COMPARING STATE AND DISTRICT TEST RESULTS TO NATIONAL NORMS: INTERPRETATIONS OF SCORING "ABOVE THE NATIONAL AVERAGE"

CSE Technical Report 308

Robert L. Linn, M. Elizabeth Graue, Nancy M. Sanders
University of Colorado
UCLA Center for Research on Evaluation,
Standards, and Student Testing

January, 1990

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred.

This paper was written as part of a research project sponsored by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST). It was prepared for presentation at the Annual Meeting of the American Educational Research Association, San Francisco, March, 1989.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

Acknowledgements

We thank the many state and district directors of testing and other staff who responded to our requests for data and took the time to complete the questionnaires and participate in the telephone interviews. Without their cooperation this study would not have been possible. A number of students worked part time on various aspects of collection, organization, computer entry, and analysis of the data and we are grateful for that assistance. We especially thank Frank Beaty, William Carlin, Ann Carrington, Sharon Catto, Jolene Dehning, Roy Grimm, and Kelly Strong in this regard. We also thank the members of the project advisory committee (David Bayless, Stan Bernknopf, Gordon Ensign, H.D. Hoover, Richard Jaeger, Daniel Koretz, and Lynn Winters) for their many helpful suggestions. Finally, we thank Eva Baker, Leigh Burstein, Stephen Dunbar, and Lorrie Shepard for their helpful comments on earlier drafts of this report.

Introduction

It has become commonplace for a state or district to report that its students are "scoring above the national average." Indeed, it has been suggested that all 50 states and most districts are reporting above average achievement test scores (Cannell, 1987). Is it really the case that all states claim that their students are performing above average on achievement tests? If so, how should such results be interpreted?

These are two of several questions that motivated a study of (a) norm-referenced test results that are being reported by states and school districts and (b) factors related to those scores. This report presents part of the findings of that study. Published reports and results of mail and telephone surveys of states and a nationally representative sample of school districts were used to document the degree to which "above average" achievement test results are being presented. Analyses of the possible influence of the changing meaning of norms are also presented. Subsequent reports will address a number of other factors that may have an impact on the achievement test scores of states and districts and on the proper interpretation of those results.

Background

Standardized achievement tests have long been used by schools to report student achievement to parents, policy makers, and the general public. In recent years, however, the attention given to test scores has increased dramatically. Low-stakes testing programs with results returned to teachers and reported in a low-key fashion to school boards and interested parents have given way to high-stakes testing programs that have direct and important effects on students, teachers, and school administrators. The increased emphasis on the use of test results for purposes of accountability has made questions of test quality and the trustworthiness of interpretations of major concern to educators and policy makers.

A common, albeit not the only or necessarily the best, way of providing the various audiences a means of interpreting test scores is to compare achievement test scores for a school building, a district, or a state to national norms. Slightly over half of the states and a substantial majority of the school districts rely on off-the-shelf, standardized achievement tests, for which normative comparisons provide a primary basis of interpretation. These comparisons take on a wide variety of forms, including the average grade equivalent score, the average normal curve equivalent score, the median percentile rank or percentile rank of the mean, the proportion of students scoring above the "national average," or more precisely, the national median, and the proportions of students with "below average, average, or above average" scores where the three categories correspond to stanines 1 thru 3, 4 thru 6, and 7 thru 9, respectively. In each of these examples, national norms provide the primary basis of comparison.

Norms, of course, are not the only basis of test score interpretation. Some states and districts rely on criterion-referenced interpretations of either publisher- or locally developed tests. In such cases, comparisons to past performance provide a key means of interpretation. For example, trends in the proportion of students passing a minimum-competency test, the proportion of students mastering specific objectives, or the average number of objectives mastered provide a means of comparing the current year's achievement with a benchmark. Trends may also be important in the interpretation of norm-referenced results, but the national norm still provides the major frame of reference for expressing the scores. Even states with locally developed or customized assessment programs sometimes also use comparisons to national norms to aid the interpretation of their achievement test results; these comparisons are obtained through special equating studies or item response theory links.

The pros and cons of normative comparisons have been discussed on many occasions. Discussions of appropriate and inappropriate normative interpretations are provided, for example, by Angoff (1971), Petersen, Kolen, and Hoover (1989), and in several introductory texts on educational and psychological measurement. Good discussions of appropriate and inappropriate uses and interpretations of norms may also be found in the technical manuals and interpretive guides provided by the publishers of the major standardized achievement tests.

Despite these discussions, normative interpretations continue to be misused and misinterpreted. The distinction that Angoff (1971) and others have made between the statistical meaning of "normative," which refers to "performance as it exists," and the use of the term to refer to "standards or goals of performance" (p. 533) is too often overlooked. The fact that norms for school averages or district averages differ markedly from norms for individual students is too often ignored or is given insufficient emphasis in interpretation. Because a school average is based on a range of student scores it necessarily falls somewhere in between the score of the highest scoring individual student and that of the lowest scoring student. Consequently, the distribution of school average scores is less variable than the distribution of individual student scores. The average achievement score that corresponds to the 70th percentile using school building norms, for example, may correspond to only the 60th percentile using norms for individual students.

It is widely believed that some tests have "easier" norms than others. If the norms of test A are easier or less stringent than those of test B, then a given level of achievement would be expected to appear better (e.g., result in a higher percentile rank or a larger proportion of students scoring above the national average) with test A than with test B. Note that the difficulty of norms is different than the intrinsic difficulty of test items. A test that asked easy questions could have hard norms because the norming sample was unusually able in the content area of the test. Conversely, a second test that asked relatively more difficult questions could have easier norms because the norming sample for the second test included a disproportionate number of low achieving students. The relative difficulty of norms for a particular school, school district, or state may also depend on the degree to which the test content matches the curriculum at the building or classroom levels.

The meaning of norms depends fundamentally on the definition of the reference population, and secondarily on the adequacy of sampling, the level of participation, and the motivation of the students in the norming sample, among other considerations. The year in which the norms were obtained is one of the important properties that define the reference population and it is clearly the case that norms become dated. If achievement is improving nationally, then the use of old norms will make a district or state appear to be doing better relative to the nation than would the use of current norms that provide a higher standard of comparison.

Although the above concerns about the use of norms are hardly new, questions about the meaning and trustworthiness of normative comparisons that states and districts are using to communicate test results to policy makers and the public have recently taken on increased importance. The increased importance is due, in part, to escalation in the stakes involved in testing. Concerns about normative comparisons were also exacerbated by the publication of a report by Dr. John J. Cannell (1987) titled "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average."

The Cannell report is based on a survey conducted by a community group, the Friends of Education, which found that "no state scores below the publisher's 'national norm' at the elementary level on any of the six major nationally normed, commercially available tests" (Cannell, 1987, p. 2, emphasis in original). Based on this finding, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press inflated and misleading reports on achievement levels" (p. 2).

Cannell was not the first to notice that states were reporting results that were above the national norm in greater numbers than would be expected based on past experience or common-sense notions of the likely relative standing of particular states. In 1984, the Southern Regional Education Board (SREB) reported that 9 of 11 SREB states with norm-referenced test results for elementary grades were at or above the national average (SREB, 1984). Two years later, "[i]n June, 1986, SREB first described this situation in which student achievement in nearly all states was reported to be at or above the national averages as the 'Lake Wobegon effect'—descriptive of Garrison Keillor's mythical town where all children are above average" (Korcheck, 1988, p. 3). However, it was the Cannell report that placed the issue in the national limelight.

The Cannell report attracted a good deal of attention in the press when it was released in the fall of 1987 and has been the focus of considerable debate and controversy among professional educators and measurement specialists ever since. There are undoubtedly a number of factors that helped focus attention on the findings. Dramatic statements regarding the findings such as those illustrated in the above quotes may be part of the reason. Interest in the report was probably enhanced also by the sharp criticisms of test publishers ("we believe inaccurate initial norms are the reason for high scores", p. 5, emphasis in original), of educators for the "integration of unchanging test questions into the curriculum" (p. 5, emphasis in the original), of those responsible for reporting student achievement ("no state publication honestly described norm-referenced testing," p. 6), of university and public educators serving as consultants to test publishers "who too often are mere sycophants, giving the commercial interests what they want" (p. 9), and of the U.S. Department of Education, "whose lack of knowledge of these tests constitutes nonfeasance" (p. 9, emphasis in original).

Even without the dramatic language and sharp criticism, however, the Cannell report raises serious questions and issues. The percentage of students reported to be scoring above the national 50th percentile in a number of states seems to defy common sense.

The Cannell report has been the focus of considerable discussion at national meetings and in professional journals concerned with issues of educational achievement and measurement. It was a major topic, for example, at the 1988 and 1989 Annual Assessment Conferences sponsored by the Educational Commission of the States. The report was featured along with six commentaries from test publishers and representatives of the U.S. Department of Education in the Summer 1988 issue of *Educational Measurement: Issues and Practice*. The report also led the U.S. Department of Education to arrange a meeting involving Dr. Cannell, representatives of major test publishers, and selected academics to discuss the findings and their implications in February, 1988.

Reviewers of the Cannell report (e.g., Drahozal & Frisbie, 1988; Koretz, 1988; Lenke & Keene, 1988; Phillips & Finn, 1988; Qualls-Payne, 1988; Stonehill, 1988; Williams, 1988) identified a number of factors, some of which were also suggested by Cannell, that might contribute to the seemingly anomalous finding that all states are above the national average. The fact that norms become dated was probably the most frequently mentioned potential explanation. Differences in the rules for exclusion of students from testing in norming and in operational testing programs was also proposed as a possible explanation by several reviewers (e.g., Drahozal & Frisbie; Koretz; Lenke & Keene; Phillips & Finn). Other suggested partial explanations included the possible effect of a closer match between the test and the local curriculum in operational testing programs than in norming samples (e.g., Koretz; Lenke & Keene; Phillips & Finn), and the possibilities that poor security, familiarity with the specific content of tests that are reused year after year, or teaching the test may inflate scores (e.g., Drahozal & Frisbie; Koretz; Phillips & Finn).

Reviewers (e.g., Drahozal & Frisbie, 1988; Koretz, 1988; Lenke & Keene, 1988; Phillips & Finn, 1988; Williams, 1988) also identified several shortcomings of the Cannell study and interpretations. The failure to distinguish between group and individual student norms in interpretations, aggregation bias that results when the percentage of <u>districts</u> with average scores above the national median is used to make inferences about the percentage of <u>students</u> with scores above the national median, and the treatment of the percentage of students at the 4th stanine or above as if it were an indicator of the percentage of students above the national average are among the misleading analyses and interpretations that were identified.

Despite these and other limitations, some reviewers concluded that Cannell's major findings are still probably correct. Stonehill (1988), for example, stated simply that "Cannell's evidence is compelling" (p. 23). Others were more circumspect. Koretz (1988), for example, noted that "Dr. Cannell's errors are to some extent beside the point...for they are not sufficient to call into question his basic conclusion" (p. 11), and Phillips and Finn (1988) stated that in the absence of "evidence to the contrary" they generally concurred with "the central finding of Dr. Cannell's report" (p. 10).

Procedure

The Cannell study provided part of the stimulus for the present study. Certainly the issues raised in that study are important ones that deserve to be investigated in greater detail. Of particular concern were the issues of aggregation bias, the sampling of districts to obtain estimates for states without statewide testing programs that provide normative comparisons to the nation, and the type of information obtained from districts. The Cannell study only asked districts whether their students were above or below the national average. More detailed district results would be more informative. Since the Cannell study did not include results for secondary schools, it was also important to expand the coverage to all elementary and secondary school grades.

Our interest, however, was in more than simply obtaining estimates of the number of states or the proportion of districts that report achievement test results that are above the national median or that have average achievement above the national mean. Such statistics are of interest, but are apt to raise more questions than they answer. It is evident that we also need to better understand the ways in which states and districts are using normative comparisons, the validity of those comparisons, and the factors that influence the results and the validity of test scores and their interpretation. Therefore, the present study was designed to collect data not only about the achievement scores that were reported by states and districts, but on a variety of related issues, including the way in which test results were used (e.g., public reporting, grade retention, school incentives), when and why the uses were initiated, how and when the tests were adopted, and policies regarding test administration, test security and the preparation of students for taking tests. The present report, however, is focused on the test results and the possible influence of changes in the stringency of norms over time. Other aspects of the project data are addressed elsewhere (e.g., Baker, 1989; Burstein, 1989; Shepard, 1989).

State Survey

Two national mail and telephone surveys were conducted. In the first survey, a letter and a data collection form (see Appendix A) were mailed to the directors of testing in all states. As can be seen in the sample copy in Appendix A, the state testing directors were asked to provide test results in reading and mathematics for all grades (K through 12) for the three most recent academic years (1985-86, 1986-87, and 1987-88).

States were asked to report the percentage of students scoring above the national 50th percentile statewide if the information was available. When it was not available, the states were asked to report state means and standard deviations in reading and mathematics as well as the scores corresponding to the 25th, 50th and 75th percentiles statewide. In addition to test score information, the states were asked to provide the name, edition, and form of the test used at each grade; the year the test was first used in the state; the year it was normed; the month of administration; and the way the scores were routinely reported (e.g., percentage of students above the national median). The number of students enrolled, the number tested, and the number for whom scores were reported were also requested at each grade for each of the three years in question.

Since much of the information we were seeking was already available in published reports, the state directors of testing were asked to send copies of reports containing the requested information. The reports served in place of completed data collection forms if the reports contained the necessary information. Since information about how scores are communicated to the public and how they are interpreted by the press was relevant to our interests, copies of press releases and newspaper articles about test results were requested.

Following the mailings, state directors of testing were contacted by telephone to arrange telephone interviews. Detailed results of the telephone interviews are presented in other reports of study results (see Shepard, 1989), hence only a brief description of the interview is presented here.

A copy of the telephone interview guide is shown in Appendix B. In addition to clarification questions about testing data requested on the data collection forms, testing directors were asked questions about test use, test selection, the alignment of curriculum with the test, about time spent on teaching tested objectives, about objectives given less time as a result of the test, about guidelines for test preparation, about typical and extreme practices in preparing students to take tests, and about test security practices and experience.

District Survey

A stratified random sample of districts designed to be representative of the fifty states was selected. The 1980 census data were used to stratify school districts by region, size, and socio-economic status (SES). The definitions of the levels of three stratification variables are provided in Table 1. As can be seen in Table l, the three stratification variables, region, size, and SES, had four, eight, and five levels, respectively. Thus a total of 160 cells were defined. The SES index, which is defined in Table l, was used to rank the school districts and then to define five strata such that approximately 15% of the students were in each of the two extreme strata (low and high), approximately 20% were in each of the adjacent strata (above and below average), and approximately 30% were in the average stratum.

Five districts were randomly selected for each cell where a sufficient number of districts was available according to the 1980 census. Five districts were available and selected for most cells; however, 15 of the cells were void and 39 of the cells had fewer than five districts. For example, there were no high SES districts with enrollments of 100,000 or more in the North/Central region and there was only one low SES district with an enrollment of 100,000 or more in the East region.

Table 1

Definitions of Stratification Variables Used to Sample School Districts

A. REGION. Region of the country was defined to have 4 strata.

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

B. SIZE. District enrollment, 1980 Census, 8 strata.

1. Less than 1,200
2. 1,200 to 2,499
3. 2,500 to 4,999
4. 5,000 to 9,999
5. 10,000 to 24,999
6. 25,000 to 49,999
7. 50,000 to 99,999
8. 100,000 or more

C. SES. Community socio-economic status index based on the 1980 census. SES equals the median family income in thousands of dollars plus 6 times the median years of education of the population 25 years old or older. SES used to define 5 strata. The labels of the strata and approximate percentage of students in each are:

1. Low (15%)
2. Below Average (20%)
3. Average (30%)
4. Above Average (20%)
5. High (15%)

The first of the randomly-ordered districts in each of the 145 non-void cells was selected for inclusion in the survey. Because achievement test results of large school districts have been the focus of considerable attention in recent years, we were particularly interested in obtaining better information about the achievement test results being reported by larger districts. Therefore, districts with enrollments of 50,000 or more were oversampled. With the oversampling of large districts, a total of 175 districts were selected for the sample. Appendix C lists the number of districts selected per cell.

After districts were selected, telephone calls were made to confirm that the district was still operating (had not, for example, been consolidated with another district since the 1980 census), to identify appropriate respondents who were responsible for the district testing program, and to obtain complete mailing addresses. Where a district no longer existed, the second listed district in the corresponding cell of the sampling design was selected as a replacement. Once addresses were obtained, letters (see Appendix D) and data collection forms were mailed.

A subsample of the districts was identified for telephone interviews, which were conducted following the mail survey (see Appendix E for a description of the procedures used to identify the interview subsample). Because telephone interviews were conducted with a subsample of the districts, two different letters requesting participation and two different data collection forms were sent to districts (see Appendix D). The same basic test data that were requested from states were also requested for all districts. Districts in the mail-survey-only subsample were also sent a brief questionnaire covering some of the interview questions about the use of test results and perceived effects of testing in the district (see Appendix D). Districts in the interview subsample did not receive a questionnaire, but were asked questions shown in the interview guide in the telephone survey (Appendix D).

Follow-up letters were sent to districts approximately three weeks and again six weeks after the initial mailing. If no response was received within three weeks after the second follow-up, attempts were made to reach respondents by telephone and urge them to respond to the survey. When district personnel declined to participate in the survey or could not be reached after repeated telephone attempts, the reason for the non-participation was recorded, and a substitute district was selected from the appropriate cell in the sampling design.

Results

States with Norm-Referenced Comparisons

A total of 35 states provided results that allowed norm-referenced comparisons for one or more grades in at least one of the three years for which data were collected (1985-86, 1986-87, and 1987-88). The remaining 15 states did not use tests with national norms. The 35 states for which norm-referenced comparisons were obtained are listed in Table 2 with an indication of the basis for the comparison and the grade levels for which test results were reported. The basis for comparisons to national norms for states that administered an off-the-shelf, norm-referenced test is obvious. However, in order to obtain estimates of the percentage of students scoring above the national median or the percentile rank of the state mean or median test score, it was sometimes necessary to convert scores from the form in which they were reported. For example, if the state reported mean grade-equivalent scores, those scores were converted to the corresponding percentile rank by reference to the test publisher's norms tables for individual pupils.

Table 2
States with Norm-Referenced Comparisons and
Grades Where at Least One Comparison is Available

State	Basis of Comparison*	Grades 1	2	3	4	5	6	7	8	9	10	11	12
Alabama	NRT	+	+		+	+		+	+		+		
Alaska	NRT	+	+	+	+	+	+	+	+	+	+	+	+
Arizona	NRT	+	+	+	+	+	+	+	+	+	+	+	+
Arkansas	NRT				+			+			+		
California	LINK			+			+		+				+
Colorado	NRT			+			+			+		+	
Delaware	NRT	+	+	+	+	+	+	+	+			+	
Georgia	NRT		+		+			+		+			
Hawaii	NRT			+			+		+		+		
Idaho	NRT						+		+			+	
Illinois	LINK			+			+		+				
Indiana	NRT	+	+	+			+		+	+		+	
Iowa	NRT	+	+	+	+	+	+	+	+				
Kentucky	NRT/LINK	+	+	+	+	+	+	+	+	+	+	+	+
Louisiana	NRT				+		+			+			
Maryland	NRT			+		+			+				
Mississippi	NRT	+			+		+						
Missouri	LINK			+			+		+		+		
Nevada	NRT			+			+						
New Hampshire	NRT				+				+		+		
New Mexico	NRT			+		+			+				
North Carolina	NRT	+	+	+			+		+				
North Dakota	NRT			+		+		+					
Oklahoma	NRT			+				+			+		
Oregon	LINK								+				
Rhode Island	NRT			+			+		+		+		
South Carolina	NRT				+	+		+		+		+	
South Dakota	NRT				+				+			+	
Tennessee	NRT		+			+		+		+			+
Texas	LINK	+		+		+		+		+			
Utah	NRT					+						+	
Virginia	NRT				+				+			+	
Washington	NRT				+				+		+		
West Virginia	NRT			+			+			+		+	
Wisconsin	NRT				+				+			+	
Number of States:	35	10	10	20	16	13	18	13	22	11	11	13	5

* NRT = Norm-Referenced Test LINK = Equated to NRT
NRT/LINK = Some years based on NRT and others on LINK

Several of the states listed in Table 2 obtained normative comparisons indirectly by linking non-normed tests or state assessment results to a norm-referenced test through the use of special equating studies or the inclusion of norm-referenced test items with known item parameters in a customized test (see, for example, Yen, Green, & Burket, 1987, for a discussion of customized tests). States for which norm-referenced comparisons were obtained indirectly through such linkages are indicated in Table 2 by the word "LINK" in the column showing the basis of comparison.

