52:368–420.

Baker, David P. (1996). Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with Comments on School Reform, Governance, and Finance. In John E. Mullens & Daniel Kasprzyk (eds.) *The School and Staffing Survey: Recommendations for the Future*. U.S. Department of Education. Pp.19–37.

Bankston, Carl L. III & Min Zhou. (1995). The Effects of Minority Language Literacy on Academic Achievement of Vietnamese Youths in New Orleans." *Sociology of Education*. 68(1, January):1–17.

Bidwell, Charles E. & Kasarda, John D. (1975). School District Organization and Student Achievement." *American Sociological Review,* 40(1):55–70.

Bidwell, Charles E. & Kasarda, John D. (1976). Reply to Hanna, Freeman and Meyer, and Alexander and Griffin. *American Sociological Review*, 41(1):152–160.

Blatchford, P. & Mortimore, P. (1994). The Issue of Class Size for Young Children in Schools: What Can We Learn from Research? *Oxford Review of Education*, 20(4):411–428.

Bollen, K.A. & Bollen, W. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Boruch, R.F. & Terhanian, G. (1996). "So What?" The implications of new analytic methods for designing NCES surveys. In Hoachlander, G., Griffiths, J.E., & Ralph, J.H. (ed.) *From Data to Information – New Directions for the National Center for Education Statistics*. Washington, DC: National Center for Education Statistics, No. 96–901.

Boyer, E. (1995). *The Basic School: A Community for Learning*. Princeton, NJ: The Carnegie Foundation.

Bryk, Anthony S. & Raudenbush, Stephen W. (1987). Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin*, 101(1):147–158.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of Educational Opportunity*. Washington, DC: Government Printing Office.

Conant, J. (1959). *The American High School Today*. New York: McGraw-Hill.

Ferguson, Ronald F. (1990). Shifting challenges: Fifty Years of Economic Change toward Black-White Earning Equality. *Challenge: A Journal of Research on Black Males* 2(1) 37–76.

Ferguson, Ronald F. (1991). Paying for Public Education: New Evidence on How and Why Money Matters. *Harvard Journal of Legislation,* 28(465):465–498.

Finn, J. & Achilles, C. (1990). Answers and Questions about Class Size: A Statewide Experiment. *American Educational Research Journal*, 27(3):557–577.

Friedkin, N.E. & Slater, M.L.. (1994). School Leadership and Performance. *Sociology of Education*. 67(2, April):139–157.

Gibson, M.A. (1989). *Accommodation without Assimilation: Sikh Immigrants in an American High School.* Ithaca, NY: Cornell University Press.

Glass, G., Cahen, L., Smith, M., & Filby, N. (1982). *School Class Size.* Beverly Hills, CA: Sage.

Gordon, E. & Yowell, C. (1994). Cultural Dissonance as a Risk Factor in the Development of Students. In R. Rossi (ed.), *Schools and Students at Risk: Context and Framework for Positive Change.* New York: Teachers College Press.

Granovetter, Mark. (1982). The Strength of Weak Ties: A Network Theory Revisited. pp. 105–130 In Marsden P.V. & N. Lin (eds.), Social Structure and Network Analysis. Beverly Hills, CA: Sage.

Hannan, Michael T., Freeman, John H., & Meyer, John W. (1976). Specification of Models of Organizational Effectiveness; Comment on Bidwell and Kasarda. *American Sociological Review*, 41(1):136–143.

Hanushek, E. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*:1141–1177.

Herriot, R.F. & Firestone, W.A. (1984). Two Images of Schools as Organizations: A Refinement and an Elaboration. *Educational Administration Quarterly*, 20(4):41–58.

Hoy, Wayne K. & Hannum, John W. (1997). Middle School Climate: An Empirical Assessment of Organizational Health and Student Achievement. *Educational Administration Quarterly*, 33(3):290–311.

Ingersoll, Richard M. (1996). Teachers Decision-Making Power and School Conflict. *Sociology of Education*, 69(2):159–176.

Kaufman, Phillip. (1996). Linking Student Data to SASS: Why, When, How. In John E. Mullens & Daniel Kasprzyk (eds.) *The School and Staffing Survey: Recommendations for the Future*. U.S. Department of Education. pp.53–65.

