DOCUMENT RESUME

ED 438 417	CE 079 706
AUTHOR	Stern, David; Byrnes, Matthew; Levesque, Karen; Lauen, Douglas
TITLE	Enabling High Schools To Assess Schoolwide Results of Reform: A Pilot Test.
INSTITUTION	National Center for Research in Vocational Education, Berkeley, CA.; MPR Associates, Berkeley, CA.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC.
REPORT NO	NCRVE-MDS-1299
PUB DATE	2000-01-00
NOTE	101p.
CONTRACT	V051A30003-99A; V051A30004-99A
AVAILABLE FROM	National Dissemination Center for Career and Technical Education, The Ohio State University, 1900 Kenny Road, Columbus, OH 43210-1090. Tel: 800-678-6011 (Toll Free); Fax: 614-688-3258; Web site: http:nccte.com.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Comparative Analysis; *Educational Change; Educational Improvement; Evaluation Methods; Evaluation Research; Feasibility Studies; *High Schools; Partnerships in Education; Pilot Projects; *School Surveys; Success
IDENTIFIERS	High Schools That Work; *Impact Studies

ABSTRACT

A procedure for assessing and comparing schoolwide change, so that groups of schools could use the resultant information for continuous improvement, was developed. The feasibility of the new assessment procedure was tested by applying the new procedure in 5 schools in the Bay Area School Reform Collaborative (BASRC) in the San Francisco Bay Area, and 27 schools in the Southern Regional Education Board's High Schools That Work (HSTW) network. Administrators and faculty from both groups of schools met with the researchers conducting the test, and with administrators and faculty from the other schools in their network, to determine whether individual schools' indicators of student achievement and engagement were improving, and explore why improvement was occurring more rapidly at certain schools. The study demonstrated that joint inquiry into comparative data is indeed feasible, and that a collegial process for analyzing comparative data that involves teachers, counselors, and principals may be a useful addition to the evolving set of institutional mechanisms designed to help schools improve. (Contains 21 references and 11 tables/figures. Appended are the 1997-98 school data report created for collected data from BASRC schools and a sample of the site report produced for each participating HSTW school.) (MN)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 438 417

NCRVE

National Center for Research in Vocational Education

University of California, Berkeley

ENABLING HIGH SCHOOLS TO ASSESS SCHOOLWIDE RESULTS OF REFORM: A PILOT TEST

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Supported by
the Office of Vocational and Adult Education,
U.S. Department of Education

CE 079 706

NATIONAL DISSEMINATION CENTER

CAREER & TECHNICAL EDUCATION

THE OHIO STATE UNIVERSITY

National Dissemination Center for
Career and Technical Education
The Ohio State University
1900 Kenny Road
Columbus, Ohio 43210-1090

This publication available from

NDCCTE Product Sales Office
The Ohio State University
1900 Kenny Road
Columbus, OH 43210-1090

800-678-6011 ext. 2-4277
Fax: 614-688-3258

Website: www.ndccte.com
Email: ndccte@osu.edu
Listserv: careertech@lists.acs.ohio-state.edu

ERIC

ENABLING HIGH SCHOOLS TO ASSESS SCHOOLWIDE RESULTS OF REFORM: A PILOT TEST

David Stern
Matthew Byrnes

University of California, Berkeley

Karen Levesque
Douglas Lauen

MPR Associates

National Center for Research in Vocational Education
University of California at Berkeley
2030 Addison Street, Suite 500
Berkeley, CA 94720-1674

Supported by
The Office of Vocational and Adult Education
U.S. Department of Education

January 2000

MDS-1299

FUNDING INFORMATION

Project Title:	National Center for Research in Vocational Education
Grant Number:	V051A30003-99A/V051A30004-99A
Act under which Funds Administered:	Carl D. Perkins Vocational Education Act P.L. 98-524
Source of Grant:	Office of Vocational and Adult Education U.S. Department of Education Washington, DC 20202
Grantee:	The Regents of the University of California c/o National Center for Research in Vocational Education 2030 Addison Street, Suite 500 Berkeley, CA 94720-1674
Director:	David Stern
Percent of Total Grant Financed by Federal Money:	100%
Dollar Amount of Federal Funds for Grant:	$4,500,000
Disclaimer:	This publication was prepared pursuant to a grant with the Office of Vocational and Adult Education, U.S. Department of Education. Grantees undertaking such projects under government sponsorship are encouraged to express freely their judgement in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official U.S. Department of Education position or policy.
Discrimination:	Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance." Title IX of the Education Amendments of 1972 states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving federal financial assistance." Therefore, the National Center for Research in Vocational Education project, like every program or activity receiving financial assistance from the U.S. Department of Education, must be operated in compliance with these laws.

ACKNOWLEDGMENTS

Our first thanks go to the high school teachers, counselors, and administrators who spent long hours compiling data for this study. We hope that this report will help repay those efforts, and will encourage the continuing search for improvement in high schools.

We would also like to thank the leadership of High Schools That Work for their cooperation and vision. It has been a particular privilege to work with Dr. Gene Bottoms, whose dedication to educational improvement has inspired so many, including ourselves, to try to do better.

Finally, we are grateful to Charles Dayton and Alan Weisberg for their help in the early stages of this study, to Gary Hoachlander for his guidance and support, and to Denise Bradby and an anonymous reviewer for helpful feedback on an earlier draft.

Whatever errors of fact or interpretation may remain in this report are, of course, the responsibility of the authors alone.

EXECUTIVE SUMMARY

Many high schools are engaged in ongoing efforts to redesign their curriculum and instructional methods in order to improve student learning. In pursuit of improvement, schools are attempting, and sometimes are required, to make more systematic use of data on student performance. Current policies that emphasize data-based school improvement and accountability are premised on the hypothesis that systematic use of data by schools will help improve student performance.

In this study, we are interested in an additional subhypothesis: that results will be better if teachers and school administrators have an opportunity to meet with their counterparts from other schools, for the purpose of comparing data and discussing how changes in instructional programs may be affecting student performance. This report describes a pilot test of whether this procedure is feasible.

Unlike the comparisons of test scores that now appear frequently in newspapers and on the Internet, and which some states are using to reward or punish schools, we did not focus on ranking schools according to their students' average level of performance. Such rankings are strongly correlated with students' socioeconomic characteristics, which schools cannot control.

Instead, to create a more level playing field, we are more interested in comparing changes over time. Our objective was to collect or compile data that would enable each school to determine whether measures of its students' achievement and engagement were improving, and whether its own students' performance was improving faster or more slowly than in other schools. We would then bring teachers and administrators from different schools together to discuss possible reasons *why* some schools were attaining faster rates of improvement than others. The purpose was to engage these educators in seeking links between changes in school practices and observed changes in student performance. Section I of this report explains the rationale for this process in more detail.

We worked with representatives from two different groups of high schools, all of whom volunteered to participate in this project. Five of the participating high schools are associated with the Bay Area School Reform Collaborative (BASRC), which provides resources for schools to undertake systematic efforts to improve teaching and learning.

ERIC

BASRC requires funded schools to collect and analyze data to measure the effects of their improvement efforts, but leaves it up to the schools to decide what kind of data to collect. BASRC high schools, therefore, do not share a common base of information. A large part of our effort with the five BASRC schools went into creating a common database showing changes in measures of student achievement and engagement. Section II describes how we worked with schools to design and build this database, and presents some of the actual data.

We also worked with 27 schools belonging to the High Schools That Work (HSTW) network, which in 1998-1999 encompassed more than 800 high schools across the nation. HSTW requires all its member schools to join in a biennial assessment that collects student test data in math, science, and reading, as well as survey data from students and teachers. The HSTW assessment usually includes only students who are defined as vocational completers, which in most cases means students who have completed three or more related vocational courses.

The primary focus of our work with the 27 HSTW schools was to test the feasibility of expanding the sample for the 1998 assessment to include a schoolwide cross-section of seniors. This kind of schoolwide sample allows comparisons among groups of students within a school, in particular between vocational completers and other students. We used the data from the 1998 expanded assessment to produce a report for each school that included comparisons among groups of students. The report also showed results for all 27 schools, so that each school could see its students' performance in relation to that of students in similar schools. Section III describes the work with HSTW schools in more detail.

Unfortunately, the availability of schoolwide data for only one year means that we could not compare changes over time for the HSTW schools in this report; however, most of these 27 schools are repeating the expanded schoolwide sampling for the HSTW assessment in 2000. This will provide the kind of trend data that should be most useful in analyzing the effects of changes in school practices. Data from the 2000 assessment will also indicate whether the additional information obtained from schoolwide sampling is worth the additional cost.

The information presented in this report is not intended to determine whether a given school is succeeding, or whether certain practices have been effective. This is not an evaluation of schools, practices, or networks. It is a study of the feasibility of collecting and

ERIC

analyzing data on schoolwide change in a way that might fuel a sustained cycle of inquiry involving people from different schools.

This project demonstrated that joint inquiry into comparative data is feasible. Teachers, counselors, and principals voluntarily participated in collecting, compiling, and analyzing the data. Many commented that the insights they gained were worth the effort. The experience suggests that a collegial process for analyzing comparative data may be a useful addition to the evolving set of institutional mechanisms designed to help schools improve.

This report is intended to be most useful for state and federal agencies, school district administrators, foundations, school network leaders, and education researchers or consultants working with groups of schools to promote continuous improvement. An Epilogue contains some additional thoughts on the relative benefits of centralized and decentralized data collection.

TABLE OF CONTENTS

Acknowledgments i
Executive Summary iii
I. Introduction: Issues in Using Data To Support School Improvement 1
II. Comparing Trend Data with Five High Schools from the Bay Area School Reform Collaborative 11
III. Comparing Schoolwide Data for 27 Schools from the High Schools That Work Network 41
IV. Epilogue: Some Thoughts on Centralized Versus Decentralized Data Strategies 55
References 57
Appendices
Appendix A: 1997-98 School Data Report Created for Collecting Data from BASRC Schools 61
Appendix B: Sample of Site Report Produced for Each Participating HSTW School 69

10

I. INTRODUCTION: ISSUES IN USING DATA TO SUPPORT SCHOOL IMPROVEMENT

Overall Purpose of the Study

The purpose of this study was to develop and test the feasibility of a procedure for assessing and comparing schoolwide change, so that groups of schools could use this information for continuous improvement. To this end, we worked closely in 1998 and 1999 with administrators and faculty in two different school networks—the Bay Area School Reform Collaborative (BASRC), located in the San Francisco Bay Area, and the Southern Regional Education Board's (SREB's) High Schools That Work (HSTW) network, which involves schools from 23 states across the nation.

All participating schools were committed to making pervasive changes in curriculum content and instructional methods, trying to prepare all students for college and careers, and systematically using data for continuous program improvement. Ultimately, 5 BASRC and 27 HSTW schools participated in the study.

We aimed to test the feasibility of involving these groups of schools in seeking answers to the following three questions:

1. Is our school improving?
2. Is it improving faster or more slowly than other schools?
3. Why?

It is important to emphasize at the outset that this study is *not* an evaluation of BASRC, HSTW, or any of the participating schools. Although this report does contain data on school performance, readers should not try to draw any inferences about the effectiveness of these schools or networks. The data presented here are intended only to illustrate the kind of information that can be obtained from the procedures we developed.

Because of differences between the BASRC and HSTW networks, our work took a different focus in each context. The study with BASRC schools concentrated on the feasibility of gathering and analyzing comparable information across schools that had not previously participated in a common data collection effort. In contrast, the HSTW schools were already required by SREB to participate in the same common assessment, which included testing student performance in reading, math, and science. In addition, HSTW schools already had opportunities to analyze their data in comparison to one another. The challenge for the HSTW schools was collecting schoolwide data because they had previously collected information only on students classified as vocational completers. For the purpose of this study, HSTW schools added an expanded sample representing a cross-section of the entire senior class. Sections II and III of this report describe our work with the BASRC and HSTW schools, respectively.

Developing Comparable Data for BASRC Schools

One of the central ideas underlying the procedure we developed in this study is that a school may be able to learn more by comparing its own results with those in other schools than by examining its own data alone. Just as home buyers, policymakers, and researchers all find it useful to compare information for different schools, the teachers and administrators who are responsible for improving school performance should also be able to gain insights by viewing other schools' results and practices in comparison to their own. Of course, teachers and administrators from different schools often exchange information, in an informal fashion.

The desirability of engaging schools in more systematic analysis of comparative data may seem obvious, but there is sometimes reluctance to do it because such comparisons have often been used to shame and blame schools where students are not performing well, while honoring schools with high-performing students. Simply comparing the level of standardized achievement test scores across schools at one point in time, as is often done, reveals more about the socioeconomic composition of students in different schools than about the effectiveness of teaching and learning practices. Educators in low-income schools find this frustrating, and have developed a strong resistance to interschool comparisons of standardized achievement test scores.

At the national level, a similar tension has existed about whether to publish comparisons among states. National data on student achievement was first collected systematically in the 1960s, with the establishment of the National Assessment of Educational Progress (NAEP). Following the publication of *A Nation At Risk* in 1983, there was renewed concern about the performance of U.S. schools and the paucity of relevant national education data. Subsequently, efforts to improve and develop national education data multiplied, encompassing the National Center for Education Statistics (NCES), the National Science Foundation (NSF), and various professional and advocacy organizations such as the Council of Chief State School Officers (CCSSO). NCES began a concerted effort to improve federal education data, revising its annual *Condition of Education* and also publishing *Elementary and Secondary Education Indicators in Brief*, among other reports; NSF focused on developing national indicators of math and science education; and CCSSO identified its own set of priority indicators. National interest in school reform in general and in the reform of math and science education in particular fueled these efforts (Smith, 1988).

The desire to make comparisons among states led policymakers to ask for national data to be broken down into smaller geographic units (Murnane & Raizen, 1988). In the early 1980s, the U.S. Department of Education published an annual "wall chart" of state education statistics, which highlighted for the first time differences in education performance among the states. Even the CCSSO, which had initially opposed state comparisons, voted in 1985 to support generating and publishing state-by-state NAEP achievement test scores. A majority of states had implemented statewide student assessment programs by the late 1980s, and several had begun developing school and state "report cards" to highlight their own performance and promote accountability among schools and school districts (Blank & Gruebel, 1993; Selden, 1994).

During the 1990s, federal initiatives required states to develop performance measures and standards for monitoring educational progress. The Carl D. Perkins Vocational and Applied Technology Education Act of 1990 (the 1990 Perkins Act) required that states receiving Perkins funds develop and implement performance measurement systems for use in local vocational education program evaluations. This emphasis on performance measurement for vocational education was reinforced by the 1998 Perkins Act amendments. In addition, the School-to-Work Opportunities Act of 1994 (STWO Act) required that states receiving school-to-work funds develop performance measures to monitor the progress of their educational systems, particularly with regard to school-to-work activities.

Because of flexibility in the 1990 Perkins Act, states adopted widely varying performance measures (Hoachlander, Levesque, & Rahn, 1992). With passage of the STWO Act and 1998 Perkins Act, however, the federal government has supported efforts to develop comparable measures across the states. Work groups of state representatives have met periodically to identify common outcomes and measurement approaches. Balancing state autonomy with interstate comparability has been a key concern. State and federal administrators, themselves, want the ability to compare progress across the states. They also recognize public pressure for such comparisons.[1] State comparisons provide an important source of information for interpreting educational progress as well as for evaluating various state-level education reform approaches.

A similar logic has led many states to collect and publish data on individual schools over time, in order to make comparisons among schools within a given state. According to Education Week (1999), 13 states in 1998 maintained websites that contained "report cards" on individual schools, with test data for more than one year in subjects that include English or math for 11th grade. Ten other states tested 11th graders in subjects that include English or math, but either do not include school report cards on their websites or else do not report data for more than one year (pp. 85-87).

Now that databases have been developed to allow policymakers and analysts to make comparisons among schools, the question remains whether schools themselves can make use of this data to guide their own self-improvement efforts. Schools find themselves awash in data that has been collected to comply with various outside requirements, yet few schools make systematic use of such data to direct or evaluate their own planning and design efforts. An NCRVE study of the effects of 1990 Perkins performance measurement requirements found that state-produced data reports were rarely used at the local level (Stecher et al., 1994). NCRVE's experience with data and schools suggests that although schools routinely collect a variety of data, it is not easy to make the leap from collecting and reporting this data for administrative purposes to using the data to improve school practices and outcomes. School administrators and teachers face many challenges and heavy demands on their time. *At Your Fingertips: Using Everyday Data To Improve Schools* (Levesque, Bradby, Rossi, & Teitelbaum, 1998) was developed by NCRVE through MPR Associates to assist schools in this endeavor. It describes in nontechnical language how schools can develop performance indicator systems to monitor progress toward their educational goals. Our work with the BASRC schools in this study employed methods from *At Your Fingertips* to help schools use data for improvement purposes, and it went beyond those methods by developing new procedures to engage groups of schools in comparative analysis of their data.

[1] The 1993 Government Performance and Results Act (GPRA) has provided additional pressure to develop comparable data at the federal level. GPRA has required federal agencies, including U.S. Department of Education agencies, to develop performance plans that indicate how their program objectives will be met and results measured.

ERIC

BASRC is an excellent context in which to explore the possibility of schools using comparative data. The BASRC process is designed to foster a cycle of inquiry in which schools continually collect and use data to determine whether their efforts to improve teaching and learning are producing the desired results. Every school that receives BASRC funding is required to participate in an annual Review of Progress, which includes intensive examination of data by school stakeholders to determine whether the school is progressing toward its stated goals. BASRC also has encouraged exchanges of information among schools by pairing schools as "critical friends" who visit and meet with each other, and through an annual Collaborative Assembly where teachers and administrators make presentations about what they are accomplishing.

In spite of these efforts, the evaluation of BASRC after its second year found that only 43% of participating high school teachers said their schools used data as a basis for decisionmaking, compared to 76% of elementary teachers (McLaughlin, Talbert, & Crowe, 1998, p. III-8). This survey question produced the biggest difference between elementary and high schools, although high school teachers also reported less progress toward schoolwide improvement than elementary teachers in response to other questions. Evidently, high schools find it particularly difficult to muster a coherent set of data to inform decisionmaking.

Our work with the BASRC schools was intended to find out whether it would be possible to engage a group of high schools in comparative data analysis that would promote their own cycles of inquiry. Comparing strategies and results among schools could reveal which schools have been making more rapid progress, and might suggest which practices account for that. Schools could thus learn from one another's experience over time. The imposition of new testing and accountability measures in California also is creating stronger external incentives to engage in this kind of comparative self-assessment.

On the other hand, there are a number of possible barriers to designing a data collection effort that allows comparisons among schools and also supports each school's own self-improvement efforts. BASRC spans a number of different school districts, among which reform objectives vary. Also, the measures that some teachers say they value most tend to be the least available and least comparable—for example, locally developed portfolios and other authentic assessments. Thus, the data elements that are collected as part of a comparability study may be less useful than those that are excluded because of their inherent lack of comparability. Section II describes how we dealt with these and other obstacles.

Developing Schoolwide Data for HSTW Schools

From the inception of HSTW, use of data for continuous program improvement has been strongly emphasized—probably more than in any other large school-reform effort. All member schools are required to participate in a biennial HSTW assessment, which provides achievement data in math, science, and language arts for seniors who are classified as vocational completers. The students also answer dozens of detailed questions about courses they have taken and whether they have experienced specific school practices. In addition, the biennial assessment surveys teachers and collects follow-up information from recent graduates. Every school receives a report of its assessment results along with several benchmarks, including national averages and results for high-achieving HSTW schools that have similar sociodemographic characteristics. HSTW teams visit each school periodically, and then use the assessment data to identify areas of needed improvement. Based on assessment results and other school data, HSTW also identifies pace-setting schools in each state, which others are encouraged to visit and observe. Finally, HSTW sponsors numerous events in which assessment results are discussed, including teleconferences, state and national workshops devoted to using data, and the annual professional development conference attended by thousands of teachers and administrators, during which both plenary and concurrent sessions draw on HSTW assessment data.

In this context where an extraordinary amount of attention has been given to data-based program improvement, the primary objective of our study was to investigate the feasibility and usefulness of gathering data from a representative sample of the entire senior class, in addition to the data HSTW schools were already collecting on vocational completers. Our supposition was that data from a schoolwide sample would enhance schools' ability to address our three basic questions: (1) Is our school improving? (2) Is it improving faster or more slowly than other schools? (3) What accounts for the different rates of progress?

At the point when this report was written, however, the HSTW schoolwide data did not permit measurement of progress or change over time. Since most HSTW schools have not previously collected HSTW assessment data from schoolwide cross-sections of students, the whole-school data collected for our study represents a 1998 baseline for computing changes in subsequent years. Most of the HSTW schools in this study will be collecting data from another schoolwide sample in the next HSTW assessment in 2000. They will then be able to compute changes between 1998 and 2000. The 2000 data will not be available in time for this report, however. This report, therefore, presents only 1998 data to illustrate the kind of additional information that can be obtained from a schoolwide sample. The purpose here is only to test the feasibility of schoolwide sampling in the HSTW assessment.

Still, the basic questions that frame our effort are posed in terms of change, not the absolute level of performance. These questions will be answerable after the 2000 assessment. Comparing schools in terms of absolute levels of performance can make some schools look bad just because their students come from low-income neighborhoods where educational performance tends to be lower than in more affluent places. Comparing change over time avoids this problem, creating a more level playing field—although changes in student populations still have to be taken into account in attempting to understand trends over time. These issues will be examined with the 2000 data, which is not yet available for this report.

Data from schoolwide samples can be used to guide changes designed to affect the whole school or only part of it. HSTW's goals are to give more students access to the advantaged academic core curriculum, to help students find a focus in a career area, and to give students the guidance and extra assistance they may need to meet higher course standards. In response to these goals, some HSTW member schools have developed schoolwide strategies such as organizing an entire school into clusters, majors, pathways, or academies in which every student belongs to one group, with a theme related to a particular industry, occupational area, or career field. In this kind of setting, only data on a schoolwide cross-section of students can give a complete picture of what is happening as a result of programmatic changes.

Schoolwide data can also help to assess the effectiveness of part-school initiatives. One of the main purposes of HSTW is to use vocational education to advance academic achievement. If schools are successful in raising student achievement through vocational-technical coursework tied to a challenging academic core, then vocational completers will score higher on HSTW achievement tests than they would have in the absence of HSTW reforms. As a result, in a school that implemented HSTW key practices the scores of seniors completing a vocational sequence in 1998 would be higher than in 1996, for example. In fact, such a positive trend has been observed in many HSTW schools, as well as the HSTW network as a whole.

It is also possible that the observed positive trend is to some extent caused by changes over time in the kind of students who become part of the vocational completer sample. HSTW reforms may improve the reputation of vocational education within a school and, therefore, attract students whose academic achievement was higher to begin with. More energetic and talented teachers may also be drawn into the vocational program. To the extent that the composition of students and teachers in these part-school programs changes favorably over time, a simple comparison of how achievement changes over time for students in the program will overstate the true effects of the program.

A hypothetical illustration is given in the box on the next page. As this extreme illustration suggests, gains in a part-school program could all come at the expense of the remainder of the school, if the upgraded vocational program merely pulls higher-achieving students and more gung-ho teachers from the rest of the school. It is important to emphasize that this is a purely hypothetical illustration of an extreme case. We do not know whether such an effect is actually occurring at all, or how large the effect is if it exists. And that is the point: We need schoolwide data to find out. Only data for a cross-section of the whole school can reveal whether such effects are occurring.[2]

Cycles of Inquiry for Schools, Networks, and Policymakers

If people are trying to decide how their own school can improve faster, they may learn by considering how other schools have done it. The explanation of why some schools improve faster than others is seldom simple, however. The process of explanation involves both quantitative analysis and reflective interpretation. The kind of process we have in mind—where people from different schools openly share and analyze one another's data—has rarely, if ever, been used in the past, but the growing interest in standards-based school reform makes some process of this kind increasingly necessary. Our purpose here is to try out this process and document the results.

[2] Even if the overall distribution of test scores remained constant, as in this hypothetical example, it would also be possible for mean test scores to increase in *both* the vocational program and the rest of the school! This would occur if the vocational program grew by adding students from the low-performing (L) group. This illustrates even more dramatically why it is valuable to have data for a schoolwide cross-section as well as for particular groups within the school.

ERIC

Hypothetical Illustration of Pure Selection Effects

Suppose that academic test data is collected each year on the senior class in a particular high school. For simplicity, suppose also that both the total number of seniors and their levels of academic achievement do not change from one year to the next. To be specific, imagine that each year 50 seniors all score at a high (H) level of 70, another 50 students all score at a medium (M) level of 50, and the remaining 50 seniors all receive a low (L) score of 30. The average academic test score for all 150 seniors is therefore 50, and this does not change.

Nevertheless, it is easy to construct a scenario in which upgrading the vocational education program raises the average score of seniors in the program, despite the fact that the distribution of scores across the school as a whole does not change. For instance, suppose initially the vocational program is a traditional automotive repair class that does not draw students who make high academic scores, so that the distribution of academic test scores for seniors in the vocational program and the rest of the school is as follows:

Vocational Students	*Mean Score*	*Remainder of School*	*Mean Score*
10L, 5M	36.7	40L, 45M, 50H	51.5

The vocational program enrolls 10 low-scoring students and 5 medium-scoring students, who have a combined average score of 36.7, compared with an average score of 51.5 for the rest of the senior class.

Now suppose the vocational program has been transformed into a transportation academy where students can fulfill course requirements for college admission in addition to learning technical skills. This attracts more of the middle- and high-scoring students:

Academy Students	*Mean Score*	*Remainder of School*	*Mean Score*
10L, 10M, 5H	46.0	40L, 40M, 45H	50.8

The mean score of vocational students is now almost 10 points higher, despite the fact that the overall distribution of scores in the school as a whole has not changed.

Collecting data on the whole senior class, therefore, gives a more accurate picture of the impact of programmatic change on achievement than collecting data only for students in the particular program.

One element of the explanatory process is quantitative statistical analysis. This could include estimating statistical models in which the dependent variables are measures of school change in student achievement, engagement, or other outcomes of interest to the school.

Predictors would include measures of change in school practices, as well as changes in the composition of the student population.

In using data to improve programs, statistical analysis can never be the end of the story. Statistical findings cannot become a basis for action unless they make sense to school decisionmakers. Teachers, administrators, parents, and others must have the opportunity to question the results and decide whether they are just statistical artifacts or clues to what is really going on. Building schools' capacity to use data is an iterative process that requires high-quality data (such as the HSTW assessment) and a long-term commitment to assessment and data collection about student course taking, teaching practice, curriculum, and continual quality improvement.

Furthermore, if schools are to learn from one another, people from different schools must meet together to examine the data and discuss what it means. To facilitate this, we organized discussions among the BASRC and HSTW schools participating in this project to try to explain observed differences in school progress. Sections II and III include some observations from these multischool conversations.

Just as the cycle of inquiry never ends for an individual school, the process of structuring incentives and supports for school improvement also continues. The basic premise that schools can improve by making more systematic use of data has itself not been subjected to any clear empirical test. Given absence of data on the effectiveness of this policy, current attempts by states, federal authorities, and school reform networks like BASRC and HSTW to promote data-based accountability and improvement for schools must be viewed as tests of a general hypothesis—that schools can be more effective in improving student performance if they pay more systematic attention to evidence about how students are performing. To this, we add a subhypothesis: that schools can learn more effectively from student performance data if they engage in certain kinds of comparative analysis with other schools. We do not have the data necessary to test either our subhypothesis or the main hypothesis in this study. As stated at the outset, this study was a pilot test of whether it is even feasible to engage schools in such analysis. If feasible, the procedure developed in this study could become part of broader institutional designs intended to promote data-based school improvement.