Although comparisons to national norms either directly or through an equating link could be obtained for a total of 35 states in all, the number of comparisons varied substantially by grade level. As can be seen in Table 2, the largest number of states with results for any single grade was 22 at Grade 8. Grade 3, with 20 states, and Grade 6, with 18 states, were used for statewide testing nearly as often as Grade 8. However, there was no grade for which normative comparisons were available for a majority of the 50 states. Test results were reported by only 10 or 11 states at Grades 1, 2, 9, and 10; only 5 states reported normative test results for Grade 12.

Where possible, estimates of the percentage of students in a state who scored above the national median were obtained separately for each grade tested in reading and mathematics. Where estimates of the percentage of students above the national median could not be obtained, the state median percentile rank or the percentile rank corresponding to the statewide mean was used. Note that here, and throughout this report, it is the individual pupil norms, rather than norms for school buildings or school districts, that were used to determine percentile ranks. For some states, estimates of both the percentage of students above the national median and the median percentile rank or percentile rank of the statewide mean were available and used.

The number of states and the number of students for which estimates of the percentage of students above the national median were obtained are reported in Table 3 by year of test administration, test content, and grade. Parallel numbers are reported in Table 4 for states where estimates of the median percentile rank or the percentile rank of the statewide mean were obtained. The latter numbers were also used to obtain weighted mean percentile ranks for the states for which those results were obtained. In many cases the number of states and number of students in Tables 3 or 4 are the same for mathematics as for reading, because of the fact that both content areas were usually tested and a single number of students tested was reported for both tests. However, there are some differences (e.g., Grade 8 in Table 3), because results were available in reading but not mathematics for a given state.

Percentage of students above national median. The combined results for states of the percentage of students scoring above the national median are summarized in Figure 1. The percentages shown in Figure 1 are weighted by the number of students tested in each grade for the states reporting data for each of the three years for which data were collected. Thus each bar in the figure represents the percentage of students in the states that provided data in this form who scored above the national median for a given school year and a given grade in either reading or mathematics. For example, the first column for Grade 1, 1985-86, is based on the 281,734 first-grade students in the 7 states (see Table 3) that reported test results in this form; it shows that 54% of those students scored above the national median in reading.

The results in Figure 1 are consistent with the general results reported by Cannell (1987) in that the overall percentage of students above the national median was greater than 50 in all of the elementary grades in both reading and mathematics for each of the three years studied. The percentage above the national median was

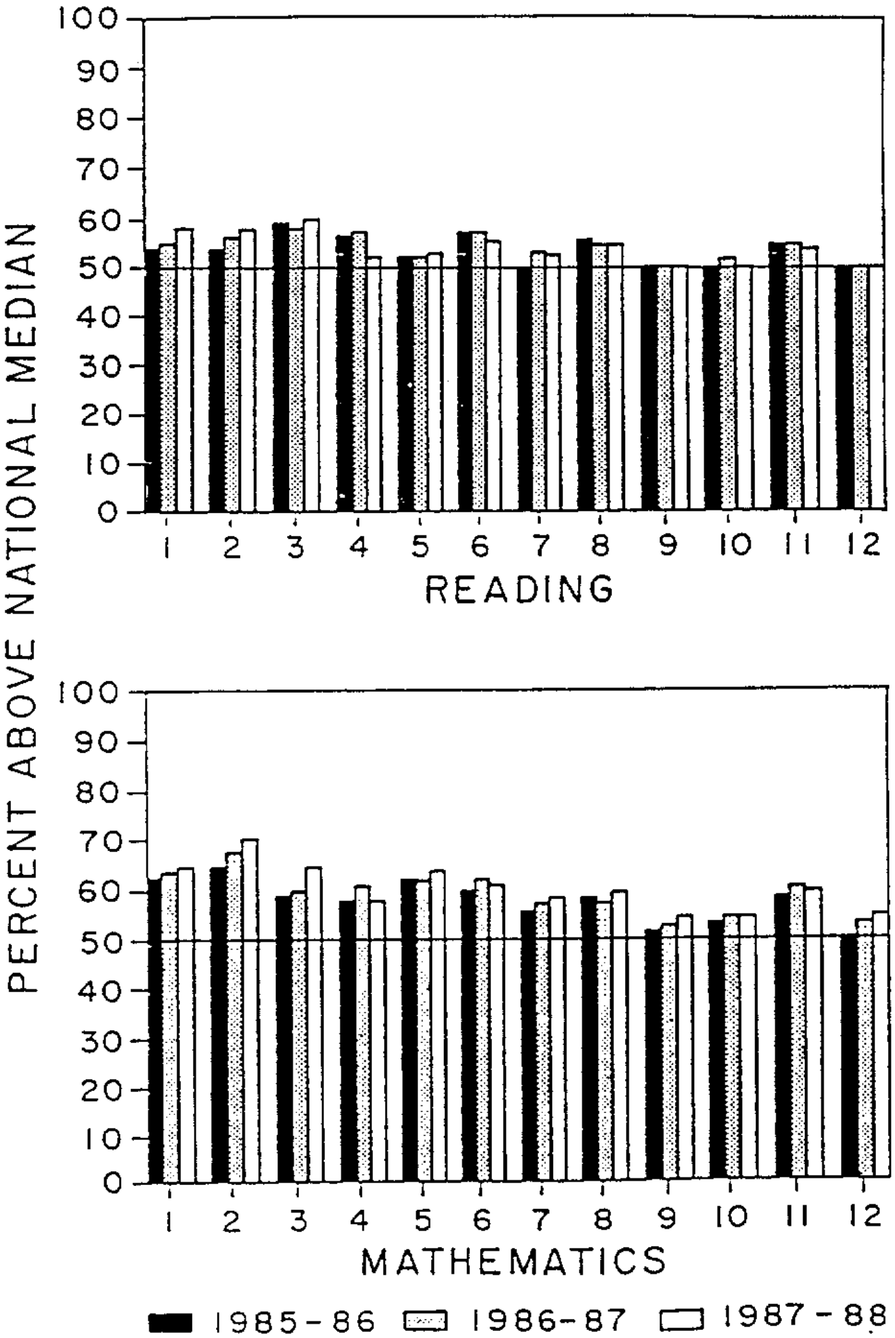
Table 3
Number of States and Number of Students Contributing to Estimates of Percentage of Students Above National Median by Year, Test Content, and Grade

	1985-86		1986-87		1987-88	
Grade	Number of States	Number of Students	Number of States	Number of Students	Number of States	Number of Students
I. Reading						
1	7	281,734	6	271,954	7	302,544
2	8	343,490	7	329,928	7	330,255
3	12	362,239	12	302,893	10	461,152
4	14	460,480	13	452,447	13	485,084
5	8	242,871	7	209,289	8	226,122
6	10	288,671	10	231,702	11	474,498
7	10	381,570	8	283,334	9	337,862
8	13	445,687	16	433,801	13	505,762
9	10	250,712	7	244,762	8	351,102
10	8	271,706	10	296,866	8	258,866
11	10	250,712	11	239,223	11	241,956
12	3	65,809	3	67,782	2	68,841
II. Mathematics						
1	7	281,734	6	271,954	7	302,544
2	8	343,490	7	329,928	7	330,255
3	11	353,612	11	293,452	9	339,089
4	14	460,480	13	452,447	13	485,084
5	8	242,871	7	209,289	8	226,122
6	9	280,053	9	222,886	10	364,093
7	10	381,570	8	283,334	9	337,862
8	13	445,687	15	424,959	12	396,574
9	7	300,728	7	244,762	8	351,102
10	8	271,706	9	287,457	8	258,866
11	10	250,712	11	239,223	11	241,956
12	3	65,809	3	67,782	2	68,841

Table 4
Number of States and Number of Students Contributing to Estimates of Percentile Rank of State Means or Medians by Year, Test Content, and Grade

	1985-86		1986-87		1987-88	
Grade	Number of States	Number of Students	Number of States	Number of Students	Number of States	Number of Students
I. Reading						
1	5	250,628	5	264,972	6	295,840
2	6	308,342	6	323,318	7	385,391
3	11	623,579	12	336,372	12	394,641
4	11	389,954	12	446,642	13	509,839
5	7	206,325	8	250,586	11	336,191
6	8	526,312	8	245,215	11	391,526
7	8	317,994	8	281,849	11	401,015
8	11	403,406	16	471,619	14	468,180
9	6	295,903	6	239,606	8	348,617
10	6	236,868	9	291,311	8	253,699
11	9	246,555	10	234,746	10	237,583
12	3	276,030	2	65,120	2	68,841
II. Mathematics						
1	5	250,628	5	264,972	6	295,840
2	6	308,342	6	323,318	7	385,391
3	11	623,579	12	336,372	12	394,641
4	11	389,954	12	446,642	13	509,839
5	7	206,325	8	250,586	11	336,191
6	8	526,312	8	215,215	11	391,526
7	8	317,994	7	244,332	11	401,015
8	11	403,406	16	471,619	14	468,180
9	6	295,903	6	239,606	8	348,617
10	6	236,868	8	253,671	8	258,722
11	9	246,555	10	234,746	10	237,583
12	3	276,030	2	65,120	2	68,841

Figure 1
Percentage of Students Scoring Above National Median
Based on States Reporting (Weighted by Number of Students)



usually greater for mathematics than for reading. Percentages were usually higher for elementary than secondary grade levels. For Grades 1 thru 6, the percentage of students scoring above the national median in mathematics ranged from a low of 58% in Grade 4 for the 1985-86 school year to a high of 71% in Grade 2 for the 1987-88 school year, whereas the corresponding range for reading was from 52% (Grade 5, 1985-86) to 60% (Grade 3, 1987-88). For Grades 7 through 12, the percentage of students scoring above the national median ranged from 49% (Grade 12, 1985-86) to 60% (Grade 11, 1986-87) in mathematics and from 48% (Grade 9, 1986-87) to 55% (Grade 8, 1985-86) in reading.

It should be noted that while the percentages displayed in Figure 1 are generally above the naive expectation of 50%, many individual students were, in fact, receiving scores that were well below the national median. If a state reported that 55% of its students had scores at or above the national median, for example, it is obviously the case that the remaining 45% of the students in the state were receiving scores below the national median.

The results in Figure 1 provide only a very global picture since they combine the data for varying numbers of states at each grade level. They do not, for example, provide an indication of the variability from state to state. Some sense of the variability can be obtained from Figures 2 and 3, which show the distributions of the percentage of students above the national median in reading and in mathematics, respectively.

The data for the most recent year available for each state were used for the distributions in Figures 2 and 3, which for most states was the 1987-88 school year. Each point in Figures 2 and 3 represents the percentage of students in a state who scored above the national median in a particular grade.

As can be seen in Figure 2, there is considerable variability from state to state. The tendency for the percentages to be greater than 50 is quite evident for the elementary grades. However, there are some cases where the percentage is substantially below 50. It should be noted that the point in Figure 2 that is most out of line with the Cannell (1987) results is the Grade 4 reading point that corresponds to a state where only 33% of the students were reported to have scored above the national median. This state introduced a statewide test in 1987-88 and hence was not included in the results reported by Cannell.

The results shown in Figure 3 for mathematics show even greater state-to-state variability than was seen for reading. Consistent with the global results in Figure 1, the tendency for the percentages to be above 50 is more evident in mathematics than in reading. Some of the percentages in Figure 3 are extraordinarily high. Note, for example, Grade 2, where one state reported that 86% of the students scored above the national median. The only two examples of a state where the percentage is below 50 for Grades 1 through 6—the 41% at Grade 4 and the 49% at Grade 6—are both for the state that introduced statewide testing in 1987-88 and therefore was not included in Cannell's state-level data collection.

Median percentile ranks or percentile rank of state means. Since the percentage of students scoring above the national median could not be estimated for all states, the median percentile ranks or percentile ranks of state means were also analyzed. Figures 4, 5, and 6, which parallel Figures 1, 2, and 3, respectively, display the results of the latter analyses. In general, the results using these percentile rank statistics are quite similar to the results using the percentage of students scoring above the national median. This is so despite the differences in the properties of the two statistics and the fact that the two sets of analyses are based on different, albeit overlapping,subsets of states.

Figure 2
Percentage of Students Reported by States to be Scoring above the National Median in Reading (Each Point Represents a State)

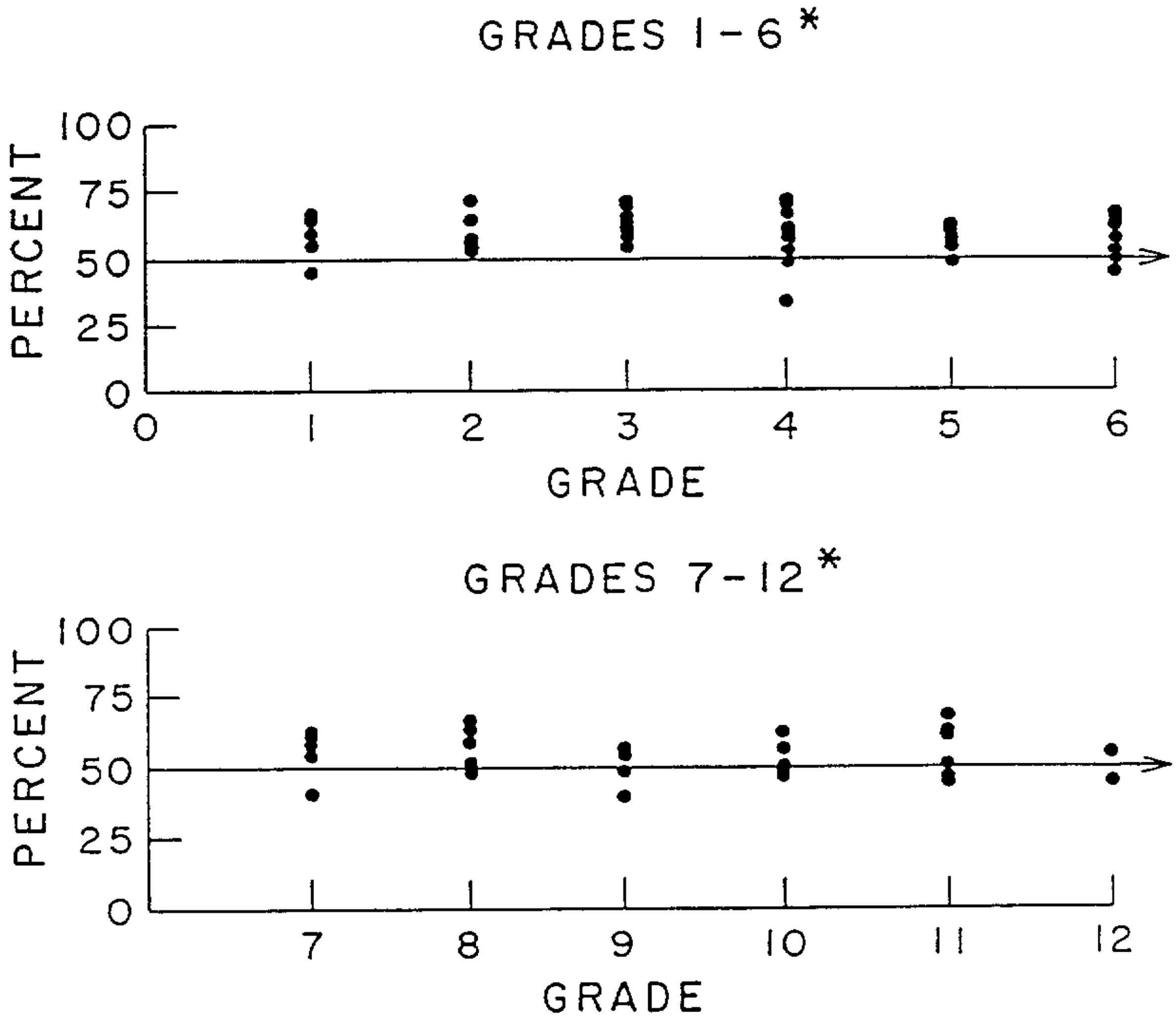


Figure 3
Percentage of Students Reported by States to be Scoring above the National Median in Mathematics (Each Point Represents a State)