Lareau, A. (1987). Social Class Differences in Family-School Relationships: The Importance of Cultural Capital. *Sociology of Education*, 60:73–85.

Lee, Valerie E. & Bryk, Anthony S. (1989). A Multilevel Model of the Social Contribution of High School Achievement. *Sociology of Education*, 62 (3):172–192.

Lee, Valerie E. & Julia B. Smith. (1993). Effects of School Restructuring on the Achievement and Engagement of Middle-Grade Students. *Sociology of Education*, 66(3):164–187.

Lee, Valerie E. & Julia B. Smith. (1995). Effects of High School Restructuring and Size on Early Gains in Achievement and Engagement. *Sociology of Education*, 68(4):241–270.

Lee, Valerie E., Smith, Julia B, & Croninger, Robert G. (1997). How High School Organization Influences the Equitable Distribution of Learning in Mathematics and Science. *Sociology of Education*, 70(2):128–150.

Lindsay, P. (1982). The Effect of High School Size on Student Participation, Satisfaction, and Attendance. *Educational Evaluation and Policy Analysis*, 4(1):57–65.

McLaughlin, D.H. & Broughman, S. (1997) *Private Schools in the United States: A Statistical Profile, 1993–94*. Washington, DC: U.S. Department of Education. National Center for Education Statistics. (NCES 97–459).

McLaughlin, D.H. & Drori, G. (1998) *School-level Correlates of Reading and Mathematics Achievement in Public Schools*. Paper presented at the National Center for Education Statistics Seminar on Issues in Assessing Achievement. Washington, DC, November, 1998.

Meyer, R. H. (1997). Value-Added Indicators of School Performance: A Primer. *Economics of Education Review*, 16:283–301.

Moore, W. & Esselman, M. (1992). Teacher Efficacy, Power, School Climate, and Achievement: A Desegregating District's Experience. Paper presented at the annual meeting of American Educational Research Association, San Francisco, April 1992.

Monk, D. (1994). Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement. *Economics of Education Review, 13*(2):125–145.

Mosteller, F., Light, R. J. & Sachs, J. A. (1996). Sustained Inquiry in Education: Lessons from Skill Grouping and Class Size. *Harvard Educational Review*, 66(4):797–842.

Mullis, Ina V.S., Campbell, J.R., & Farstrup, A.E. (1993). *NAEP 1992 Reading Report Card for the Nation and the States*. Washington, DC: U.S. Department of Education, National Center for Education Statistics (No. 23–ST06).

Mullis, Ina V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). *NAEP 1992 Mathematics Report Card for the Nation and the States*. Washington, DC: U.S. Department of Education, National Center for Education Statistics (No. 23–ST02).

National Education Goals Panel. (1996, June). *Profile of 1994–95 State Assessment Systems and Reported Results*.

Newmann, F., Rutter, R., & Smith, M. (1989). Organizational Factors That Affect School Sense of Efficacy, Community, and Expectations. *Sociology of Education, 62*:221–238.

Oakes, Jeanie. (1989). What Education Indicators? The Case for Assessing the School Context. *Education Evaluation and Policy Analysis*, 11(2):181–199.

Oxley, D. (1994). Organizing Schools into Small Units: Alternatives to Homogeneous Grouping. *Phi Delta Kappan, 75*(7):521–526.

Pallas, A. (1988). School Climate in American High Schools. *Teachers College Record*, 89:541–553.

Pittman, R. & Haughwout, P. (1987). Influence of High School Size on Dropout Rate. *Educational Evaluation and Policy Analysis, 9*(4):337–343.

Portes, Allejandro & McLeod, Dag. (1996). Educational Progress of Children of Immigrants: The Roles of Class, Ethnicity, and School Context. *Sociology of Education*, 69(4):255–275.

Portes, Alejandro & Schauffler, Richard. (1993). Language and the Second Generation. In *The New Immigration*, edited by Ruben G. Rumbaut & S. Pedrazgo. NY: Wadsworth.

Purkey, S. and Smith, M. (1983). Effective Schools: A Review. *The Elementary School Journal, 83*(4):427–452.