ERIC

II. COMPARING TREND DATA WITH FIVE HIGH SCHOOLS FROM THE BAY AREA SCHOOL REFORM COLLABORATIVE

In order to investigate the feasibility and usefulness of gathering and comparing data across schools, we worked with a small set of reforming high schools that are part of the Bay Area School Reform Collaborative (BASRC). BASRC was formed in the spring of 1995 as part of the Bay Area's Hewlett-Annenberg Challenge, which provided $50 million to Bay Area public schools to be matched by public and private funds over a five-year period. BASRC's vision of reform sees schools, districts, and stakeholders focused squarely on issues of student learning supported by a culture of ongoing inquiry and broad accountability (BASRC, 1999).

In 1999, BASRC included 86 Leadership Schools, each of which was awarded annual grants of up to $150 per student for three to five years, after completing a rigorous, evidence-based, peer-reviewed portfolio process. In addition to the Leadership Schools, BASRC included 135 Membership Schools, 60 districts, numerous support organizations, funders, and a number of community partners; these organizations receive no funding, but participate in BASRC-sponsored events. BASRC requires that its funded schools think about and use data to support their reform work, a focus that was consistent with the objectives of this study.

A number of BASRC Leadership Schools were approached to participate in this study. Four BASRC Leadership Schools agreed to participate because they believed they could use the data to help inform their reform work. A fifth school, which is a BASRC member but not a Leadership School (and therefore receives no BASRC funding), also agreed to participate. Virtually all of these schools are engaged in some work on the development of academic standards, and some are beginning to assess student performance on standards. All four Leadership Schools included a focus on career-related curriculum in their reform work.

ERIC

Method

As a first step in gathering comparable data that are also useful to individual schools, we attempted to identify common data elements that the schools agreed were important. At a BASRC meeting of about 20 Bay Area high schools held in the fall of 1997, the schools identified student *achievement* and *engagement* as the two key performance areas for which they would like to gather and use data to document progress. Subsequently, we focused on collecting data in these two broad areas.

The goal was to select for this study a set of indicators of achievement and engagement that were both meaningful to educators and also readily available from the schools. Through brainstorming and a review of the relevant literature, we began developing lists of possible indicators of student achievement and engagement. The initial list was meant to be exhaustive, and did not necessarily take into account whether such data would be readily available from the schools. For example, scores on student portfolios were included on the initial list of achievement indicators, even though few schools are making extensive use of portfolios.

A review of the literature on measures of achievement and engagement turned up few surprises. For student achievement, standardized achievement tests such as the Scholastic Aptitude Test (SAT) and the Comprehensive Test of Basic Skills (CTBS) were found to be the most widely used and accepted measures.[3] Standardized achievement tests are useful for assessing students' declarative knowledge of a subject; however, they are not designed to match the curriculum being taught in any particular locality or state (Murnane & Raizen, 1988). Course-taking patterns are another indicator of student achievement, particularly when considering higher-level math and science courses (Madigan, 1997; Mullis, Jenkins, & Johnson, 1994). A course title such as Algebra 2 may mean different things in different schools or districts, however. Other measures of achievement, which also vary from place to place, include student progress on locally or state adopted standards, student portfolios, school or district proficiency exams, and satisfaction of college entrance requirements.

A review of the literature on measures of student engagement showed that the most commonly used indicators are direct observations of students' behaviors in the classroom.

[3] Although the SAT is sometimes regarded as an indicator of student achievement, it has been validated only as a measure of aptitude for college, not of what students learn in high school. It should not be confused with the Stanford Achievement Test, which is a high school (and middle and elementary school) achievement exam.

For example, one study attempting to link teacher practices to student engagement used observations of the students every three minutes to determine their engagement levels (Ebmeier & Ziomek, 1983). Other possible indicators of engagement are attendance, including tardies and cuts, and student survey data on rate of homework completion, participation in extracurricular activities, study habits, and other proxies for student engagement with school.

Following this preliminary work, we took lists of the most promising indicators to the five participating BASRC schools. We interviewed administrators at each school to determine whether they could provide data on each of the identified indicators. Second, we interviewed groups of teachers (and, at one school, groups of parents and students) to obtain their opinions about the quality and usefulness of various indicators.

The five schools varied slightly in what they reported they could provide, but all said they were able to provide data reflecting the more traditional measures of student achievement, such as GPA and SAT scores. Most were in the process of developing standards, but no school could provide quantitative data that measured student attainment of these standards. For measures of engagement, all schools said they were able to provide similarly common measures such as attendance and dropout rates.

In the focus groups, teachers' opinions about which measures were most important varied. Almost all seemed to be in favor of developing standards and benchmarks for measuring academic achievement. Most also supported the use of student surveys for measuring engagement. Unfortunately, conducting such surveys was not possible with the resources available for this study. In addition, teachers saw value in considering rates of change over time, both for evaluating themselves and comparing their school to others. Unfortunately, the indicators that teachers believed to be most useful and important, such as GPA scores for individual students *by marking period* or accurate attendance data *by class period*, were not currently being compiled by the schools.

Based on educators' reports about which data were both meaningful and available, we selected specific data elements for inclusion in a school survey. The final measures that were selected for the spring 1998 survey are listed in Table 1. A similar set of measures, revised and updated, was included on the spring 1999 survey. A copy of the spring 1999 survey is attached as Appendix A.

ERIC

For student achievement, the indicators that met the initial feasibility and usefulness criteria included GPA, satisfaction of the "a-f" course requirements for admission to California public universities (explained later in this section), SAT scores, and AP test scores. Although proficiency test scores were not directly comparable because schools historically used different assessment instruments, the schools were interested in tracking them for their own purposes or comparing their pass rates to those of other schools.[4] To measure student engagement, the indicators that were selected included daily and period-by-period attendance rates, transfer rate, and dropout rate.

Table 1. Data Elements on the Spring 1998 Survey of 5 BASRC High Schools*

School Context
- Student enrollment, fall 1996
- Demographic characteristics, fall 1996

Student Achievement
- School proficiency test information, 1996-1997
- Grade point averages by grade level, 1995-1996 and 1996-1997
- Number and percent of seniors meeting each of the University of California's course entrance ("a-f") requirements, 1996-1997
- SAT and ACT information, 1995-1996 and 1996-1997
- Advanced Placement test information, 1996-1997

Student Engagement
- Attendance as reported to the California Department of Education on April 15, 1995-1996 and 1996-1997, including excused absences
- Average attendance rate, 1995-1996 and 1996-1997
- Period-by-period attendance rate, 1995-1996 and 1996-1997
- Transfer rate, 1996-1997
- One-year dropout rate, 1996-1997
- Four-year dropout rate, 1996-1997

* A copy of the spring 1999 survey is attached as Appendix A.

In order to calculate change scores for comparison across schools, the data for GPA, SAT scores, and attendance rates were requested for both the 1995-1996 and 1996-1997 school years. When the study continued in 1999, we requested similar information for the 1997-1998 school year. Schools were also given the opportunity to provide any other performance data they wished to track over time, although in the end the task of collecting the data asked for proved to be more time consuming than expected, and none of the schools decided to report additional data.

[4] All students in California are required to pass a district approved or developed test of basic skills in order to graduate. Students often begin taking this "proficiency" test as early as the 10th grade.

For test score and GPA data, schools were asked to provide the mean (or average), the standard deviation, and the 25th and 75th percentile scores. Definitions of the terms "mean," "standard deviation," and "percentile scores" were provided on the survey, and support was offered to any school staff who had questions about exactly what data were being requested. Requesting measures of dispersion such as standard deviations and percentile ranks gives a more complete depiction of the data than just the mean alone. In addition, since some data were ultimately collected at three points in time, schools could compare not only the mean level over time, but also changes in the dispersion of data over time. This might be an important analytical tool if, for example, a school wanted to not only raise the mean level, but also decrease the disparity between the highest and lowest achieving students.

Data collection and analysis for the study were augmented by three workshops for representatives from the participating schools. At two of the workshops, participants were guided through an analysis of their data, and discussed implications for their own schools as well as the potential for cross-school comparisons and information sharing. At the third workshop, participants were given instruction on how to use the Excel spreadsheet program to compile and analyze data at their school sites. Particular attention was given to downloading information from the school's student record database into Excel worksheets.

Survey Results

Student Populations Served by the Schools

The participating BASRC schools serve very different student populations (see Table 2). Three of the schools are located in the suburbs; two are urban. Included are an urban charter school with 482 students from a wide variety of backgrounds and an urban school that serves 2,307 mostly Black and Hispanic students. The suburban schools include two with relatively small minority populations and one with a racially and ethnically diverse student population. The proportion of students on AFDC in 1997-1998 ranged from 60% and 16% in the urban schools to below 7% in the suburban schools. Three of the schools have sizable populations of limited-English proficient (LEP) students; two have relatively small LEP populations.

ERIC

Table 2. Characteristics of BASRC Schools

School ID	1	2	3	4	5
Urban/Suburban	Suburban	Urban	Urban	Suburban	Suburban
1997 Enrollment	1,508	2,307	482	2,066	2,105
% AFDC	6.8	59.3	18.0	2.1	5.3
% LEP	5.1	27.9	25.1	2.3	14.9
% Free Lunch	16.8	53.2	56.6	6.5	19.7
% Asian	6.3	13.4	33.6	0.7	13.2
% Black	2.7	38.8	15.8	2.0	18.9
% Hispanic	15.7	44.8	32.0	25.8	24.6
% White	70.4	0.7	10.6	68.9	33.3
% Other	4.9	2.3	8.0	2.6	10.0

Student Achievement Measures

GPA and SAT Trend Data

In order to determine whether schools are improving, it is most important to look at trends over time. The two spring surveys aimed to collect schoolwide data for three points in time—the 1995-1996, 1996-1997, and 1997-1998 school years—to attempt to measure whole-school change. Two achievement measures collected were GPA and SAT data. For each measure, schools were asked to compute means, standard deviations, and 25th and 75th percentile values; to disaggregate the data by grade level; and to provide the number of students included in each calculation. Initially, it was believed that school personnel could collect and compute this information during the time agreed upon. It quickly became clear, however, that the data were difficult to obtain, often came in differing formats (e.g., paper, diskette, central or local computerized database), and that school personnel did not have the proper technology and/or expertise to perform the calculations within the allotted time.[5]

For example, although all schools had originally reported that they had access to SAT and GPA data, ultimately only two schools were able to provide the requested SAT data, and three schools were able to provide GPA data. In the case of the SAT data, project staff were able to augment the available information by obtaining some trend data from the California Department of Education's website on the Internet. Using the GPA data to examine trends proved difficult, since the first survey asked only for grade point averages for each grade level for the 1995-1996 and 1996-1997 school years, without specifying that the GPAs be calculated only for coursework taken during each particular year rather than cumulative GPAs. Although the second survey was more specific, there was still some

[5] For a more detailed account of these difficulties, see the "Lessons Learned" section on page 32.

question as to whether the guidelines were followed by all participating schools. As a result, the two points in time cannot be compared.

Grade Point Average

As a measure of student achievement, GPA is probably the most widely used and immediate indicator. GPA and individual grades are the only measures of achievement that are provided regularly to parents, via report cards, over the course of each school year. Parents and students are able to measure performance in specific subject areas as well as overall and, if they are so inclined, devote time and energy to those subject areas in which the student is not performing well or has not shown improvement. Individual student GPAs can also be compared to those of other students in order to provide a measure of improvement relative to a particular cohort.

At first glance, GPA is also an attractive measure for the school to use when attempting to measure its own improvement over time, both internally and in comparison to other schools. Nearly all schools use a GPA system of recording student performance and the majority use a 4-point scale. Due to the regular reporting of grades to students, parents, and district offices, GPA records are kept at the school and updated frequently. As such, GPA data are up-to-date and thought to be accessible. GPA is also a data element in which school personnel usually have some confidence and with which they are familiar. As our attempts to collect GPA data from the schools revealed, however, using GPA as a measure of growth over time is far more complex than one might initially assume, both in terms of availability and relevance.

As has already been mentioned, all schools in the study were confident that they could provide GPA data in a variety of formats. They believed it was one data source over which they had a modicum of control. Unfortunately, like many of the data that were being collected for district reporting purposes, GPA proved difficult to disaggregate in a manner that was conducive to analysis. The study raised questions about both the feasibility and usefulness of GPA as a measure of student achievement.

In discussions during the site visits and workshops, school staff were skeptical about using GPA as a measure of school effectiveness. Teachers and administrators pointed to the well-known variability in teacher grading patterns and differences in course difficulty as two confounding influences on the reliability of GPA as a schoolwide measure. A 3.0 GPA at one school does not necessarily indicate the same absolute level of academic achievement as a 3.0 from a different school. Not only are school GPAs not tied to any common standard, but the standard used at individual schools was said to vary widely based on teacher preference—so much so that it was common for two teachers teaching the same course to have completely different standards and expectations. Participants also questioned the value of the aggregate GPA numbers from their own schools, citing grade inflation and creative accounting methods (not including students with 0.0 GPAs, for instance) as additional concerns. These discussions of the problems with GPA provided guidance for collecting and analyzing such data in the future, in individual schools, or in groups of schools doing comparative analysis.

If schools could collect GPA by grade-level cohort (cumulative only for a particular school year or semester, for instance) and then calculate means, standard deviations, and percentile ranks for each cohort, they might be able to do several things. First, by collecting such data over the course of several years, schools could monitor the performance of a particular cohort from semester to semester, noting changes that occurred and asking questions that might shed light on what school practices, if any, were responsible for the changes. If, for instance, the standard deviation grew by a statistically significant amount from one year to the next, the school might look to course-taking patterns, or attempt to further disaggregate the data by subject, to see if a particular area—science or math perhaps—was accounting for the lion's share of the change. By comparing the performance of several cohorts in this manner, schools could look for trends over time that might indicate other areas needing improvement. Were a school to collect GPA data in a manner that was conducive to the kinds of analysis discussed throughout this report, questions and concerns like those raised by school personnel could be examined more closely and most likely be answered more definitively than is now possible without appropriate data collection procedures.

Whether this kind of GPA data collection and analysis can become a reality depends on schools' capacity and will. As with other measures (e.g., attendance, transfer data), the data are already consistently collected at the school site but not in a form that can be used for meaningful analysis. Making full use of GPA data to guide school improvement would require, at a minimum, equipping someone at the school site with the hardware, software, and know-how to retrieve individual data and report it in a consistent way over time and across schools. Beyond that, increasing the validity and reliability of course grades as measures of student achievement would require extensive communication among teachers to reduce variability in grading practices.

Scholastic Aptitude Test Results

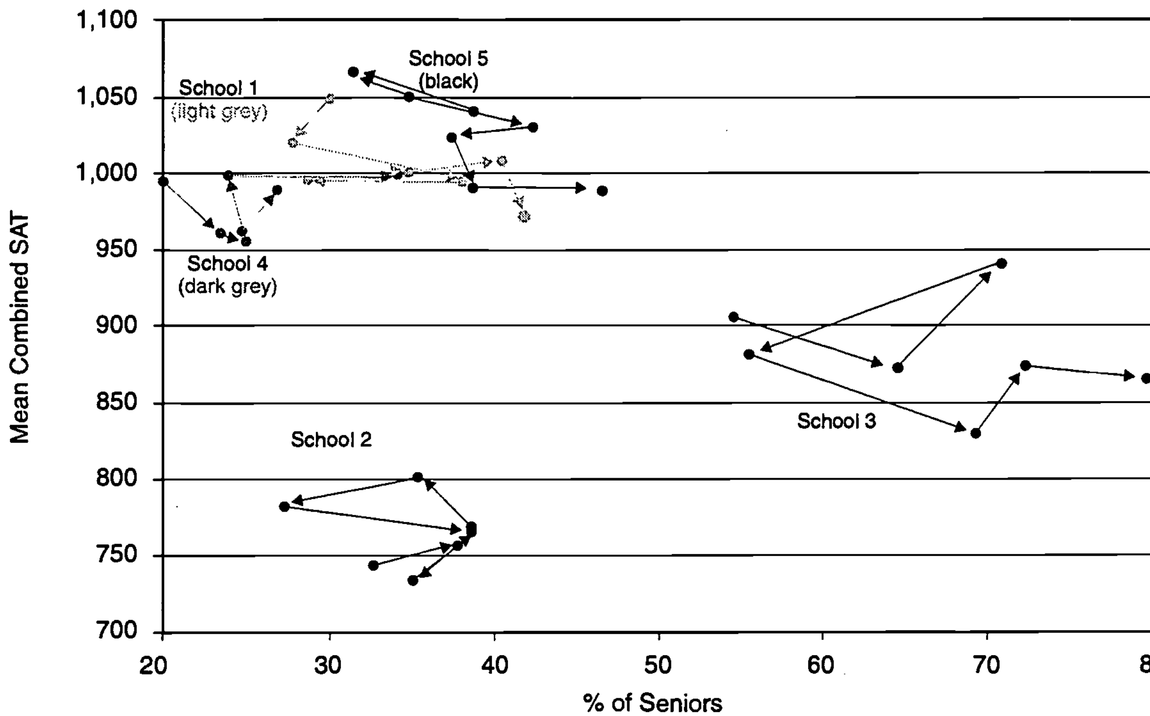
SAT scores also proved difficult for schools to report; however, we were able to use information provided by the Educational Testing Service (ETS) to the State of California to examine the relationship between mean SAT scores and the percentage of seniors taking the test each year from 1990 to 1996 (see Table 3). Over the course of these seven years, all of the participating BASRC schools demonstrated an increase in the percentage of seniors who took the SAT.

Table 3. Percentage of Seniors Taking SAT-1, and Mean Combined Math and Verbal Score, 1990-1996, by School

	1990	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*
School 1							
% Seniors	29.9	27.8	38.0	28.6	33.9	40.3	41.7
Mean Score	1,049	1,022	996	998	1,001	1,009	973
School 2							
% Seniors	32.5	37.8	35.1	38.6	35.3	27.2	38.6
Mean Score	745	760	735	767	802	784	765
School 3							
% Seniors	54.5	64.6	70.8	55.5	69.4	72.3	79.5
Mean Score	907	873	943	883	832	876	867
School 4							
% Seniors	24.7	26.8	20.0	23.4	24.9	24.0	34.6
Mean Score	964	991	996	963	957	1,000	1,002
School 5							
% Seniors	38.7	31.3	34.7	42.3	37.3	38.6	46.5
Mean Score	1,042	1,067	1,053	1,032	1,025	993	991

Figure 1 is a graph of the data in Table 3. The annual data points for each school are connected by arrows showing the direction of movement over time. Every school shows a movement toward the right, indicating the increasing trend in the percentage of seniors taking SAT exams. In this figure, points that are farther to the right and nearer the top of the graph are preferable because they reflect both higher percentages of seniors taking the test and higher mean scores. It is interesting to compare Schools 4 and 5, which are both suburban and similar in size. School 5 has had substantially higher percentages of seniors taking SATs and has been scoring as well or better than School 4, despite the fact that School 5 enrolls more LEP students, and more students who are eligible for subsidized lunch. It might be interesting for School 4 to learn more about how School 5 achieves these results.

Figure 1. Year to Year Changes in Percentage of Seniors Taking the SAT-1, and in Mean Combined Math and Verbal Score, 1990-1996, by School



It would also be interesting for other schools to know how School 4 made such striking gains in 1996, raising the percentage of seniors taking the tests from 24 to 34.6%, while registering a slight gain in mean score from 1,000 to 1,002. This may just be a statistical fluke: Figure 1 shows many instances where a positive change in one year is reversed in the following year. This kind of data analysis allows each school to examine its own recent history, pinpoint years when significant improvement occurred, and ask what was being done "right" or differently in those years. Data of this kind can support a cycle of inquiry in which each school may learn from its own experience, and from one another's.

Results on State-Mandated Achievement Tests

All California schools were required in the 1997-1998 school year to begin administering the Stanford Achievement Test-9 (SAT-9)—not to be confused with the Scholastic Aptitude Test (SAT)—in 2nd through 11th grades. This provides another data source with which to make cross-school comparisons. Data from the state's website are displayed in Table 4 and Figure 2.[6] Specifically, for the graduating classes of 1999 and 2000 at each school, the numbers shown are the percentages of students who scored at or above the nationwide 50th percentile on the SAT-9 exam in math and reading, in 1997-1998 and 1998-1999.

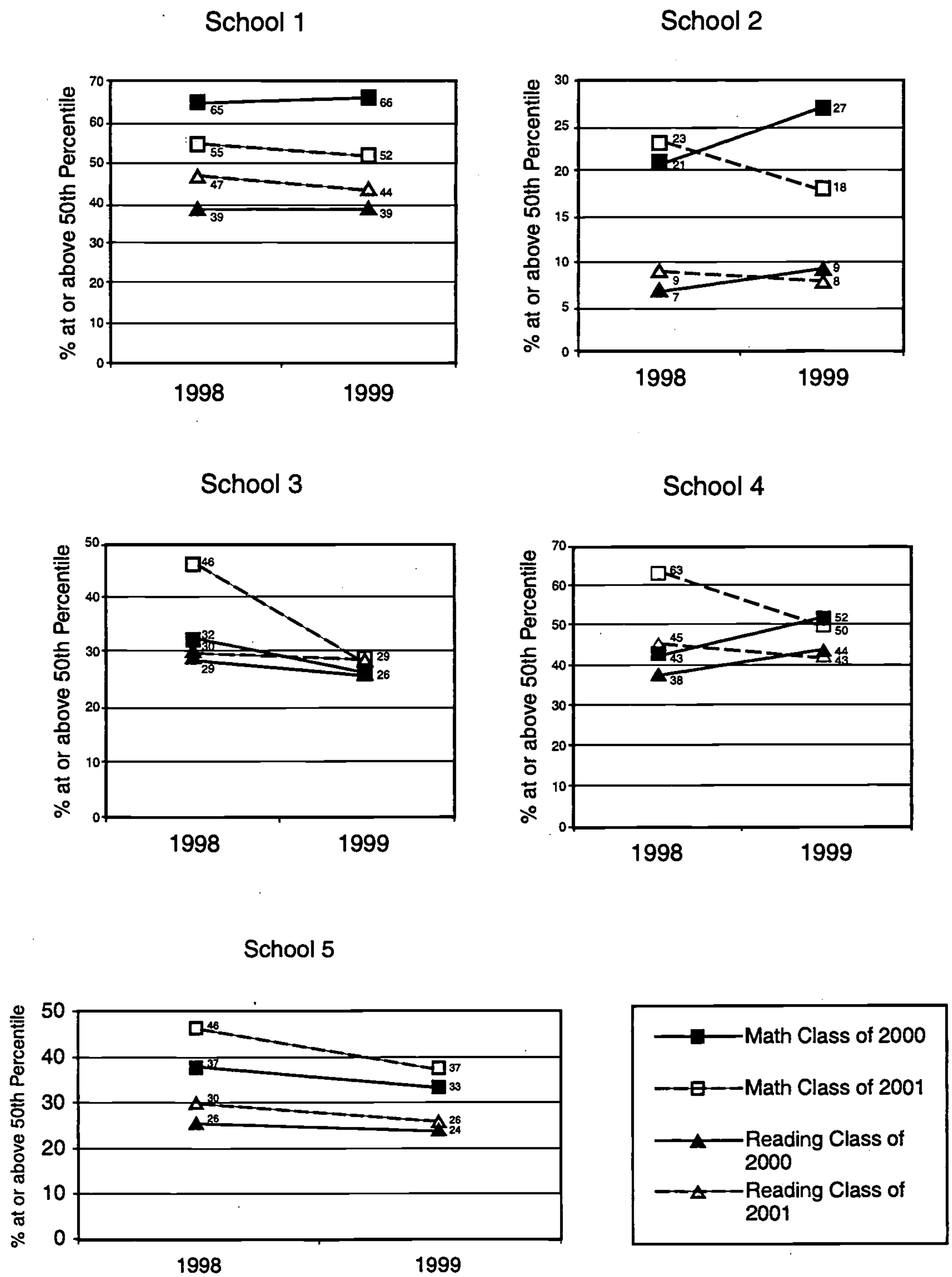
Figure 2 shows a general decline in this indicator for all five schools, except the class of 1999 in Schools 2 and 4. In the School 4 cohort, the number of students at or above the 50th percentile increased from 43 to 52, while at School 2 the same cohort increased from 21 to 27. The value of this cross-school comparison is derived from the insights that might be drawn from the two out of five schools that showed improvement from one year to the next. Though these schools are very different demographically, their insights into practice might reveal similarities that could be replicated at other schools. They might also have taken very different approaches that might pertain more readily to schools that share a similar background.

Table 4. Percentage of Students Scoring At or Above the 50th Percentile on SAT-9 Tests in Math and Reading, 1998 and 1999, by School

	Math		*Reading*	
	Class of 2000	Class of 2001	Class of 2000	Class of 2001
School 1				
1998	65	55	39	47
1999	66	52	39	44
School 2				
1998	21	23	7	9
1999	27	18	9	8
School 3				
1998	32	46	29	30
1998	26	29	26	29
School 4				
1998	43	63	38	45
1999	52	50	44	43
School 5				
1998	37	46	26	30
1999	33	37	24	26

[6]These are averages for all students tested, including LEP students; however, School 3 was in a district that did not test LEP students.

Figure 2. Percentage of Students Scoring At or Above the 50th Percentile on SAT-9 Tests in Math and Reading, 1998 and 1999, by School



SAT-9 data is relatively easy for schools to obtain. It is available on the Internet, and it is disaggregated by gender, LEP/non-LEP, and socioeconomic status (SES). It is a relatively simple task to download a file from the state website into an Excel file. Using the file transfer procedures we taught in our Excel workshop, school staff could incorporate these data into their own cycle of inquiry. That being said, the data is also quite complex and the measure might not be a good "fit" for all schools. With five subject areas—reading, math, language, science, and social science—it might be difficult for schools to match on-site curriculum with the subject-matter items on the SAT-9 tests. A poor showing on the test may say more about the fit between the two than about the quality of instruction at the school. While this may present problems, it may also raise questions about the relevance and difficulty of the courses being offered in the school, which would serve a useful purpose in aiding the process of reform.

This leads to difficult questions about the nature and direction of school improvement programs in relation to efforts to measure them. While this data is attractive for the ease with which it is gathered and manipulated, does it accurately and equitably measure the performance of the individual schools? In using the SAT-9 data to measure student performance, a school should have the option of first deciding whether or not the assessment measures the knowledge and skills on which they have placed the greatest emphasis. If they accept that the assessment matches their mission, then a school must be certain that its focus and curriculum dovetails with the assessment. Even so, the fact that the state has mandated that a certain test be given annually and has put in place rewards and sanctions to ensure that it is a high stakes test, gives individual schools less philosophical space within which to maneuver.