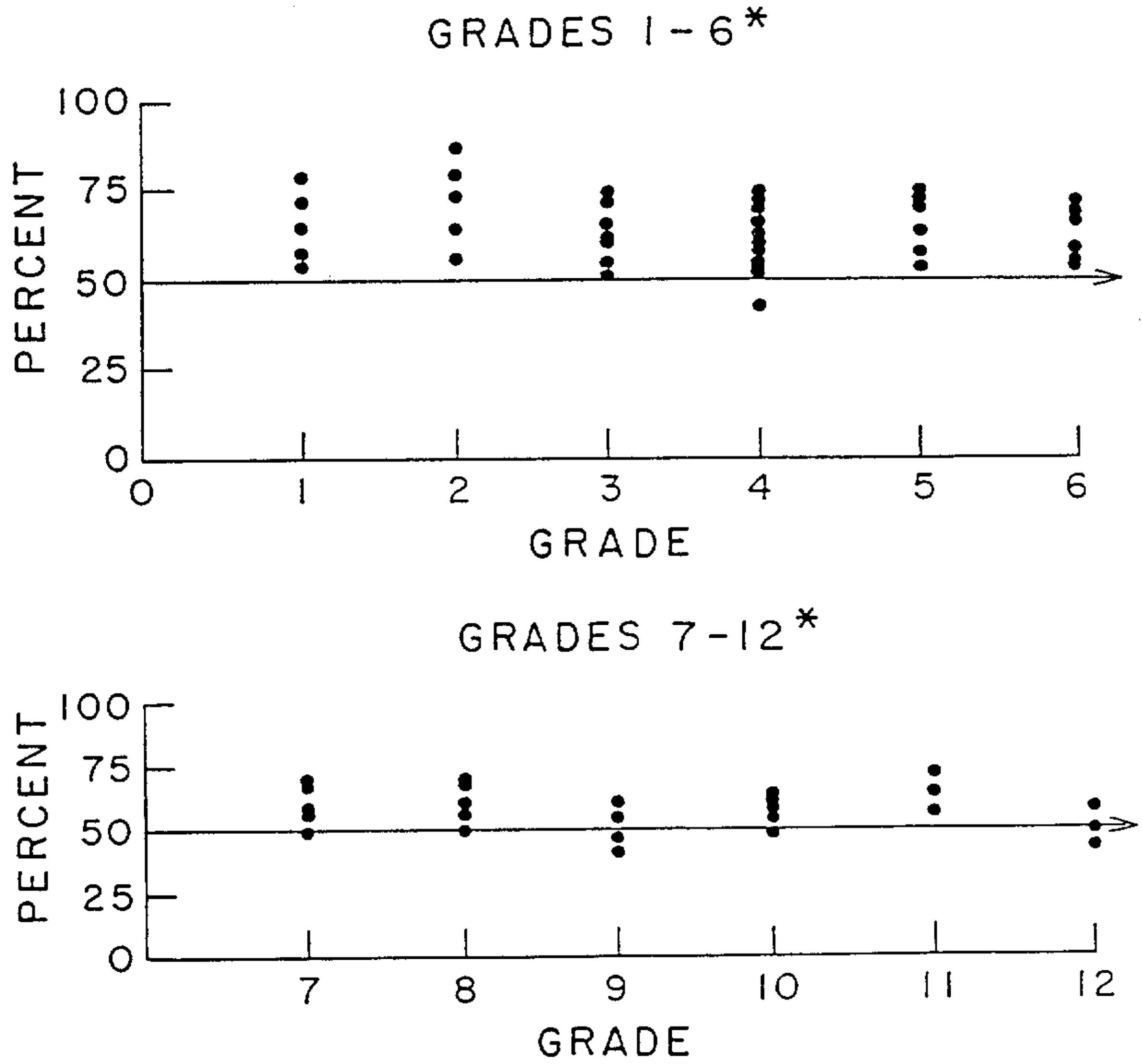


Figure 4
Weighted Mean of State Percentile Ranks

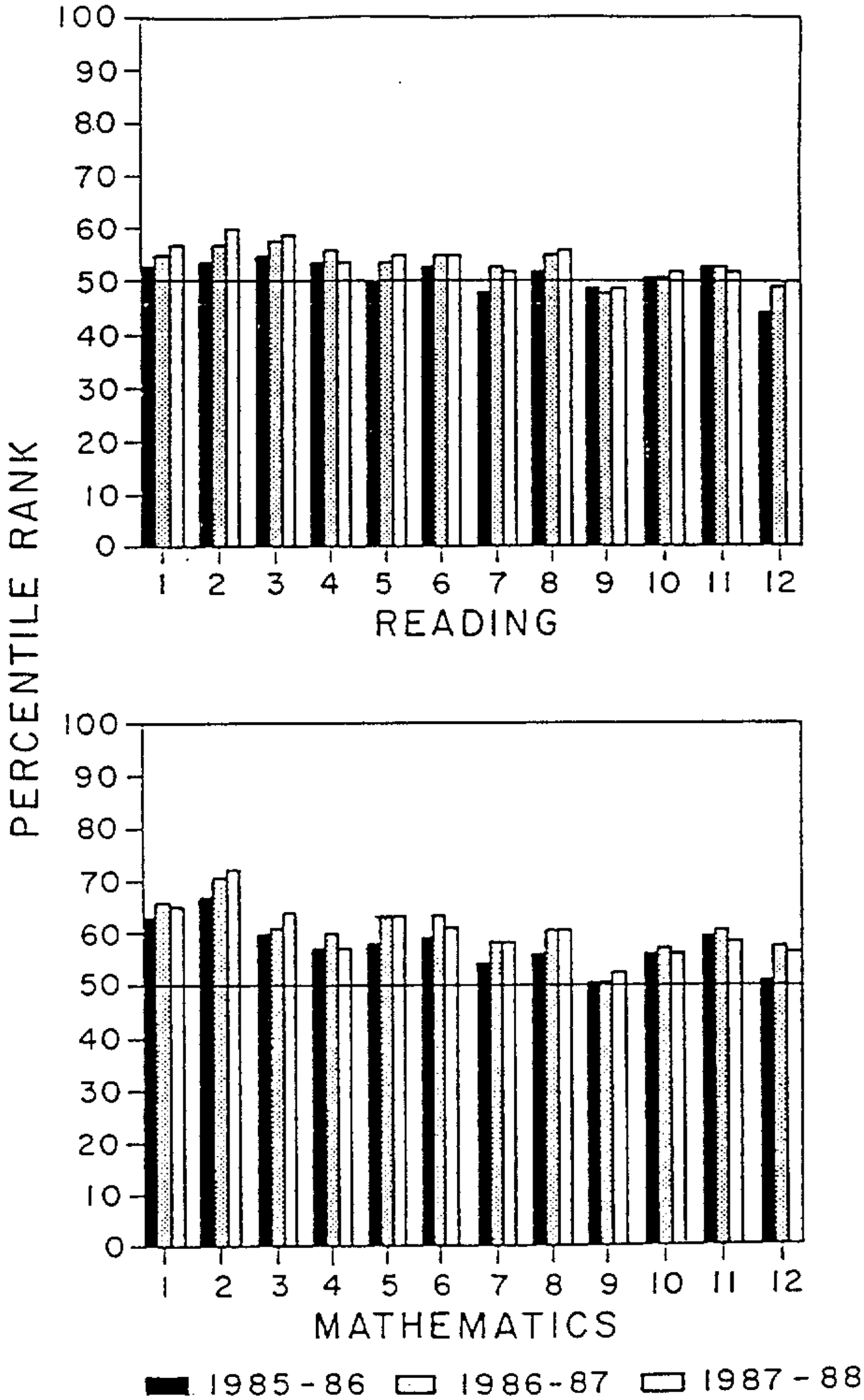


Figure 5
State Median Percentile Rank or Percentile Rank of
State Mean Test Score in Reading (Each Point Represents a State)

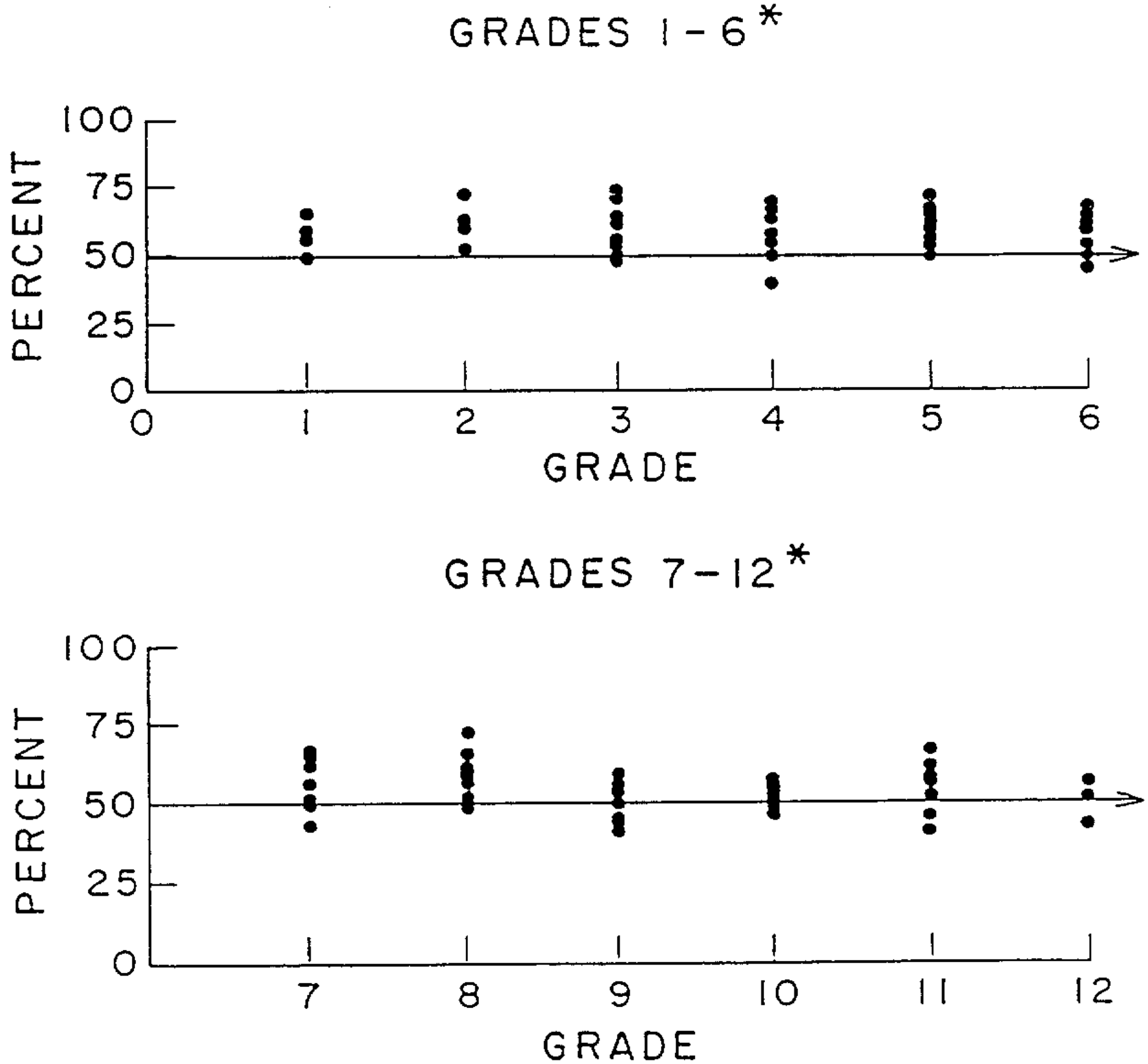
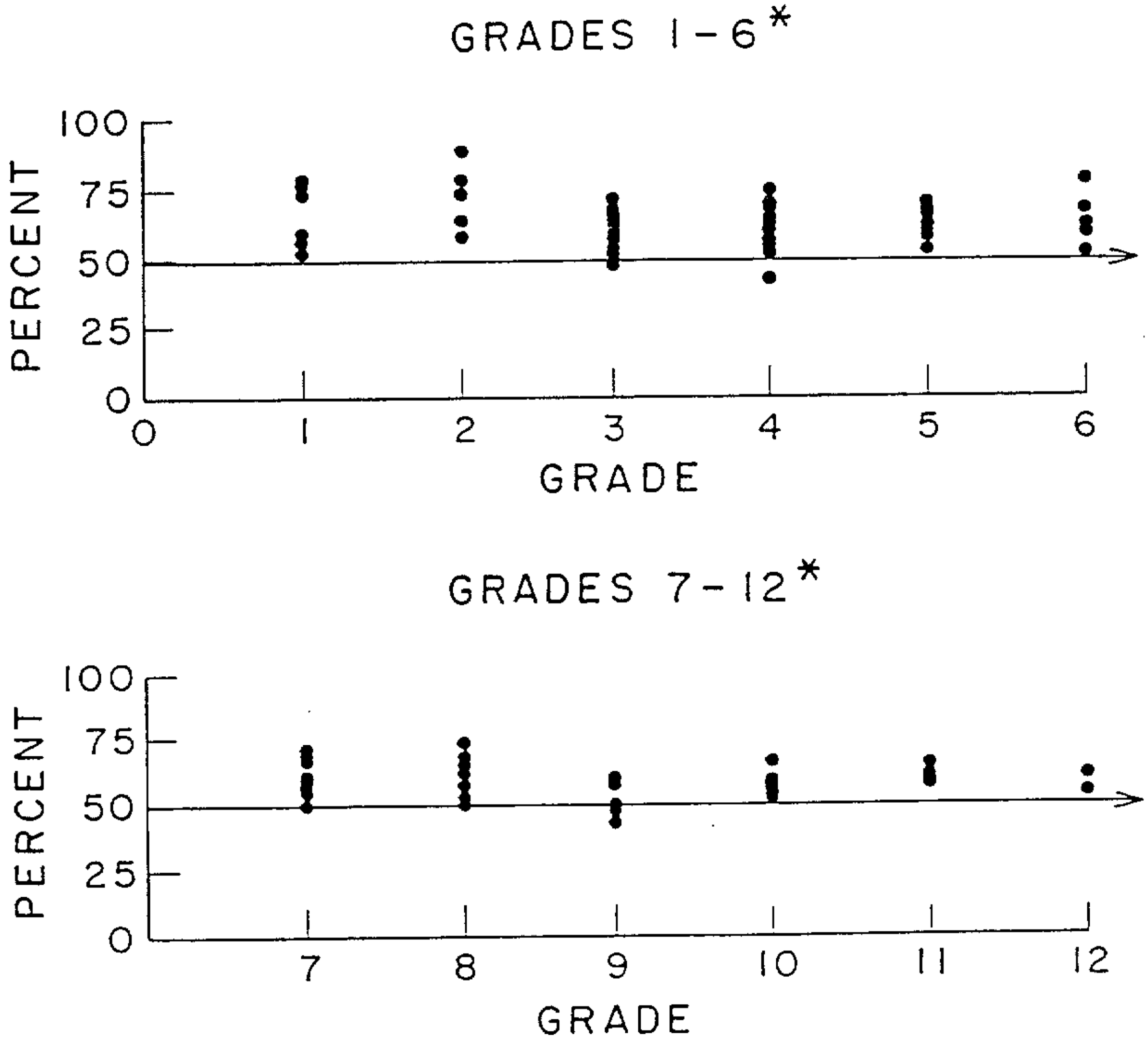


Figure 6
State Median Percentile Rank or Percentile Rank of
State Mean Test Score in Mathematics (Each Point Represents a State)



The conclusions (a) that most states are reporting results above the national average, (b) that the discrepancy is greater in mathematics than in reading, and (c) that the discrepancy is generally greater in the elementary grades than in the secondary grades do not depend on the use of a particular metric (e.g., the percentage of students above the national median). The same conclusions are supported by the use of the median percentile rank for each state or the percentile rank of the state mean.

Normative Comparisons Based on District Results

Data were obtained from 153 districts, or 87%, of the target of 175 districts. Appendix F provides a listing of the region, size, and SES of each of the 153 districts that returned questionnaires, provided reports on their testing programs, or completed telephone interviews. Districtwide norm-referenced test results were available for 148 of the 153 districts. For the remaining 5 districts, districtwide normative comparisons could not be obtained for the reasons indicated in Appendix F (e.g., only criterion-referenced results were available).

Also shown in Appendix F are the grades where norm-referenced test results were reported for each district. The grades where the largest number of districts reported norm-referenced test results are Grades 3, 4, 5, 6, and 8, in which test results were obtained for between 118 and 123 districts. As was shown in Table 2, those grades, with the exception of Grade 5, were also popular choices for statewide norm-referenced testing.

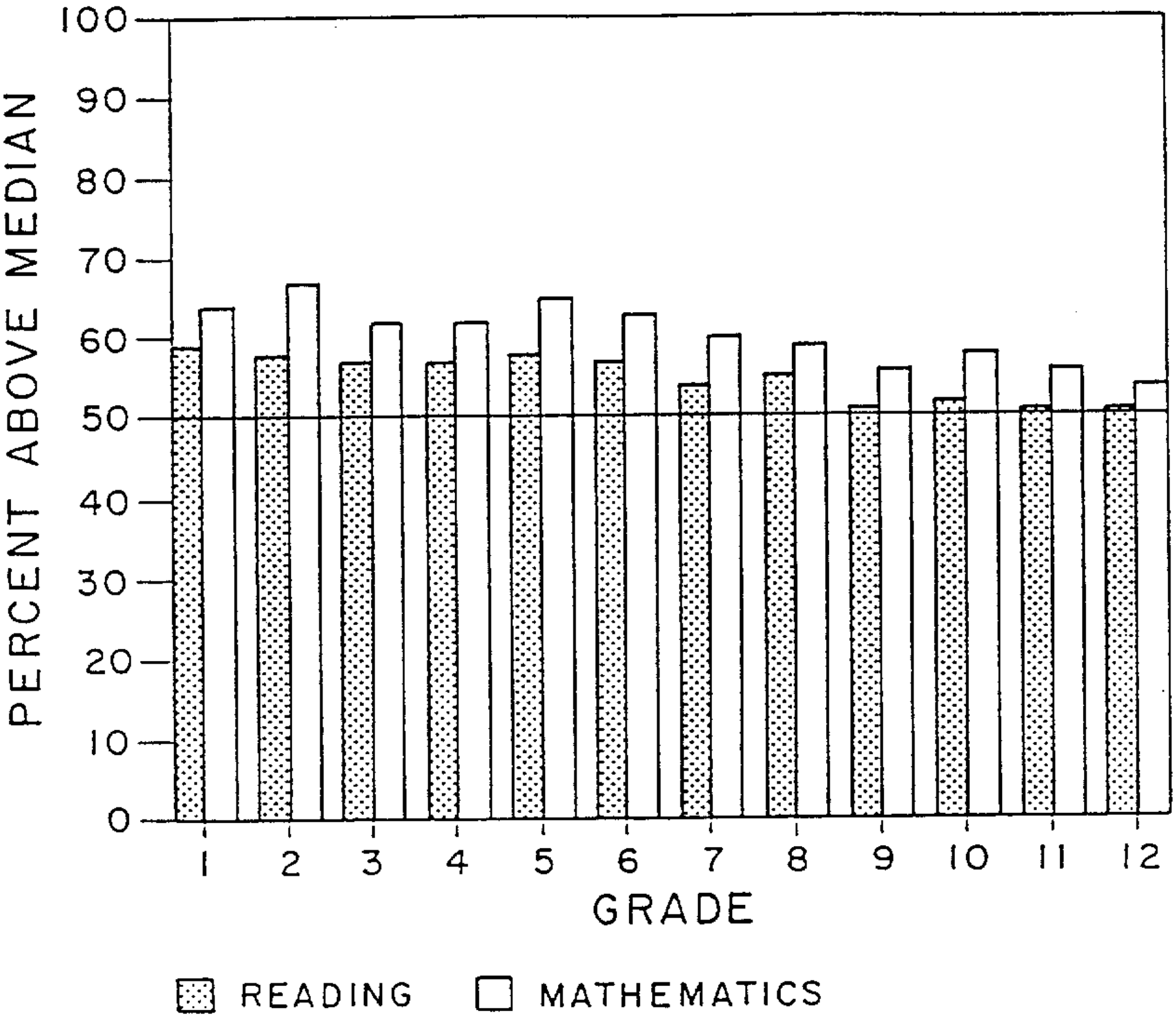
As was done for states, estimates of the percentage of students in a district who scored above the national median were obtained for each grade tested in reading and in mathematics whenever possible. Where these estimates could not be obtained, the district median percentile rank or the percentile rank corresponding to the district mean was used.

Estimates, based on the district data, of the percentage of students scoring above the national median in reading and mathematics for Grades 1 through 12 are plotted in Figure 7. The percentages plotted in Figure 7 are weighted by district size, region, and SES and thus are estimates of the percentage of students nationwide at a given grade that scored above the national median in reading or in mathematics. The number of districts on which these estimates are based varies by grade. The number of districts reporting data that could be used for the estimates in Figure 7 was 57, 77, 89, 87, 88, 85, 70, 84, 61, 52, 49, and 21 at Grades 1 through 12, respectively.

As can be seen, the estimated percentage of students scoring above the national median is consistently above 50%. For Grades 1 through 6, at least 57% of the students are estimated to have scores above the national median in reading. For mathematics, at least 62% of students are estimated to be above the national median Grades 1 through 6. In Grades 9 thru 12 the estimates of 51% or 52% for reading are closer to 50%; however, with the exception of Grade 12 with an estimate of 54%, the percentage of students estimated to have scores above the national median in mathematics is 56% or higher in every grade. Although 56% is obviously greater than 50%, it is still the case that nearly half the students (44%) received score reports below the national median when 56% scored above the median.

Figure 8 presents results that are parallel to those in Figure 7, that is, based on the data from districts where estimates of median percentile ranks or the percentile ranks of the district means were obtained. The weighted means of these percentile rank statistics are based on substantially fewer districts at each grade (the number of districts equaled 17, 27, 34, 29, 31, 27, 26, 29, 15, 16, 15, and 4 at Grades 1 through 12, respectively). Nonetheless, the results in Figure 8 lead to conclusions

Figure 7
Estimated Percentage of Students Scoring Above National Median
Based on District Results Weighted by Region, District Size and SES



that are essentially the same as those based on the estimated percentage of students above the national median. With the exception of Grade 12, where the number of districts reporting data in this form was extremely small, all of the weighted means are greater than 50. The results for the elementary grades are higher than those for the upper grades and the results for mathematics are higher than those for reading.