Raudenbush, Stephen, Rowan, Brian, and Kang, Sang Jin. (1991). A Multilevel, Multivariate Model for Studying School Climate with Estimation via EM Algorithm and Application to U.S. High-School Data. *Journal of Educational Statistics*, 16(4):295–330.

Reese, C.M., Miller, K.E., Mazzeo, J., and Dossey, J.A. (1997*) NAEP 1996 Mathematics Report Card for the Nation and the States.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Roeber, E., Bond, L., and Braskamp, D. (1997*) Annual Survey of State Student Assessment Programs: Fall 1996.* Washington, DC: Council of Chief State School Officers.

Roeber, E., Bond, L., and Connealy, S. (1998*) Annual Survey of State Student Assessment Programs: Fall 1997.* Washington, DC: Council of Chief State School Officers.

Ross, J.A. (1992). Teacher Efficacy and the Effect of Coaching on Student Achievement. *Canadian Journal of Education*, 17(1):51–65.

Rowan, Brian, Chiang, Fang-Shen, & Miller, Robert J. (1997). Using Research on Employees' Performance to Study the Effects of Teachers on Students' Achievement. *Sociology of Education*, 70:256–284.

Scott-Jones, D. (1984). Family Influences on Cognitive Development and School Achievement. *Review of Research in Education, 11*:259–304.

Sirotkin, K.A. (1980). Psychometric Implications of the Unit-of-Analysis Problem (with Examples from the Measurement of Organizational Climate). *Journal of Educational Measurement*, 17:245–282.

Slavin, R. (1989). Class Size and Student Achievement: Small Effects of Small Classes. *Educational Psychologist, 24*(1):99–110.

Sui-Chu, Esther Ho & J. Douglas Willms. (1996). Effects of Parental Involvement on Eighth-Grade Achievement". *Sociology of Education*, 69(2):126–141.

Tarter, John C., Sabo, Dennis, & Hoy, Wayne K. (1995). Middle School Climate, Faculty Trust, and Effectiveness: A Path Analysis. *Journal of Research and Development in Education*, 29 (1):41–49.

Taylor, Dianne L. & Tashakkori, J. (1995). Decision Participation and School Climate as Predictors of Job Satisfaction and Teachers' Sense of Efficacy. *Journal of Experimental Education*, 63(3):217–230.

Tienda, Martha. (1984). Language, Education, and the Socioeconomic Achievement of Hispanic Origin Men. *Social Science Quarterly*, 65:519–536.

Turner, R. & Camilli, G., (1988). *The influence of salary schedule variables on teacher applicant pools, retention, and advanced degrees, and on student achievement.* Paper presented at the Annual Meetings of the American Educational Research Association, New Orleans, LA.

Wehlage, G., Rutter, R., Smith, G., Lesko, N., & Fernandez, R. (1989). *Reducing the Risk: Schools as Communities of Support*. New York: The Falmer Press.

Wu, G., Royal, M., & McLaughlin, D. (1997). *Working Paper: Development of a SASS School-Level Student Achievement Subfile, Using Existing State Assessment and State NAEP Information.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 97–44).

Appendix A

Results of SEM: Measurement Level

A measurement model is required to link the latent variables in the structural equation model to SASS measures. Three tables provide the information necessary for understanding the latent variables.

The first table gives a reliability estimate for the teacher-based indicators; that is, a measure of the tendency of teachers at the same school to give the same responses. The estimate is one minus the ratio of (a) the sum of within-school variances, divided by the total number of teachers responding, to (b) the variance of school means. Values substantially less than .5 (such as average class size, classroom control perceptions, and years of experience) indicate that more of the variance in the indicator is within schools than between schools, and alternative indicators should be given greater weight in the model.

The second table gives intercorrelations between the indicators. For example, the lowest intercorrelations (.28 and .29) are for the average class size with the other two indicators of the class size factor and between average years of teachers' experience and percent of teachers with a masters degree, all at the elementary school level. These intercorrelations need not be substantial for SEM analyses, because the estimation procedure should identify the weighting of the indicators that most effectively accounts for variance in other measures. Nevertheless, values substantially less than .50 indicate that most of the variance in the indicators is not in common across the latent trait.