Also rarely discussed are issues such as test administration. During the course of our study, one member of the study staff happened to be in a local public high school (not one of the study schools) on the day that the SAT-9 test was administered. He was told that attendance that day was particularly low because students felt the test had little relevance or importance to their own lives. The staff member also witnessed a wide range of test administration procedures, from those which went strictly according to the test instructions to those in which a classroom full of students had not yet begun taking the test 30 minutes after it was supposed to have been started.

ERIC

University Course Requirements

For the first data survey, schools were also asked to collect information about the percentage of students meeting the so-called "a-f" requirements for admission to the University of California (UC) or California State University (CSU). These are part of the eligibility requirements designed to identify the top one-eighth of the state's graduating high school seniors for UC, and the top one-third for CSU, as prescribed by the California master plan for higher education. Minimum eligibility requirements include completing the following a-f courses:

a. Two years of history/social science, including one year of U.S. history or one-half year of U.S. history and one-half year of civics or American government; and one year of world history, cultures, and geography

b. Four years of college preparatory English, including frequent and regular reading of classic and modern literature and writing

c. Three years of math (four recommended), including elementary and advanced algebra and two- and three-dimensional geometry

d. Two years of laboratory science (three recommended), in at least two of these three areas: biology, chemistry, and physics

e. Two years of a language other than English (three recommended), with at least two years in the same language

f. Two years of college preparatory electives from at least two of the following areas: history, English, advanced math, laboratory science, language other than English, social science, and visual and performing arts.[7]

Applicants must have a GPA of at least 3.3 in these courses, or else sufficiently high SAT or ACT scores to compensate for a lower GPA.[8]

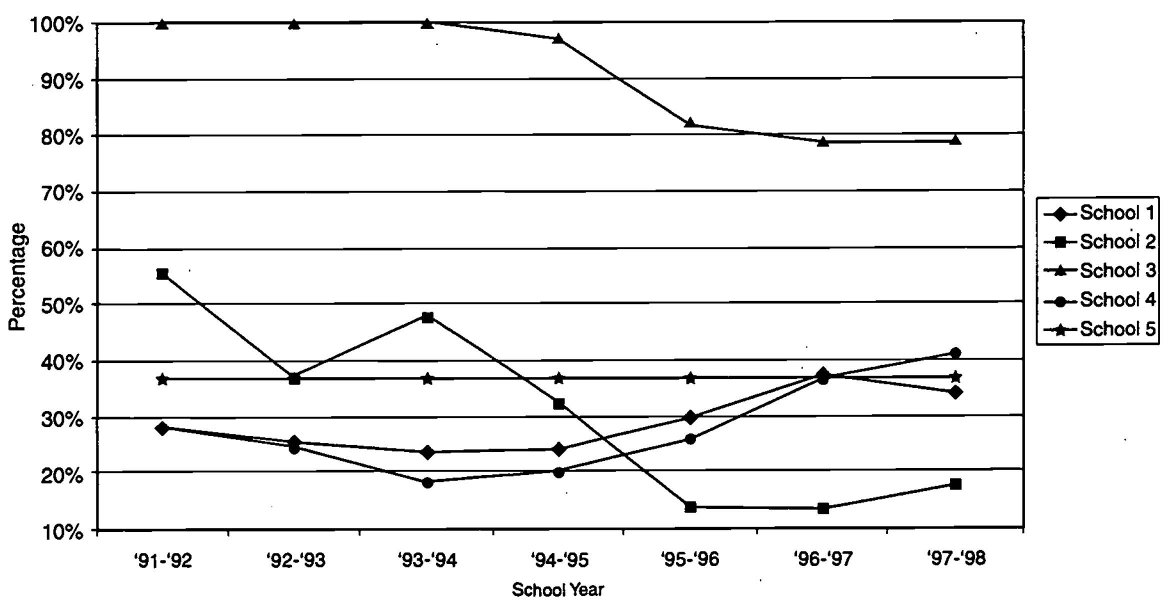
Through the state Department of Education, we were able to collect seven-year trend data for the percentage of graduates meeting all of these requirements.[9] As can be seen in

[7] Visual and performing arts will soon become a separate requirement, known as requirement "g".
[8] For applicants with GPA less than 3.3 in the a-f courses, the minimum GPA ranges from 3.29 with a combined SAT I (math and verbal aptitude) score of 570, as far down as 2.82 if the applicant has the maximum possible combined SAT I score of 1,600.
[9] Schools were asked to disaggregate by course type, but only two schools were able to provide data in this form.

Figure 3, School 3 evidently tailored its curriculum to these course requirements, so that for three years all seniors had fulfilled them. Figure 3 also shows that the percentages of students who met the requirements in Schools 1 and 3 fell substantially between the 1993-1994 and 1995-1996 school years. It turns out that a-f course requirements became more strict during this period, and that Schools 1 and 3 had not yet adjusted their curricula accordingly. Representatives from the schools also expressed their belief that the quality and rigor of the a-f required courses varied considerably from school to school, despite the universities' desire to maintain common standards for course content.

On the other hand, Schools 4 and 5 increased their proportions of seniors meeting a-f requirements from 1994-1995 through 1997-1998. It could be instructive to find out how they did that.

Figure 3. Percentage of Seniors Meeting All A-F Requirements, by School, 7-Year Trend



Student Engagement Measures

Student Attendance

While trend data on attendance would be quite useful to any whole-school reform effort, collecting accurate, dependable attendance figures proved beyond the capacity of project staff and data collectors. When we started down the path by collecting attendance data for the first survey, we soon found that the results would not be analytically useful. At the first data workshop, a slide was shown with attendance data showing significant increases in two of the schools (1 and 4) and very little change in two others (2 and 5). These findings from the first survey, when presented to the school staff and data collectors, sparked a discussion about how much faith one should place in increases as dramatic as these, which indicated that school staff did not have a great deal of confidence in the attendance data in general. In order for the attendance rate to increase from 89% to 93% in one year, the average student would have to come to school seven more days during the school year (assuming a 180-day school year).[10] In order for School 4's attendance increase to be accurate, the average student would have had to miss no more than four days of school during the entire school year, a figure that stretched the bounds of credulity among participating educators. Another possibility is that these data were collected in somewhat different ways during this two-year period of time. As a result of these apparent inaccuracies and the inordinate amount of time required to collect meaningful attendance data, we decided not to attempt to study attendance trend data.

The obvious importance of accurate attendance data to any whole-school reform effort makes the difficulties in collecting it even more vexing. Student absenteeism is one of the greatest challenges facing schools with majority enrollments of low SES students. While attendance can be seen as a direct reflection of student engagement, it also influences the effects of a policy, program, or curriculum on student achievement, since interventions cannot be effective unless students are exposed to them.

Although effective systems are in place for collecting and recording attendance at every school, the actual number of students in daily attendance may tend to be over-reported due to its importance in obtaining school funding from the state. In California, a student who is reported as absent for all but one period during the school day is nevertheless counted for fiscal purposes as having been present that day.

Unfortunately, detailed attendance data is not always retained at the school after it is reported to the district. Schools receive regular summary reports back from the district, but these reports usually lack the detail necessary to guide program improvement. Furthermore, some districts purge detailed attendance data on a yearly basis, making it impossible to compare trends over time for different groups of students.

[10] School staff speculated that a teacher strike during the 1995-1996 school year negatively affected student attendance.

Enrollment by Grade Level

The pattern of enrollment by grade level can be interpreted as a rough indicator of student engagement. Although more refined measures of engagement would be preferable, enrollment by grade level has the advantage of being one of the few data elements that was readily available for more than one year at every participating school site. Table 5 shows the numbers of students enrolled at each grade level in fall 1995, 1996, and 1997.[11] Data for 1995 and 1997 are plotted in Figure 4. A steeper profile in Figure 4 means that more students have dropped out or transferred out, relative to the number who have transferred in, from one grade level to the next. If the profile becomes less steep between 1995 and 1997, it suggests that the school's power to hold on to its students and attract new transfers has increased.

Table 5. Fall Enrollment in Each Grade Level for 1995, 1996, 1997, by School

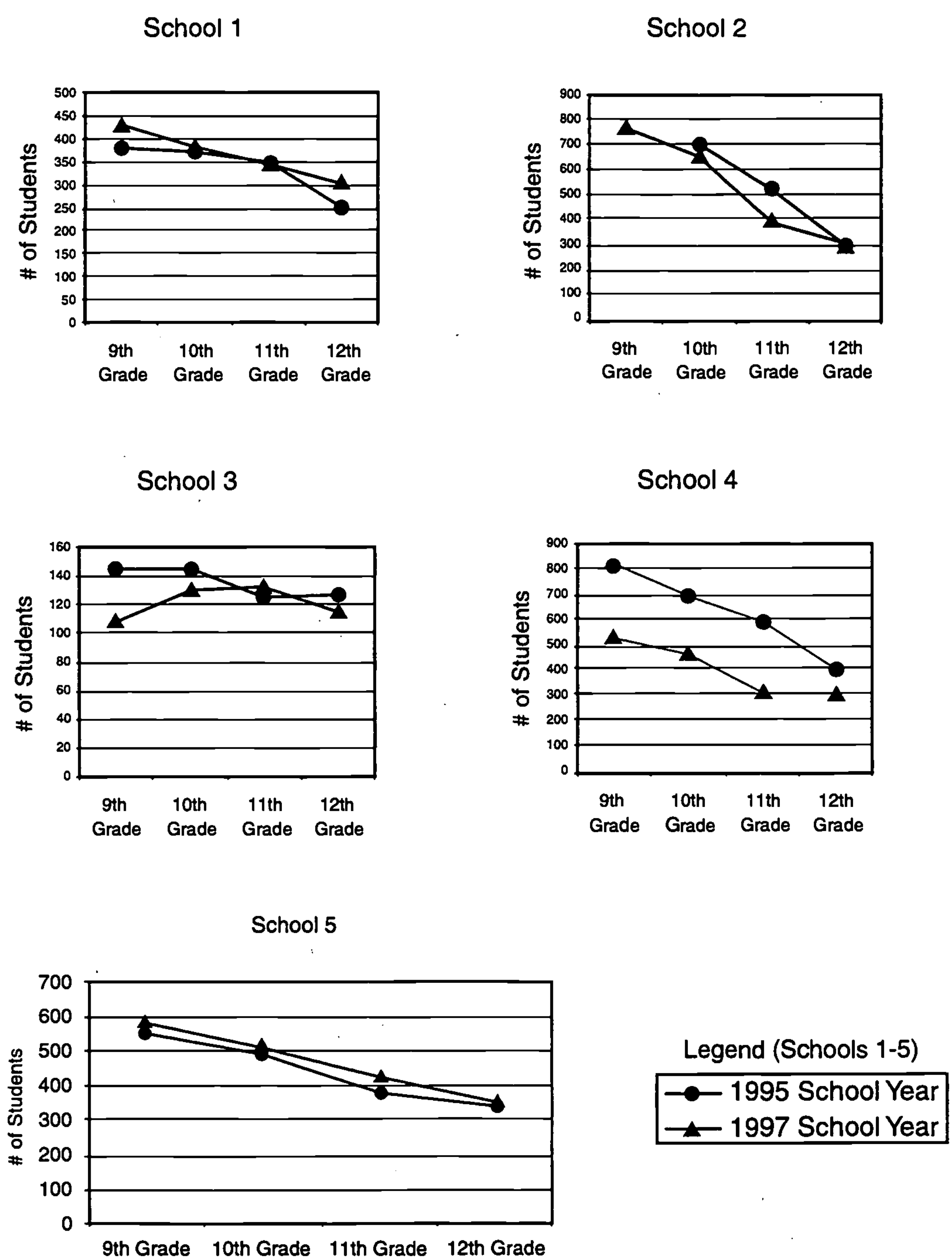
	9th Grade	*10th Grade*	*11th Grade*	*12th Grade*
School 1				
1995	383	374	346	254
1996	380	351	351	291
1997	430	385	349	309
School 2				
1995		698	529	293
1996		727	436	328
1997	767	656	399	309
School 3				
1995	145	145	126	127
1996	127	133	140	124
1997	110	132	133	116
School 4				
1995	815	695	594	408
1996	562	533	641	444
1997	531	473	321	316
School 5				
1995	559	495	387	346
1996	601	482	422	324
1997	594	512	437	359

A school's profile might also become less steep if the number of entering freshmen decreases. This could occur if the school-age population in the school's attendance area is declining. To the extent that such a decline is the result of a demographic wave affecting all

[11] Adding the numbers from Table 5 on 1997 enrollments by grade level gives a lower number than the total 1997 enrollment reported for each school in Table 2 because some students in each school are not assigned to a particular grade level.

schools, it would cause the profile for all schools to become less steep. Comparing changes in the steepness of profiles for a number of schools makes it possible to observe changes that are *not* due to general demographic shifts in the school-age population.

Figure 4. Fall Enrollment in Each Grade Level, 1995 and 1997, by School



In Figure 4, the school that showed the most striking changes between 1995 and 1997 were Schools 3 and 4, where the ratio of seniors to freshmen was substantially higher in 1997 than in 1995. This might suggest an increase in the school's holding power; however, the change resulted from a reduction in the numbers of students in 9th-11th grades, not an increase in the number in 12th grade. Since this kind of change occurred only in Schools 3 and 4, it cannot be attributed to some general demographic shift. It has something to do with Schools 3 and 4, but the data do not tell us what. As with other findings, this one is the beginning of an inquiry, not the end.

Transfer Trends

A clearer but still limited indicator of a school's *lack* of holding power is the fraction of students who transfer out. Although some transfers are completely unrelated to the quality of the school—those due to parents changing jobs, for example—other transfers are initiated by students or parents based on their judgment of how well a school is performing. As with other indicators of engagement or achievement, *changes* in the out-transfer rate are likely to be more informative than the level of those rates, especially if a trend continues for several years. Three of the five BASRC schools provided data on the numbers of students transferring out of each grade in both years, 1996-1997 and 1997-1998. The three schools that provided this information happen to be the suburban schools: Schools 1, 4, and 5. The data are shown in Table 6 and Figure 5.[12]

Figure 5 shows that out-transfer rates declined for all grades in Schools 1 and 4 but increased in School 5. The decline was especially pronounced in School 4; this may be another indication that School 4's holding power was growing stronger in this period. Again, this comparison should provoke these schools to seek an explanation. Was School 4 becoming more effective in holding on to students?

[12] Students who transfer *into* the school are not accounted for in these data. In other words, the out-transfer rates are gross, not net.

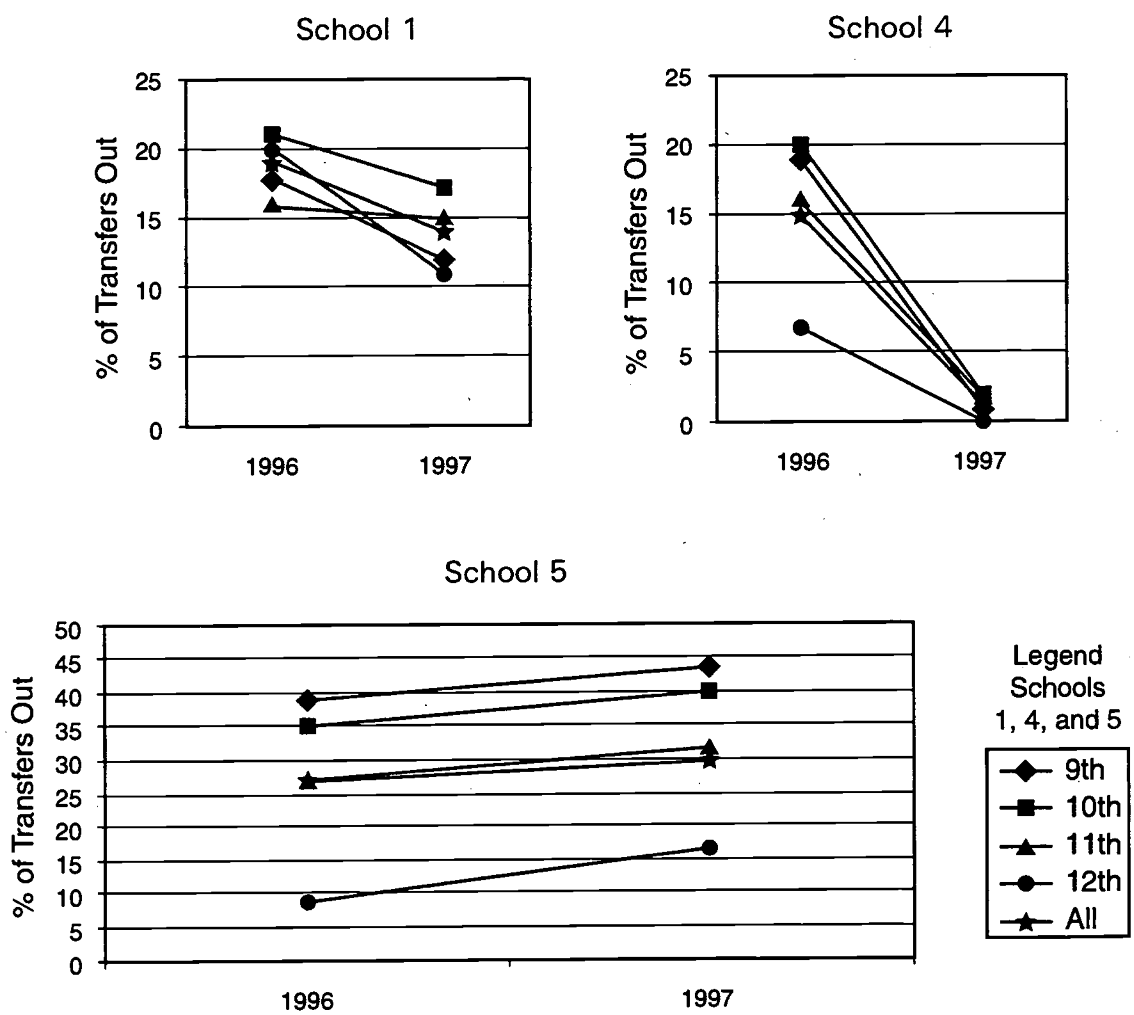
Table 6. Student Transfers Out, As a Percentage of Class Enrollment, for Schools 1, 4, and 5 in 1996 and 1997

	9th	*10th*	*11th*	*12th*	*All*
School 1					
1996	18	21	16	20	19
1997	12	17	15	11	14
School 4					
1996	19	20	16	7	15
1997	1	2	2	0	1
School 5					
1996	39	35	27	9	27
1997	44	40	32	17	30

Within a large district, changes in the rate of out-transfers (and in-transfers) could reveal how schools' relative position is changing in the hierarchy of desirability. One member of the study team observed that in a large local school district there were two comprehensive high schools that were considered desirable by parents and students and a number of others that represented decreasing levels of effectiveness and safety. A cross-school comparison of changes in transfer rates, including an analysis of the reasons for transfer, would shed some light on changing perceptions of relative school quality among parents.

ERIC Full Text Provided by ERIC

Figure 5. Student Transfers Out as a Percentage of Class Enrollment by Grade Level, for Schools 1, 4, and 5 in 1996 and 1997



41

Lessons Learned

Schools Lack Sufficient Capacity To Compile and Organize the Data They Collect

It took much longer than anticipated for schools to complete and return the data survey, which was a nonstandard report that required gathering data from disparate sources and necessitated calculations that hadn't been computed previously. The survey was first sent in mid-April of 1998, and was anticipated by the schools as a result of meetings with administrators and teacher focus groups. Schools were asked to complete the survey by late May. The fact that the survey coincided with the end of the school year contributed to the delay in receiving responses. The many other demands on the time of school staff at the end of the year made nonmandatory data collection a lower priority than other pressing concerns. As a result, all schools waited until the school year was over before investing the time required to fill out the survey. By then, many of the relevant staff members, particularly attendance clerks and counselors, had left for summer vacation.

By mid-July, only two of the five schools had returned the survey. The others completed the survey only after the summer, since the individuals best able to assemble the needed data were on vacation during the summer months. No school was able to provide all of the information requested, despite some help provided by project staff and reimbursement to the schools for the time spent assembling the data. To collect even incomplete information took data collectors a considerable amount of time. One data collector spent 86 hours filling out the survey. (See copy of the spring 1999 survey in Appendix A.)

In developing the survey, we attempted to request certain information in a form that was consistent with that required of schools by the State of California, such as the student attendance reports that generate school revenue. This proved difficult to do, however, particularly in the area of attendance. We found that data collectors were often not able to locate the appropriate records of what was submitted to the state. In general, the systems in place for collecting and reporting attendance data were geared toward funding or student management purposes, as opposed to school improvement purposes. Although period-by-period attendance could be reported for individual students, no school was able to report this kind of information for a whole grade level or the whole school.

In the end, data from the survey was supplemented with information from district offices and the state Department of Education. Some of the data gathered for this study were easier to collect via the Internet than from school staff at the schools themselves. Much of the data at the district and state levels, however, were not stored in a manner that allowed for computation of percentile ranks, standard deviations, or disaggregation by grade level.

Schools were able to obtain complete or nearly complete information on four survey measures: (1) enrollment, (2) Advanced Placement scores, (3) transfer rate, and (4) the one-year dropout rate. Attendance data proved to be the most difficult data to collect. The reasons for this varied by school. In some cases, data collectors did not coordinate with the appropriate person in the school. In other cases, school information systems could not be queried in ways that would produce the particular statistics requested by the survey. In general, student achievement measures were more accessible than measures of engagement.

There is a good deal of data on standardized achievement test results available at the schools, although this information was often kept on paper instead of in electronic form. This meant that data entry or hand calculations were required to perform the analyses. Academic year 1996-1997—the most recent year for which we asked for standardized achievement test results—was the last year of a hiatus in state-required testing. Because of this, the schools in our study varied greatly on what standardized achievement tests they administered, and at what grade level, for the two years for which we asked for data. With the new California testing program now in place, all schools are now using the same SAT-9 test. SAT-9 test results are now readily available through the state's website. One of the study schools is also continuing to give the CTBS—the test they have used for several years—in order to have continuous trend data.

Most schools also provided information on proficiency test results. The State of California requires that districts choose or create a proficiency test that all students must pass in order to graduate. While some schools develop their own tests for this purpose, others simply set a percentage standard on the achievement test battery they administer to all students. For example, they might set 70% correct on the 10th grade verbal and math tests as the proficiency level. If students do not achieve this level in the 10th grade, they continue taking the tests until they do. Most schools offer special tutoring or mini-classes to help students achieve proficiency. Although there are usually some students who do not

ERIC

pass the proficiency test until near the end of their senior year, the number actually denied graduation is quite small.

Calculating means, standard deviations, and percentile ranks for GPA and SAT data proved to be quite difficult for school staff. Individual GPA data were typically maintained in order to produce student transcripts and calculate class rank. Computing mean GPA for a cohort or the entire school was not something typically done. Schools can order reports from the College Board that describe the distribution of SAT scores for their students, but none of the five schools had taken this step.

In addition to lack of time, schools lacked the technical capacity required to provide the information quickly. The software packages used by some schools do not calculate standard deviations or percentiles, and the school employees who completed the survey did not know of any way to obtain these measures. To produce the relevant information, we visited the school site and helped to find out what calculations were possible. While most schools had the hardware and software to do all of the relevant calculations, student record databases had to be queried in such a way as to output text files that could then be read and manipulated in a spreadsheet package such as Excel. This helped school data collectors to avoid having to perform computations manually on a calculator.

In an effort to increase the capacity of school staff to collect and analyze relevant data, we held an Excel workshop in April 1999, which was designed to teach the basics of downloading student records from the SASI database into Excel spreadsheets for analysis. The information disseminated at the meeting was met with much enthusiasm, as school data collectors learned far more convenient ways to collect and analyze the data that they had struggled to obtain in 1998. Unfortunately, although these methods have the potential to ease data collection and analysis at the school level, facility with them can only be gained through continued use, which most collectors did not have time for in their capacity as part-time data collectors during the 1999 school year. So, although participants were happy to have learned an easier way to collect and analyze their data, few were able to effectively use what they had learned while collecting data for the 1999 survey. As has been seen in other aspects of the study, facilitating data collection and analysis is not only a matter of providing tools, but also of creating capacity and agency. If data collection and analysis are seen as part-time events that occur in spurts only when time is available, then the results will most likely fall short of what is required for meaningful analysis.

Although schools have plenty of data at their fingertips, no one is usually charged with organizing and analyzing the information to see what it might mean for improving instruction and student outcomes. Schools generally do not have a systematic way of collecting and storing data for ongoing analysis. Much of the data we would like to have examined had passed through the school at some point in some form, but had not been saved. Once the data has passed on to the district office, it is often virtually unretrievable.

Difficulty in compiling data for schoolwide decisionmaking stems from the fragmented nature of data collection and reporting systems in schools and districts. The official data collectors in the schools—attendance administrators, guidance counselors, academic deans, assistant principals, and principals—usually collect and report their data in a unique format to a separate agency at the district level. More often than not, only one or two people in the school have the knowledge and expertise to operate specific data collection systems. In the area of attendance, for instance, the district office is usually able to provide various reports back to the schools, if specifically requested; however, it is usually only the attendance administrator who knows how to request such reports. If this administrator is not interested and involved in the plan for improving a school's use of data for program improvement, this information may be difficult to obtain in a timely manner.

External Reporting Requirements Constrain Schools' Ability To Collect Data for Their Own Purposes

Most of the schools' ongoing efforts to measure student achievement is driven by district and state reporting requirements. In contrast, our data collection efforts were motivated by what teachers and school administrators told us they considered important. Despite our desire to collect data that school staff wanted, we were constrained to gather what was available as a result of district or state requirements. It was not appropriate, for instance, to ask a school to change how they collected and organized GPA data for a given year if what we were proposing did not fit in with the collection and reporting requirements of the district. Having reporting requirements drive data collection and analysis procedures creates an atmosphere in which innovation and change can occur only within given parameters.

A good example is attendance, which turned out to be the most difficult data to collect and perhaps the most unreliable data provided during our study. In most schools, attendance is recorded by each teacher during every class period of the school day. Attendance

ERIC

information is reported to the district on a daily basis and is used to determine the level of a school's funding, based on the number of students attending. Despite the regularity of the data collection and the existence of a fairly sophisticated system of accounting, this information is not generally stored or tracked at the school site. Although one might be able to find out about the attendance patterns of a particular student by looking at his or her report card at the end of a marking period, there is no routine analysis of attendance by class, gender, or ethnic group, for example. With the introduction of a few additional procedures at the local level, schools might be able to analyze attendance data for various groups of students without having to deviate from district-mandated collection or reporting requirements. This would help to guide local school reform.

Meeting district and state requirements also consumes most of the time available to those who might collect and analyze data at the school level. During any given school year, only so much extra time can be devoted to collecting data that is not somehow required. Negative attitudes toward data on the part of teachers and other school personnel, fostered by years of forced conformity to reporting requirements, also undermine efforts to create whole-school data collection and analysis systems.