In addition to providing overall estimates of student performance levels, the district results provide a basis for investigating between-district variability and characteristics of districts associated with level of performance. Estimates of the percentage of students who scored above the national median in reading and mathematics were obtained for a majority of the districts that returned test results. Distributions of these percentages for districts were inspected at each grade level in both content areas. Since the complete distributions for all grades are rather voluminous, distributions for only one grade are presented and discussed in detail. Summaries of the distributions for other grades are provided and complete distributions for Grades 1 through 12 are included in Appendix G. Grade 3 was chosen for illustrative purposes since it was the earliest of the grades that were most frequently tested and reported by districts in the sample.

A total of 123 districts reported norm-referenced test results for Grade 3. Eighty-nine of those districts provided data that could be used to estimate the percentage of students scoring above the national median in reading and mathematics. The remaining districts reported data that could be used to obtain the median percentile rank or the percentile rank of the district mean, but did not provide a basis for obtaining the percentage of students scoring above the national median.

Distributions of district percentages of students scoring above the national median are illustrated by the stem-and-leaf plots in Figure 9. The "stem" corresponds to the tens digit of the percentage of students in a particular district that scored above the national median. The "leaf" reports the units digit for a district's percentage. The results for each district are depicted by a leaf (i.e., a single digit under the leaf column), that is associated with a particular stem which gives the tens digit for each leaf in that row. For example, one district reported that 93% of its students scored above the national median in reading and one district reported that 94% of its students scored above the median. Those two districts are depicted in the upper-left-hand corner of Figure 9 by the 34 under the leaf column next to a stem of 9. The lowest percentage above the median for reading that was reported by a district was 15%. The results for that district are indicated by the leaf of 5 next to a stem of 1 toward the bottom of the stem-and-leaf diagram for reading.

As can be seen in Figure 9, a majority of the districts reported that 50% or more of their students scored above the national median in reading (61 of 89 districts) and mathematics (69 of 89 districts). Only 16 of the 89 districts reported that less than 40% of their students scored above the national median in reading, but there were 12 districts that reported that three-fourths or more of their students scored above the national median. In mathematics the results show even larger numbers of districts that reported a substantial majority of their students above the median.

In order to summarize the distributions of district percentages of students reported to have scored above the national median, the 10th, 25th, 50th, 75th, and 90th percentiles of the distributions were obtained. For Grade 3, those percentiles are reported at the bottom of the two columns of Figure 9. (Parallel results for the other grades are presented in Appendix G.) These figures indicate, for example, that 10% of the districts reported that 32% or fewer of their third-grade students scored above the national median in reading. On the other hand, the 90th

Figure 8
Means of District Percentile Ranks Weighted by Region, District Size, and SES

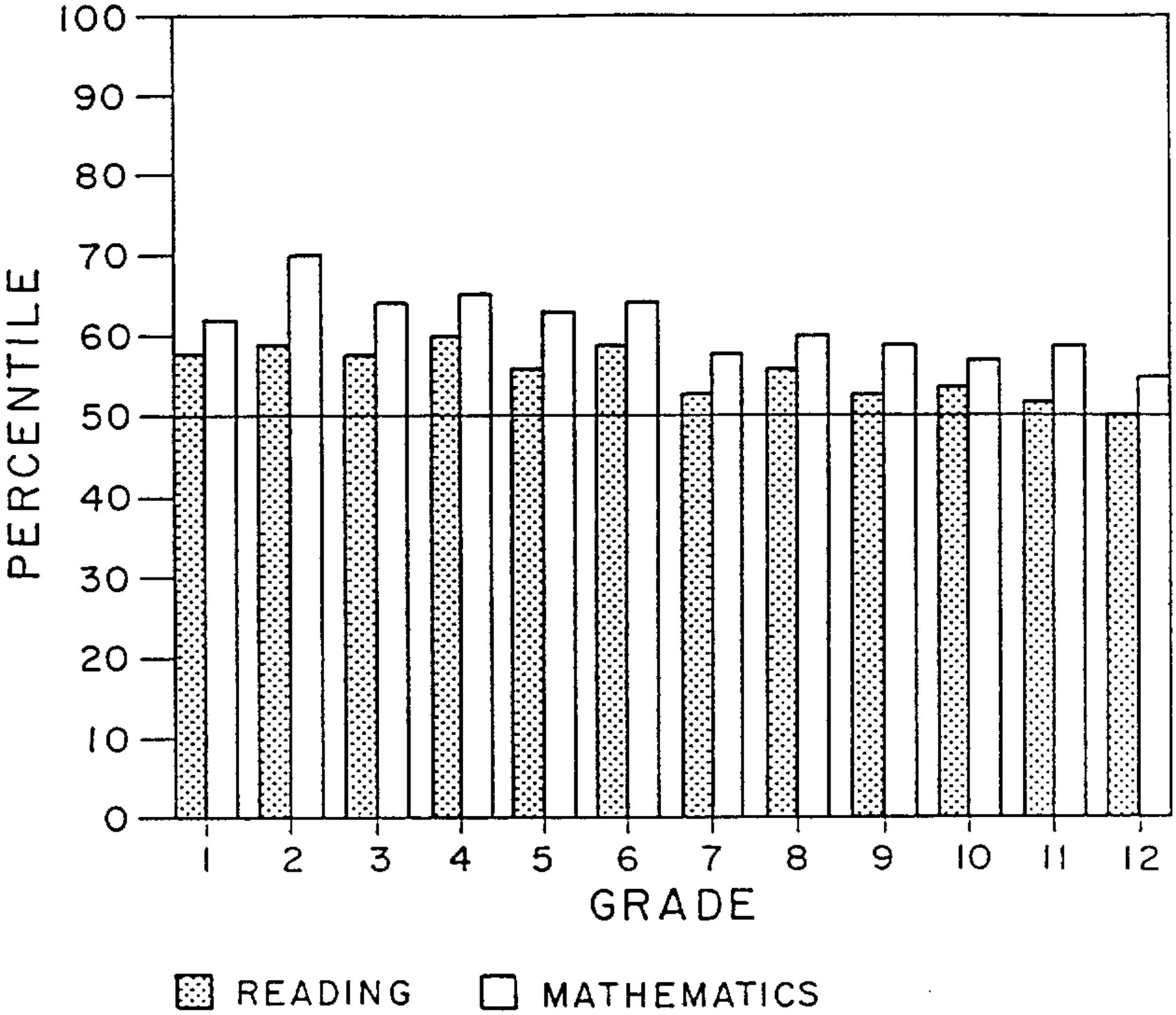


Figure 9
Stem-and-Leaf Distribution of the District Percentages of Students Scoring above the National Median at Grade 3

Reading

Stem	Leaf	Count
9 :		0
9 :	34	2
8 :	558	3
8 :	12	2
7 :	56799	5
7 :	0122344	7
6 :	677777789	9
6 :	00111224444444	14
5 :	5566677899	10
5 :	001233344	9
4 :	556889	6
4 :	001223	6
3 :	69	2
3 :	012223344	9
2 :	89	2
2 :	14	2
1 :	5	1
1 :		0

P90 = 78
P75 = 67
P50 = 58
P25 = 45
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :	123	3
8 :	7899	4
8 :	012224	6
7 :	88	2
7 :	000112244	9
6 :	556778888899	12
6 :	000123344444	12
5 :	55567788999	11
5 :	1222333444	10
4 :	556667899	9
4 :	0224	4
3 :	69	2
3 :	334	3
2 :		0
2 :	0	1
1 :		0
1 :	1	1

P90 = 82
P75 = 70
P50 = 61
P25 = 52
P10 = 42

percentile of 78 indicates that 10% of the districts reported that over three-fourths of their third-grade students scored above the national median in reading.

The five selected percentiles (10th, 25th, 50th, 75th, and 90th) of the district distributions of the percentage of students scoring above the national median were computed for all twelve grades. Those percentiles are shown in the box-and-whisker plots displayed in Figures 10 and 11 for reading and mathematics, respectively. Looking, for example, at the Grade 1 box-and-whisker plot for reading in Figure 10, it can be seen that the 10th percentile for the 57 districts reporting data at Grade 1 was 35, indicating that 1 district in 10 reported that 35% or less of its students scored above the national median. From the remaining percentiles for the Grade 1 reading results it can be seen that one district in four reported 45% or less of its students scored above the national median, half the districts reported 55% or less, three districts in four reported 66% or less, and nine districts in ten reported 81% or less.

From an inspection of Figure 10, it can be seen that districts at the 50th percentile reported that more than half (54% to 58%) of their students scored above the national median in reading in Grades 1 thru 8. Only at Grade 10 did a district at the 50th percentile report that slightly less than half (48%) of its students scored above the national median in reading. For the elementary grades, the tendency to have more than half of the students in a district scoring above the national median is much stronger in mathematics (Figure 11) than in reading (Figure 10). In Grades 1 thru 6, for example, the 25th percentile is equal to or above 50. In other words, three-quarters of the districts had more than half their students scoring above the median. Moreover, half the districts had 59% or more of their students above the national median in mathematics for Grades 1 thru 8.

The percentage of districts that had more than half of their students scoring above the national median should not be interpreted as a direct indication of the percentage of students across districts who were scoring above the median. It would be possible, for example, for a substantial majority of districts to have more than half their students above the median while less than half of all students across districts were above the median. Nonetheless, it is clear that it is more common for a district to report test results that are "above average" than ones that are "below average."

The district results provide support for the general finding that it is more common to have students scoring above the national median than it is to have them scoring below the median. However, there are more exceptions to this rule, particularly in reading, than were suggested by the Cannell study, which reported that 169 of 188 districts were "above average." Five districts refused to provide the information and only 14 districts were classified as "below average" in the Cannell study.

Cannell's results were based on a telephone survey of the largest districts in the sixteen states where statewide results were unavailable. Districts were "asked if their elementary (1-6) total battery scores were above, at, or below the national average" (Cannell, 1987, p. 22). A district was called above average if four of six grades were above the national norm, and scores on reading, language, and math were used in cases where total battery scores were unavailable.

That the frequency of districts with scores below the median suggested by Figures 10 and 11 is greater than that suggested by the Cannell results is attributable largely to the difference in definitions. For example, one district that was classified as above average based on the Cannell study reported that for Grades 2 through 6 the percentages of students scoring above the national median in reading during the 1986-87 school year were 56, 47, 35, 44, and 48, respectively. While this district would appear to be "below average" based on these reading test results, it would

Figure 10
Box-and-Whisker Plots Showing the Percentage of Students Reported to be Above the National Median in Reading by Grade for Districts at the 10th, 25th, 50th, 75th, and 90th Percentiles for the District Distributions

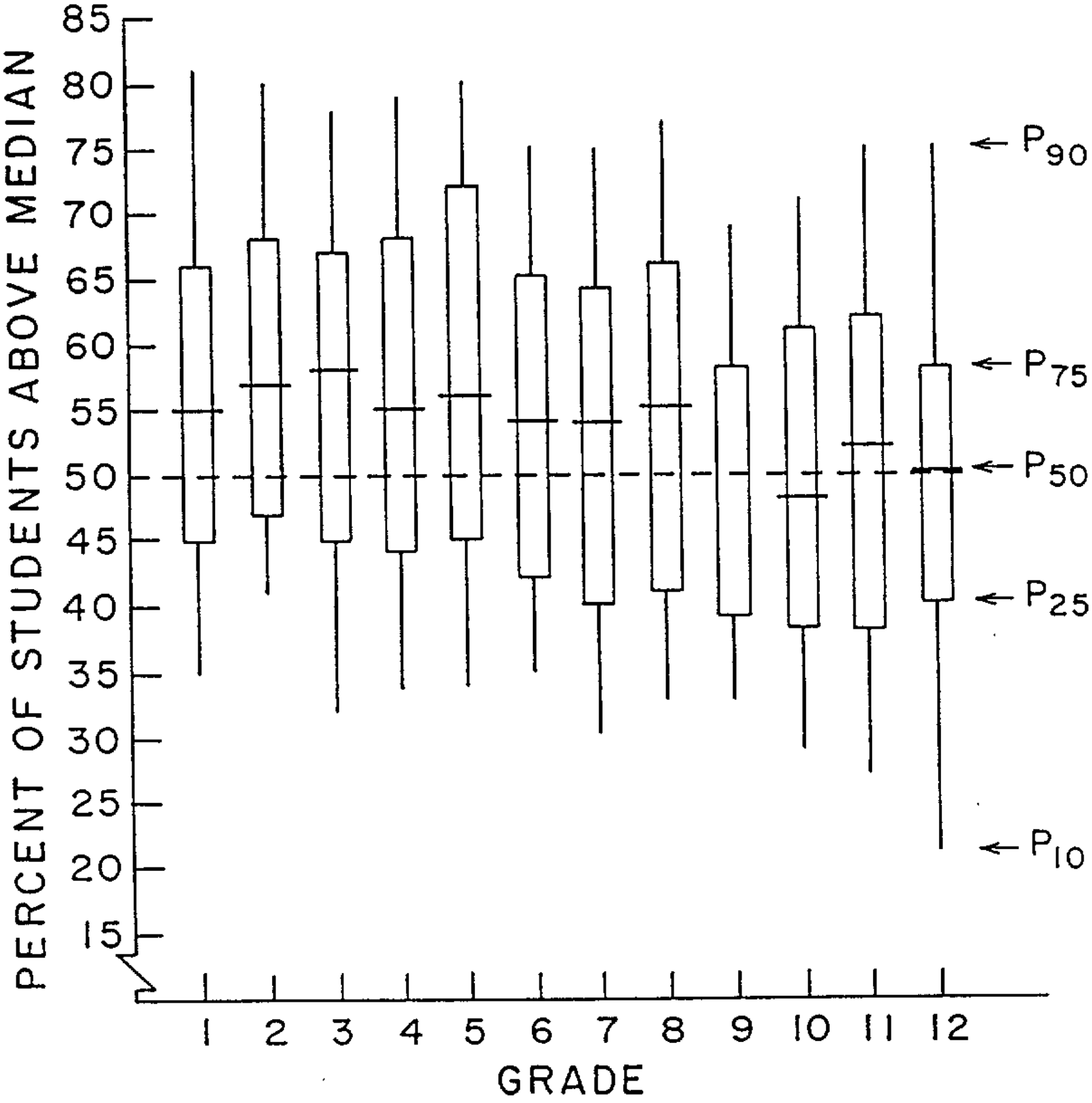
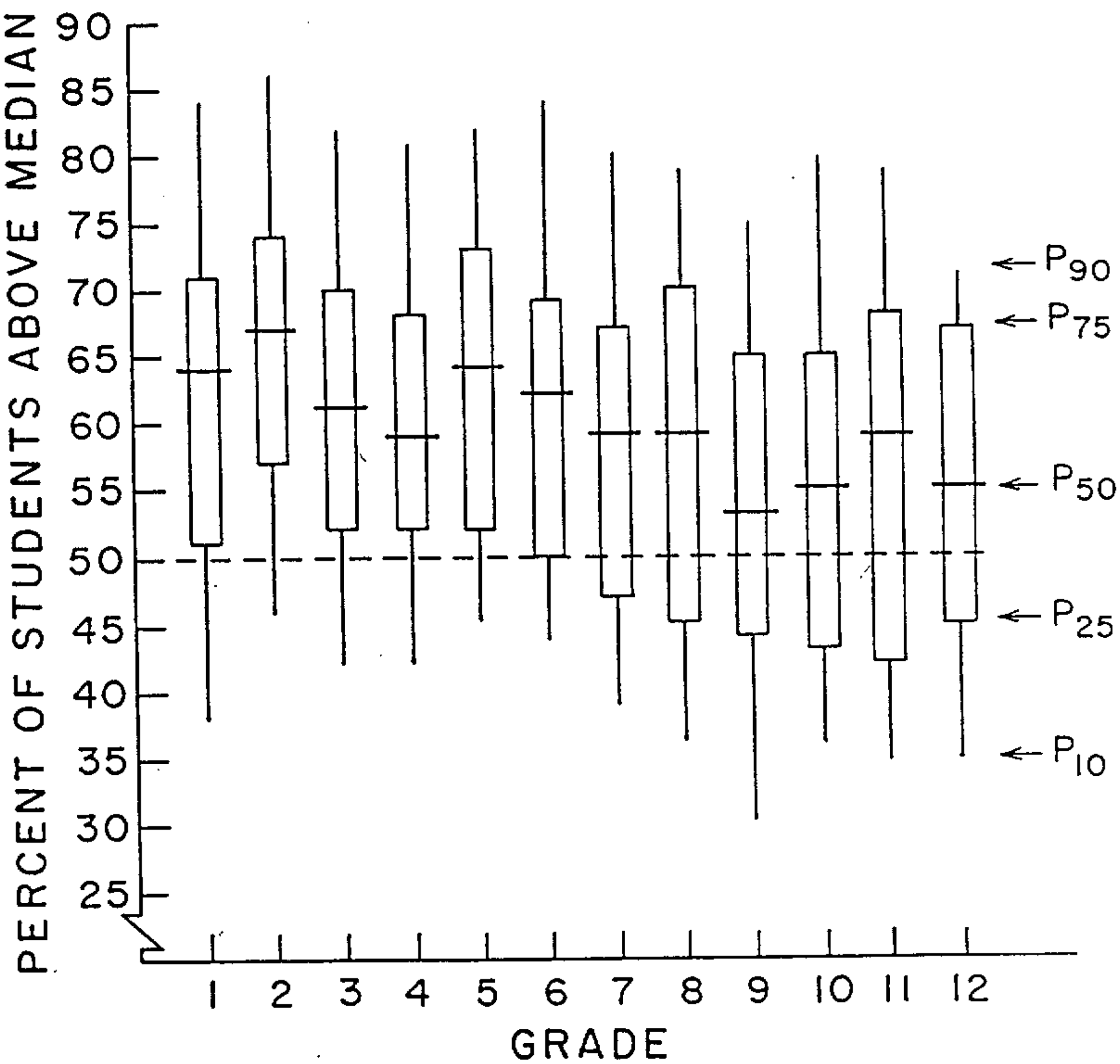


Figure 11
Box-and-Whisker Plots Showing the Percentage of Students Reported to be Above the National Median in Mathematics by Grade for Districts at the 10th, 25th, 50th, 75th, and 90th Percentiles of the District Distributions



appear to be clearly "above average" based on the corresponding percentages for mathematics (64, 64, 54, 60, and 68, for Grades 2 through 6, respectively). In general, districts reported a larger percentage of students above the national median when using total battery or mathematics scores than when using reading scores.

Summary of State and District Results

Clearly it was the exception rather than the rule for a state to report that its students, particularly its elementary school students, were performing below the national average. Although it was somewhat more common for a district than a state to report that less than half of its students were scoring above the national median, a substantial majority of districts reported that their students were performing above average (i.e., more than 50% of the students were reported to be above the national median). The tendency for students to score above the national median was especially strong in mathematics for Grades 1 thru 8. Nonetheless, it should be noted that some districts reported that substantially less than 50% of their students scored above the national median. At Grade 3, for example, 1 district in 10 reported that a third or less of its students scored above the national median in reading.

Achievement Trends and Dated Norms

Although the state and district results are generally consistent with the Cannell and earlier SREB findings which reported that achievement test results are more often above than below the national norm, they provide no real indication of the reasons that led to this result. As was discussed earlier, a wide variety of factors have been suggested as possible explanations of the apparently high test results that are being reported by states and districts. General improvement in student achievement, at least at the elementary grades, is clearly one possibility. When there are upward trends in achievement, old norms are easier (i.e., they provide a lower standard of comparison) than new norms, and thus a state or district whose students score at the current national average would score above the average defined by dated norms.