The third table contains the SEM measurement parameter estimates, which were obtained simultaneously with the structural equation parameter estimates. For example, the latent poverty variable is set to be in the same units as the free-lunch eligible fraction, by presetting its coefficient to 1.0; and at the elementary school level, $.65^2$, or 42 percent, of the variance in the free lunch eligible fraction is attributed to the latent poverty trait.

Table A1.—Reliability coefficients for teacher-based components of school-level factors, by school level. (Estimated fraction of sample mean variance that is between schools)

		Elementary	Middle	Secondary	All Schools
Class Size	Average	.31	.37	.44	.42
	Satisfied?	.45	.47	.54	.51
Climate	Climate 1	.70	.74	.77	.82
	Climate 2	.78	.78	.86	.85
Normative Cohesion	Cooperation	.51	.46	.47	.50
	Clear Norms	.54	.55	.61	.60
Teacher Control	Classroom Control	.36	.38	.50	.41
	Influence on School Policies	.61	.62	.62	.63
Teacher Qualification	Years Experience	.26	.14	.27	.24
	Masters	.41	.39	.44	.41
Poverty	Problem?	.82	.79	.82	.80
Minority Conflicts	Problem?	.73	.79	.84	.80

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table A2.—Intercorrelations of components of school-level factors, by grade level

		Elementary	Middle	Secondary	All schools
		Class size factor			
Average class size	Student/teacher ratio	.28	.50	.57	.45
Average class size	Class size satisfaction	.29	.45	.49	.40
Student/teacher ratio	Class size satisfaction	.43	.42	.55	.48
		Climate factor			
Climate 1	Climate 2	.61	.72	.50	.72
		Normative cohesiveness			
Cooperation	Clear norms	.66	.67	.66	.69
		Teacher control and influence			
Classroom control	School policy influence	.50	.49	.49	.48
		Teacher qualifications			
Years of experience	Masters degree	.29	.30	.37	.32
		Poverty			
Percent free-lunch eligible	Poverty a problem	.55	.51	.47	.49
		Racial/ethnic minorities			
Percent minority	Race conflicts a problem	.49	.45	.43	.45
		Language minorities			
Percent limited English proficient	Percent in ESL classes	.80	.75	.70	.77

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Table A3.—School-level factors SEM measurement model estimates

Indicator	Elementary	Middle	Secondary
Read.Achieve	=1.0 (.97) Achieve + .23e	=1.0 (.94) Achieve + .33e	=1.0 (.91) Achieve + .37e
Math.Achieve	=.86 (.83)* Achieve + .55e	=.95 (.89)* Achieve + .45e	=.97 (.88)* Achieve + .54e
Pct free lunch eligible	=1.0 (.66) Poverty + .75e	=1.0 (.66) Poverty + .75e	=1.0 (.61) Poverty + .79e
Teachers see poverty a problem	=1.19 (.77)*Poverty -.98 (-.39)* T_Quest + .52e	=1.14 (.71)* Poverty -.88 (-.43)* T_Quest + .55e	=1.37 (.85)* Poverty - .25 (-.08) T_Quest + .50e
Pct. limited English proficient	=1.0 (.87) Language + .49e	=1.0 (.89) Language + .46e	=1.0 (.85) Language + .53e
Pct. in ESL instruction	=1.04 (.89)*Language+ .45e	=.93 (.81)* Language + .59e	=.97 (.82)* Language + .57e
Pct. minority	=1.0 (.83) Race + .55e	=1.0 (.86) Race + .51e	=1.0 (.85) Race + .56e
Teachers see racial tension a problem	=.71 (.59)* Race -1.10 (-.43) T_Quest + .68e	=.61 (.50)* Race -1.10 (-.54)* T_Quest + .68e	=.59 (.50)* Race -.97 (-.31)* T_Quest + .84e
Perceived classroom control	=1.0 (.57) T.Control + .82e	=1.0 (.59) T.Control + .81e	=1.0 (.73) T.Control + .68e
Perceived policy influence	=1.53 (.88)* T.Control+.48e	=1.40 (.83)* T.Control + .56e	=.92 (.67)* T.Control + .75e
School size	=1.0 (1.0) School.size +.0e	=1.0 (1.0) School.size +.0e	=1.0 (1.0) School.size + .0e
Average class size	=1.0 (.39) Class.size +.92e	=1.0 (.61) Class.size + .79e	=1.0 (.72) Class.size + .70e
Student/teacher ratio	=1.82 (.72)* Class.size+.69e	=1.28 (.79)* Class.size+ .62e	=1.11 (.80)* Class.size +.59e
Teachers see class size as satisfactory	=1.50 (.58)*Class.size + 1.0 (.38) T_Quest + .72e	=.87 (.50)* Class.size + 1.0 (.54) T_Quest + .73e	=.83 (.61)* Class.size + 1.0 (.32) T_Quest + .43e
Clarity of norms	=1.0 (1.0) Norm.coh + .0e	=1.0 (1.0) Norm.coh + .0e	=1.0 (1.0) Norm.coh + .0e
Cooperation	=.65 (.65)* Norm.coh +.76e	=.66 (.65)* Norm.coh + .76e	=.65 (.65)* Norm.coh + .76e
School behavioral climate 1	=1.0 (.89) School.climate -.84 (-.33)* T_Quest + .30e	=1.0 (.91) School.climate -.73 (-.37)* T_Quest + .18e	=1.0 (.94) School.climate -.82 (-.27)* T_Quest + 21e
School behavioral climate 2	=.97 (.87)* School.Climate -.98 (-.39) * T_Quest + .30e	=.92 (.84)* School.Climate -.85 (-.43) * T_Quest + .34e	=.89(.84)* School.Climate -1.34 (-.44)* T_Quest + 38e