Comparing Data for Schools in Different Districts Requires Special Attention to Definitions

Although we gave data collectors detailed instructions in how to fill out the surveys, it was not possible to be absolutely sure that the resulting data were, in fact, comparable in the strict sense of the word. Besides some computational mistakes and software knowledge gaps that were evident, some practical decisions were no doubt made in the course of compiling the data that might compromise comparability. It should be remembered that this was a two-time data collection that required many computations, and participating school staff devoted a significant amount of time and effort to produce the requested data. We are grateful for the thought and effort spent responding to the surveys.[13]

Unanticipated complexity in GPA data provided one set of lessons about the difficulty of ensuring comparability. When calculating GPA by grade level, it became apparent that schools reported two kinds of data: (1) mean grades received by students in a single year, by

[13] Although the data collectors were offered compensation for their efforts, not all of them sought reimbursement for their time, indicating that the monetary incentive was not always the primary draw.

grade level; and (2) cumulative GPA, by grade level (i.e., the mean of all grades received by students to date, by cohort). Since we were not clear about which schools used which definition, the only GPA data from our survey that is strictly comparable is freshman GPA because for this grade level, cumulative GPA and average GPA for courses taken in one year are the same. In addition, schools made different decisions about which students to include in the calculations. Some collectors may have excluded students who transferred out, or dropped out, or recorded a GPA of 0.0 for some other reason; some may have left these students in.

Dropout rates are also indicators that deserve some degree of caution in interpretation due to the financial implications of taking a student off the books. The state Department of Education has strict rules about when students should be classified as dropouts, but there is no doubt some room for interpretation. Without an audit, it would be difficult to tell whether schools have all counted dropouts in exactly the same way.

We presumed that data obtained from the Internet was highly comparable because each data element was compiled and reported by a single source; however, some problems with the initial SAT-9 achievement test data were reported by the test maker, serving as a reminder that even professional data analysts make mistakes. The high accountability stakes associated with much of the test data available on the Internet make it likely that any problems will eventually be ferreted out and corrected. Data users should consider, however, how reliable the source of any online performance data may be.

Uniform definitions are most crucial when the purpose is to compare the level of performance across schools at a given point in time, especially where consequences are attached to performance. Consistency across schools may be a little less important when examining trends in performance over time, if change is measured in terms of percentages or standard deviations[14]—although even in this case different definitions might produce different trajectories. The accuracy of trend data depends on the same data definitions and procedures being applied consistently over time, even if a school is interested only in analyzing its own data without comparison to other schools.

[14] This study planned to investigate the usefulness of standardized change scores for comparing dissimilar data elements; however, it proved to be the case that the least comparable data were also the least available. For example, few schools had implemented portfolio assessments or were administering proficiency exams other than the SAT-9; therefore, these measures were not included in the data survey. The remaining data elements were relatively comparable and did not warrant calculating standardized change scores.

ERIC

Involving Staff from Different Schools in Comparing Their Data Is Feasible and Possibly Instructive

The BASRC study was also concerned with whether it is possible to gather data that is simultaneously comparable across schools and useful to individual schools. Results are encouraging. Despite the many difficulties encountered during data collection, and the sometimes qualified nature of the comparisons, teachers and administrators who participated in the workshops and shared their experiences and results were consistently positive in their reactions to the experience. Participants reported that the feeling of community at the workshops, where they shared information and insights about their schools, helped to keep them focused and enthusiastic about the prospects for further data collection and analysis. Participants also felt that the cross-school data comparisons led to frank discussions about specific, often mutual, problems in the schools and were both illuminating and helpful to ongoing reform efforts.

In addition to what they said and wrote on workshop evaluation forms, participants also gave evidence of changing practices. At one school, the principal reported presenting some of the data collected for the study to several stakeholder groups. Staff from the four schools that participated in the Excel workshop intended to use their newly acquired skills in the future for purposes unrelated to the study. One school's BASRC coordinator also reported plans to institutionalize a data collector/analyst role by creating a half-time paid position for this purpose. At another school, the principal has involved the school's leadership council in a discussion about using data and has created a "data box," where all data generated about the school is to be deposited in order to facilitate locating and using the data.

The analysis in this report, which goes beyond what was presented at project workshops, yielded additional preliminary findings that could add further momentum to a joint cycle of inquiry among these schools. Some of the results seem to indicate faster rates of improvement in certain schools in certain years. Again, we must emphasize that the analysis here must not be considered as an evaluation of the participating schools, either individually or as a group. Instead, the results presented here are only a starting point for further inquiry. Given that some schools seemed to improve faster than others in some respects, it would be instructive to find out whether these positive developments occurred as a result of some deliberate action on the schools' part. For example:

ERIC

- How did School 4 sharply increase the percentage of seniors taking SAT tests in 1996, without reducing average scores? (Figure 1)
- How did Schools 2 and 4 produce gains in SAT-9 reading and math scores for certain cohorts of students in 1999? (Figure 2)
- How did Schools 1 and 4 achieve steady increases from 1994-1995 to 1997-1998 in the percentage of seniors meeting a-f course requirements? (Figure 3)
- How did Schools 1 and 4 reduce the percentage of students transferring out, and raise the ratio of seniors to freshmen? (Figures 4 and 5)

School 4 clearly stands out in these comparisons. Were these positive trends a result of improvements in School 4's curriculum, teaching methods, or other instructional strategies?[15] If so, other schools might try to emulate what School 4 did. This illustrates the potential for additional learning from these cross-school comparisons.

The enthusiasm with which this study was received by the participating schools should be regarded with some caution. It must be remembered that these schools volunteered to take part in the study. A few other schools declined the invitation, saying they were too overwhelmed to participate in additional data collection activities. Furthermore, within the participating schools, the teachers and administrators who were willing to work with us presumably had a greater interest in data analysis to begin with. We did not attempt to expand the discussion schoolwide within any of the participating high schools.

Nevertheless, we conclude from this study that it is feasible to collect information which is both comparable across schools and useful to individual schools. Moreover, the study suggests that such data, when shared in a spirit of nonthreatening inquiry, can prompt insights for individual schools that they could not obtain from their own data alone. It appears that this kind of collegial, comparative inquiry can enhance the process of continuous improvement. By easing schools into sometimes difficult discussions about differing performance, this kind of comparative analysis can foster a sense of community based on shared professional commitment.

[15] Among the other possible explanations would be change in the socioeconomic composition of students at School 4. We did not collect the data to test that, but it would not be difficult to do.

III. COMPARING SCHOOLWIDE DATA FOR 27 SCHOOLS FROM THE HIGH SCHOOLS THAT WORK NETWORK

Background

In 1998, we began working with 30 schools from the High Schools That Work (HSTW) network to gather baseline data on whole-school change. HSTW began in 1987 as a program of the Southern Regional Education Board (SREB), and in 1998 included more than 800 schools in 23 states. HSTW has three major goals[16]:

1. Raise the math, science, communication, problem-solving, and technical achievement of more students to the national average and above.

2. Blend the essential content of traditional college-preparatory studies—mathematics, science, and language arts—with quality vocational and technical studies by creating conditions that support school leaders, teachers, and counselors in carrying out key practices.

3. Advance state and local policies and leadership initiatives necessary to sustain a continuous school-improvement effort.

To achieve these goals, HSTW recommends ten key practices for schools, describes five key supporting conditions, and defines a recommended curriculum. As stated on the website, "The centerpiece of High Schools That Work is a curriculum that blends the essential content of college-preparatory mathematics, science, language arts, and social studies courses with modern vocational-technical studies in grades nine through 12." The SREB-recommended curriculum includes the following:

- At least four credits in English courses with the content and performance standards of college-preparatory English

- At least three credits in math courses with the content and performance standards of college-preparatory Algebra I, geometry, Algebra II, and trigonometry

- At least three credits in science, including two credits in courses with the content and performance standards of college-preparatory biology, chemistry, physics, or applied physics

- At least three social studies credits in courses with the content and performance standards of college-preparatory courses

[16] According to the website, www.sreb.org/Programs/hstw/about/Brochure/brochure99.html.

- At least four credits in an academic or a vocational-technical major
- At least two credits in related academic and technical fields, including at least one-half credit in a basic computer course covering word processing, database entry, presentation software, and use of the Internet and e-mail.

As noted in Section I, HSTW has placed extraordinary emphasis on using data to guide school improvement. Every two years, seniors who are classified as "vocational completers" take a special HSTW achievement test in math, reading, and science. Data from this assessment, along with results from student and faculty surveys and a follow-up of recent graduates, are compiled by the Educational Testing Service (ETS) in a report to each school.

For this study, schools were asked to depart from the usual HSTW practice of testing vocational completers only. As part of the spring 1998 assessment, they were also asked to assess a representative schoolwide sample of seniors.[17] Our main purpose in the HSTW part of this study was to test the feasibility of this schoolwide sampling procedure. After schools repeat this procedure in the 2000 assessment, it will be possible to draw conclusions about the advantages and disadvantages of schoolwide sampling.

Method

Here we describe the method for selecting sites and collecting data for the expanded assessment. The data were analyzed and then shared with the schools in the form of individualized site reports and data workshops involving groups of schools. In addition, six site visits were conducted to assess the implementation of reform practices at the visited schools and to inquire about how school staff were using the site reports.

Site Selection

SREB invited 30 schools, from 21 states, to participate in this study. As can be seen in Appendix B, the 27 schools that participated vary in sociodemographic composition and levels of student achievement. Some are rural; others are urban or suburban. All were considered to be relatively intensive implementers of HSTW key practices, but some were new to the network while others had been members for a number of years. In addition, SREB sought to recruit schools that would be especially interested in assessing a schoolwide sample.

[17] Previously, a handful of schools had opted to assess all of their seniors, or a random sample of them.

The HSTW Assessment

The HSTW assessment consists of three parts: (1) a student survey, (2) a course experience survey, and (3) a standardized achievement test of reading, math, and science. The student survey consists of 125 questions concerning what was expected of students in high school, the nature of what they were taught, what they were asked to do, the level of effort they had to make to meet school standards, and the like.[18] In addition, basic demographic information (e.g., race, parental education, and information about the home environment) is also asked in this part of the assessment. The course experience survey compiles information from transcripts about the student's history of high school courses from 9th-12th grades. This portion of the assessment is filled out either by students, under the supervision of testing coordinators, or by the school's counseling staff. Although the survey does not collect information about when courses were taken (which would allow for analysis of course sequences), it does provide a detailed picture of the types of courses for which students received credit.

The Expanded Sample

As part of their normal assessment, HSTW sites test either all their vocational completers (mostly seniors) or a random sample of these completers. For this study, schools were also asked to draw a representative sample from their entire senior class. ETS guidelines require that a minimum of 60 students be tested in a school to produce reliable results. Consequently, in many of the small schools participating in the study, all seniors were assessed. In the larger schools, however, two random samples were drawn: one of vocational completers and one to represent the entire senior class. ETS used the vocational completer sample in their analyses, and NCRVE used the whole-school sample. With input from SREB and ETS, we provided detailed sampling instructions to the schools for drawing these two samples.

Eligible students[19] were listed alphabetically and numbered consecutively. Two samples were then chosen according to random number lists we provided. Each student who was selected was classified in one of three categories: (1) the vocational sample only,

[18] See *High Schools That Work* 1999 Assessment Description and Content with Sample Questions, *HSTW* Consortium of the States, Southern Regional Education Board, Prepared by the Educational Testing Service, Princeton, New Jersey.

[19] Eligible students were those who did not have an Individual Education Plan (IEP) and normally participated in standardized achievement assessments.

(2) the all-senior sample only, or (3) both samples. School staff were then asked to assign the appropriate sample code to each student.

Site Reports

After analyzing the assessment data, individualized site reports were prepared to summarize each school's whole-school test results. A reproduction of one of the site reports from this study, stripped of school names to protect confidentiality, is included as Appendix B. The original versions were in color and relied mostly on graphs, accompanied by text highlights, to summarize each school's results to make the site reports as accessible as possible to a wide range of interested parties.

The intent of the site reports was to spark discussion, reflection, and further investigation about student outcomes and the practices implemented to achieve them. Consequently, a selective, rather than comprehensive approach was taken, and a small number of measures was highlighted. In general, graphs focused on subgroup comparisons and quartile depictions of test scores—methods of analysis that school staff were unlikely to have performed on their own.

To place vocational completer results in the larger context of the entire school's performance, data were subgrouped by students' curriculum concentration or "program": college preparatory only, vocational completers (including those who also completed the college preparatory curriculum), and students completing neither curriculum. Sometimes, the vocational completer group was separated into vocational completers only and those who also completed a college preparatory curriculum. We used the HSTW recommended curriculum as an approximate definition of a college preparatory program, although admission requirements for many bachelor-degree programs include additional courses in math, science, or foreign language.

It is important to emphasize again that the site report in Appendix B cannot be construed as an evaluation of the effectiveness of "School W," any of the other schools, or the HSTW network. An actual evaluation would require, among other things, observation of schools at more than one point in time, and the correlation of student performance results with school practices. Information about results achieved by HSTW, and a description of practices in most-improved schools, are available from the HSTW brochure on the Web.

A statistical analysis of associations between changes in HSTW school practices and student performance trends over time has been conducted by Kaufman, Bradby, and Teitelbaum (2000).

Workshops

To help schools better understand their data, we conducted workshops for school representatives in January and July of 1999. During these workshops, educators were encouraged to participate in a cross-school dialogue about data results. In the January workshop, school-by-school results were discussed with the educators who attended, in an effort to encourage staff to share their thoughts about trends in performance. Results for vocational completers were placed in context by comparing them with results for other groups. Presenters also emphasized analyzing the distribution of test scores, rather than simply average levels of performance. Participants were encouraged to discuss the practical significance of perceived differences and to think critically about the usefulness of data results and what they indicated about student learning. In a discussion about the performance of vocational completers, participants examined school input, practice, and outcome indicators to help them consider the relationships among different indicators. Participants were also encouraged to keep in mind important demographic differences among the schools when interpreting cross-school test score results.

The July workshop provided an opportunity for school staff to better understand their site reports; gain perspective on their school's data by examining other schools' results; and learn how to interpret data about the study's overall findings, including the use of regression analysis. The site reports were used to help teachers and administrators focus on several basic questions about data analysis—for example, "Who is included in the statistic?" "What do the data say in concrete terms?" and "How well do you seem to be performing?" Because these site report data were standardized across schools, the mechanics of running a cross-school data workshop were greatly simplified.

Site Visits

During the spring of 1999, we conducted six site visits to determine how the site reports were actually being used and to obtain some descriptive information about the efforts schools were undertaking to improve student achievement. To maintain confidentiality, this report does not provide site-specific information. The visited schools had the following characteristics:

- A small town school in the South with about one-quarter of the student population made up of minorities
- A suburban school in the West with a predominantly white student population
- An urban school in the West with a predominantly minority population
- A rural school in the Midwest with mostly white students
- A small town school in the Midwest with about one-third of the student population made up of minorities
- A rural school in the South with a mostly white population

Results

Participation in the Expanded Assessment

After we received the data file from ETS, it became apparent that the sample coding had not always been done properly. For this reason, and because no codes were assigned to students in samples consisting of an entire senior class, we verified with each school the sample codes assigned to participating students. The data were then recoded and cleaned to create usable analytic files.

Three schools were eventually excluded from the study: (1) one school did not receive the testing materials in time and did not conduct the assessment; (2) one school's data was not scored by ETS, the test publisher; and (3) another was excluded because of invalid sampling.[20] The remaining 27 schools represented 20 different U.S. states. Eight of the schools chose to assess all seniors. The remaining 19 schools chose random samples of their seniors for testing.

[20] In this last case, the sampling was done in a way that could not be fixed after the fact. The testing coordinator in this school randomly chose nonvocational students for the study sample, rather than sampling from the entire senior class.

ERIC

Conversations with a few of the testing coordinators revealed some troubling anecdotes. In one school, the testing coordinator claimed that a teacher of one of the more academically advanced students tried to persuade the student to skip the assessment because it was for vocational students, not for students "like him." Another testing coordinator reported that nonvocational students did not take the assessment seriously and that she could not vouch for the results. These types of comments were rare, however. By and large, testing coordinators reported that students scheduled to take the assessment showed up and did their best.

Site Reports

The site reports produced for this study generated much enthusiasm in the schools, and staff reported discussing the results with various stakeholders. Many school representatives said they thought the site reports were attractively presented and useful. Schools used them in presentations to faculty, school boards, and community members. Staff particularly liked the cross-school quartile graphs (Appendix B, Figures 5a-5c) because these placed an individual school's performance in a broader context and showed the full distribution of test scores, with NAEP proficiency levels as benchmarks.

Our analytical framework calls for schools to analyze *change* in student performance, but we have only 1998 baseline data available for the schoolwide sample at this point. Nevertheless, some results in the site report illustrate how the schoolwide sample permits comparisons among groups of students that might prompt further inquiry in the schools. For example, School W's results show that mean scores for college preparatory students were typically the highest, with vocational mean scores generally quite close to those of the "neither" group (Appendix B, Figure 1); however, further subgrouping of the data revealed that vocational completers who also completed a rigorous academic curriculum performed better than their vocational peers who did not take such a challenging curriculum. In school W's case, vocational completers who took the higher-level academic curriculum scored virtually the same as college preparatory students (Appendix B, Figure 2). To further emphasize the fact that those who took more rigorous academic coursework generally scored higher on standardized achievement tests, Figure 4 in Appendix B depicts test score differences between students who did and did not meet HSTW's recommended curriculum goals. This kind of comparison could reinforce School W's efforts to enroll more of its vocational students in the HSTW recommended curriculum.

Following this line of thinking, the site report in Appendix B also illustrates how School W may learn from cross-site comparisons. In particular, staff at School W may wonder about School B's impressive results as depicted in Appendix B, Figure 5a, which shows that School B has high reading scores and relatively little variance. School W may notice that School B is somewhat similar to them in that they are both "urban fringe" schools and have predominately white student populations (as seen in Appendix B, Table 1). School W is in the Northwest and School B is in the South. Despite their impressive test scores, however, School B's parental education levels are lower than School W's (35% of School B parents have a four-year college degree versus 49% of School W parents). The biggest differences, however, are in the curricular tracks. Whereas in School W only 19% of students completed a college preparatory curriculum,[21] 92% of students in School B did so. Furthermore, in School W, one-quarter of the students completed neither a vocational concentration nor a college preparatory curriculum. School B has no students in this "neither" classification. This course-taking information may cause School W to hypothesize that course-taking patterns explain why School B's scores are both high and narrowly distributed. These insights could help spur changes in School W, although clearer inferences could be drawn if we also had data showing trends over time.

Workshops

Educators participating in the workshops appreciated the technical assistance that we provided for analyzing and interpreting the assessment data. As one participant said, "Presentation of data was done in a way to help us understand what it meant. This will help us use data and will help us in our school improvement efforts." Educators were also generally positive about having the chance to discuss and compare results with the other schools. In particular, workshop participants liked the "consultancy" process that was used to structure dialogues between schools. During this process, pairs of schools prepared five-minute presentations for each other, then took turns listening to and providing feedback on the presentations. This exercise helped staff gain perspective on their results through the viewpoints of people from another school.

---

[21] Includes the college preparatory only group—8%—and those students who were both vocational and college preparatory—11%—for a total of 19%.

Site Visits

The site visits were conducted to confirm that participating schools were implementing key HSTW practices and to learn more about these practices, to discuss with school staff how they typically used data, and to determine how the schools were using the site reports.

Through their association with HSTW, the six schools we visited had all undergone significant reform efforts. In two cases, this meant implementing career majors and senior projects for all students. Both of these efforts were studied carefully before implementation. In one school, every certified staff member was an advisor with between three to eight students; advisors help students select a senior project topic, write a research paper, coordinate a job shadow related to the project, and prepare the presentation. In another school, students often told visitors that the senior project was the one thing about which they were most proud.

Converting to a block schedule and creating career academies were the major reform efforts in another school. This school now has Health and Humanities Academies, in which students take most of their classes together with a core group of teachers. In the Health Academy, juniors participate in work-based learning opportunities in acute care, long-term care, and community health care facilities, and seniors work in paid internships. Students in this academy feel they are getting a more meaningful education than other students because they learn important skills (e.g., CPR and first aid) as well as medical terminology that is useful in the world of work. They believe they also have a better chance of getting into college to continue their training than other students.

In another school, reform efforts centered on eliminating the general track by raising graduation requirements. Students are now required to take three years of math and science instead of two, and grading is done on an A-B-C-Fail basis, which staff believe holds students accountable for high-quality work. As a result of these efforts, more students are now making the honor roll (60% have a B average or above), discipline referrals have improved, and the school has developed more of an academic atmosphere than a social one. One teacher said that teacher morale is better after eliminating the general track, stating, "setting expectations higher makes it more satisfying to teach because I feel like I am making more of a contribution to students."

Innovative programs were the focus at another large, overcrowded school in a small town that had recently been in the news because of racial tensions, fights, and gang activity. In an effort to lower the dropout rate (freshman classes were typically around 600 students; graduating classes were only 300), this school created a freshman academy to ensure that students who needed remediation in basic academics received the individualized attention they needed. Another of their efforts to keep students engaged in school was to generously support an innovative broadcasting program and the award-winning teacher who ran it.[22] This teacher created, with his students, an integrated yearbook on CD-ROM that included a virtual tour of the high school, 21 videos, and yearbook pictures. He structured class like a production studio and produced daily student video announcements, weather forecasts, and sports broadcasts; as well as prom videos, wedding rehearsals, and telethons to raise money for the program. Because of this program's innovative approach to hands-on learning, students of all academic backgrounds—including those from a college preparatory background—were attracted to broadcasting, and some graduated as vocational completers.

All of the visited schools expressed appreciation for the data provided by HSTW and the NCRVE study. Some felt that HSTW gave them better data about their students than any other school, district, or state assessment. They felt the HSTW assessment was a better reflection of what students should be able to know and do than other tests. One principal believed the test scores gave him more reliable information than grades because he felt teachers sometimes graded on factors besides academic achievement (such as attendance and attitude); however, many teachers were skeptical of standardized achievement tests and felt much better about performance assessments such as senior projects. Many respondents appreciated the student and teacher data that gave them a window into classroom practices. These results sometimes were a rude surprise, however. In some cases, negative findings were dismissed as anomalous; in others, they were taken seriously.

As is often the case when discussing student achievement, school staff raised the challenges they faced when trying to improve outcomes such as test scores. One principal was relatively pessimistic about his ability to raise test scores. He felt that systemic efforts like implementing academic standards and benchmarks would take many years to affect test scores. This principal also faced resistance from his math teachers about changing the math curriculum and teaching practices.[23] In other schools, teachers faced parent apathy. In one

[22] This teacher has an extended contract to support summer curriculum development and revision.

[23] Classroom observations in this school revealed fairly traditional drill-based instruction methods.

farming community, a parent said to a teacher in a conference with his daughter, "I didn't graduate from high school and I am doing just fine!"

Urging students to set their sights higher than their parents' was also difficult in another heavily immigrant community with low parental education levels and low expectations for students. Respondents in this school noted that the threat of bad grades on report cards was an empty one in some cases because some parents were not aware that the school issued report cards. They also noted that some parents were only semiliterate in their native languages, so even translating school notices from English into these native languages was of little use. Some of these schools faced mobility rates of more than 30%, a factor that made test score comparisons with other schools with lower mobility rates particularly frustrating.[24]

Despite these and other challenges, HSTW has given these schools a coherent reform agenda which they are seriously pursuing. Each of them now displays the beginnings of a reform history they can be proud of and improve upon. Data from the schoolwide sample in 2000 should help to clarify where they are making the most progress, and where to focus further efforts.

Lessons Learned

Schoolwide Sampling Is Feasible

Largely because most site coordinators already had experience drawing student samples for the biennial HSTW assessment, they did not have a lot of difficulty drawing the requested whole-school samples for this study; however, the need to draw two separate samples complicated matters, ultimately eliminating one of the 30 schools from the analysis because of faulty sampling, and making it necessary to verify the sample coding for the remaining schools. In order to eliminate errors and allow for analysis of response rates in the 2000 assessment, study staff drew samples from lists of eligible students provided by the schools.

[24] This rate also means that only 70% of the students were at the school for the entire school year, a fact that makes raising student achievement quite difficult.

As in the BASRC study, we found that a new data collection effort may require several iterations to work out the bugs. Although great attention may be paid to providing clear and detailed instructions the first time around, it is nearly impossible to foresee all contingencies and to predict how school staff will interpret both what is and what is not stated. Providing adequate data collection instructions requires learning on the part of both research and school staff as well as an iterative fine-tuning of instruments and procedures. Instructions must also be clear enough to allow for staff turnover from year to year.

Schoolwide Data Is Useful for Individual Schools

Initial indications are that whole-school information is useful for several reasons. First, such data makes it possible to compare the performance of subgroups within schools. Such comparisons will be more informative when they can use future assessments of schoolwide samples to compute trends for different subgroups over time. Will the performance gap between vocational completers and college-preparatory students narrow as more vocational students complete the HSTW recommended curriculum? Schoolwide sampling is necessary to answer this kind of question within an individual school.

Schoolwide data also makes it possible to compare different groups in terms of their sociodemographic characteristics, such as parental education and poverty levels, which tend to be correlated with test scores. With data for vocational completers only, it is not possible to consider sociodemographic differences in interpreting test score results.

Schoolwide data generates interest among a broad group of stakeholders. Participants in this study told us that they had discussed findings with faculty, school boards, and various community members. Reporting to all stakeholders is especially important in schools that are engaged in schoolwide program improvement. Only schoolwide data can accurately measure the results of such efforts. Again, measuring progress depends on obtaining data for multiple points in time; however, setting whole-school baseline levels of performance is a necessary first step in analyzing the effect of reform efforts on schools in the study. The fruition of this attention to whole-school data will be best assessed as future waves of data become available.

Finally, as explained in Section I, use of schoolwide data can avoid certain biases and statistical illusions that might possibly occur as a result of changes in the composition of students identified as vocational completers or other categories.

Cross-School Comparative Data and Discussions Were Well-Received

Overall, participating school staff were very positive about the different opportunities for cross-school comparisons, including parts of the site report and the two data workshops. In some high-stakes accountability situations, cross-school comparisons can produce hostile reactions from educators. In the context of this study, however, such comparative information was provided to encourage professional inquiry and a sense of community, and no specific consequences were attached to performance. The study created a safe environment for schools to contemplate and discuss cross-school information. An effort was also made to include school background information, as well as outcome data, in order to make explicit obvious differences across the schools that might affect baseline data. Consequently, participating staff appeared to avoid defensive responses to negative information and candidly qualified some positive results that may have been due to external or unintentional factors.

The enthusiasm with which this study was received by the participating schools should be regarded with some caution. It must be remembered that one reason why SREB invited these schools to participate was that they thought the schools would be interested. Also, since NCRVE paid for the expanded assessment, the participating schools did not have to allocate any of their HSTW funds to collecting the whole-school data.

Still, some of the teachers and administrators participating in this study had to spend a considerable amount of extra time and effort arranging for the schoolwide sample, taking part in workshops, and interacting with study staff on site visits. The fact that many said and wrote that the exercise was worthwhile indicates that they are interested in the kinds of comparisons possible with data from a schoolwide sample. The additional fact that 26 of 27 schools opted to repeat the schoolwide sampling procedure in 2000 is further evidence of their interest. With data from two points in time, it will be possible to learn a good deal more.

ERIC

IV. EPILOGUE: SOME THOUGHTS ON CENTRALIZED VERSUS DECENTRALIZED DATA STRATEGIES

BASRC and HSTW are both multidistrict reform networks that help schools use data to inform school improvement. In both networks, data are not used to accredit schools or punish them, and quantitative performance measures are not used as the sole basis of progress towards reform goals.