Using the aggregate results for districts, the district percentages of students scoring above the median in reading and in mathematics were related to the age of the norms used by districts at each grade (i.e., the number of years between the date of the test administration by a district and the date of the test norming by the publisher). Table 5 lists the number of districts that provided information on the year that the norms in use were obtained and the percentage of students scoring above the median for Grades 1 through 12. Also shown in Table 5 are the mean age of the norms used by districts, the mean change in the percentage of students scoring above the median for each additional year since the norms were obtained, and the estimated mean change in the percentage that resulted from the use of old norms rather than current norms.

As can be seen in Table 5, the average district that returned data was using norms that were four or five years old. Although most districts were using the most recent norms available from the publisher for the test being used, there was still an average of four or five years between the date of test administration by the district and the date of norming because publishers typically have collected norms only about every seven years. With a single exception, the percentage of students scoring above the median increased in both reading and mathematics with each additional year since the norms were obtained. The exception was for reading at Grade 10. By using norms that were four or five years old rather than current norms, assuming the latter were available, the percentage of students scoring above the median was estimated to be higher in all but Grade 10 in reading and in every grade for mathematics. For Grades 1 through 8 the expected increase ranges from

Table 5
Changes in District Percentages of Students
Above the National Median with Increasing Age of Norms

Grade	Number of Districts	Mean Age of Norms (Years)*	Mean Change in Percentage Above Median per Year		Estimated Mean Change (Old Minus Current Norms)	
			Reading	Math	Reading	Math
1	46	4.7	1.3	1.7	6	8
2	63	4.8	1.0	1.9	5	9
3	73	5.1	1.2	1.7	6	9
4	70	4.3	1.3	1.4	6	6
5	73	5.2	1.4	1.9	7	10
6	69	4.5	1.0	2.3	5	10
7	61	4.8	0.5	2.2	2	11
8	70	5.1	1.7	2.2	9	11
9	49	4.7	0.5	2.3	2	11
10	42	4.7	-0.3	1.1	-1	5
11	42	5.0	1.1	2.3	6	12
12	14	5.4	0.2	1.2	1	6

* Mean age of norms is the average number of years between the date of test administration and the date that the norms used to report district results were collected by the publisher.

2% to 9% in reading and from 6% to 11% in mathematics. Taking differences of the latter magnitude into account would largely eliminate the tendency for these districts to report results that are above the national median.

Trends over Several Years for Selected States

The district results in Table 5 show that there is a relationship between the age of norms used and the level of achievement test scores for the districts in this sample. These results are cross-sectional, and there may be a variety of other district characteristics associated with the age of norms for the test used as well as the level of student achievement. Therefore, these results do not provide a sufficient basis for concluding either that older norms are easier than newer norms or that achievement has been going up.

Figures 1 and 4, which were considered earlier, present achievement test results for three years. Neither of these figures provides a very clear indication that achievement scores went up or down during the three years for which data were collected. There is some suggestion from both of these figures that scores went up in Grades 1, 2, and 3. However, the direction of change is not only unclear at most other grades, but would be difficult to interpret in any event because the subset of states for which data were obtained changed somewhat from year to year. Furthermore, three years is too short a time interval to assess long-term trends.

Though not a specific part of the data collection design, results included in the state assessment reports for some of the states made it possible to look at trends for longer time intervals. Achievement trends for four states are summarized in Figure 12.

The upper-left-hand quadrant of Figure 12 shows a plot of the percentage of students in one state (State A) scoring above the national median in reading and mathematics at Grade 4 for each of the past six school years. During this interval a single test form of a single edition of a test was administered each year and results were based on comparisons to the 1980-81 national norms provided by the test publisher. As can be seen, the first year the test was administered, 1982-83, the percentage of students scoring above the national median was well below 50 for both reading (41%) and mathematics (44%). During each of the following five years these percentages increased, most notably in mathematics. In 1987-88, 57% of the students scored above the national median in reading and 68% scored above the national median in mathematics.

Similar results using the alternative statistic of the percentile rank in the individual pupil norms corresponding to the statewide mean test score are shown for another state (State B) in the upper-right-hand quadrant of Figure 12. As in the previous example, the results are shown for a six-year period during which a single form of a single edition of a test was administered each year. Comparisons were to norms obtained in 1978 in this case. Although the trend for State B was less steep than the one for State A and was based on a different metric, there was a clear upward trend during the six years in both reading and mathematics.

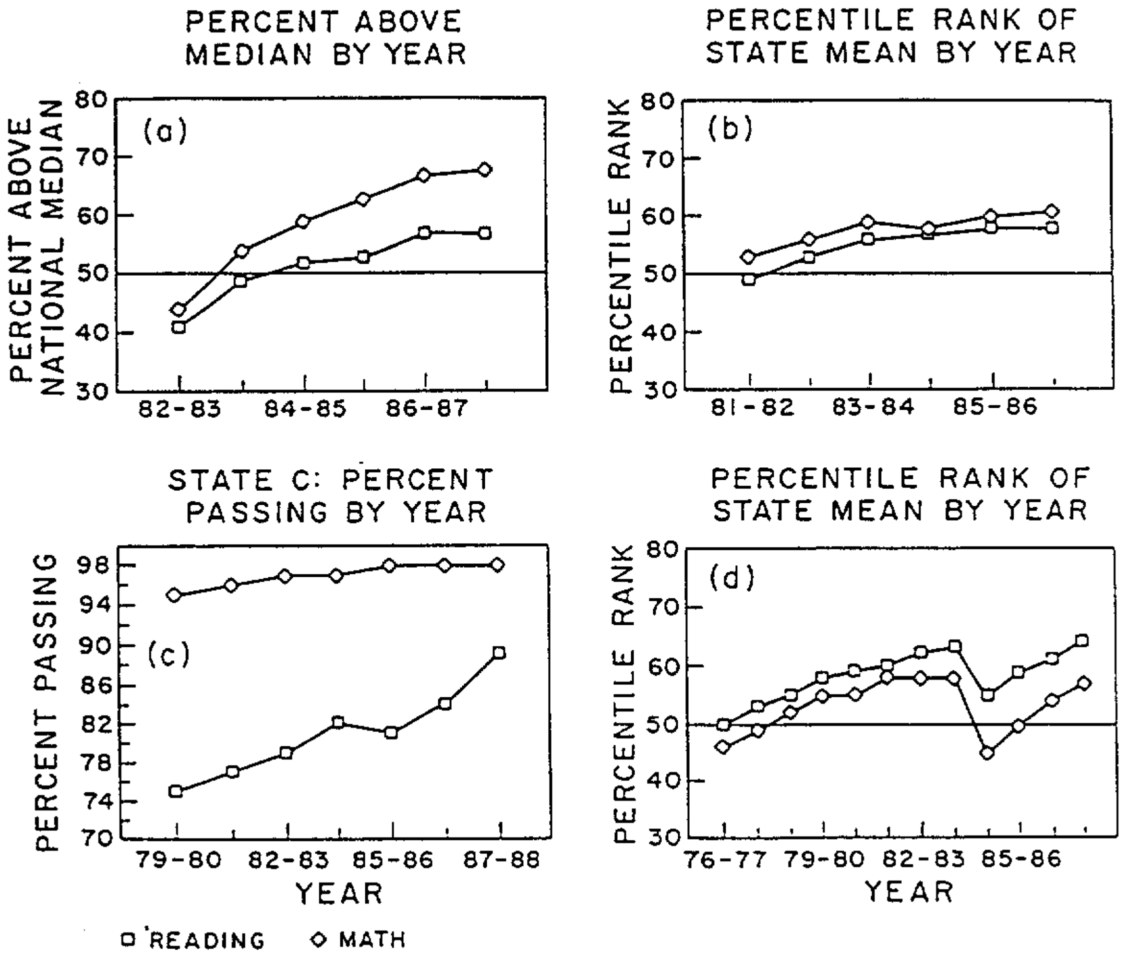
The third example, State C, shown in the lower-left-hand quadrant of Figure 12, uses an entirely different metric than has been considered so far. The plots for State C show the percentage of students passing statewide minimum-competency tests in reading and mathematics for each of seven years. In mathematics the percentage passing was 95 in the first year and gradually increased to 98% over time. For reading, where there was more room for movement, the increases between the first and most recent years of test administration were more substantial.

The final plot shown in the lower-right-hand quadrant of Figure 12 displays the percentile ranks of the state means in reading and mathematics based on individual pupil norms for Grade 3 in State D. The State D results not only span the longest time interval, twelve school years, but include a change in test editions within the period of time that was covered. A single form of a single edition of a test was used for the eight years starting in 1976-77 and running through 1983-84. The pattern for those first eight years was reasonably similar to the ones shown for the other three states in Figure 10. There was a consistent upward trend during those years.

The feature of the plot for State D that most clearly sets it apart from the plots for the other three states in Figure 12 is the sharp decline shown in percentile rank between the 1983-84 and 1984-85 school years, followed by increases over the next three years to bring 1987-88 results back to approximately where they were in 1983-84. As was previously indicated, during the 1984-85 school year the new edition of the test was introduced and the same form of that edition was administered in each of the last four years covered in the plot of results for State D. Thus the sharp decline corresponds to the introduction of the new test edition.

The sharp decline in performance relative to national norms that State D experienced when the new edition of the test was introduced is not unique. Figures 13 and 14, for example, show the results for two large school districts that introduced new editions during the 1987-88 school year. As can be seen, both districts experienced large declines in the percentage of students scoring above the national median between 1986-87 and 1987-88.

Figure 12
Trends in Reading and Mathematics Achievement for Four States



There are several possible interpretations of the trend results shown in Figures 12, 13, and 14. The most straightforward interpretation of the trends in Figure 12 is that achievement in reading and mathematics for the grades in question improved rather steadily in all four states. The dip when a new edition was introduced in State D could simply reflect general increases in student performance across the nation, which made the more recent norms associated with the newer edition more stringent than the norms associated with the older edition of the test. This same interpretation could also explain the dips in performance levels associated with a new test edition for the two districts shown in Figures 13 and 14.

An alternative interpretation of these results, however, is that increases in test scores simply reflect increasing familiarity with a given test form and more focused instruction on the content of that specific form. By administering the same form of a test for several years teachers are apt to become increasingly familiar with the specifics of the test content and alter instructional emphases to better match the content of the test. As indicated by Mehrens and Kaminski (1988) and by Shepard (1989), test familiarity might influence instruction in a wide variety of ways, ranging from practices that would generally be considered sound uses of test results (e.g., identifying and working on objectives where students show weaknesses) to those that most educators consider unethical (e.g., teaching the specific items on a test just prior to test administration).

It is not possible to distinguish whether the trends in Figures 12, 13, and 14 were due to improvements in achievement, to increased familiarity with the tests, or to some alternative explanation, solely from the results presented in those figures. However, other data can be brought to bear on the issue. In particular, the questionnaire and interview results which are discussed in other reports based on this project (e.g., Shepard, 1989) speak to some of these issues. Only the question of whether norms are changing in difficulty with time as a result of increases in student achievement nationally will be considered here.

Achievement Trends and Changes in the Difficulty of Norms

National changes in achievement levels obviously lead to differences in the meaning of norms. During a period of declining performance such as the nation experienced in the 1960s and the first part of the 1970s (Harnischfeger & Wiley, 1975; Koretz, 1986; 1987), newer norms provide a less stringent standard of comparison than older norms. Koretz (1987), for example, estimated that during the period of the much publicized test score decline (roughly the early or mid 1960s to the mid 1970s) "the average decline in grades six and above was large enough that the typical (median) student at the end of the decline exhibited the same level of achievement as was shown before the decline by students at the 38th percentile" (p. 2). Thus a state or district using old norms in the mid 1970s could have appeared to be well below the national average when in fact their students were scoring at the then current national average. On the other hand, when performance on achievement tests is increasing, newer norms become harder and the use of old norms can make a state or district that would have only average or below average scores in terms of current national norms appear to be above average. Clearly, national trends in achievement tests scores have importance for understanding normative comparisons.

Although increases in test performance have not received as much attention as the decline of the 1960s and 1970s, several sources of evidence suggest that achievement test scores have been going up. National Assessment of Educational Progress (NAEP) reports (e.g., Dossey, Mullis, Lindquist, & Chambers, 1988; NAEP, 1985) indicated that there were some increases in reading and mathematics between the early or mid 1970s to the mid 1980s. Based on his review of NAEP and data from several other tests, Koretz (1987) concluded that the decline in test scores ended

Figure 13
Percentage of Students Above National Median for District A
Before and After a Change of Test Editions (New Edition in 1987-88)

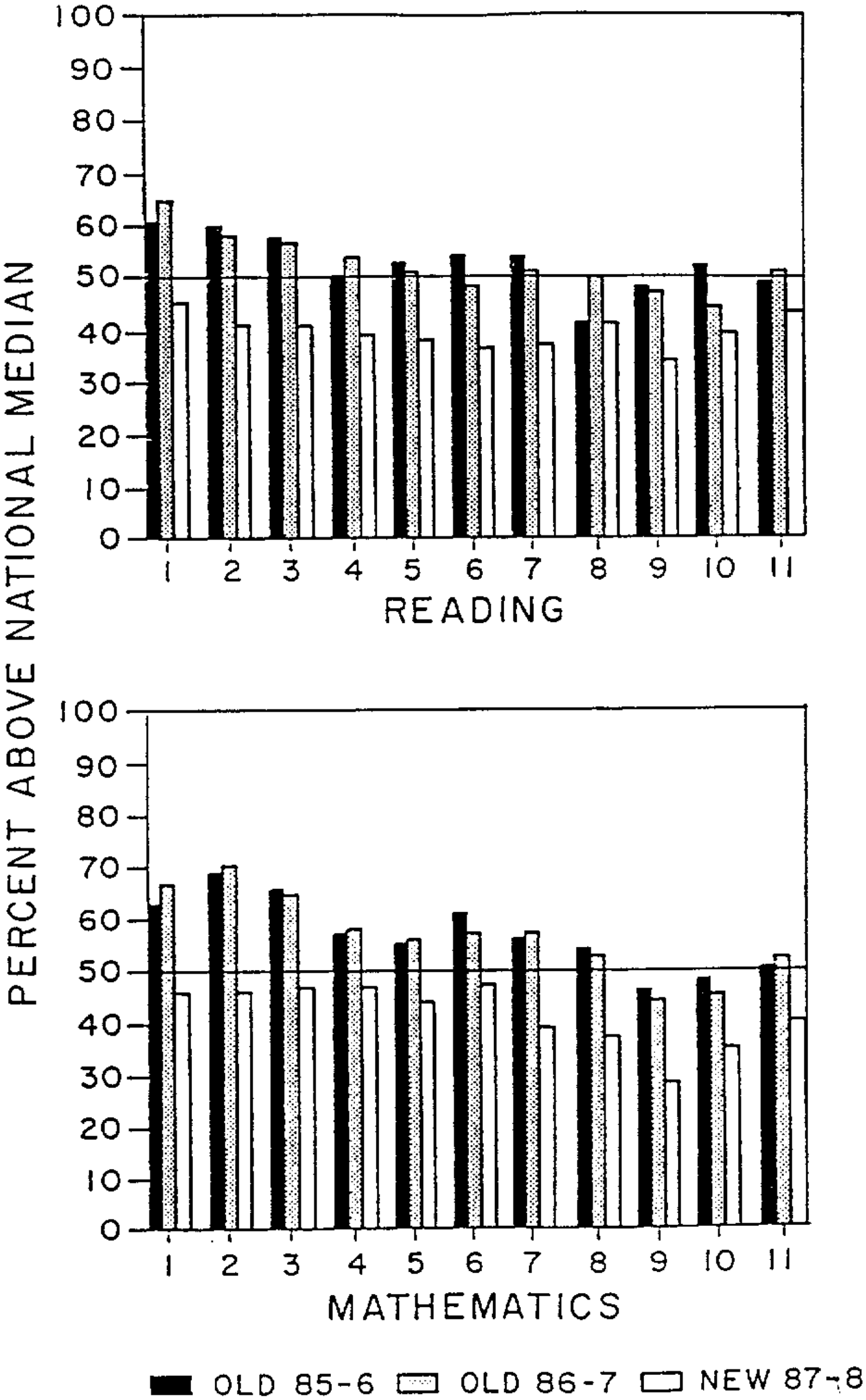
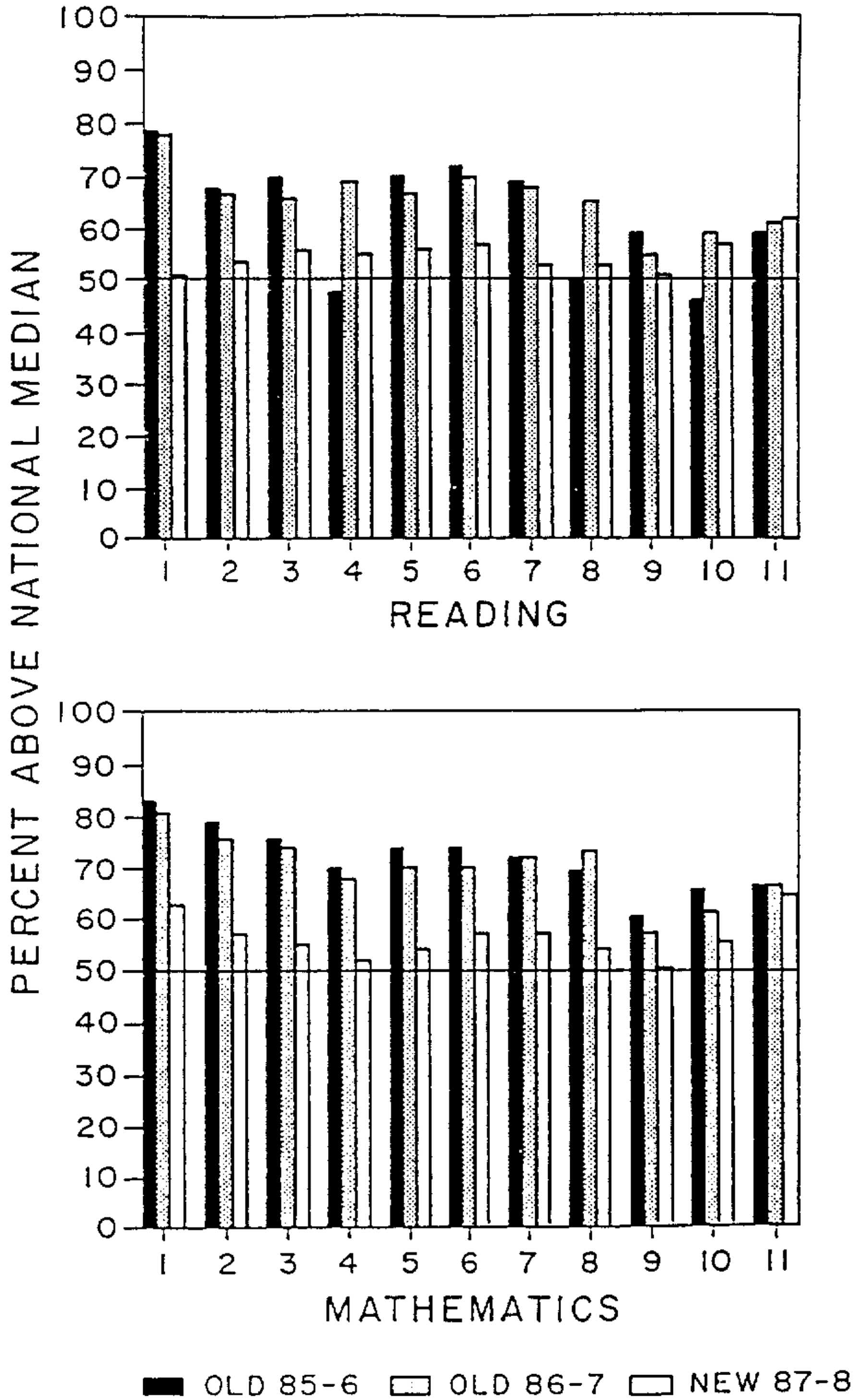


Figure 14
Percentage of Students Above National Median for District B
Before and After a Change of Test Editions (New Edition in 1987-88)



with cohorts of students that entered school in the late 1960s and that subsequent cohorts of students "produced a sharp rise in scores on most, but not all, tests. In the majority of instances in which scores increased, the rise has been steady—with each cohort tending to outscore the preceding one—and often roughly as fast as the decline" (p. 2).