NOTES: Data in table include: Unstandardized coefficient or factor loading (standardized coefficient or factor loading) significance level, factor name, and magnitude of error term. Indicators with an unstandardized factor loading of 1.0 are the reference indicators. Error terms of .0 were set to this value in order to identify the model.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Appendix B

Factors Associated with School Climate and Achievement in Public Schools, Reversing the Causal Order between these Two Factors: Results from SEM Analyses

A separate SEM analysis was carried out in which the school climate trait was omitted from the equation for achievement and the school achievement trait was included in the equation for school climate.

Independent effects	Dependent factor	Elementary (n=1124)	Middle (n=496)	Secondary (n=595)
School size	**Student achievement**	+0.10	+0.01	+0.39*
Class size		–0.25*	–0.27	–0.37*
Normative cohesion		–0.09*	–0.02	–0.00
Teachers' qualifications		–0.01	–0.13	–0.12
Teachers' influence		+0.02	+0.02	–0.05
r^2		0.69	0.85	0.81
School size	**School climate**	–0.05	–0.06	–0.22*
Class size		–0.20	–0.26	–0.09
Normative cohesion		+0.22*	+0.37*	+0.34*
Teachers' influence		+0.01	+0.11	+0.02
Student achievement		–0.29	–0.03	–0.11
r^2		0.68	0.75	0.70
	Statistical summary measures			
	GFI (AGFI)	0.95 (0.92)	0.91 (0.85)	0.93 (0.88)
	Π2 (d.f.)	606 (130)	552 (130)	506 (130)

NOTES: Entries in table are standardized gamma coefficients. The same factors and equations were included as represented in table 7, except for the reversal of achievement and school climate. One additional elementary school, deleted in later analyses, was included in this analysis.

(*) P<.05 based on repeated half-sample standard deviations.

r^2 values include effects of three background factors in addition to factors shown.

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

Appendix C

Table C1.—Correlations between grade 3/4 state assessment and NAEP school means in reading in 1993–94, for selected states.

State	Correlation coefficient for school means
Alabama	.72
California	.78
Delaware	.32
Florida	.85
Georgia	.66
Hawaii	.65
Kentucky	.37
Louisiana	.84
Massachusetts	.82
Maine	.63
Michigan	.73
Montana	.50
New Hampshire	.65
New York	.39
Rhode Island	.86
Tennessee	.63
Texas	.72
Washington	.73
Median	.70

SOURCES: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1993–94 (School and Teacher Questionnaires) and State Reading and Mathematics Assessment Scores for Public Schools in 20 States.

United States
Department of Education
ED Pubs
8242–B Sandy Court
Jessup, MD 20794–1398

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (A)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)