The data strategies of the two networks do differ in one important respect. BASRC takes a decentralized approach, leaving the measures up to the school to identify, compile, and analyze. HSTW, on the other hand, takes a more centralized approach, by designing student and teacher questionnaires and relying on the ETS to design a student assessment that vocational students take every other year. Because of its decentralized approach, data from individual schools in the BASRC network are more difficult to compare, whereas HSTW schools are encouraged to compare their performance to national averages, high-performing schools, and schools that are sociodemographically similar.

Although the purpose of our study was not to compare these two data strategies, in the course of our work we have formed some opinions about the pros and cons of the two approaches, and we will share them here.

BASRC's decentralized approach places less emphasis on standardized achievement tests and comparability, allowing schools the freedom to define which data elements they wish to be held accountable for in their annual Reviews of Progress. There are several advantages to this approach. School staff are allowed to determine at the school site which data elements are most useful in measuring progress toward school reform goals. This allows staff to tailor data to fit reform and, over time, could lead to a more positive attitude about collecting and analyzing data and using it to motivate reform. This approach also allows for selection of more valid measures of student achievement than standardized achievement test scores, for example.

This decentralized strategy also has some disadvantages. School staff may not have the time to collect data about some of the more valid measures they may identify. Noncentralized and nonstandardized assessments—like senior projects, for example—require a great deal of time to develop and score. In order to get schoolwide measures of achievement

ERIC Full Text Provided by ERIC

on these types of assessments, school staff need to enter individual results into a computer and then analyze these data. It is rare to find a school staff with the time and know-how to do this. Consequently, it is difficult in the short run for schools to use nontypical data to drive reform. In the long run, it is perhaps possible, given appropriate resource allocation, professional development, and consistency of purpose. In the meantime, if schools want to use data to drive reform, they must rely on typical data such as standardized achievement test scores, grade point averages, enrollment, and attendance.

HSTW's more centralized strategy gives greater emphasis to standardized achievement tests and comparability. Their effort involves a mandatory, biennial assessment of vocational completers based on NAEP-like tests, a course experience questionnaire and background survey for students taking the assessment, a teacher survey, and a graduate follow-up survey. This approach reduces the burden on schools by contracting with the ETS to centralize the data compilation and reporting functions for the more than 800 HSTW schools in the network. Assessment results are linked with course taking and student background survey data, allowing for a multitude of meaningful comparisons. The assessment, although standardized, also includes some open-response items, measures some critical thinking and problem-solving abilities, and is based on NAEP frameworks. Assessment results are compared to national averages, NAEP performance levels, and schools of similar racial/ethnic and parental education levels and are generally comparable across sites. Reporting of results is more extensive and detailed than any other state or school reform network effort we have come across, with site reports currently running at about 150 pages.

A drawback of the HSTW approach is that assessments are not custom-tailored to fit each school's particular interests—although schools are always free to collect any additional data on their own. Furthermore, as we have explained in this report, results from testing only vocational completers are of limited use in gauging the progress of whole-school reform.

The opinions expressed in this Epilogue are only conjectures. To determine whether a centralized or decentralized data strategy is more effective would require systematic data comparing the progress of the two networks over time, which we do not have. It would be highly desirable to collect such information and make this kind of comparison in order to guide the further development of strategies that use data to promote school improvement.

REFERENCES

Bay Area School Reform Collaborative (BASRC). (1999). *Annual report, Fall 1999.* San Francisco: Author.

Blank, R. K., & Gruebel, D. (1995). *State indicators of science and mathematics education 1995: State-by-state trends and new indicators from the 1993-94 school year.* Washington, DC: Council of Chief State School Officers.

Blank, R. K., & Gruebel, D. (1993). *State indicators of science and mathematics education 1993: State-by-state trends and new indicators from the 1991-92 school year.* Washington, DC: Council of Chief State School Officers.

Ebmeier, H. H., & Ziomek, R. L. (1983). *Student academic engagement rates.* Berkeley: University of California, Berkeley, Education Library. (ERIC Document Reproduction Service No. ED 241 599)

Education Week. (1999). *Quality counts '99: Rewarding results, punishing failure.* Bethesda, MD: Editorial Projects in Education, Inc.

Hoachlander, E. G., Levesque, K. A., & Rahn, M. L. (1992). *State systems for accountability in vocational education* (MDS-491). Berkeley: National Center for Research in Vocational Education, University of California, Berkeley.

HSTW Consortium of the States, Southern Regional Education Board (SREB). (1999). *High Schools That Work 1999 assessment description and content with sample questions.* Princeton, NJ: Educational Testing Service.

Kaufman, P., Bradby, D., & Teitelbaum, P. (2000). *High Schools That Work and whole school reform: Raising academic achievement of vocational completers through reform of school practice* (MDS-1295). Berkeley: National Center for Research in Vocational Education, University of California, Berkeley.

Levesque, K. A. (in press). *Vocational education in the United States: Toward the year 2000.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Levesque, K. A., Bradby, D., Rossi, K., & Teitelbaum, P. (1998). *At your fingertips: Using everyday data to improve schools* (MDS-1018). Berkeley: National Center for Research in Vocational Education, University of California, Berkeley.

Madigan, T. (1997). *Science proficiency and course taking in high school: The relationship of science course-taking patterns to increases in science between 8th and 12th grades.* Suitland, MD: Bureau of the Census, U.S. Department of Commerce.

Maryland State Higher Education Commission. (1996). *Relationship between high school and college performance by Maryland students: Student outcome and achievement report.* Berkeley: University of California, Berkeley, Education Library. (ERIC Document Reproduction Service No. ED 404 976)

McLaughlin, M. W., & Talbert, J. E. (1999). *Assessing results: The Bay Area school reform collaborative, year three.* San Francisco: Bay Area School Reform Collaborative.

McLaughlin, M. W., Talbert, J. E., & Crowe, B. (1998). *Assessing results: The Bay Area School Reform Collaborative, year two.* Stanford, CA: Center for Research on the Context of Teaching, Stanford University.

Mullis, I. V. S., Jenkins, F., & Johnson, E. G. (1994). *Effective schools in mathematics: Perspectives from the NAEP 1992 assessment.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Murnane, R. J., & Pauly, E. W. (1998, March). Educational and economic indicators. *Phi Delta Kappan, 69*(7), 509-513.

Murnane, R. J., & Raizen, S. A. (Eds.). (1988). *Improving indicators of the quality of science and mathematics education in grades K-12.* Washington, DC: National Academy Press.

ERIC

Selden, R. W. (1994). How indicators have been used in the U.S.A. In K. Riley & D. Nuttall (Eds.), *Measuring quality: Education indicators: UK and international perspectives.* London/Washington, DC: Falmer Press.

Smith, M. S. (1998, March). Educational indicators. *Phi Delta Kappan, 69*(7), 487-491.

Stecher, B. M., Hanser, L. M., Hallmark, B., Rahn, M. L., Levesque, K., Hoachlander, E. G., Emanuel, D., & Klein, S. G. (1995). *Improving Perkins II performance measures and standards: Lessons learned from early implementers in four states* (MDS-732). Berkeley: National Center for Research in Vocational Education, University of California, Berkeley.

U.S. Department of Education, Office of Educational Research and Improvement. (1988, September). *Creating responsible and responsive accountability systems*. Washington, DC: U.S. Government Printing Office.

APPENDIX A

1997-1998 School Data Report Created for Collecting Data from BASRC Schools

ERIC

1997-98 SCHOOL DATA REPORT

Please provide the following information about your school for the 1997-98 school year. **Please return to MPR Associates by May 21, 1999.** Fax to (510) 849-0794. Mail to 2150 Shattuck Ave, Ste. 800, Berkeley, CA 94704. If you have any questions or need assistance, please contact Doug Lauen (510) 849-4942 or Matt Byrnes (510) 642-2050.

School Name:	**Key Definitions**
______________________	1. **Mean** – The average of a set of scores or other statistics, obtained by adding the scores together and dividing by the number of scores.
Name and title of person filling out form:	2. **Minimum and Maximum** – The lowest and highest scores in a set of scores
______________________	3. **25th percentile (25th %)** – The score below which 25 percent of scores in a set of data fall.
______________________	4. **Median or 50th percentile** – The score below which 50 percent of scores in a set of data fall.
	5. **75th percentile (75th %)** – The score below which 75 percent of scores in a set of data fall.

I. Enrollment and Student Demographics

1. Total number of students enrolled in:

9th Grade	10th Grade	11th Grade	12th Grade	Entire School

Source: CBEDS report; www.cde.ca.gov/demographics/reports/#senrol

2. Percent of students who are:

Asian	Hispanic	Black	White	Other	LEP	On AFDC

Source: high school profile at www.ed-data.k12.ca.us

II. Student Achievement

1. School proficiency test information—please provide all relevant information regarding your school's proficiency test(s) in 1997-98, including at least the following (if you only score your test pass/fail, you will only be able to fill out the first 6 columns):

Name of Test	Test Date	Grade Level*	Number of Students Taking Test	Number Passing	Percent Passing	Mean Score	Min Score	25th % Score	Median Score	75th % Score	Max Score

* If more than one grade level of students took the same proficiency exam, please use a separate row to record the results for each grade level.

2. Grade point averages. If possible, please do *not* report cumulative GPA, but rather the grade point averages of only those courses taken during the 1997-98 school year. Please report GPA on a 4-point scale, without counting an extra point for honors and/or AP classes. If there is no way for you to calculate GPA in the way we request, please provide cumulative GPA and note on this form that you are providing cumulative instead of yearly data.

	Number of Students included in calculation*	Mean GPA	Minimum GPA	25th Percentile GPA	Median GPA	75th Percentile GPA	Maximum GPA
All Students							
9th Grade							
10th Grade							
11th Grade							
12th Grade							
Asian							
Hispanic							
Black							
White							
Other							

* Although you may have reported similar numbers for Total Enrollment, the numbers reported here may be different because CBEDS data are reported in the fall semester, and this GPA data should be for the entire 97-98 school year.

4. Number and percent of 1997-98 seniors meeting each of UC's a-f requirements (refer to the attachment on p.6 if you are unfamiliar with the a-f requirements).

	Number	Percent
Requirement a—History/Soc. Science		
Requirement b—English		
Requirement c—Mathematics		
Requirement d—Lab. Science		
Requirement e—Foreign Language		
Requirement f—College Prep. Electives		
Met all requirements		

5. Advanced Placement (AP) test information from 1997-98 school year—for each test given by your school, please list the following information:

Name of AP test	Number of students taking the test	Numbers of students who received a					% of test takers who scored a 3 or above*	% of 11th and 12th graders who scored a 3 or above**
		1	2	3	4	5		

* To compute this indicator, sum the numbers of students receiving a 3 or higher and divide by the total number of students who took that test.

** To compute this indicator, sum the numbers of students receiving a 3 or higher and divide by the sum of the 11th and 12th grade enrollment figures you reported in I.1.

If your director of guidance cannot help you with AP information, you can go to www.cde.ca.gov/ ftpbranch/retdiv/epic/. You will need to locate your county, district and school. This will only give you an aggregate for all tests.

6. Scholastic Aptitude Test (SAT) information from 1997-98 school year

	SAT I Math	SAT I Verbal	SAT I Composite
Number of seniors taking test			
Percent of seniors taking test			
Mean Score			
Minimum Score			
25th Percentile			
Median			
75th percentile			
Maximum Score			

Source: guidance office, student record system (e.g., SASI or OSIRIS), or, as a last resort, your high school profile at www.ed-data.k12.ca.us (this will only provide mean scores and percent of seniors tested).

7. If there are any other student achievement measures that you would like to track for your own purposes (for example, writing samples, portfolios, assessments of standards, postsecondary plans or follow-up data, etc.), please feel free to attach information summarizing those data for 1997-98.

III. Student Engagement

1. Average Daily Attendance for 1997-98. For this portion, please report the ADA that you reported to the state on April 15, 1998. Then, record the portion of that number that is excused absences on the second row. Subtract the second row from the first row to get your ADA not including excused absences, and report these results on the last row.

	1997-98 school year	*1998-99 school year (if available)*
ADA as reported to the state on April 15		
Portion of ADA that is excused absences		
ADA, not including excused absences		

2. Attendance rate. Please do the following to calculate your school's attendance rates for the **1997-98 school year:**

(a) For each school day of the school year, find the number of students in each grade level who attended school, for all or part of the day. Take the sum of these numbers.
(b) For each school day of the school year, find the number of students in each grade level who were enrolled on that day. Take the sum of these numbers.
(c) Divide (a) by (b) to get your *total attendance rate.*

	(a) Total student attendance days	(b) Total student enrollment days	(c) Total Attendance Rate (a÷b)
1997-98 school year			
9th Grade			
10th Grade			
11th Grade			
12th Grade			
All Students			

4. Transfer rate—number and percent of students in each grade who transferred from your school to another school ("out"), or from another school to your school ("in"), during the 1997-98 school year.

	Number Out	Percent*	Number In	Percent*
9th Grade				
10th Grade				
11th Grade				
12th Grade				
All Grades				

*Calculate these percents by using the enrollment you reported in I as the denominator.

5. Dropout rate—number and percent of students in each grade who dropped out of school during the 1997-98 school year.

	Number	Percent
All backgrounds		
Asian		
Hispanic		
Black		
White		
Other		

Source: your high school profile at *www.ed-data.k12.ca.us*

6. If there are any other student engagement measures that you would like to track for your own purposes (for example, results of student surveys, suspension rates, etc.), please feel free to attach information summarizing these data for 1997-98.

DESCRIPTION OF UC'S A-F REQUIREMENTS

a. History/Social Science—Two years required. Two years of history/social science, including one year of U.S. history or one-half year of U.S. history and one-half year of civics or American government; and one year of world history, cultures, and geography.

b. English—Four years required. Four years of college preparatory English that include frequent and regular writing, and reading of classic and modern literature. Not more than two semesters of 9th-grade English can be used to meet this requirement.

c. Mathematics—Three years required, four recommended. Three years, including elementary and advanced algebra and two- and three-dimensional geometry. Math courses taken in the 7th and 8th grades may be used to fulfill part of this requirement if your high school accepts them as equivalent to its own courses.

d. Laboratory Science—Two years required, three recommended. Two years of a laboratory science providing fundamental knowledge in at least two of these three areas: biology, chemistry, and physics. Laboratory courses in earth/space sciences are acceptable if they have as prerequisites or provide basic knowledge in biology, chemistry, or physics. Not more than one year of 9th-grade laboratory science can be used to meet this requirement.

e. Language Other than English—Two years required, three recommended. Two years of the same language other than English. Courses should emphasize speaking and understanding and include instruction in grammar, vocabulary, reading, and composition.

f. College Preparatory Electives—Two years required. Two units (four semesters) in addition to those required in "a-e" above, chosen from at least two of the following areas: history, English, advanced mathematics, laboratory science, language other than English (a third year in the language used for the "e" requirement, or two years of another language), social science, and visual and performing arts.

ERIC

APPENDIX B

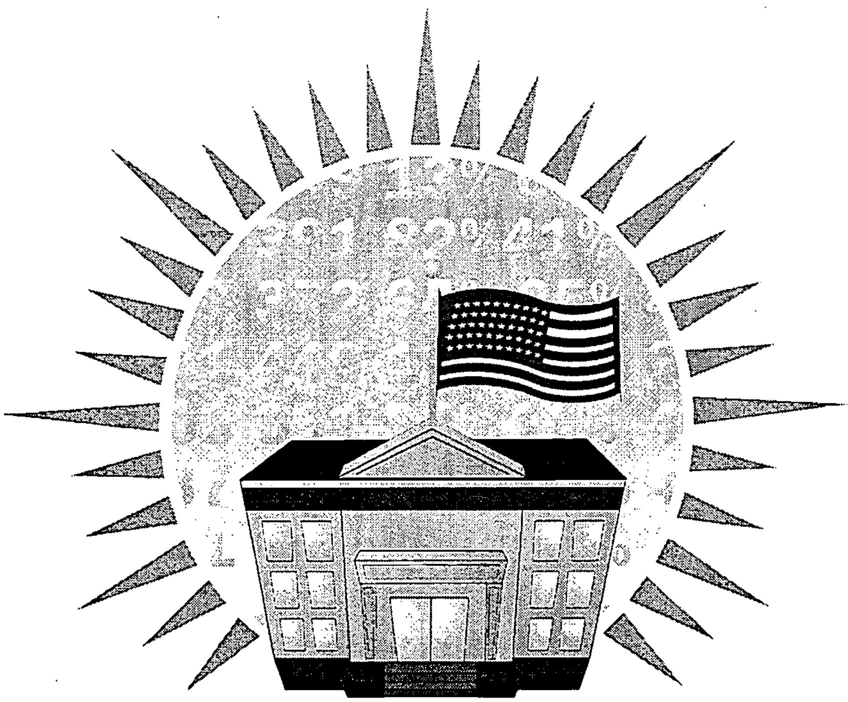
Sample of Site Report Produced for Each Participating HSTW School

Note: Information in the following site report cannot be construed as an evaluation of the effectiveness of "School W," any of the other schools, or the HSTW network. An actual evaluation would require, among other things, observation of schools at more than one point in time, and the correlation of student performance results with school practices. Information about results achieved by HSTW, and a description of practices in most-improved schools, are available from the HSTW brochure on the Web at <www.sreb.org/Programs/hstw/about/Brochure/brochure99.html>.

For a statistical analysis of associations between changes in HSTW school practices and student performance trends over time, see Kaufman, Bradby, and Teitelbaum (2000).

ERIC
Full Text Provided by ERIC

APPENDIX B



USING WHOLE-SCHOOL DATA FOR SCHOOL IMPROVEMENT

. . . .

HIGHLIGHTS FROM THE
1998 WHOLE-SCHOOL ASSESSMENT OF SENIORS

HIGH SCHOOL "W"

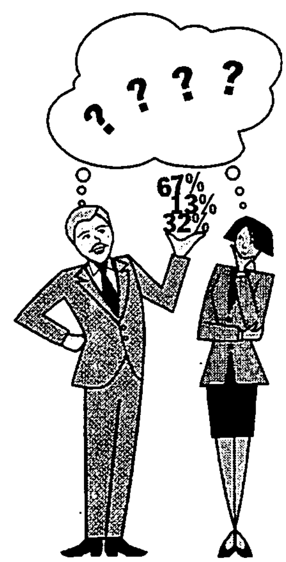
Produced by
The National Center for Research in Vocational Education
and
MPR Associates, Inc.

In Cooperation with
The Southern Regional Education Board
and
Educational Testing Service

Supported by
The Office of Vocational and Adult Education
U.S. Department of Education

Using Whole-School Data for School Improvement

This report outlines some key findings from last year's whole-school assessment of seniors. This profile provides information on an expanded sample of students participating in *HSTW's* assessment of academic achievement. Specifically, the report consists of "whole-school" data—information about either your entire 1998 senior class or a representative sample of that senior class.



This report includes results from your school and 26 other participating *HSTW* sites. Like your school, the other schools included here are committed to whole-school reform. Information about these schools is included to assist you in interpreting the figures and tables.

Our hope is that these results will spark discussion, reflection, and further investigation about student outcomes and the practices implemented to achieve them. We look forward to participating in these discussions in the next few months and encourage you and others from your school to attend SREB's 13th Annual Staff Development Conference in Atlanta on July 7–10 to examine data and share best practices.

Many figures include the following benchmarks to help you interpret the performance of your 1998 senior class:

		NAEP Proficiency Levels		
	HSTW Goal	Basic	Proficient	Advanced
Reading	279	269	304	348
Mathematics	295	250	300	350
Science	295	250	300	350

A number of figures refer to *HSTW's* curriculum goals. Please keep these goals in mind as you are interpreting your school's results:

HSTW's Recommended Curriculum

- 4 years of college preparatory English;
- 3 years of mathematics with at least 2 years of Algebra or higher math; and
- 3 years of science with at least 2 years of Biology, Chemistry, or Physics.

A "college prep student" is one who completes all three components of *HSTW's* recommended curriculum.

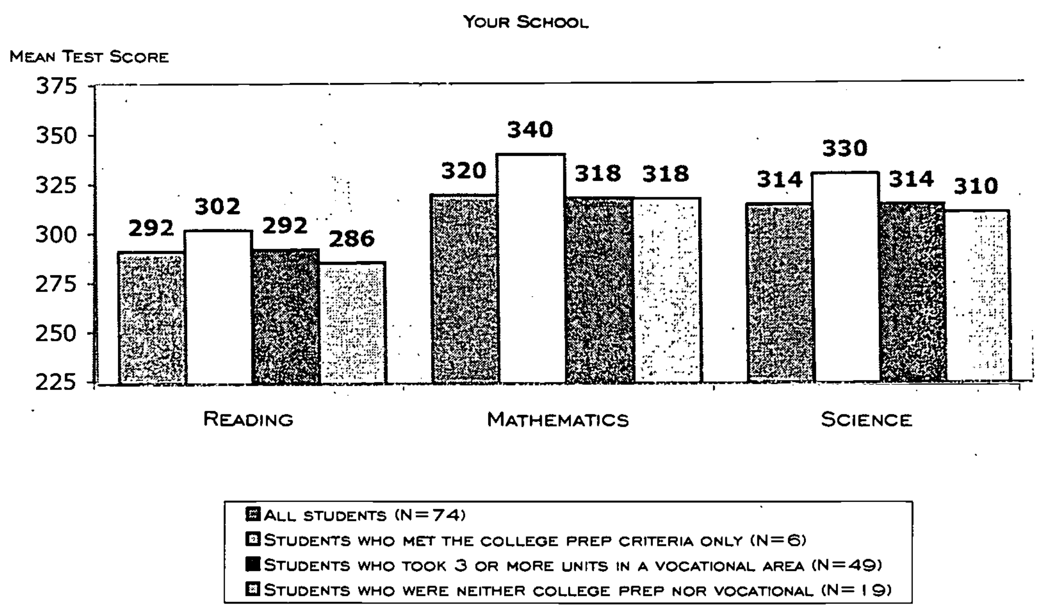
Unless otherwise noted, a "vocational completer" is a student who completed 3 or more Carnegie units (credits) in a single vocational program area.

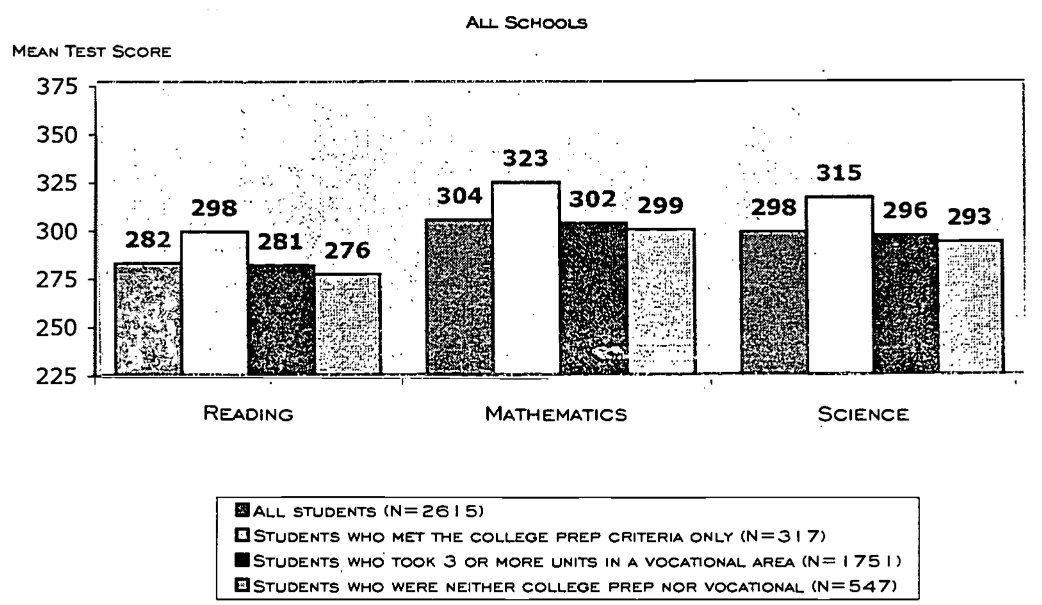
If you have any questions about this profile or would like to have additional analyses run, please contact Doug Lauen of MPR Associates at dlauen@mprinc.com. If you do not have access to email, please contact him by fax (510) 849-0794 or phone (510) 849-4942.

For simplicity's sake and methodological reasons, this report does not include statistical testing. If you would like to know whether any of the differences reported here are statistically significant, please contact Doug Lauen.

ERIC

Figure 1: 1998 Mean Test Scores by High School Program





See explanatory notes and highlights on next page.

78

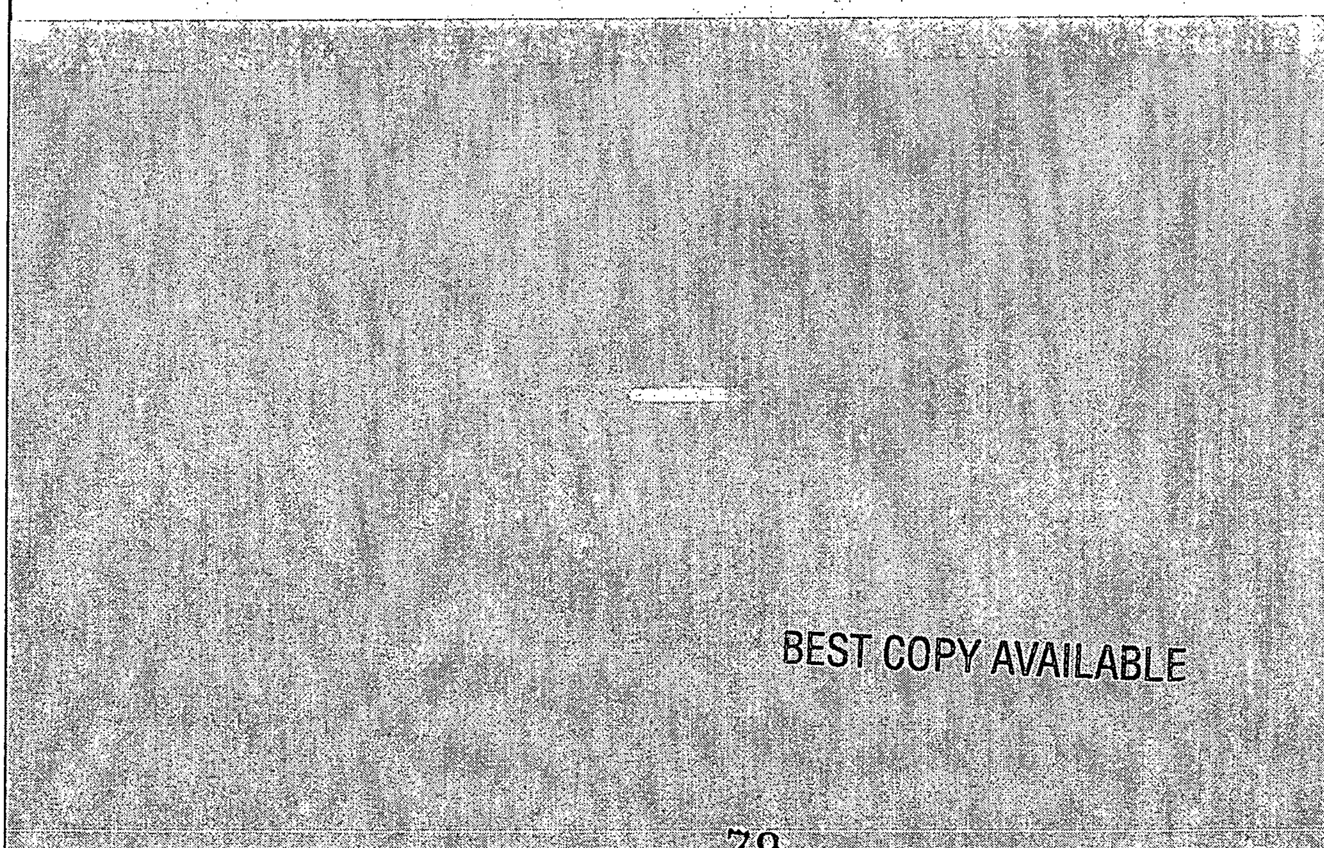
BEST COPY AVAILABLE

1998 Mean Test Scores by High School Program—Continued

This figure groups all students who participated in the 1998 assessment into categories based on the types of coursework they completed during high school. Students who completed 3 or more credits in a single vocational program area were categorized as vocational. Students who completed the *HSTW*-recommended coursework in all three academic areas (English, mathematics, and science) were classified as college prep. Students who met neither the vocational completion nor college prep criteria were classified as neither. Students who met *HSTW's* criteria for both vocational completion and college prep coursework were included in the vocational group. The N in the figure is the number of students in each curriculum category in your school. NOTE: Because of differences in scaling, direct comparisons should not be made across subject areas (reading, mathematics, and science). For example, a score of 300 in reading does not necessarily equal a score of 300 in mathematics.