Norming studies conducted periodically for standardized tests also provide evidence regarding trends in national achievement. When a new edition of a standardized test is introduced, it is customary not only to collect new normative data for the new edition but also to equate the old and new editions of the test. The equatings make it possible to estimate the extent to which achievement has increased or decreased over the years between the norming of the two editions. In some cases, new norms are collected for a previously normed edition of a test, which again provides a means of comparing national performance on the test at two points in time.

Several test publishers reported increases in achievement based on the results of their norming studies. CTB/McGraw-Hill (1987), for example, noted when the norms for Form E of the California Achievement Tests (CAT) were reported and compared to the norms for the CAT Form C to which Form E was equated that "the CAT E norms are more difficult than the CAT C norms. This seems to indicate that students in 1984-85 were achieving at a higher level than in 1977, when CAT C was normed" (p. 3-4). Increases in performance were reported when Form G of the Iowa Tests of Basic Skills (ITBS) was published. "Between 1977-78 and 1984-85, the improvement in ITBS test performance more than made up for previous losses in most test areas. Composite achievement in 1984-85 was at an all time high in nearly all test areas" (Hieronymus & Hoover, 1986, p. 148). Increases in performance have also been reported for the Stanford Achievement Test (SAT7) (Wiser & Lenke, 1987) and the Comprehensive Tests of Basic Skills (CTBS) (Rothman, 1988) and increases can be inferred from comparisons of the norms for the Metropolitan Achievement Tests (MAT6) (Psychological Corporation, 1988) and norms for equivalent scores on the previous edition of the MAT (Prescott, Balow, Hogan, & Farr, 1978; 1986).

Table 6 provides a summary of the changes in the percentile rank of achievement test scores that were at the national median at one of the two times that norms were obtained for the six most used standardized achievement tests. The numbers are estimates of the changes in national percentile rank in reading and mathematics between the two norming years indicated at the head of each column of the table. Also shown for comparative purposes are estimated changes in national percentile ranks based on NAEP.

As is indicated in the footnotes to Table 6, the numbers in each column of Table 6 are derived from different sources and involve different types of comparisons. In the case of the CTBS, the comparison is between 1981 norms and estimates of 1987 norms for the same test form based upon a weighting of user data. The Stanford results are based on 1981-82 and 1986 norming studies for the same test form. The other published test comparisons involve norming studies for successive editions of the test battery. However, the numbers in Table 6 all have a similar interpretation. A positive number indicates that performance was higher when measured at the more recent of the two norming years indicated at the top of each column. For example, the number 14 shown for reading achievement on the California Achievement Tests (CAT) in Grade 2 indicates that an equated Form C or Form E score that would have placed a student at the national 50th percentile using the 1977 Form C norms would lead to a national percentile rank of only 36 using the 1984-85 Form E norms. The 14 shown in Table 6 is the difference between the percentile ranks of 50 in 1977 and 36 in 1984-85.

Table 6
Estimated Changes in National Percentile Rank of Achievement Scores at the National Median at One Point in Time

I. Reading Achievement

Source/Years Being Compared*

Grade	CAT[1] 77 to 84-5	CTBS[2] 81 to 87	ITBS[3] 77-8 to 84-5	MAT[4] 77-8 to 84-5	SRA[5] 78 to 83-4	Stanford[6] 81-2 to 86	NAEP[7] 74-5 to 83-4
1	28	7	9	20	-3	11	
2	14	10	12	5	1	4	
3	12	2	11	13	1	6	3
4	11	8	12	5	-1	2	
5	14	5	11	7	2	2	
6	11	8	12	6	-3	2	
7	16	6	11	9	-2	2	0
8	11	5	10	7	-4	1	
9	15			9	2	3	
10	8			-5	2	0	
11	4			-3	-2	4	2
12	1			-5	-7	3	

II. Mathematics Achievement

Grade	77 to 84-5	81 to 87	77-8 to 84-5	77-8 to 84-5	78 to 83-4	81-2 to 86	77-8 to 85-6
1	16	18	3	12	10	15	
2	14	22	5	9	3	10	
3	13	13	5	15	-6	9	4
4	11	14	9	7	-2	8	
5	13	17	8	11	3	8	
6	13	17	8	10	0	7	
7	15	15	10	2	1	6	5
8	18	11	10	5	0	7	
9	14			0	1	4	
10	8			4	4	4	
11	5			7	-2	4	0
12	2			6	-4	5	

*Footnotes on following page

Footnotes for Table 6

[1] Differences in California Achievement Tests (CAT), Form E (1984-85 norms) percentile ranks and corresponding CAT, Form C (1977 norms) percentile ranks of 50 (CTB/McGraw-Hill, 1987, Table 38, p. 3-35).

[2] Differences in Comprehensive Tests of Basic Skills (CTBS), Form U percentile ranks in 1981 and those required to have a percentile rank of 50 on the CTBS in 1987 (based on November, 1988, CTB-McGraw-Hill press release, "CTB/McGraw-Hill Studies Show Students Achieving at Higher Levels in Basic Skills", see also, Rothman, 1988, p. 20). The 1987 norms are estimated from weighted user data.

[3] Differences in Iowa Tests of Basic Skills (ITBS), Form G (1984- 85 norms) percentile ranks and corresponding ITBS, Form 7 (1977-78 norms) percentile ranks of 50 (Hieronymus & Hoover, 1986, Table 6.31, p. 153).

[4] Differences in Metropolitan Achievement Tests (MAT6), Survey Forms L and M (1984-5 norms) and corresponding MAT, Forms J and K (1977-78 norms) percentile ranks of 50 (Psychological Corporation, 1988; Prescott, Balow, Hogan, & Farr, 1978; 1986).

[5] Differences in SRA Achievement Series, Forms 1 and 2 (1983-84 norms) percentile ranks and corresponding SRA Achievement Series Forms 1 and 2 (1978 norms) percentile ranks of 50 (Science Research Associates, 1979; 1986).

[6] Differences in Stanford 7 Plus (1986 norms) percentile ranks and corresponding Stanford Early School Achievement Test, 2nd edition; Stanford Achievement Test, 7th edition, and Stanford Test of Academic Skills (TASK), 2nd edition (1981-82 norms) percentile ranks of 50 (Gardner, Madden, Rudner, Karlsen, Merwin, Callis, & Collins, 1983; 1987).

[7] Differences for the National Assessment of Educational Progress (NAEP) are based on age (9, 13, and 17) rather than grade (3, 7, and 11) cohorts. For reading, the differences are between the 1983-84 assessment percentile ranks and the corresponding 1974-74 assessment percentile rank of 50 (NAEP, 1985). For math, the differences are between the 1985-86 assessment percentile ranks and the corresponding 1977-78 percentile rank of 50 (NAEP, 1988; frequency distributions provided by Beaton).

With the exception of the SRA Achievement Series, the differences for Grades 1 thru 8 are all positive, indicating that more recent norms are more stringent than older norms for five of the six tests. For Grades 10 through 12 the differences are generally smaller than those shown for the earlier grades and two of the four tests with results for the high school grades have some differences that are negative, indicating a decline in performance and therefore easier recent norms in those instances.

The changes in percentile ranks shown in Table 6 are based on various time intervals between norming studies. More direct comparison can be made by dividing the changes in percentile ranks in Table 5 by the number of years between the norming studies to obtain estimates of yearly changes in percentile ranks. Such yearly changes in percentile ranks for Grades 1 thru 8 are presented graphically in Figures 15 and 16 for reading and mathematics, respectively.

In general, the results in Figures 15 and 16 are fairly consistent with those based on the analyses of the district data that were reported in Table 5. The estimates of yearly changes derived from the district data are greater than those

Figure 15
Estimated Yearly Changes in Reading Percentile Rank:
Publisher Results at the Median

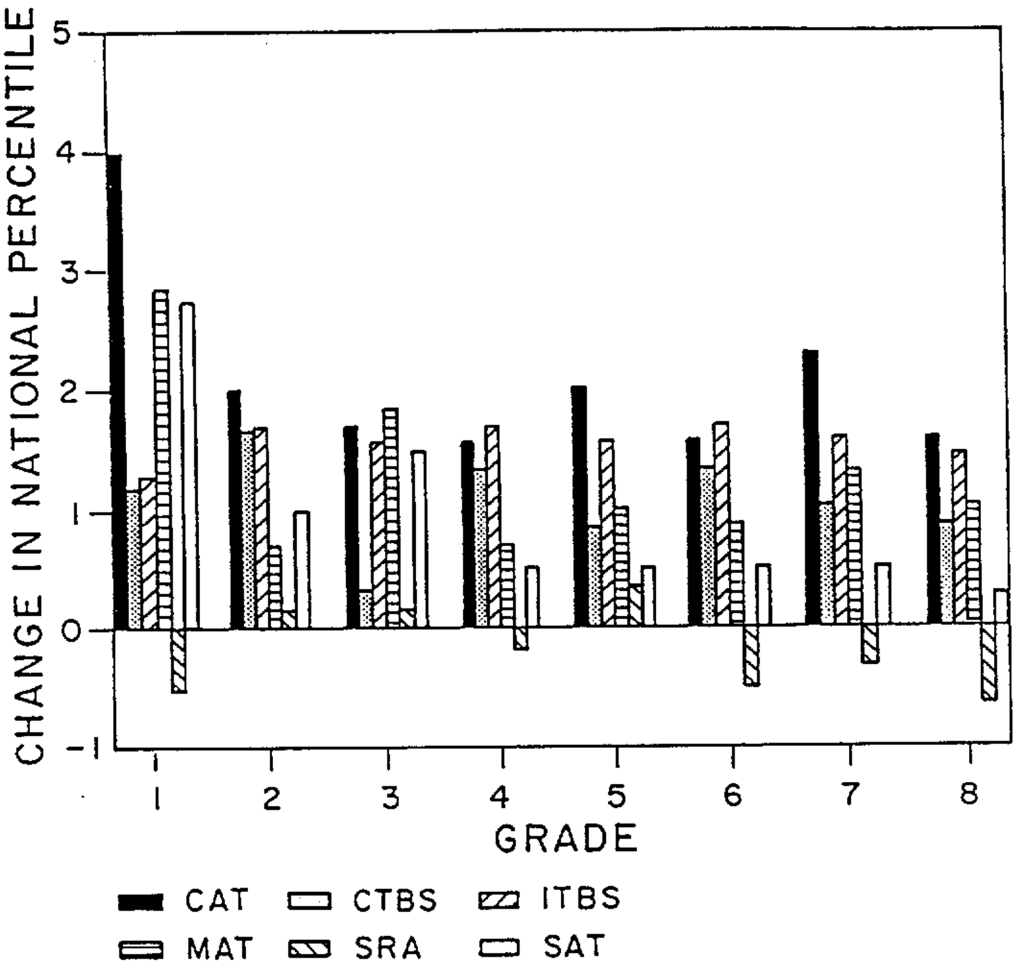
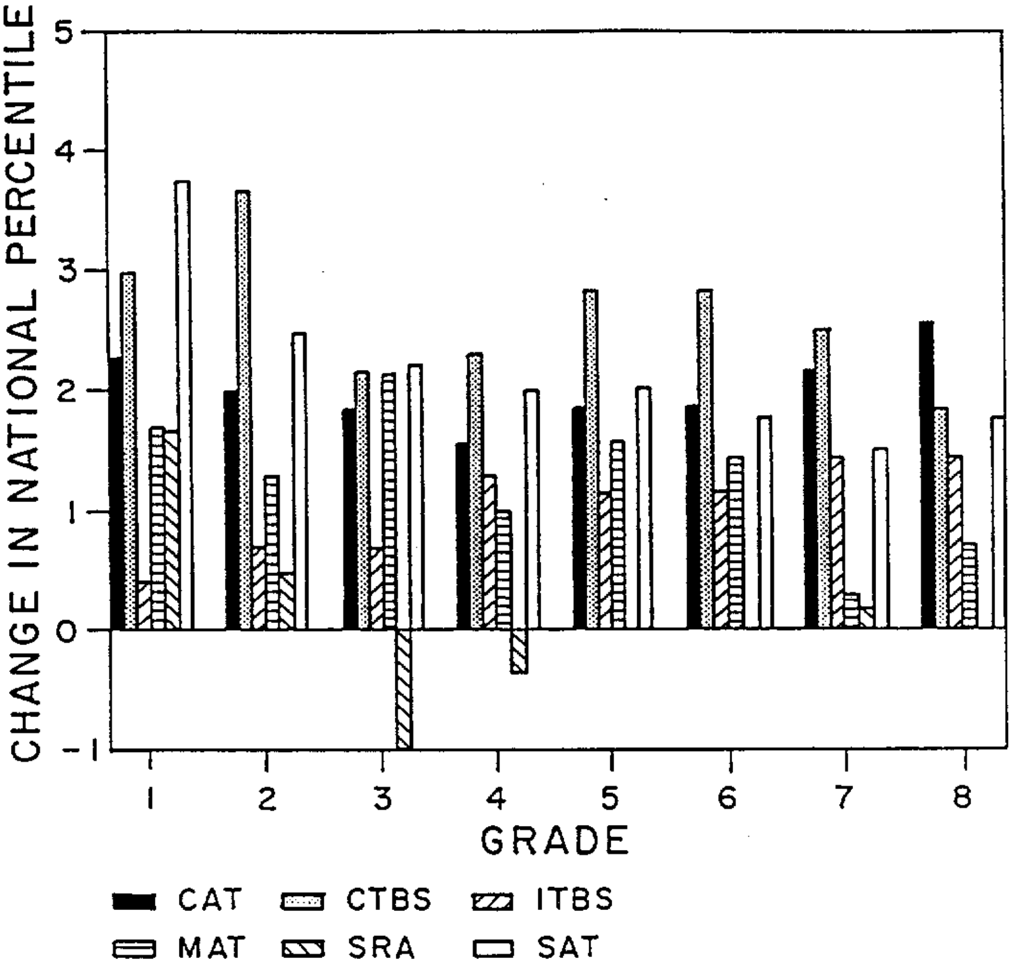


Figure 16
Estimated Yearly Changes in Mathematics Percentile Rank:
Publisher Results at the Median



shown in Figures 15 and 16 for some tests but smaller than those for other tests. The Table 5 estimates of changes in norm-referenced performance that would be expected as a result of a change in the date of the norms, however, are of the same order of magnitude as those shown in Figures 15 and 16.

Although the NAEP trend results are based on age cohorts rather than grade cohorts, the NAEP results represent the best available independent means of estimating national changes in achievement. Changes in percentile ranks estimated from NAEP results between the 1974-75 and 1983-84 assessments for reading and between 1977-78 and 1985-86 for mathematics are plotted in Figures 17, 18, and 19 for 9-, 13-, and 17-year-olds, respectively. Also shown in these figures are the changes for the six norm-referenced tests at the modal grades for 9-, 13-, and 17-year-olds, that is, Grades 3, 7, and 11.

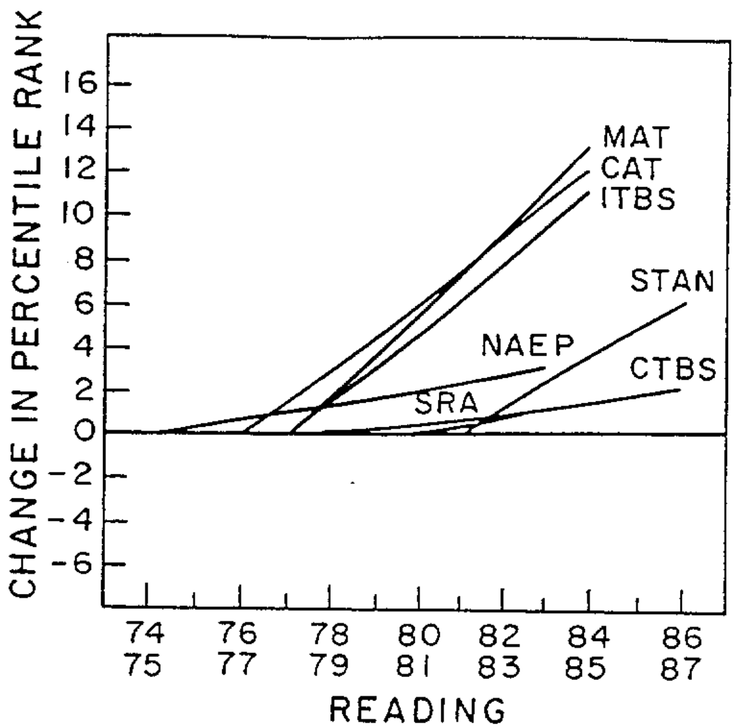
As can be seen in these figures, the different data sources vary a good deal in the magnitude of change in performance. The NAEP results suggest either some increase in performance (ages 9 and 17 in reading and ages 9 and 13 in mathematics) or no change during the interval in question. The increases indicated by NAEP are smaller than those shown by some, but not all, of the standardized tests. Comparing the publisher Grade 3 results with NAEP age 9 results (Figure 17), it can be seen that four of the six standardized tests show larger gains in reading and five of the six show larger gains in mathematics than would be estimated by NAEP. At age 13 (Figure 18) NAEP shows no change in reading and two of the standardized tests (SRA and Stanford) indicate only small changes at Grade 7, but the remaining four tests suggest more substantial increases in performance. In mathematics, two standardized tests suggest smaller changes at Grade 7 than NAEP obtained for 13-year-olds, one standardized test shows a change similar to the one obtained by NAEP, and the remaining three standardized tests show larger gains in performance. At Grade 11 or age 17 (Figure 19), relatively little change is indicated by any of the data sources for reading and relatively small and inconsistent changes are indicated for mathematics.

Of course, the dates of the first and second normings are not the same for all the tests and the tests differ in content coverage and in the specifics of the samples on which the norms were based. Nonetheless, the different data sources give rather different answers in some cases to the question of the degree to which test performance has increased during the past decade. The discrepancy between increases suggested by NAEP and most of the standardized tests raises questions about the possibility that artifacts may inflate the norm-referenced test results.

One possible artifact is that the norms obtained for a standardized test may be biased because of differential participation rates in norming studies by school districts according to whether the districts were already using the standardized test being normed (Baglin, 1981). If school districts that are already using a standardized test are more likely to participate in the norming of a new edition of the test than districts using another publishers test, and if districts that are using a given test generally have curricula that match more closely the objectives of both the new and old editions of that test or emphasize those objectives because the test is used, then the norms could be more difficult. In other words, such an influence would run counter to the observed tendency for states and districts to report that more than 50% of their students score above the national median.

To investigate the latter possibility, Wiser and Lenke (1987) compared the performance of user and non-user groups when the 1986 norms for the Stanford were obtained. They found that "users performed as well or better than non-users in all subject areas through Grade 6." For Grades 7 through 12 the results were more mixed, with users performing better in some subject areas at some grades but non-users performing better for other combinations.