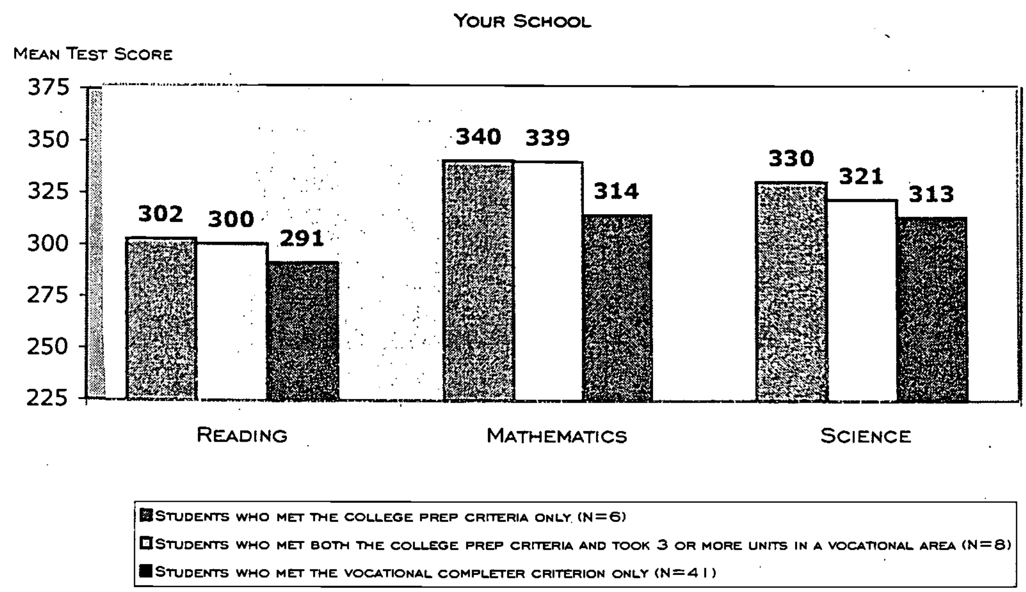
Note: Due to the small sample size of college prep students, care should be taken when interpreting these data.

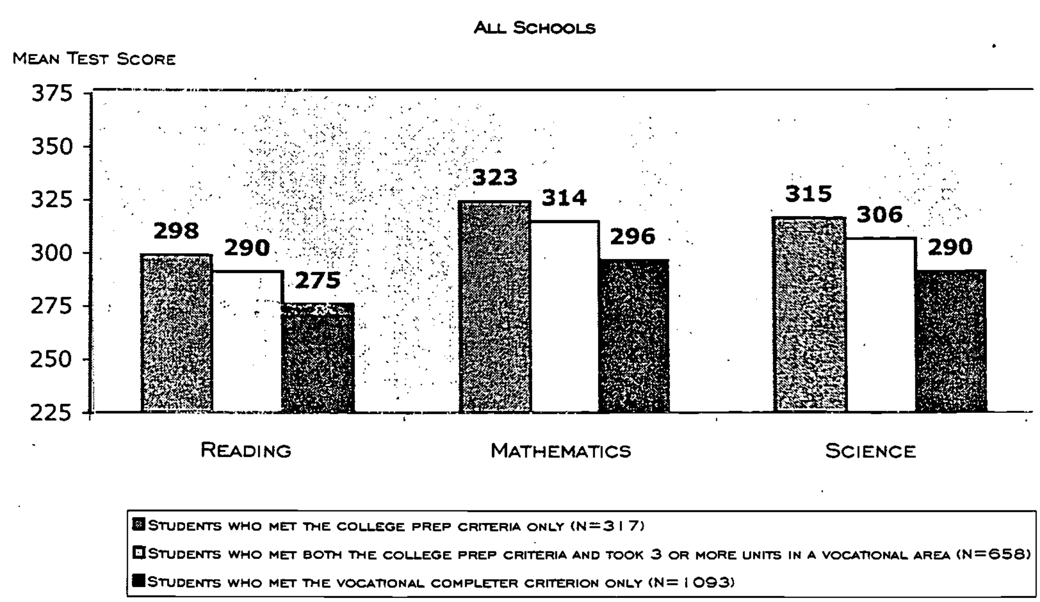
- A total of 74 High School "W" students participated in the *HSTW* assessment's whole-school sample in Spring 1998. Of these students, 49 (or 66%) were vocational completers. Six (8%) students met the college prep criteria and 19 (26%) met neither the college prep only nor vocational completer criteria.
- Vocational completers scored lower than did the college preparatory group on all three reading, mathematics, and science tests. Vocational completers scored at about the same level as students who were neither vocational nor college prep completers.
- The lower figure, which represents all students in all 27 schools in the study, shows that vocational mean scores are typically below the college prep mean scores and are very close to the mean scores for the "neither" category.



BEST COPY AVAILABLE

Figure 2: 1998 Mean Test Scores for College Prep Students and Different Types of Vocational Completers





See explanatory notes and highlights on next page.

80

1998 Mean Test Scores for Different Types of Vocational Completers—Continued

This figure compares the mean test scores for vocational completers who also completed a college prep curriculum with those who did not complete a college prep curriculum. The figure also includes the mean for students who only met the college prep criteria. The N in the figure is the number of students in each curriculum category in your school. NOTE: Because of differences in scaling, direct comparisons should not be made across subject areas (reading, mathematics, and science). For example, a score of 300 in reading does not necessarily equal a score of 300 in mathematics.

Note: *Due to the small sample sizes of college prep students and of vocational completers who also met the college prep criteria, care should be taken when interpreting these data.*

- Of the 49 vocational completers in School "W" who participated in the assessment, only 8 (or 16%) also completed a college prep curriculum. In reading and mathematics, these students scored higher than other vocational students and at almost exactly the same level as the college prep students. In science, these students scored about halfway between college prep students and other vocational students.
- The lower figure, which depicts results from all students in all 27 schools in the study, also shows that vocational completers who also met the college prep criteria scored higher than vocational completers who did not meet these criteria. These findings reinforce the *HSTW* belief that encouraging vocational students to complete challenging academic coursework in high school will significantly raise their academic achievement levels.

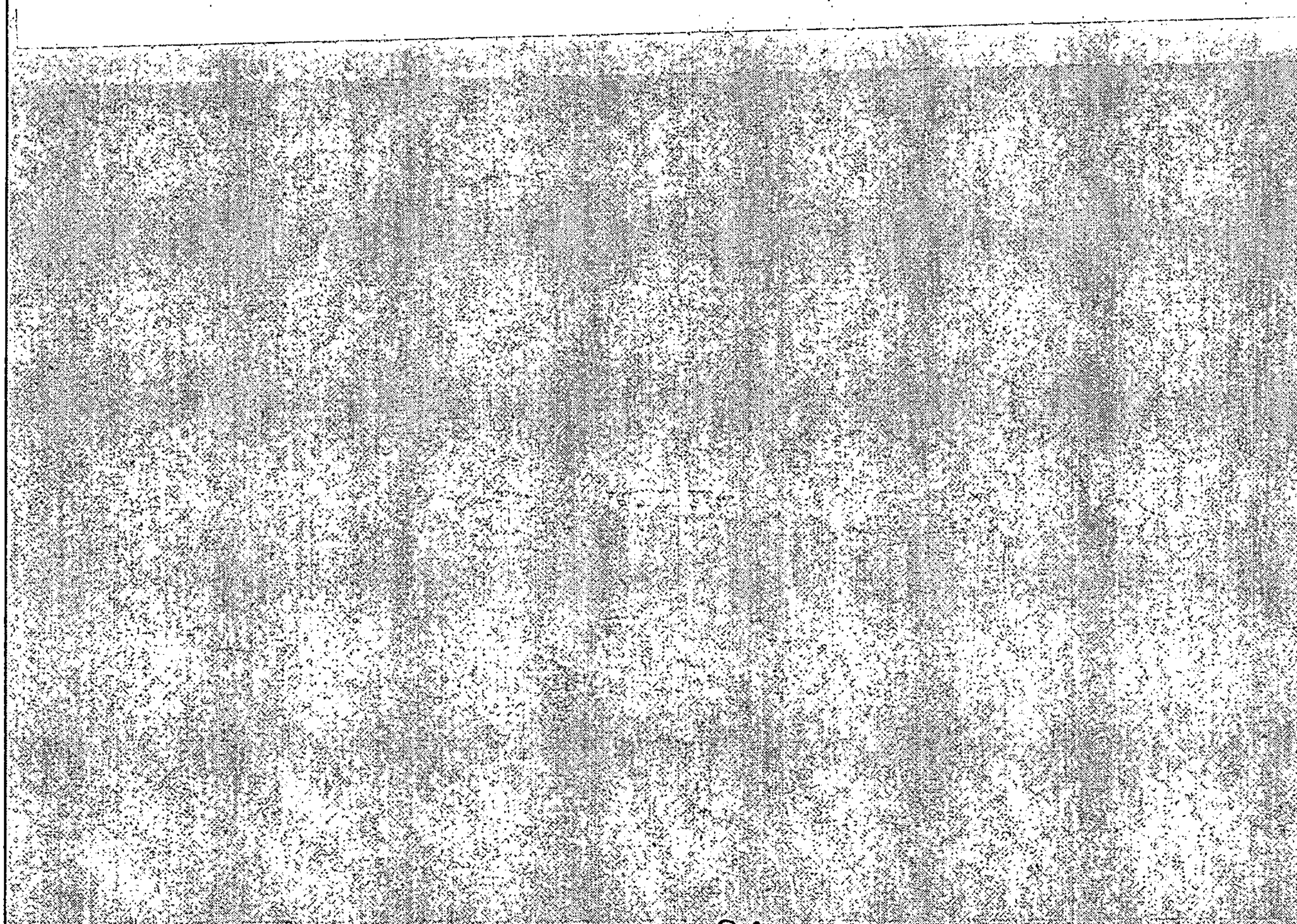
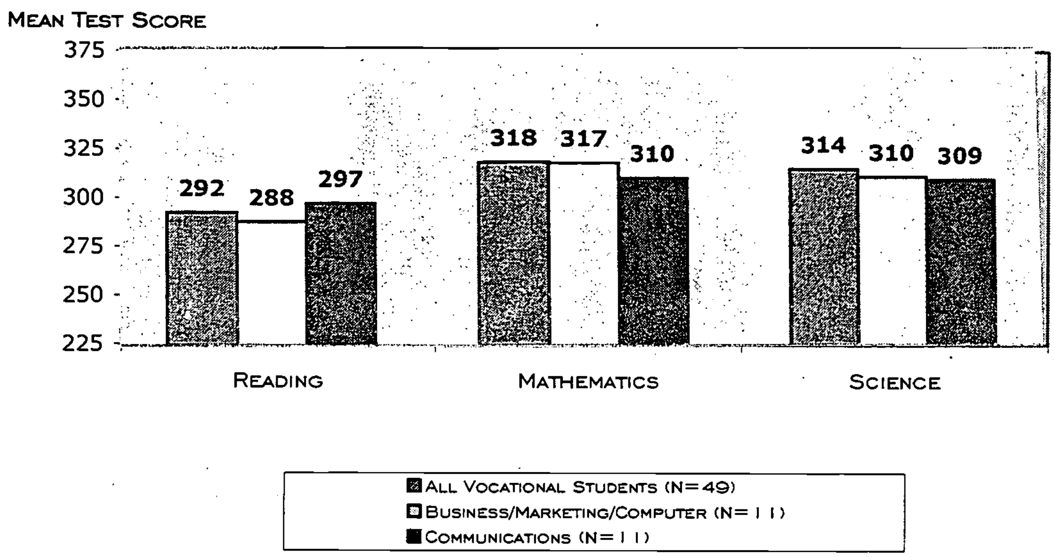


FIGURE 3: 1998 MEAN TEST SCORES FOR SELECTED VOCATIONAL AREAS THAT HAVE 10 STUDENTS OR MORE

(YOUR SCHOOL ONLY)

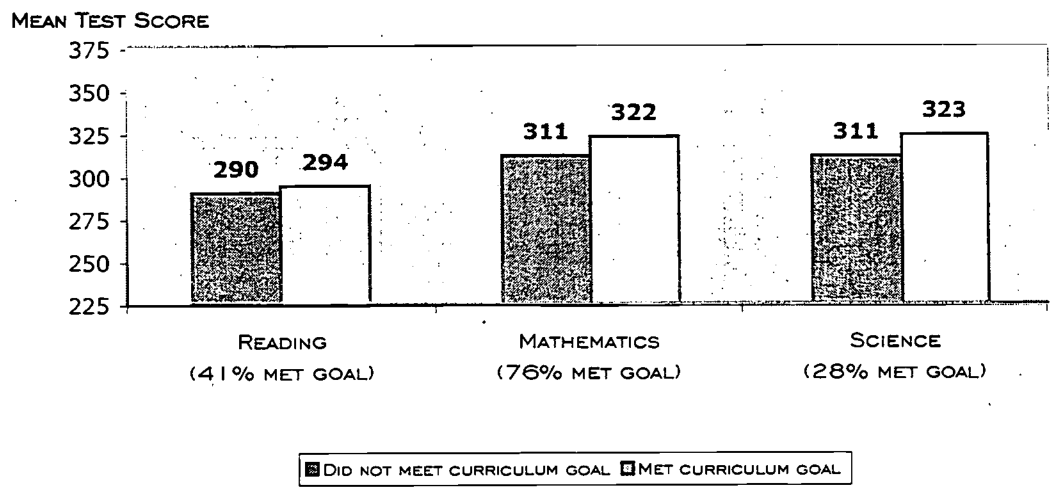


This figure compares the mean test scores of students in vocational areas with 10 or more assessed seniors. For comparison purposes, the mean for all vocational completers is also included. Vocational completers who also completed a college prep curriculum are included in the figure. The N is the number of students who concentrated in the vocational program area indicated. NOTE: Because of differences in scaling, direct comparisons should not be made across subject areas (reading, mathematics, and science). For example, a score of 300 in reading does not necessarily equal a score of 300 in mathematics.

- As seen in Figure 1, 49 School "W" vocational completers participated in the *HSTW* assessment's whole-school sample in Spring 1998. Of these, 11 (22%) reported concentrating in the Business/Marketing/Computer areas, and 11 (22%) in Communications.
- There was little difference in average science test scores for completers in the two indicated vocational areas. However, vocational completers in the Communications area, perhaps because of the emphasis placed on language arts, outperformed Business/Marketing/Computer completers. Unfortunately, Communications completers performed less well on the mathematics test.

82

FIGURE 4: 1998 MEAN TEST SCORES BY ACADEMIC CURRICULUM GOAL ATTAINMENT
(YOUR SCHOOL ONLY)

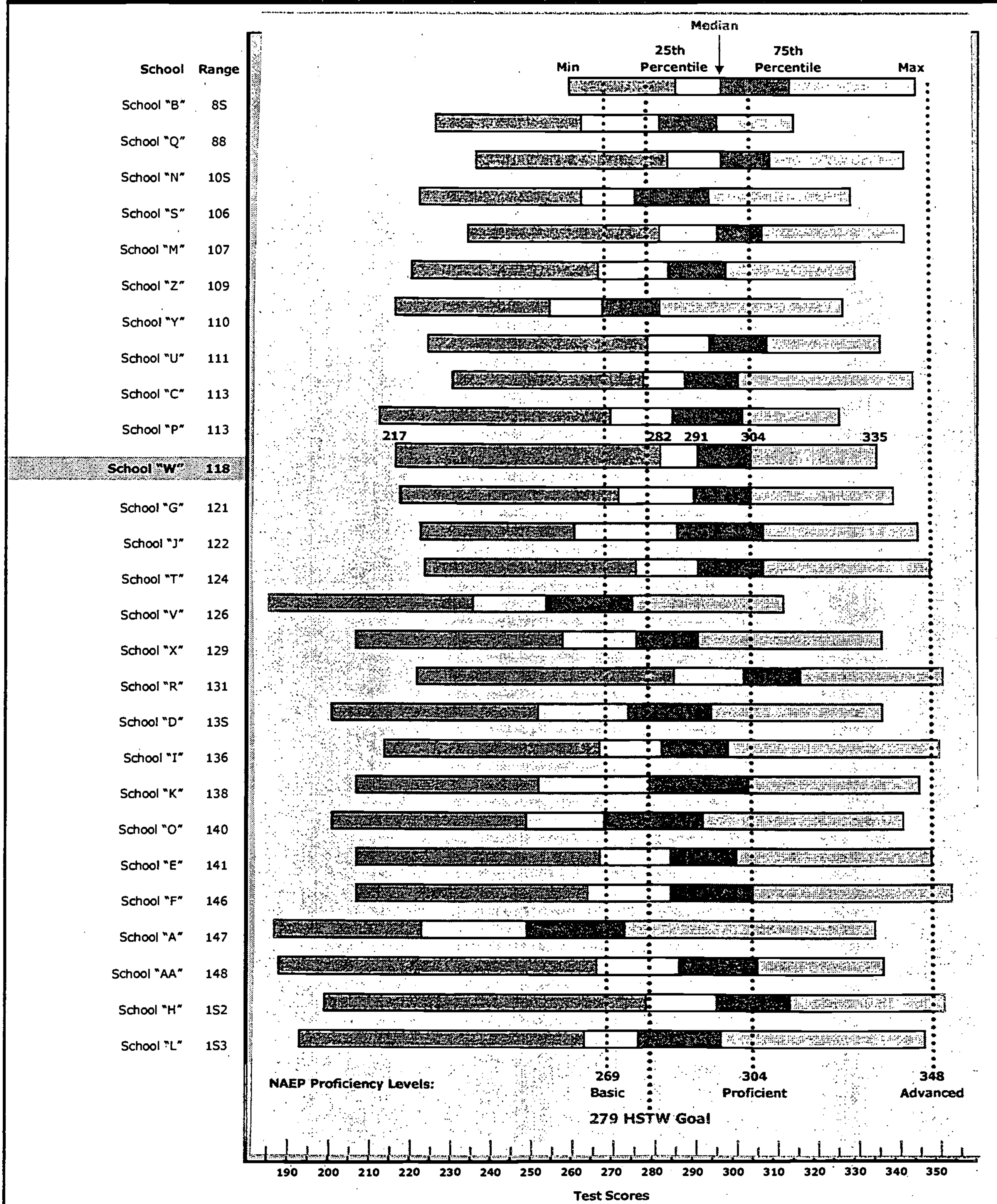


This figure compares the mean test scores for all assessed students who completed *HSTW's* recommended academic curriculum with the scores of those who did not complete this curriculum. These comparisons are made regardless of students' vocational completion status. The percentages below the subject area labels are the proportions of students who met the curriculum goals in that particular subject. For example, a 50% below the "Reading" label would mean that 50% of the assessed students met *HSTW's* goal of 4 years of college preparatory English. NOTE: Because of differences in scaling, direct comparisons should not be made across subject areas (reading, mathematics, and science). For example, a score of 300 in reading does not necessarily equal a score of 300 in mathematics.

☛ Among School "W" students who took the *HSTW* assessment in Spring 1998, 41% completed *HSTW's* recommended curriculum in English. The corresponding figures for mathematics and science are 76% and 28%. In mathematics and science, students who completed *HSTW's* challenging academic criteria outperformed their peers in all three subject areas by more than 10 points. In reading, the performance difference was more modest, at 4 points.

BEST COPY AVAILABLE

Figure 5a: 1998 Distribution of Reading Test Score Quartiles for All Schools in NCRVE Study



See explanatory notes and highlights on next page.

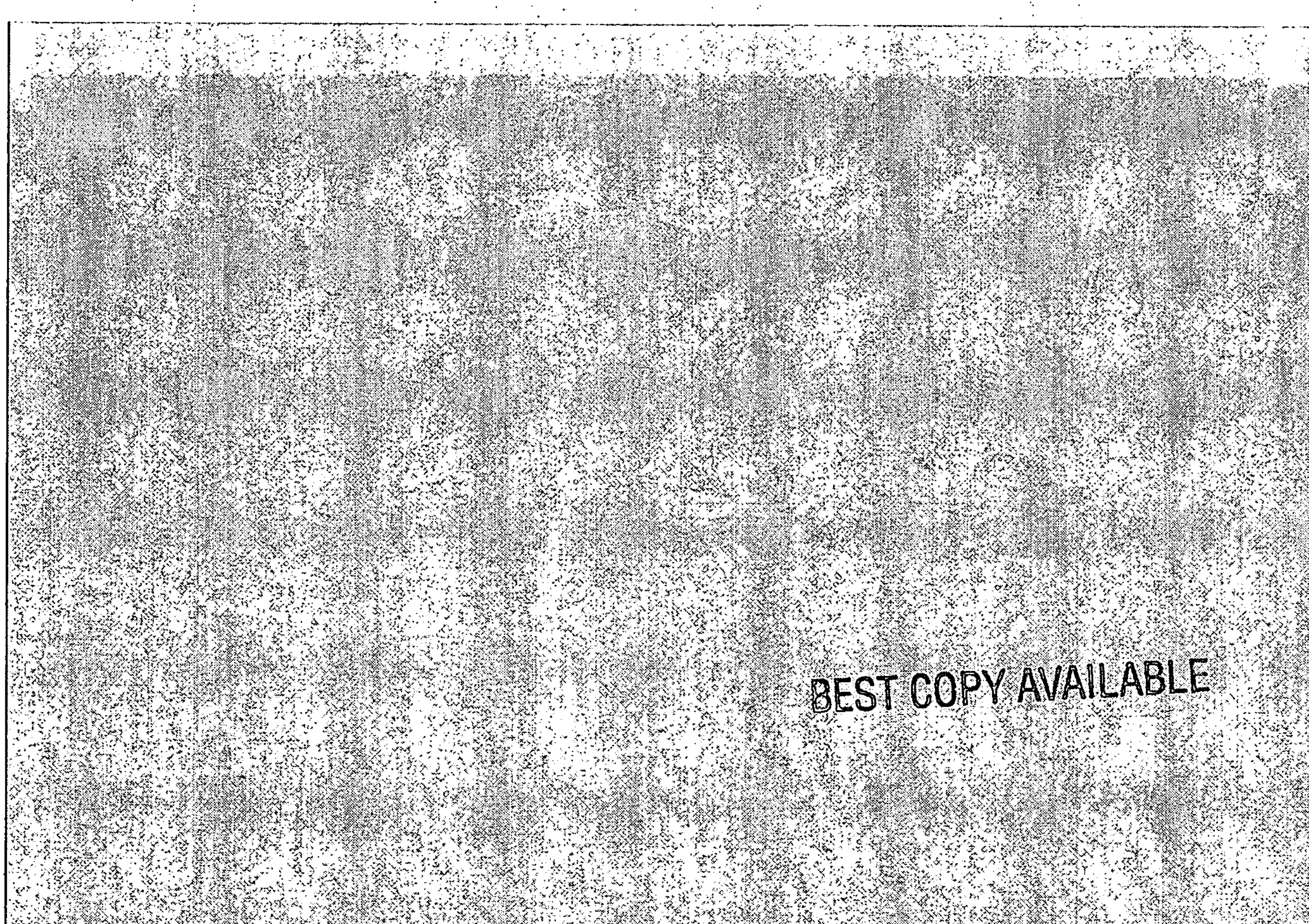
84

BEST COPY AVAILABLE

1998 Distribution of Reading Test Score Quartiles for All Schools—Continued

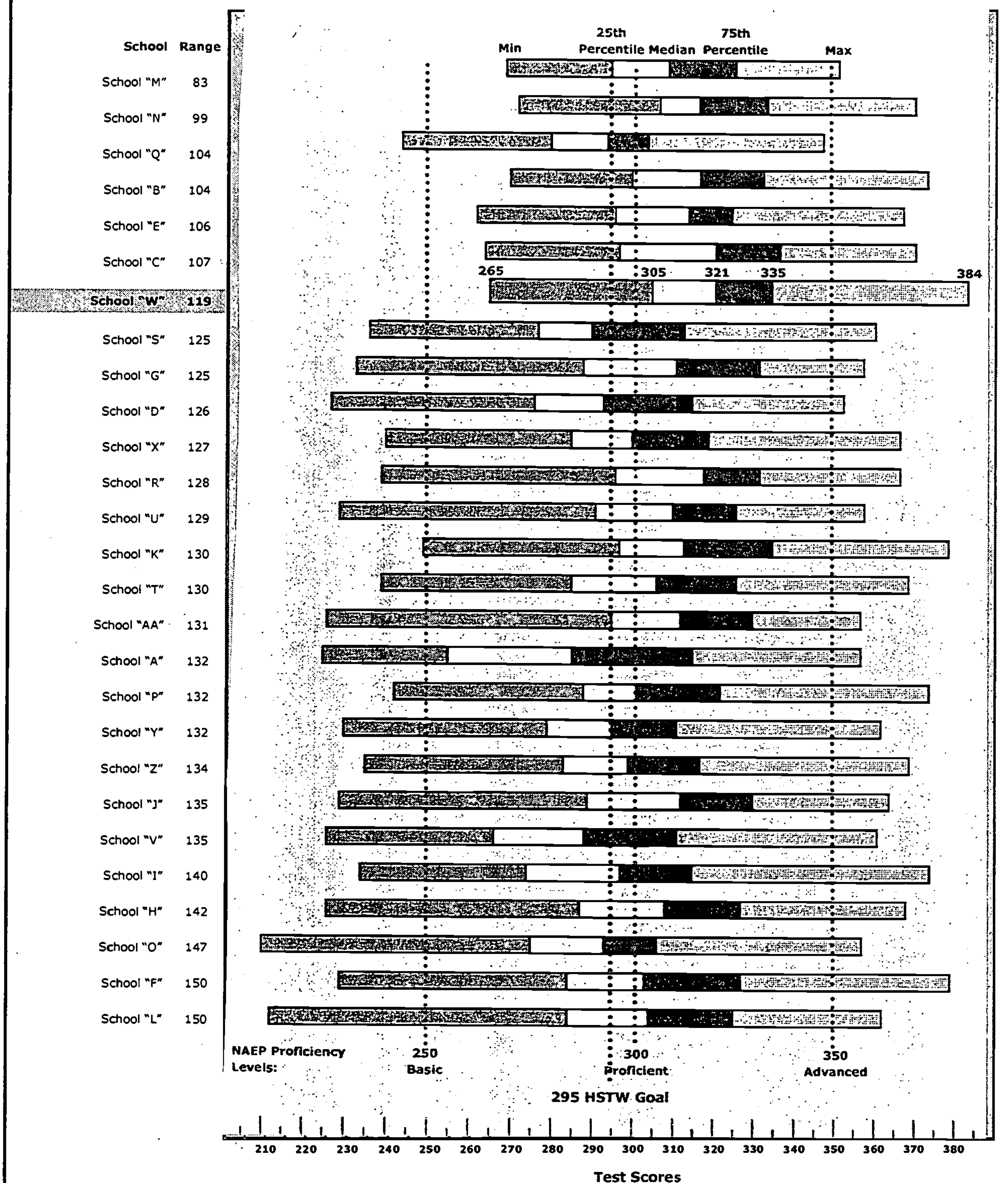
Schools are arrayed in ascending order based on the range of their test score data. The range is the difference between the highest and lowest score in the school. The scores for each school are divided into quartiles. Each quartile represents 25% of the scores. The 25th percentile marks the first quartile, the median (or 50th percentile) marks the second, the 75th the third, and the maximum the fourth. The median represents the halfway point between the highest and lowest score. In other words, half of the scores are above the median and half are below. For your reference, dotted lines corresponding to the *HSTW* Goal and NAEP proficiency levels are included to benchmark the performance of your students.

- School "W"'s median reading test score was above the *HSTW* goal of 279, which means that more than 50% of its seniors met the *HSTW* goal. A difference of 118 points separated the highest and lowest reading test scores for the school. This test score range was about average for schools in the study.
- About one-quarter of students performed below the *HSTW* Goal and a substantial portion were below the NAEP basic reading proficiency level. About one-half performed between the basic and proficient level, and one-quarter performed above the proficient level.
- A factor that probably contributed to the wide dispersion of reading scores and the substantial proportion of students achieving below the basic level is that, as seen in figure 4, only 41% of School "W"'s seniors met *HSTW's* English curriculum goal.



BEST COPY AVAILABLE

Figure 5b: 1998 Distribution of Mathematics Test Score Quartiles for All Schools in NCRVE Study



See explanatory notes and highlights on next page.

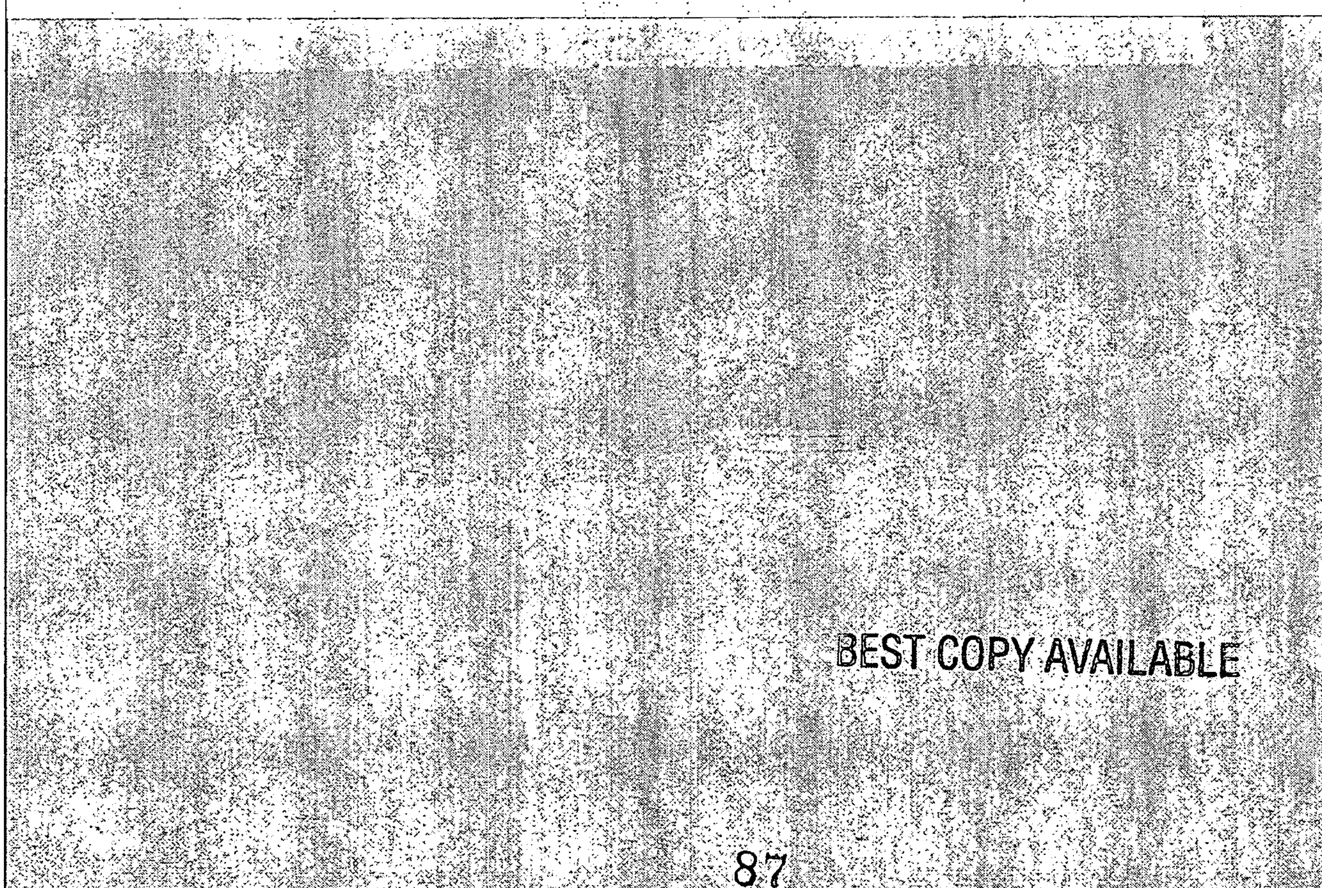
86

BEST COPY AVAILABLE

1998 Distribution of Mathematics Test Score Quartiles for All Schools—Continued

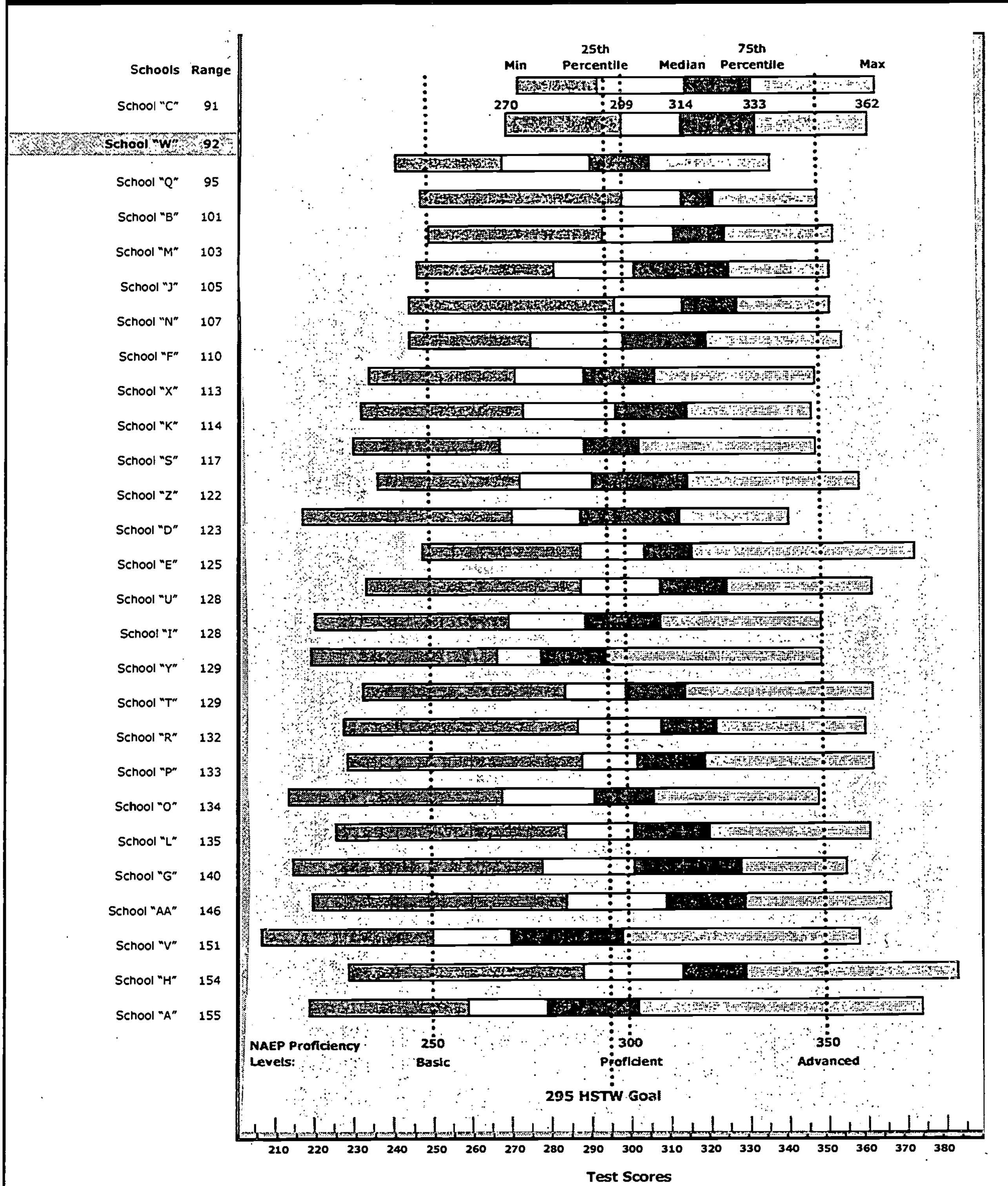
Schools are arrayed in ascending order based on the range of their test score data. The range is the difference between the highest and lowest score in the school. The scores for each school are divided into quartiles. Each quartile represents 25% of the scores. The 25th percentile marks the first quartile, the median (or 50th percentile) marks the second, the 75th the third, and the maximum the fourth. The median represents the halfway point between the highest and lowest score. In other words, half of the scores are above the median and half are below. For your reference, dotted lines corresponding to the *HSTW* Goal and NAEP proficiency levels are included to benchmark the performance of your students.

- More than three-quarters of School "W"'s seniors met the *HSTW* mathematics goal of 295.
- The range of School "W"'s mathematics scores is relatively narrow, with 119 points separating the highest and lowest scores.
- More than three-quarters of School "W"'s students performed at the NAEP proficient level in mathematics, and no students fell below the basic level. Compared with reading scores (where one-quarter were proficient and a substantial portion were below basic), it appears that School "W" is doing a relatively good job of preparing its students in mathematics.
- The fact that students did relatively better in mathematics than reading may be related to the fact that as seen in figure 4, 76% of seniors met *HSTW's* mathematics curriculum goal.



BEST COPY AVAILABLE

FIGURE 5C: 1998 DISTRIBUTION OF SCIENCE TEST SCORE QUARTILES FOR ALL SCHOOLS IN NCRVE STUDY



See explanatory notes and highlights on next page.

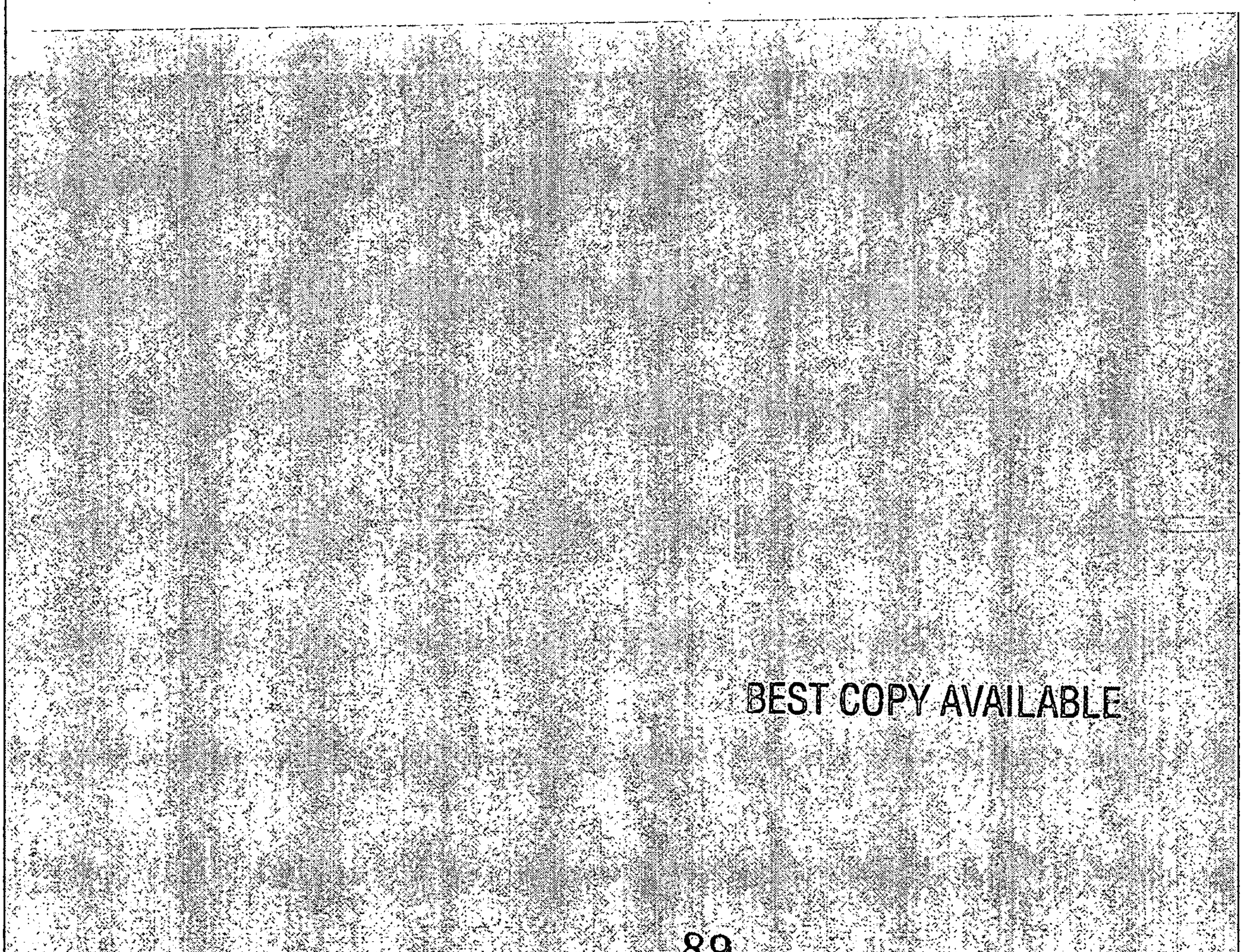
BEST COPY AVAILABLE

88

1998 Distribution of Science Test Score Quartiles for All Schools—Continued

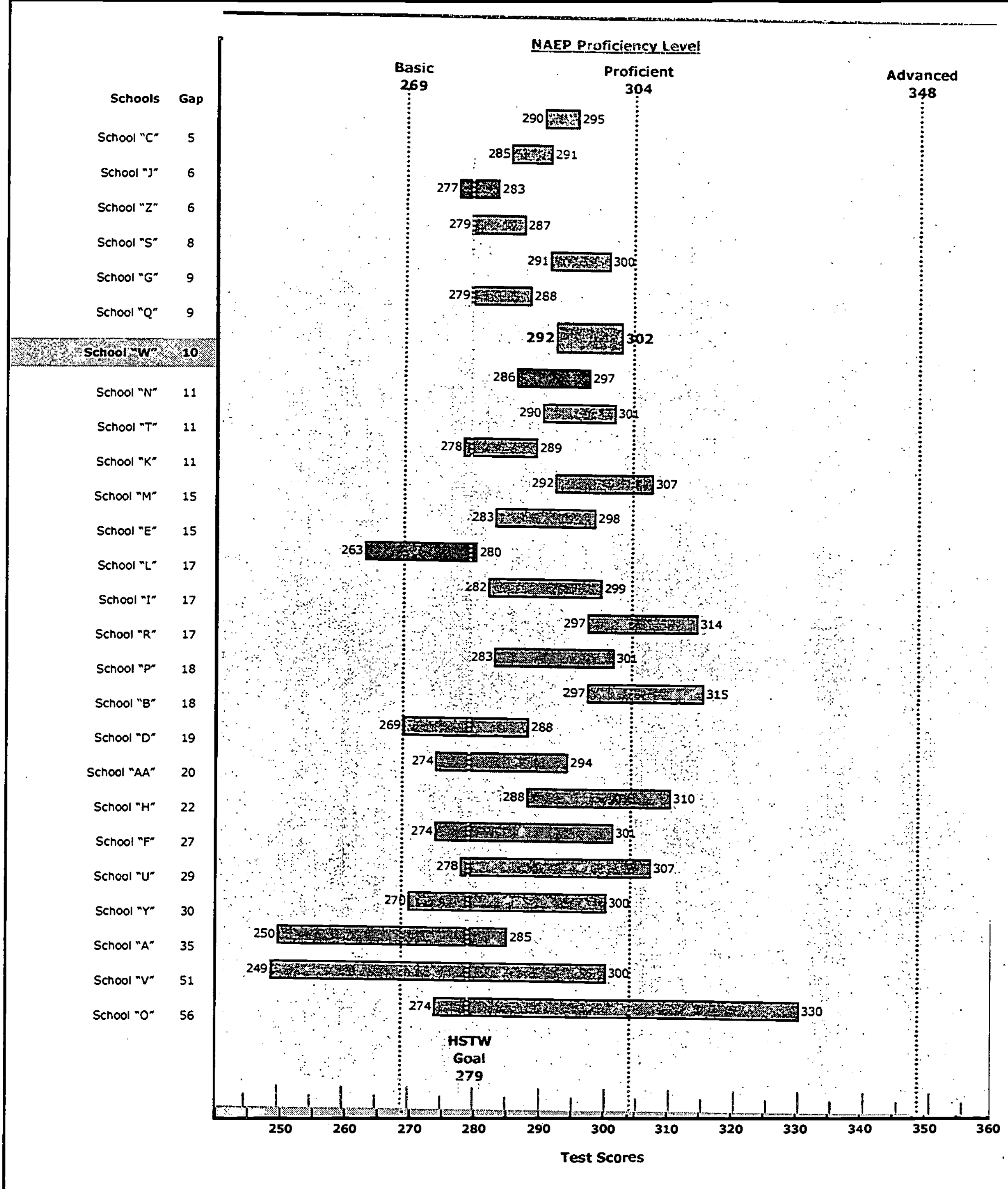
Schools are arrayed in ascending order based on the range of their test score data. The range is the difference between the highest and lowest score in the school. The scores for each school are divided into quartiles. Each quartile represents 25% of the scores. The 25th percentile marks the first quartile, the median (or 50th percentile) marks the second, the 75th the third, and the maximum the fourth. The median represents the halfway point between the highest and lowest score. In other words, half of the scores are above the median and half are below. For your reference, dotted lines corresponding to the *HSTW* Goal and NAEP proficiency levels are included to benchmark the performance of your students.

- In science, about three-quarters of School "W" students met or exceeded the *HSTW* goal and the NAEP proficient benchmark. No students fell below the basic level and some portion achieved at the advanced level.
- The range of School "W" scores was 93, which was one of the lowest ranges for schools in the study.
- As seen in figure 4, only 28% of students met the *HSTW* science curriculum goal. This suggests that if more students completed *HSTW's* recommended curriculum, perhaps even fewer students would fail to reach the *HSTW* goal.



BEST COPY AVAILABLE

Figure 6a: 1998 Mean Reading Test Score Gap Between Vocational and College Prep Students



Note: The following six schools have fewer than five students in the college prep category: School "J," School "K," School "L," School "O," School "Q," and School "S." Extreme caution should be used when interpreting these schools' results.

See explanatory notes and highlights on next page.

90

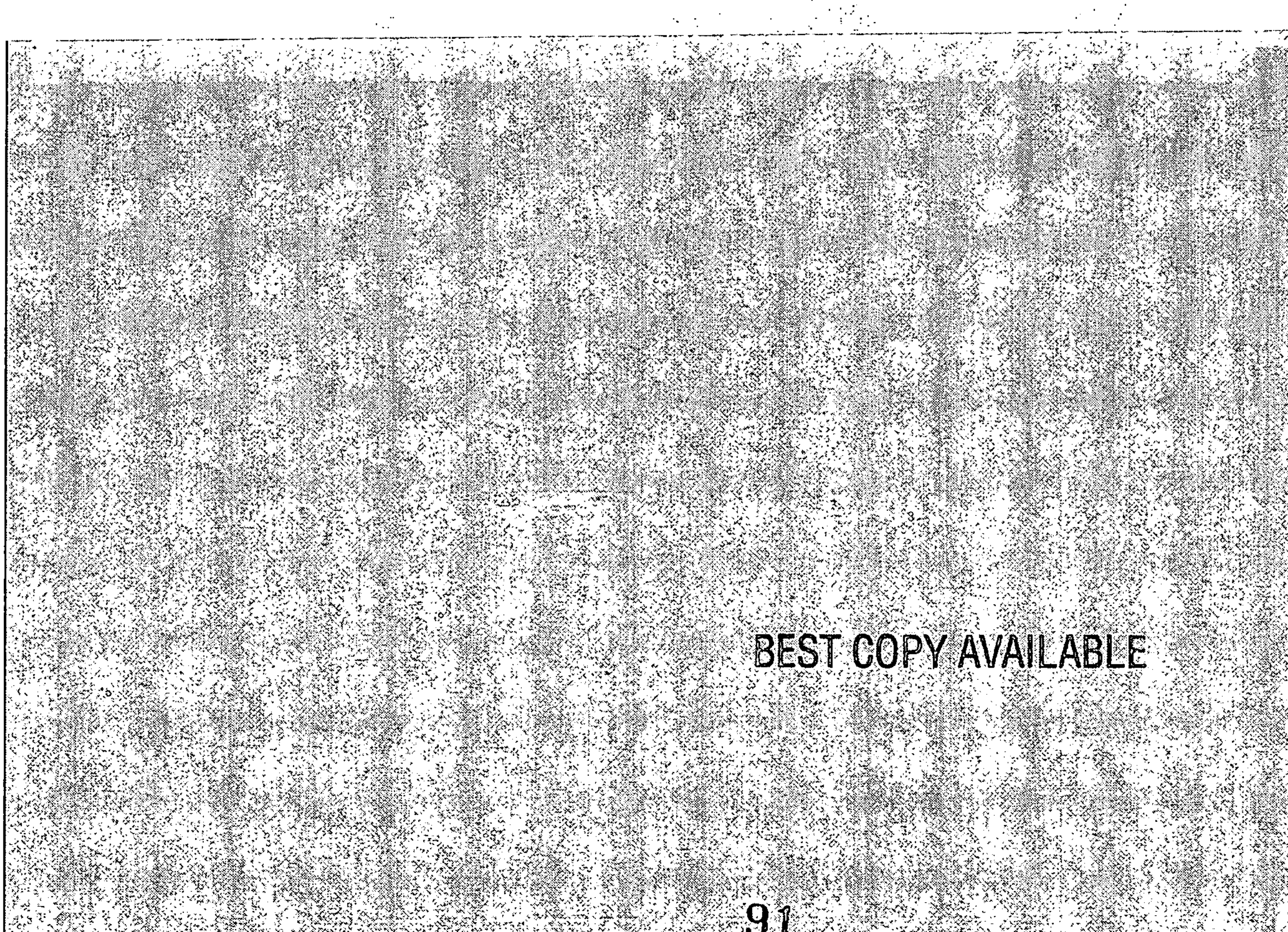
BEST COPY AVAILABLE

1998 Mean Reading Test Score Gap Between Vocational and College Prep Students—Continued

This figure allows you to compare the reading test score gap between vocational and college prep students in your school with that in other schools. Students who met both the college prep and vocational completer criteria are included in the vocational group. Schools are arranged in ascending order by the size of the mean score difference. The wider the bar, the larger the test score gap between vocational completers and college prep students. To help you benchmark the performance of your students against external standards, the *HSTW* achievement goal and NAEP proficiency levels are also included. In general, the vocational mean is below the college prep mean and appears on the left-hand side of each bar. In some schools, though, the vocational mean is above the college prep mean. In these instances, the bar is colored red. School "X" is not included in this graph because all students are both college prep and vocational completers.

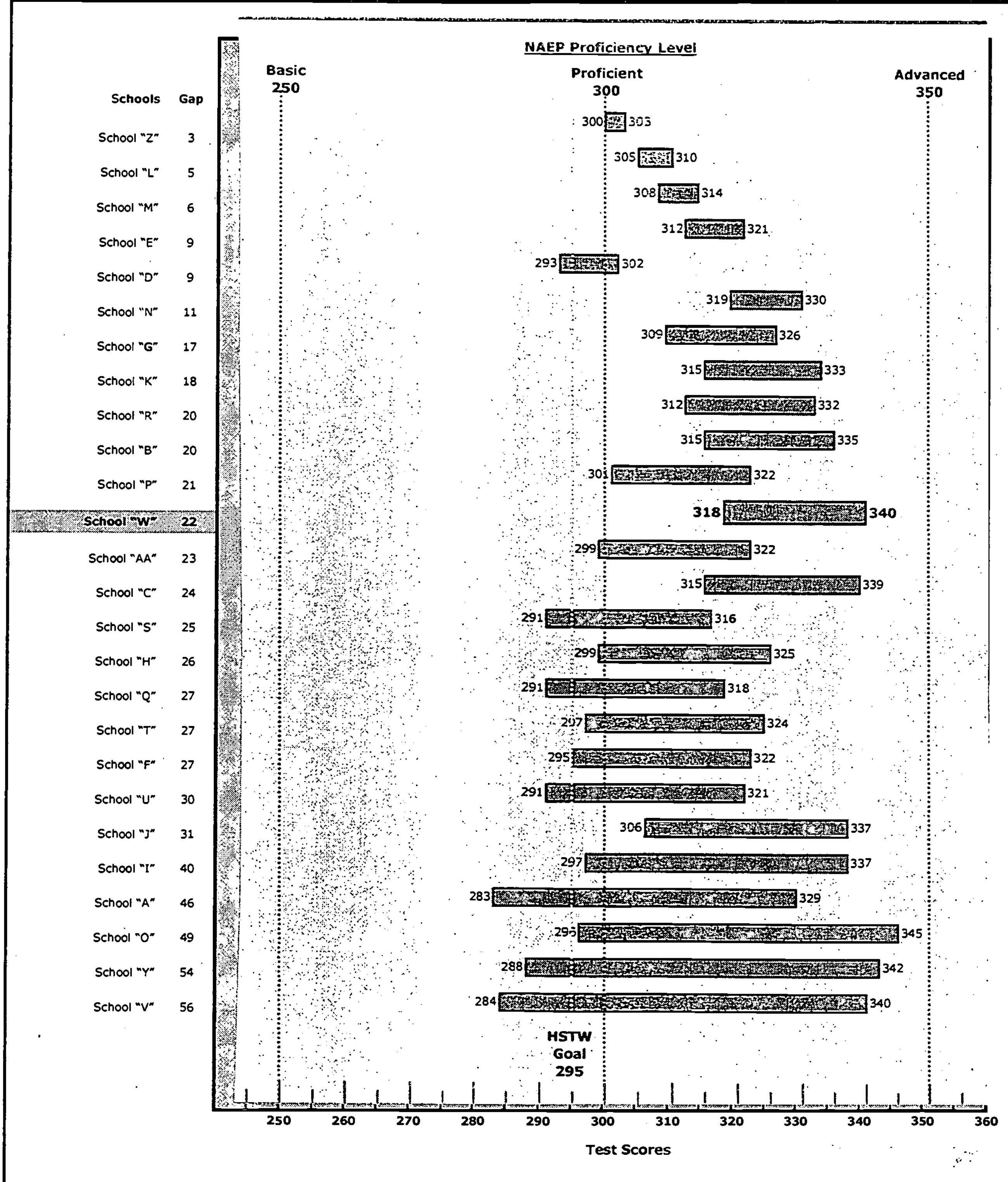
Note: Due to the small sample size of college prep students, care should be taken when interpreting these data.