Figure 17
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 3 (NAEP, Age 9)



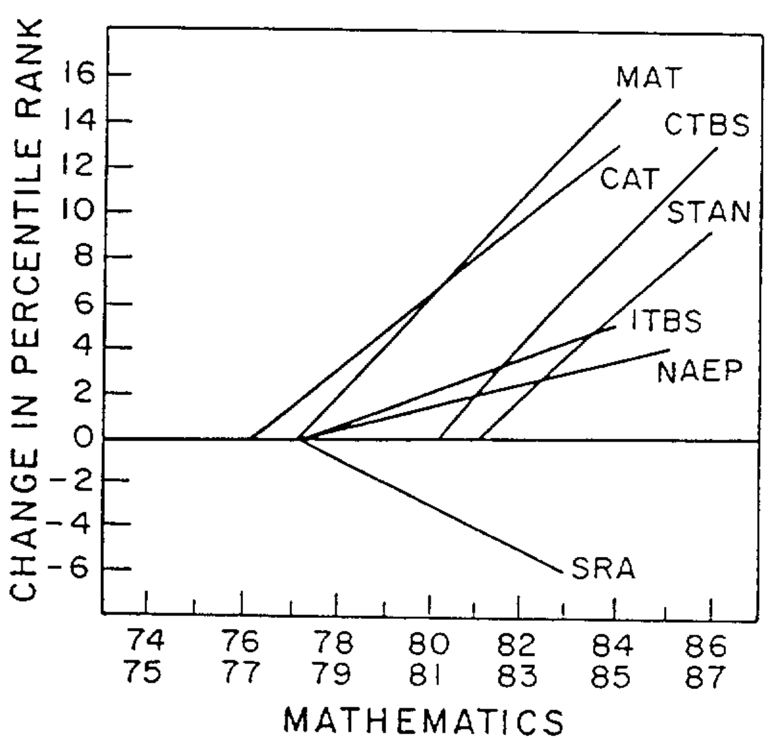


Figure 18
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 7 (NAEP, Age 13)

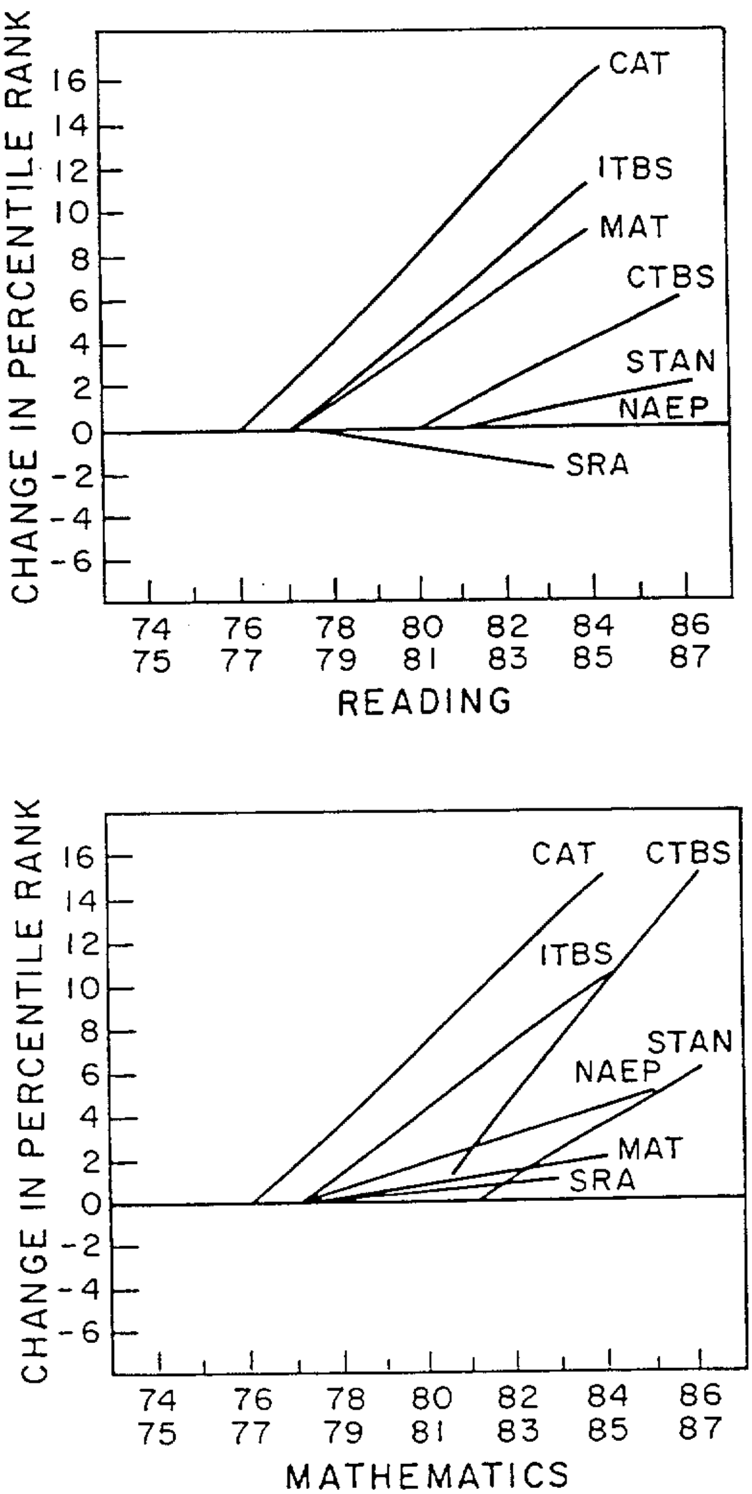
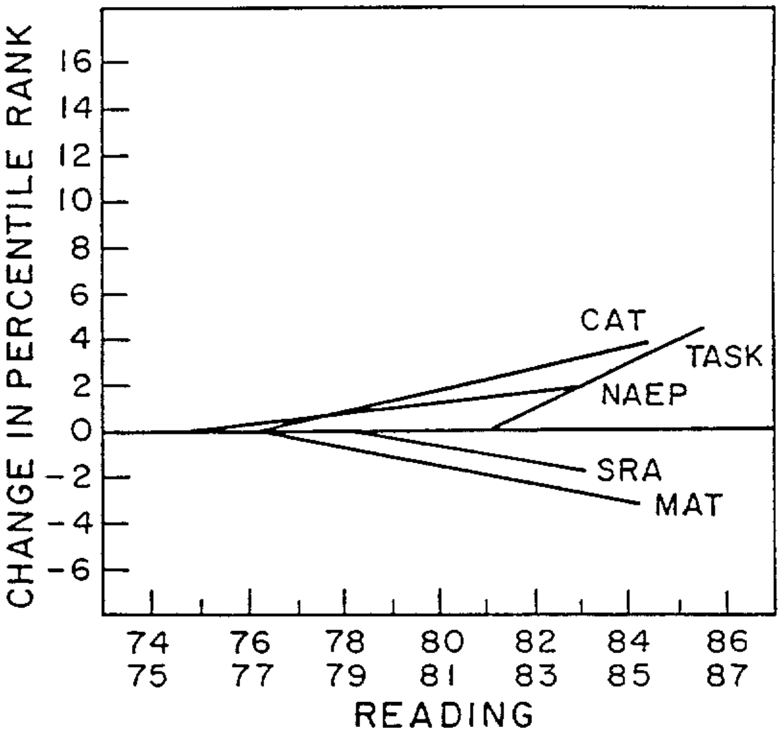
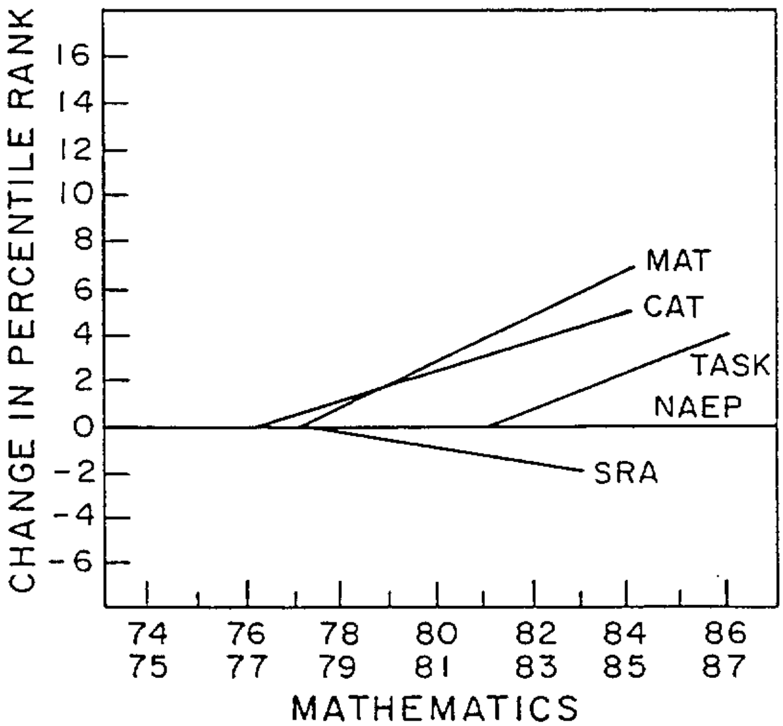


Figure 19
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 11 (NAEP, Age 17)





Wiser and Lenke noted that the comparison of particular interest in their results was between the 1986 non-users and the 1982 norming sample. Since the Stanford 7 was a new edition at the time of the 1982 norming, the participants in the norming sample had not previously used the edition and were comparable in that sense to the 1986 non-user sample. The 1982 sample and the 1986 non-user samples were also matched on school ability as measured by the Otis-Lennon School Ability Test. Thus, a comparison of the 1982 and 1986 non-user results provides an estimate of the change in achievement that is uncontaminated by the familiarity that users have with the particular edition of the test.

We used the scaled score means and standard deviations reported by Wiser and Lenke (1987) to calculate two estimates of the changes in average test scores in terms of 1982 standard deviation units for total reading and total mathematics. The first estimate is simply the mean for the full 1986 norming sample (users and non-users) minus the 1982 mean, all divided by the 1982 standard deviation. The second estimate is the 1986 mean for non-users only minus the 1982 mean, all divided by the 1982 standard deviation. The two sets of standardized differences are summarized in Table 7.

Table 7
Estimated Standardized Average Changes in Achievement Test Scores on the Stanford from 1982 to 1986 (Based on Wiser & Lenke, 1987)

	Reading		Mathematics	
Grade	Total Group[a]	1986 Non-users[b]	Total Group	1986 Non-users
1	.17	.10	.34	c
2	.13	.04	.18	.10
3	.13	.12	.15	.12
4	.03	-.01	.12	.12
5	.03	-.02	.17	.16
6	.03	-.02	.10	.06
7	.03[d]	.03[d]	.08	.06
8	.00[d]	-.08[d]	.10	.11
9	.08[d]	.03[d]	.05	.07
10	.05	.05	.04	.03
11	.10	.11	.03	.05
12	.13	.14	.05	.08

[a]The mean for the full 1986 norming sample (users and non-users) minus the 1982 mean all divided by the 1982 standard deviation.

[b]The mean for the 1986 non-users only minus the 1982 mean all divided by the 1982 standard deviation.

[c]Not available.

[d]Reading Comprehension.

For Grades 1 and 2 the non-user group data results in estimates of the gain in achievement in reading between 1982 and 1986 that are substantially smaller than the estimates based on the total norming sample. The gain in reading achievement appears to be about 40% smaller (i.e., 100x(.17-.10)/.17) at Grade 1 and about 70% smaller at Grade 2 with non-user data than with the data from the total norming sample. This difference is consistent with the premise that familiarity with a test form leads to inflated estimates of achievement gains. However, large differences in estimates based on non-user and total norming sample data such as those for reading in Grades 1 and 2 are <u>not</u> found consistently.

The non-user estimates of standardized gains in reading achievement are smaller for the total-norming-group estimates in Grades 1 through 6 and Grades 8 and 9, albeit by only a trivial amount at Grade 3. The two sets of estimates are the same to two decimal places in Grades 7 and 10, and the non-user estimates are actually larger than those based on the total norming sample at Grades 11 and 12. For mathematics, non-user estimates of achievement gains are 20% or more lower than total group estimates only at Grades 2, 3, 6, and 7, while they are larger by an equal percentage or more at Grades 9, 11, and 12.

Overall, the Wiser and Lenke results suggest that increasing familiarity with a particular test form may explain part of the apparent growth in norm-referenced test performance. The generally higher scores obtained by non-users in 1986 than were obtained in the 1982 norming of the then new edition of the test, however, suggest that there also has been some more generalized improvement in performance, particularly in mathematics.

Results recently reported by Hoover (1989) for the Iowa Tests of Basic Skills (ITBS) suggest that much of the increase in performance on a test form may occur on the first operational administration of the form. From user data weighted to estimate national performance, Hoover estimated that approximately 55% of the students scored above the 1984-85 national median across Grades 3 through 8 on the Battery Composite when Forms G and H were first administered operationally in 1985-86. In the second and third years of operational administration the average percentage of students across Grades 3 thru 8 who scored above the 1984-85 national median increased to 59% (1986-87) and then to 60% (1987-88).

The gains from the first year to the second and third years of operational use reported by Hoover may be attributable to a combination of real gains in achievement and increasing familiarity with a test form. The relatively large gain in the first year that the test was used operationally, however, may be due to a combination of several additional factors such as (a) the selection of a test that was most closely aligned with the state or district curriculum, (b) greater emphasis on the importance of good test performance when the test was used operationally than when it was normed, and (c) the exclusion of a larger fraction of less able students in operational test administrations than in norming studies. Indirect support for the latter explanation comes from Hoover's finding that only about 6%, rather than the expected 10%, of the students scored below the 10th percentile during the first year of operational administration of Forms G and H of the ITBS. High scores (at or above the 90th percentile), on the other hand, occurred at the expected rate of 10% in the first year of operational test use.

Discussion

Weighted estimates from the district sample suggest that at least 57% of the students in Grades 1 through 6 are obtaining scores above the national median on norm-referenced reading tests. The corresponding figure for mathematics is 62%. The comparable figures for Grades 7 through 12 are lower, but still somewhat greater than 50%. The state results are quite consistent with the district estimates. Thus, the results of the present study provide additional support for the general finding by Cannell and by the SREB that for the elementary grades almost all states and the majority of districts are reporting norm-referenced achievement test results that are above the national median.

While supporting Cannell's general finding that it is more common for a state or district to obtain test results that are "above the national average," our analyses lead us to conclusions that are different, and certainly less sensational, than the ones he reached. To begin with, it is important to put the "above average" findings in context. Many students are receiving scores that are "below average" even in districts or states that are reporting that substantially more than 50% of their students are "scoring above the national average." When a district reports that 57% of its students obtained reading scores that are at or above the national median, for example, the other 43% of the students obviously scored below the median. It should also be emphasized that although most districts report results that are "above the national average," there are still many districts throughout the nation that are reporting results that are below average. One out of 10 districts in our sample, for example, reported that a only about a third of its students at a given grade scored above the national median in reading.

Cannell (1987) concluded that norm-referenced achievement tests are producing inflated reports from states and districts on the achievement of their students. But the finding that more than half the students are scoring above the national median that was obtained when the norms were established does not necessarily imply that the results are inflated. There are many factors that may lead to the general finding, but it seems clear that the use of "old" norms is one of the major factors that contributes to the abundance of "above average" scores.

The evidence reviewed provides strong support for the conclusion that norms obtained for Grades 1 through 8 during the late 1970s or early 1980s are easier on most tests than are more recent norms. Consequently, a state or district where the average student scores at the <u>current</u> national average will be accurately reported to be above a national average that is defined by norms that are several years old. It appears that a substantial fraction of the "Lake Wobegon" phenomenon may be attributable to the use of old norms. It should be noted that the use of "old" norms is not purposeful on the part of school districts or states; they generally use the most recent norms available. Since standardized tests are usually normed every seven years, the most recent norms available will be, on average, 3.5 years old in most school years.

Concerns about dated norms have led to suggestions that publishers should produce current annual norms (e.g, Cannell, 1988; Phillips and Finn, 1988) and publishers are now attempting to do this by obtaining weighted estimates of national results from user data (e.g., Rothman, 1988). As Shepard (1989) has pointed out, however, annual norms based on user data potentially have several serious defects. If users differ from nonusers in ways other than those reflected by the demographic variables used for weighting, then user-based annual norms may be worse than dated norms where there is at least an understood frame of reference. In particular, if test familiarity leads to higher test performance, a state or district that changes publishers and administers a several-years-old test form for the first time will be at a disadvantage when results are compared to user norms (Shepard, 1989).

The alternative of conducting special national norming studies every year, or even every other year, is not a realistic or desirable possibility. Norming is not only expensive, but the quality of the results is very dependent on voluntary participation of schools and well-motivated students. Current participation rates in norming studies conducted roughly every six or seven years by a publisher are already far lower than would be desired. More frequent attempts to norm tests would surely lower the participation rates still further and thereby degrade the quality of the norms. Finally, it should be noted that although more recent norms provide a more stringent standard of comparison when scores are going up as they have been during the last decade, they would provide a less stringent standard during periods of decline in scores such as that experienced between the mid 1960s and the mid 1970s. Thus, we do not believe that the use of annual norms is an appropriate or effective way to deal with problems caused by dated norms.

In any reporting of test scores emphasis needs to be given to the changing meaning of norms and the age of the norms that are used. It obviously is not sufficient to report that "students in state X are scoring above the national average" without clearly indicating the year in which the norms were obtained. Simply noting the year of the norms is not enough, however. An explanation of the implications of shifting norms also needs to be provided along with an indication of what is known about recent trends in the stringency of national norms.

There is ample evidence that scores on norm-referenced tests have been going up in Grades 1 through 8 in recent years. But the more important question is: Has student achievement improved in recent years? Unfortunately, the answer to the latter question is equivocal.

Achievement test scores are of interest to the degree that they enable valid inferences to be made about broader achievement domains. But little attention has been given to the issue of the degree to which valid generalizations about broad achievement domains can be made from state or district test results.

Comparisons of the changes in norms of standardized tests with estimates of changes in achievement based on NAEP results suggest that test norms may be changing more rapidly than is student achievement as measured by NAEP. The Wiser and Lenke (1987) findings that apparent increases are generally smaller for non-users than for users of a given test series suggest that part of the apparent growth in achievement based on norm-referenced test results may be due to increased familiarity with a particular form of a test. Only part of the apparent gain can be explained in this way, however.

The differences between the gains in performance indicated by NAEP and by norm-referenced tests, and between Wiser and Lenke's total norming sample and their non-users suggest at the very least that caution is needed in interpreting gains in norm-referenced test scores as reflections of the amount of improvement that has taken place in achievement, more broadly defined. More direct assessments of the degree of generalizability of results to other tests and to other indicators of student achievement are greatly needed.

Hoover's (1989) finding that only about 6% of the students scored below the 10th percentile in the first year of operational administration of Forms G and H of the ITBS suggests that roughly a third to a half of the difference between the percentage of students scoring above the national median and the naive expectation of 50% may occur in the first year of use and may be due to what happens with the least able students. This suggests that greater emphasis in reporting needs to be given to the lower end of the score distribution and to the students who are excluded from testing when results are reported by states or districts. It may be quite appropriate, indeed desirable, to exclude students with limited English proficiency or students receiving particular types of special education services from a norm-referenced test administration. Such students should not be ignored, however, when district or state achievement results are reported: At minimum, the number of such students and the reasons for exclusion from testing should be reported.

The practice of using a single form of a test year after year poses a logical threat to making inferences about the larger domain of achievement. Scores may be raised by focusing narrowly on the test objectives without improving achievement across the broader domain that the test objectives are intended to represent. Worse still, practice on nearly identical or even the actual items that appear on a test may be given. As Dyer aptly noted some years ago, "if you use the test exercises as an instrument of teaching you destroy the usefulness of the test as an instrument for measuring the effects of teaching" (1973, p. 89).