- School "W"'s gap between vocational and college prep students in reading was only 10 points, one of the lower ranges for schools in the study.
- Both the vocational and college prep mean scores were above the *HSTW* goal.
- The college prep mean was almost at the same level as the NAEP proficient level. However, the vocational mean scores were 13 points below the NAEP proficient level, leaving room for improvement for this group.



BEST COPY AVAILABLE

FIGURE 6B: 1998 MEAN MATHEMATICS TEST SCORE GAP BETWEEN VOCATIONAL AND COLLEGE PREP STUDENTS



Note: The following six schools have fewer than five students in the college prep category: School "J," School "K," School "L," School "O," School "Q," and School "S." Extreme caution should be used when interpreting these schools' results.

See explanatory notes and highlights on next page.

92

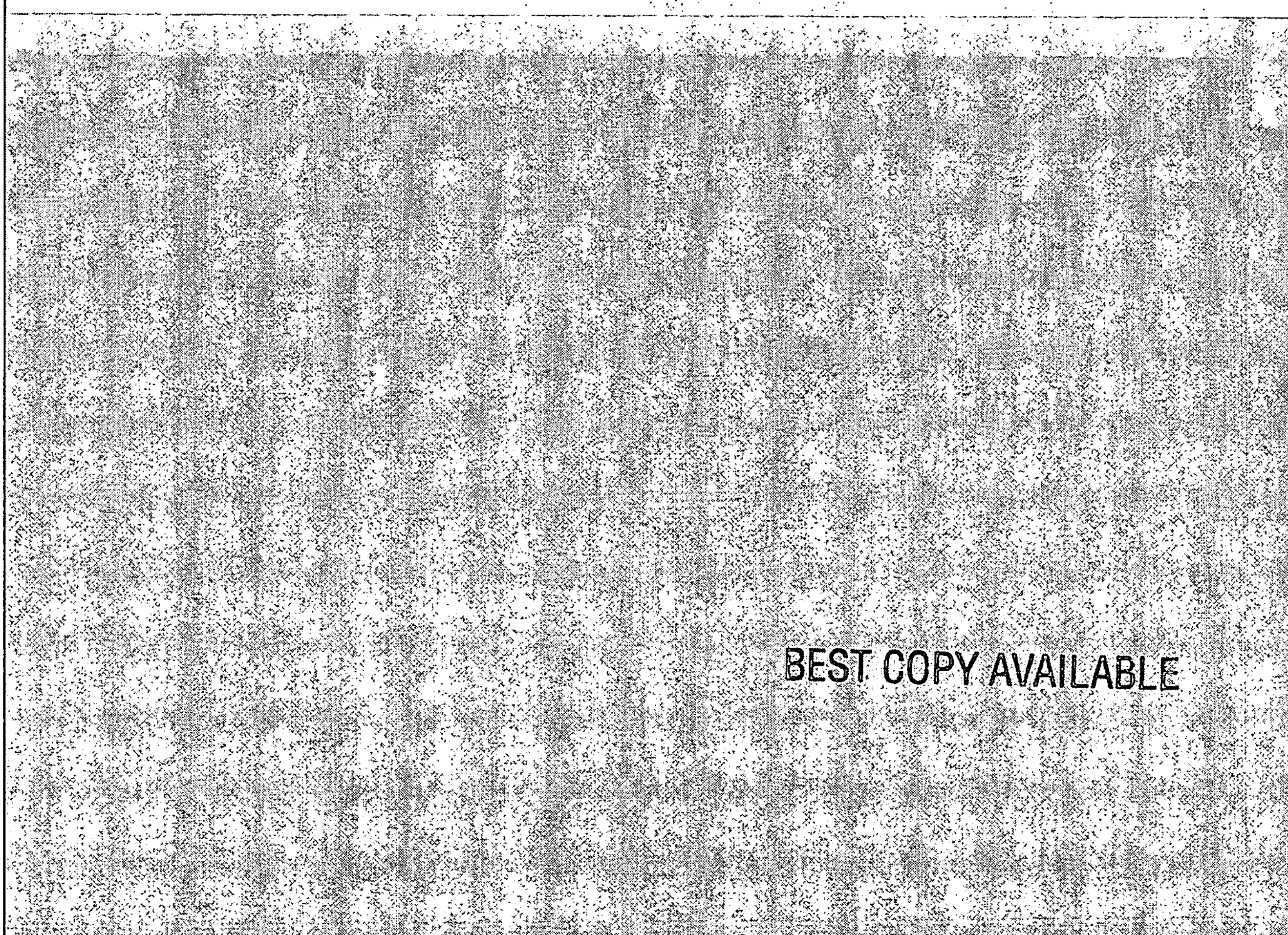
BEST COPY AVAILABLE

1998 Mean Mathematics Test Score Gap Between Vocational and College Prep Students—Continued

This figure allows you to compare the mathematics test score gap between vocational and college prep students in your school with that in other schools. Students who met both the college prep and vocational completer criteria are included in the vocational group. Schools are arranged in ascending order by the size of the mean score difference. The wider the bar, the larger the test score gap between vocational completers and college prep students. To help you benchmark the performance of your students against external standards, the *HSTW* achievement goal and NAEP proficiency levels are also included. Because the vocational mean is below the college prep mean, it appears on the left-hand side of each bar. School "X" is not included in this graph because all students are both college prep and vocational completers.

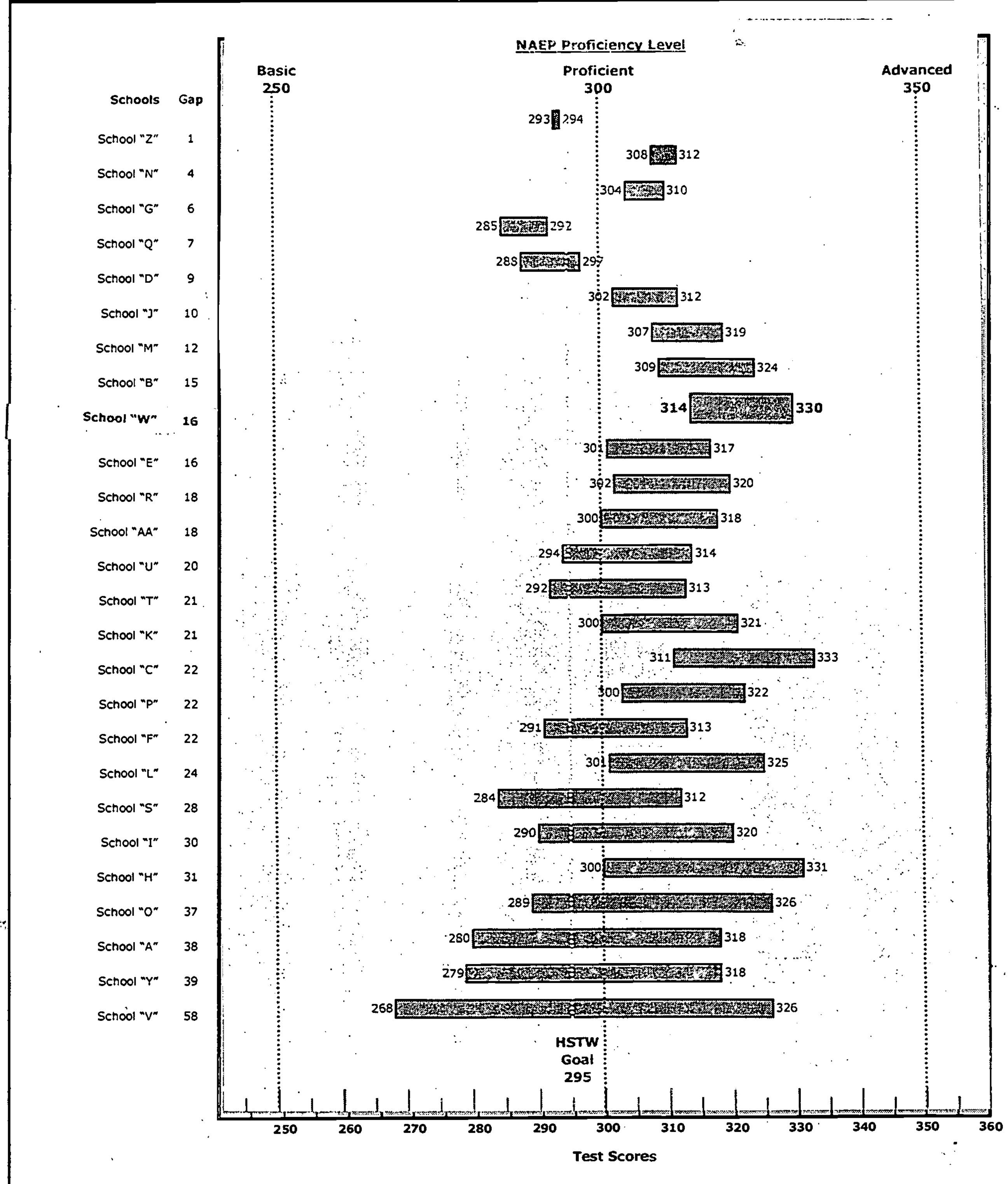
Note: Due to the small sample size of college prep students, care should be taken when interpreting these data.

- School "W" exhibited an average gap in mathematics test scores between its vocational and college prep students. There was a difference of 22 points between the average mathematics test scores for the two groups.
- Both the vocational and college prep groups performed, on average, above the NAEP proficient mathematics level and met the *HSTW* goal of 295. However, there was room for improvement for the vocational group, since this group was less than halfway to achieving "advanced" proficiency (325 is the halfway point).



BEST COPY AVAILABLE

Figure 6c: 1998 Mean Science Test Score Gap Between Vocational and College Prep Students



Note: The following six schools have fewer than five students in the college prep category: School "J," School "K," School "L," School "O," School "Q," and School "S." Extreme caution should be used when interpreting these schools' results.

See explanatory notes and highlights on next page.

94

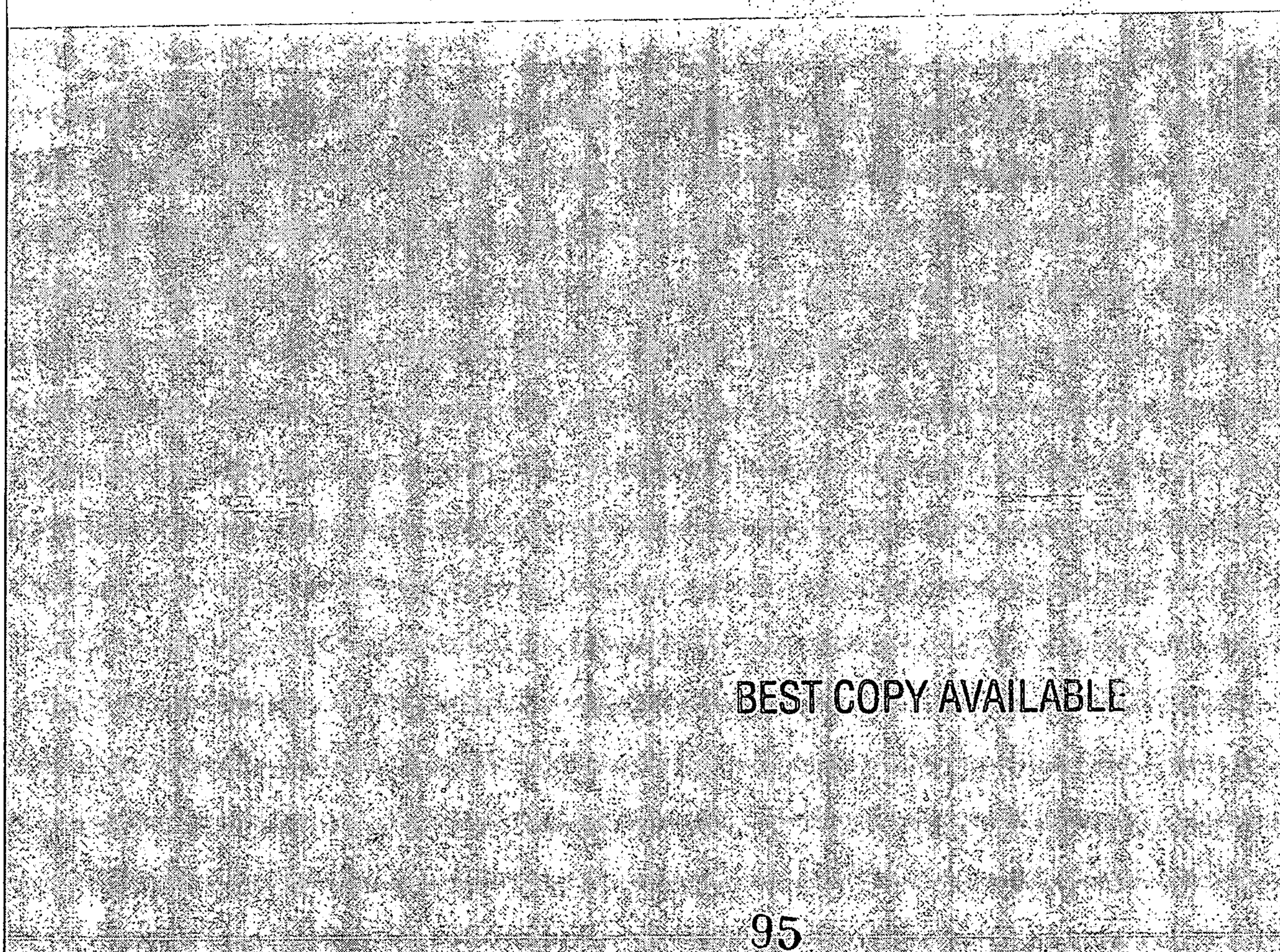
BEST COPY AVAILABLE

1998 Mean Science Test Score Gap Between Vocational and College Prep Students—Continued

This figure allows you to compare the science test score gap between vocational and college prep students in your school with that in other schools. Students who met both the college prep and vocational completer criteria are included in the vocational group. Schools are arranged in ascending order by the size of the mean score difference. The wider the bar, the larger the test score gap between vocational completers and college prep students. To help you benchmark the performance of your students against external standards, the *HSTW* achievement goal and NAEP proficiency levels are also included. In general, the vocational mean is below the college prep mean and appears on the left-hand side of each bar. In some schools, though, the vocational mean is above the college prep mean. In these instances, the bar is colored red. School "X" is not included in this graph because all students are both college prep and vocational completers.

Note: Due to the small sample size of college prep students, care should be taken when interpreting these data.

- ☛ School "W"'s gap between vocational and college prep students in science was 16 points, just above average for schools in the study.
- ☛ Both mean scores are above the *HSTW* goal of 295.
- ☛ The college prep mean (330) and vocational mean (314) are also both above NAEP's proficient level of 300, but well short of the advanced level of proficiency.



BEST COPY AVAILABLE

Table 1: Selected School and Student Characteristics

School	State	Urbanicity	Total Number of Students Assessed	High School Program: Percent Vocational Only	Percent Both Vocational and College Prep	Percent College Prep Only	Percent Neither Vocational nor College Prep	Parental Education: Percent with at Least One Parent with a Bachelor's Degree	Race/Ethnicity: Percent White	Percent Black	Percent Hispanic	Percent Asian/Pacific Islander or Other
School "A"	OK	small town	78	51	9	6	33	43	67	14	15	4
School "B"	GA	urban fringe	68	7	79	13	0	35	82	13	3	2
School "C"	PA	small town	77	44	30	13	13	40	87	7	3	4
School "D"	SC	rural	103	43	27	23	7	24	67	33	0	0
School "E"	GA	urban fringe	154	60	23	7	10	24	47	35	3	16
School "F"	WV	urban fringe	126	27	8	32	33	32	99	0	1	0
School "G"	KS	small town	95	41	19	7	33	48	64	0	27	10
School "H"	VA	urban fringe	98	36	16	23	24	40	86	12	1	1
School "I"	NC	small town	97	53	28	7	12	28	29	55	3	13
School "J"	KS	rural	48	35	6	6	52	41	83	0	11	7
School "K"	HI	urban	46	33	7	9	52	40	2	0	5	93
School "L"	WV	urban fringe	179	55	32	2	11	40	93	5	1	1
School "M"	KY	rural	70	26	57	13	4	20	99	1	0	0
School "N"	TX	rural	87	44	47	7	2	38	18	0	82	0
School "O"	LA	rural	79	66	3	1	30	25	60	37	3	1
School "P"	KY	small town	99	63	17	6	14	24	98	2	0	0
School "Q"	SC	small town	51	39	39	4	18	49	10	90	0	0
School "R"	IN	small town	146	45	14	15	25	33	98	0	1	1
School "S"	AR	urban	77	51	8	4	38	45	16	83	1	0
School "T"	MD	rural	112	44	6	25	25	33	86	12	2	0
School "U"	AL	rural	100	24	1	13	62	38	96	3	0	1
School "V"	FL	urban fringe	75	76	9	12	3	49	77	9	12	1
School "W"	WA	urban fringe	74	55	11	8	26	46	89	1	3	7
School "X"	DE	small town	184	25	75	0	0	22	77	20	1	2
School "Y"	HI	urban	103	29	4	8	59	30	5	0	3	92
School "Z"	NC	rural	109	36	33	18	13	45	42	58	0	0
School "AA"	MA	urban fringe	80	9	36	46	9	56	99	0	0	1

BEST COPY AVAILABLE

Table 2: Number and Percentage of Students and Mean Mathematics Score, by Selected Student and Family Characteristics

	Number Your School	Number All Schools	Percent* Your School	Percent* All Schools	Mean Mathematics Score Your School	Mean Mathematics Score All Schools
High School Program						
All students	74	2706	100	100	320	305
All vocational students	49	1788	66	66	318	303
Vocational students only	41	1121	55	41	314	296
Both vocational and college prep students	8	667	11	25	339	314
College prep students only	6	331	8	12	340	324
Neither vocational nor college prep students	19	587	26	22	318	300
Race/Ethnicity						
White	64	1829	89	70	321	309
Black	1	449	1	17	0	289
Hispanic	2	160	3	6	-	303
Asian/Pacific Islander	2	166	3	6	-	305
American Indian/Alaskan Native	3	29	4	1	-	301
Award Winner (applies to vocational students[1] only)						
Yes	13	556	27	31	332	326
No	36	1232	73	69	313	292
Took Mathematics in Senior Year						
Yes	25	1812	34	68	326	308
No	48	865	66	32	317	299
Parents' Highest Education[2]						
Less than high school	2	233	3	9	-	296
High school diploma	14	733	20	29	315	301
Some college	22	730	31	29	311	307
4-year college	32	850	46	33	330	312
Frequency a Family Member Asks About the Student's Schoolwork						
Almost every day	34	1382	47	52	318	305
About once a week	24	652	33	24	320	306
About once a month	6	179	8	7	335	308
Hardly ever, or never	9	462	12	17	317	303
Not Counting Vacations, Hours Worked per Week in a Part-Time Job						
None	19	982	26	37	521	306
15 hours or less	22	479	30	18	322	310
16 to 20 hours	19	518	26	19	314	304
21 to 30 hours	11	491	15	18	327	304
More than 30 hours	2	203	3	8	-	293
Hours of Television Watching per Day						
3 hours or less	65	2071	89	77	319	307
4 hours or more	8	604	11	23	326	297
Student's Plan for the Year Following High School Graduation[3]						
Not attending school, working/homemaker	16	571	22	21	313	291
Apprenticeship/Training program/ Technical and Business school	7	26	10	9	306	295
Attend 2-year college	19	509	26	19	320	304
Attend 4-year college	19	1086	26	41	337	316
Military/Other	12	271	16	10	310	300

—Too few cases for a reliable estimate.
*Due to rounding, percentages may not add to 100.
[1]Vocational students, including those who completed a college prep curriculum.
[2]The highest educational level achieved by either parent.
[3]What is the one thing that is likely to take the largest share of your time in the year after you leave high school?

ERIC

TABLE 3: PERCENTAGE AND NUMBER OF TEACHERS WHO REPORTED VARIOUS JOB-RELATED INFORMATION BY SCHOOL

TEACHERS	YOUR SCHOOL Percent	YOUR SCHOOL Number	ALL SCHOOLS Percent	ALL SCHOOLS Number
Who, during the past 3 years, had 21 or more hours of staff development about raising expectations and getting students to achieve higher standards through using extra help, applied learning, and team teaching	3	2	10	94
Who have master's/PhD/Professional or educational specialty degrees	53	41	50	576
Who teach only one or no class that is outside their field	90	69	93	1055
Who expect high-quality products and performances from students	28	22	40	452
Whose school's primary goal is to help all students master the essential content taught in college/help all students complete a program that prepares them for both employment and further learning/prepare all students for further learning	35	77	39	445
Who strongly agreed that they were encouraged to revise their curriculum so that they can teach more rigorous content in their course(s) to career-bound students	29	22	33	376

TABLE 4: PERCENTAGE AND NUMBER OF TEACHERS AND STUDENTS ACCORDING TO THEIR APPLIED LEARNING AND HOMEWORK ACTIVITIES

	YOUR SCHOOL Teacher1	YOUR SCHOOL Student2	ALL SCHOOLS Teacher	ALL SCHOOLS Student
APPLIED LEARNING IN MATHEMATICS/SCIENCE CLASSES				
Mathematics				
Number of mathematics teachers/students	8	74	175	2698
Percentage who used mathematics to solve a real-world problem[3]	88	30	79	28
Percentage who worked on an extended project[4]	75	16	48	21
Percentage who made a presentation[5]	88	16	78	12
Percentage who completed a joint assignment for mathematics and vocational teachers[6]	13	7	9	11
Science				
Number of science teachers/students	6	74	133	2698
Percentage who completed a lab assignment to solve a real-world problem[7]	83	37	76	39
Percentage who worked on an extended project[8]	83	16	62	21
Percentage who made a presentation[9]	100	36	79	27
Percentage who completed a joint assignment for science and vocational teachers[10]	0	17	18	15

HOMEWORK	YOUR SCHOOL	ALL SCHOOLS
Percentage of teachers who assigned 2 or more hours of homework per week	27	27
Percentage of students who spent 2 or more hours on homework per day	7	20

[1]Percentage of mathematics/science teachers who reported they required the students to do the following activities in class at least once or twice per semester.

[2]Percentage of students who reported they were required to do the following activities in their mathematics/science classes at least monthly or several times a year.

[3]Teacher: Used mathematics to solve a real-world problem found in the community or worksite of the vocational class.
Student: Completed a special mathematics project that required using mathematics in ways that most people would use mathematics in a work setting.

[4]Teacher: Worked on an extended, major project that lasted a week or more.
Student: Completed a mathematics assignment or project based on my own work experience or vocational class.

[5]Teacher: Stood before class to make a presentation or give a demonstration about an assignment.
Student: Stood before the class and made an oral presentation about a special mathematics project using visuals or other props.

[6]Teacher: Completed a joint mathematics assignment for mathematics and vocational teachers for which they received a grade in both classes.
Student: Completed a joint mathematics assignment for my mathematics and vocational teachers for which I received a grade in both classes.

[7]Teacher: Completed a science lab assignment in which they had to use science to address a problem found in the community or in a work setting.
Student: Completed a science lab assignment in which I used science to address a problem found in my community or a work setting.

[8]Teacher: Worked on an extended, major project that lasted a week or more.
Student: Completed a science assignment based on my own work experience or vocational class.

[9]Teacher: Stood before class to make a presentation or give a demonstration about an assignment.
Student: Stood before the class and reported on a completed science project using laboratory equipment, visuals, or other props.

[10]Teacher: Completed a joint science assignment for science and vocational teachers for which they received a grade in both classes.
Student: Completed a science project jointly assigned by a science and vocational teacher for which I received a grade in both classes.

ERIC Full Text Provided by ERIC

TABLE 5: TEACHER DATA FOR ALL SCHOOLS

SCHOOL	GENDER		AGE			SUBJECT AREA			EDUCATION		
	Male	Female	30 or under	31–50	over 50	Vocational/ Technical	Core Academic	Others	High School Diploma/ Business or Technical Certificate/ A.S.	Bachelor's Degree	Master's/ PhD/ Professional degree/ Educational Specialty
School "A"	44	56	16	53	32	42	53	5	16	58	26
School "B"	24	76	18	65	18	35	59	6	6	53	41
School "C"	64	36	14	58	28	43	37	20	8	47	45
School "D"	26	74	11	74	16	37	63	0	0	37	63
School "E"	32	68	28	52	20	44	52	4	8	56	36
School "F"	47	53	16	58	26	16	57	27	3	37	61
School "G"	43	57	37	57	7	40	53	7	0	83	17
School "H"	38	62	17	57	25	22	64	15	6	39	55
School "I"	41	59	19	64	17	31	43	26	8	67	25
School "J"	47	53	13	60	27	31	56	13	0	60	40
School "K"	52	48	24	52	24	33	67	0	0	29	71
School "L"	45	55	3	67	30	9	53	38	0	33	67
School "M"	55	45	27	59	14	59	41	0	5	32	64
School "N"	39	61	29	59	13	19	63	17	0	78	22
School "O"	38	62	19	63	19	26	44	30	4	52	44
School "P"	44	56	17	64	19	51	46	2	12	14	74
School "R"	43	58	13	64	23	18	51	31	0	13	88
School "S"	24	76	15	55	31	31	57	11	2	63	35
School "T"	46	54	15	62	23	28	54	18	5	43	52
School "U"	21	79	16	53	32	41	54	5	8	38	54
School "W"	49	51	23	60	17	18	47	34	3	44	53
School "Y"	13	87	17	38	44	30	66	4	0	35	65
School "Z"	30	70	17	53	30	33	50	17	7	73	20
School "AA"	38	63	13	66	22	6	53	41	0	41	59

Note: The following schools had no teacher data: School "Q," School "V," and School "X."

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

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").