Current accountability pressures place great emphasis on test scores. It is unlikely that any single test, no matter how well constructed, normed, and validated, can withstand the pressures to serve both as an instrument of instruction and as an instrument for measuring the effects of instruction. Making valid inferences about broad achievement domains from test scores has always been a challenging and difficult undertaking, but it is made all the harder by current demands for accountability and the use of standardized test results as primary indicators of accountability.

References

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Baglin, R.F. (1981). Does "nationally" normed really mean nationally? *Journal of Educational Measurement, 18*, 97-107.

Baker, E.L. (1989). *What's the use? Standardized tests and educational policy* (Report to OERI, Grant No. G-86-0003). Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Burstein, L. (1989). *Looking behind the "average": How are states reporting test results?* (Report to OERI, Grant No. G-86-0003). Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Cannell, J.J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average* (2nd ed.). Daniels, WV: Friends for Education.

Cannell, J.J. (1988a). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice, 7*(2), 5-9.

Cannell, J.J. (1988b). The Lake Wobegon effect revisited. *Educational Measurement: Issues and Practice, 7*(4), 12-15.

CTB/McGraw-Hill. (1987). *Technical Report: California Achievement Tests, Forms E and F, Levels 10-20.* Monterey, CA: Author.

CTB/McGraw-Hill. (1988). *CTB/McGraw-Hill studies show students achieving at higher levels in basic skills* (CTB/McGraw-Hill press release). Monterey, CA: Author.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., & Chambers, D.L. (1988). *The mathematics report card. Are we measuring up? Trends and achievement based on the 1986 national assessment.* Princeton, NJ: Educational Testing Service.

Drahozal, E.C., & Frisbie, D.A. (1988). Riverside comments on the Friends of Education report. *Educational Measurement: Issues and Practice, 7*(2), 12-16.

Dyer, H.S. (1973). *Recycling the problems in testing. Proceedings of the 1972 Invitational Conference on Testing Problems.* Princeton, NJ: Educational Testing Service.

Gardner, E.F., Madeen, R., Rudman, H.C., Karlsen, B., Merwin, J.C., Callis, R., & Collins. (1983). *Stanford Achievement Test Series: Multilevel norms booklet: National.* San Antonio: The Psychological Corporation.

Gardner, E.F., Madeen, R., Rudman, H.C., Karlsen, B., Merwin, J.C., Callis, R., & Collins. (1987). *Stanford 7 Plus: Multilevel norms booklet: National.* San Antonio: The Psychological Corporation.

Harnischfeger, A., & Wiley, D.E. (1975). *Achievement test score decline: Do we need to worry?* Chicago: ML-Group for Policy Studies in Education.

Hieronymus, A.N., & Hoover, H.D. (1986). *Iowa Tests of Basic Skills Forms G/H: Manual for school administrators, Levels 5-14.* Chicago: The Riverside Publishing Co.

Hoover, H.D. (1989, June). *Reactions to "Lake Wobegon" one year later: Results from a replication of the Cannell study.* Symposium presented at the 1989 Assessment Conference of the Educational Commission of the States, Boulder, CO.

Korcheck, S.A. (1988). *Measuring student learning: Statewide student assessment programs in the SREB states.* Atlanta: Southern Regional Education Board.

Koretz, D. (1986). *Trends in educational achievement.* Washington, DC: Congressional Budget Office.

Koretz, D. (1987). *Educational achievement: Explanations and implications of recent trends.* Washington, DC: Congressional Budget Office.

Koretz, D. (1988). Arriving at Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8-15, 46-52.

Lenke, J.M., & Keene, J.M. (1988). A response to John J. Cannell. *Educational Measurement: Issues and Practice, 7*(2), 16-18.

Mehrens, W.A. & Kaminski, J. (1988, April). *Using commercial test preparation materials for improving standardized test scores: Fruitful, fruitless or fraudulent?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

National Assessment of Educational Progress. (1985). *The reading report card: Progress toward excellence in our schools: Trends in reading over four national assessments, 1971-1984.* Princeton, NJ: Educational Testing Service.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming, and equating. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.

Phillips, G.W., & Finn, C.E. Jr. (1988). The Lake Wobegon effect: A skeleton in the testing closet? *Educational Measurement: Issues and Practice, 7*(2), 10-12.

Prescott, G.A., Balow, I.H., Hogan, T.P., & Farr, R.C. (1978). *Metropolitan Achievement Tests Forms JS and KS, Complete Survey Battery: Teacher's manual for administering and interpreting.* San Antonio: The Psychological Corporation.

Prescott, G.A., Balow, I.H., Hogan, T.P., & Farr, R.C. (1986). *MAT6, Metropolitan Achievement Tests Forms L and M: Survey battery multilevel national norms booklet.* San Antonio: The Psychological Corporation.

The Psychological Corporation. (1988). *MAT5 to MAT6 Survey: Corresponding scaled scores of the fifth and sixth editions of the Metropolitan Achievement Test Survey Battery.* San Antonio: The Psychological Corporation.

Qualls-Payne, A.L. (1988). SRA Response to Cannell's article. *Educational Measurement: Issues and Practice, 7*(2), 21-22.

Rothman, R. (1988, November, 16). Pupil gains not just mythical, test publisher says study shows. *Education Week*, p. 20.

Science Research Associates. (1979). *SRA Achievement Series Forms 1 and 2: Norms and conversion tables.* Chicago: Science Research Associates, Inc.

Science Research Associates. (1986). *SRA Achievement Series Forms 1 and 2: Answer keys, norms, and conversion tables.* Chicago: Science Research Associates, Inc.

Shepard, L.A. (1989). *"Inflated test score gains": Is it old norms, or teaching the test?* (CSE Technical Report 307). Los Angeles: UCLA Center for the Study of Evaluation.

Southern Regional Education Board. (1984). *Measuring educational progress in the South: Student assessment.* Atlanta: Southern Regional Education Board.

Stonehill, R.M. (1988). Norm-referenced test gains may be real: A response to John Jacob Cannell. *Educational Measurement: Issues and Practice, 7*(2), 23-24.

Williams, P.L. (1988). The time-bound nature of norms: Understandings and misunderstandings. *Educational Measurement: Issues and Practice, 7*(2), 18-21.

Wiser, B., & Lenke, J.M. (1987, April). *The stability of achievement test norms over time.* Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.

Yen, W.M., Green, D.R., & Burket, G.R. (1987). Valid normative information from customized tests. *Educational Measurement: Issues and Practice, 6*(1), 7-13.

Appendix A

Sample Letter and Data Collection Form for Directors of State Testing Programs

July 22, 1988

States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP

Dear States Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the Office of Educational Research and Improvement (OERI). This study was sitmulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. The interpretation of normative comparisons was called into question by Cannell's finding that "no state scores below the publisher's 'national norm' at the elementary level on any of the six major nationally normed, commercially available tests" (p. 2 of second edition of Cannell Report). The value of assessment results was further challenged by Cannell's conclusion that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of Educational Measurement: Issues and Practice, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have a better understanding of the magnitude and prevalence of the apparently high achievement results reported by Cannell. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information that will provide a better data base for determining not only what proportion of students score above determining not only what proportion of students score above the 50th percentile according to national norms, but other important characteristics of the test results such as changes in means over time and the variability in scores. We also need to obtain information on the way in which test results are currently used (e.g., public reporting, grade retention, school incentives, etc.), when these uses were instituted, and planned changes in the use of test results. Finally, we are seeking information about policies regarding test security and guidelines on preparation of students for taking tests.

A CRESST staff member will be contacting you by phone to seek your assistance and to arrange for a time for a phone interview with an appropriate person on your staff. The information that will be requested is outlined on the enclosure. We will send you more detailed worksheets between now and the time of the telephone interview to help organize the requested information.

In many cases, the information that we are seeking may be provided in reports that have previously been prepared. Thus we request that you send us copies of any reports that give summaries of district results that have been published within the past three years. Copies of press releases and newspaper articles about the test results would also be useful. If you send us reports and press releases as quickly as possible, we will use the reports to extract as much of the requested information as possible. We will call you to ask questions after we have "done our homework".

Please send reports to: Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Thank you for your consideration. We will phone you within the next two weeks to answer questions and to try to arrange a time for a telephone interview. A return postcard is enclosed so that you can indicate the name, phone number, and best times for us to try to contact the appropriate person for the telephone interview.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for the Study of Research on Evaluation Standards, and Student Testing

Explanation of Information Requested

Column	Information requested
1	Testing year
2	Grade levels tested K - 12.
3	Name of test used for statewide assessment e.g., CTBS, MAT, name of locally developed test.
4	Edition of the test used at each grade level, e.g., 1982.
5	Form of the test used at each grade level.
6	Year when test was first used.
7	Norming year of test used for reporting scores.
8	Month in which tests were administered.
9	Type of scores reported, e.g., percent correct, percentile rank, NCE. n.b. If you have more than one type of score, please provide one form of data in the preferred order as follows: Percentile Rank Grade Equivalents NCE Stanines Percent Correct ...
10	Number of students enrolled: the total number of students by grade statewide.
11	Number of students tested.
12	Number of students' scores reported: If not all scores are used to compute rankings or other statewide test results, enter the number of students' scores used to compute the achievement data.
13	Reading %: The percent of students scoring above the national 50th percentile statewide.
14	Math %: The percent of students scoring above the national 50th percentile statewide.

n.b. If neither reading nor math data requested in 12 and 13 are available, please provide the most appropriate composite scores and indicate the nature of these on the form.

If the data requested in columns 13 or 14 (percent of students scoring above the national 50th percentile) are not available, please provide as much of the following as possible (columns 15 - 20 on the Alternate Information Sheet):

Column

15 Reading statewide mean.

16 Reading statewide standard deviation.

17 Math statewide mean.

18 Math statewide standard deviation.

19 Reading score at each percentile: The score at the 25th percentile statewide.

- at the 50th percentile statewide.
- at the 75th percentile statewide.

20 Math score at each percentile: The math score at the 25th percentile statewide.

- at the 50th percentile statewide.
- at the 75th percentile statewide.

Type of scores: If the type of scores reported in columns 13-20 are not the same as those indicated in column 9, please indicate the type of scores used to compute the percentiles, mean, and standard deviations.

Statewide Testing Information

State Name: ______________

Person Supplying Information: ______________

Title: ______________

1 Testing Year	2 Grade	3 Test Name	4 Edition	5 Form	6 Year First Used	7 Norming Year	8 Testing Dates	9 Type of Scores
1985-1986								
1986-1987	K							
1987-1988								
1985-1986								
1986-1987	1							
1987-1988								
1985-1986								
1986-1987	2							
1987-1988								
1985-1986								
1986-1987	3							
1987-1988								
1985-1986								
1986-1987	4							
1987-1988								
1985-1986								
1986-1987	5							
1987-1988								
1985-1986								
1986-1987	6							
1987-1988								

1985-1986								
1986-1987	7							
1987-1988								
1985-1986								
1986-1987	8							
1987-1988								
1985-1986								
1986-1987	9							
1987-1988								
1985-1986								
1986-1987	10							
1987-1988								
1985-1986								
1986-1987	11							
1987-1988								
1985-1986								
1986-1987	12							
1987-1988								

Please Refer to Explanation of Information Requested - Attached

10	11	12	13	14
Number of Students Enrolled	Number of Students Tested	Number of Students' Scores Reported	Reading: % of Students above National 50%ile	Math: % of Students above National 50%ile

Alternate Information Available

State Name: ____________________

Person Supplying Information: ____________________

Title: ____________________

Testing Year	Grade	15 Reading Mean	16 Reading Standard Deviation	17 Math Mean	18 Math Standard Deviation	19 Reading Score at each percentile 25	19 Reading 50	19 Reading 75	20 Math Score at each percentile 25	20 Math 50	20 Math 75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

1985-1986											
1986-1987	7										
1987-1988											
1985-1986											
1986-1987	8										
1987-1988											
1985-1986											
1986-1987	9										
1987-1988											
1985-1986											
1986-1987	10										
1987-1988											
1985-1986											
1986-1987	11										
1987-1988											
1985-1986											
1986-1987	12										
1987-1988											

Appendix B

Interview Guide

code

District

State

Interviewer

date

Person(s) Interviewed

name

title

name

title

Background information: Number of schools in district

Size (range)

Center for the Study of Research on Evaluation, Standards, and Student Testing, Robert L. Linn, School of Education, University of Colorado at Boulder

Part I: District Testing Data (to be recorded on the forms provided)

YEARS TESTED 1. Are districtwide test results available for:

_____1987-88

_____1986-87

_____1985-86 If none, then the most recent year: _____

If there is no districtwide testing, ask only 12, 13, 19 - 22, and 26 for large districts.

ENROLLMENT BASIS 2. What is the basis for the enrollment figures used to give the number of students in each grade? (e.g., ADA= Average Daily Attendance)

ENROLLMENT SOURCE 3. What office provides the enrollment figures? [name of person and phone number if easily available]

TESTED = REPORTED 4. Is the number of students tested the same as the number of students that are included in the reported test results?

Yes_____ No_____

If no, how does the number included in the reported test results differ from the number tested?

probe: special education

SAMPLING PLAN

5. Were all eligible students in the grade tested or is a sampling plan used?

___ universal testing by grade ___ sampling plan

Please describe any sampling procedures used.

TESTING EXCLUSIONS

6. What rules are used to determine students who are excluded from testing?

request: copies of any written policies that describe these rules

% EXCLUDED

7. How many students (or what percent of the students) are excluded using these rules?

MAKE-UP TESTING

8. What are the policies for make-up testing (for students who are absent)?

request: if in writing

[Ask the following only if needed:]

LOCALLY CONSTRUCTED TEST

9. If a specially constructed test is used, is it linked to a norm-referenced test? If so, what is the name and edition of the norm referenced test?

REPORTING NATIONAL COMPARISONS

10. If the percent of students above the 50th percentile is unknown, please describe the way in which scores are reported and comparisons are made to the national norm.

LOCAL FACTORS IN TEST SCORES

11. Are any factors of schools or the characteristics of their students taken into account in reporting test scores?

(e.g., percent minority, percent eligible for free lunch, Chapter I)

[BEGIN TAPE RECORDING]

Part II: Testing Policies and Perceptions

USES AND IMPORTANCE

12. What are the uses of test results?

___ -local district and school instructional and evaluation decisions

___ -reporting to parents about individual student progress or school programs

___ -School Board attention (And if so, how have Board members used test results-- to increase testing programs or other forms of accountability?

___ -state or local politician use of scores in campaigning or proposing legislation

___ -changing general funding levels for schools

___ -targeted funds or mandating programs such as remediation

___ -superintendent, principal, or teacher performance rating or jobs

___ -media coverage and community awareness

*** How important are test scores in your district?

/______________/______________/______________/______________/

extremely very moderately slightly not important

REFORMS

13. Have major educational reforms been introduced in your district in the past five years?

request: Would you briefly describe these or send us written descriptions that are available?

TEST SELECTION

14. Who selected the standardized test(s) being used? (If locally developed, how was the content selected?)

probe: committee composition, e.g., teachers, parents,...

CURRICULUM ALIGNMENT

15. Have there been efforts to assure that the curriculum and the test are aligned?

If so, please describe those efforts.

TIME ON SPECIFIC OBJECTIVES

16. Do you think that teachers spend more time teaching the specific objectives on the test(s) than they would if the tests were not required?

How much more time?

IMPORTANT OBJECTIVES GIVEN LESS TIME

17. To what extent do you think important objectives are given less time or emphasis because they are not included on the test?

What kinds of objectives are neglected?

INFORMAL GUIDELINES ABOUT TEST PREPARATION

18. Do you or members of your staff provide informal guidelines about test preparation? What kind of advice do you give schools about how to prepare students to take tests?

probes:

- length of time to practice
- minimum and maximum recommended time for practice
- whether to use items in a specific format for practice

TECHNICAL ASSISTANCE ABOUT TEST PREPARATION

19. What kind of technical assistance or materials do you provide to schools about test preparation?

request: Would you send us copies of the materials or descriptions of the assistance?

probes:
- practice tests
- testwiseness packages
- curriculum domain materials but not specific test items
- amount of these activities

TYPICAL PRACTICES OF TEST PREPARATION

20. Can you describe typical practices of test preparation?

probes:
- If they say, one school does X, ask how common this is, or how many other schools do the same.
- Do schools use the materials and assistance you provide?
- What else do they do beyond what you recommend?

EXTREME PRACTICES OF TEST PREPARATION

21. Can you describe <u>extreme</u> cases of test preparation?

probes:

If they describe a worst case, ask what they would think of as a best case. (as well as what is more typical, above)

Examples of cases which violate your recommendations?

TEST ADMINISTRATION AND SECURITY POLICIES

22. Do you have written policies regarding test administration and security procedures? If not, do you have informal guidelines?

<u>request</u>: written policies

WHO ADMINISTERS OR HAS TESTS OR KNOWS TESTS

23. Who administers the tests?
Do teachers in some schools have copies of the tests prior to test administration?

How familiar are teachers with the specific items on the tests?

probes:
teachers administering same test over years
principals or teachers having test files

DETECT ANOMALIES

24. Do you have any formal procedures for detecting anomalies in the data?

<u>request</u> copies

probes:
check for missing test booklets
computer detection of significant numbers of erasures
" " of extraordinary gains from one year to the next
check numbers of students tested against enrollment

TYPICAL AND EXTREME PRACTICES

25. Can you give examples of both typical and extreme testing practices?

Have you withheld score reports because of suspected cheating?

probes:

good practices: consistent, successful make-up testing

examples of cheating-

- teachers filling in answers
- extending time limits for tests
- teaching specific items on the test
- discrepancies in numbers of students tested

[Ask the following only in districts designated as 7's or 8's- large districts]

REACTIONS TO CANNELL REPORT

26. What are your reactions to the Cannell report and its conclusions?

FACTORS IN ACHIEVEMENT TRENDS

27. What do you think are the primary factors that contribute to the recent trends in achievement test scores in your district?

probes:

educational reforms
norms (unrepresentative or old)
pressure on teachers to have high scores

Closing:

When finishing and thanking them for their time, review the things which you may have requested in writing.

Checklist of Requested Written Information

___ testing data on years not yet received (e.g., all three years 1985-1988)

___ testing data such as distribution measures

___ #3- name and phone of office or person with enrollment figures

___ #6- Rules for testing exclusions

___ #8- Policies for make-up testing

___ #13- Educational reforms in the state

___ #19- Technical assistance or materials for test preparation

___ #22- Test administration and security policies

___ #24- Procedures for detecting anomalies

The address for mailing is:

Dr. Robert Linn 303-492-8280 (Bob)
University of Colorado or -2124 (Nancy)
School of Education or -3108 (Lorrie)
Campus Box 249
Boulder, CO 80309

If you have missing answers and have to schedule another call, please indicate that in the telephone log.

Appendix C

Districts Available by Cells of Sampling Design

Appendix C

Number of Districts Available by Cells in Sampling Design

Region	District Size	SES Level	Number of Districts Available
East	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	2
		Below Average	4
		Average	0
		Above Average	1
		High	1
	50,000 to 99,999	Low	1
		Below Average	2
		Average	1
		Above Average	1
		High	0
	100,000 or more	Low	1
		Below Average	2
		Average	0
		Above Average	2
		High	1

Appendix C (page 2 of 4)

Region	District Size	SES Level	Number of Districts Available
North/ Central	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	1
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	0
		Below Average	2
		Average	5
		Above Average	5
		High	4
	50,000 to 99,999	Low	1
		Below Average	3
		Average	2
		Above Average	0
		High	0
	100,000 or more	Low	0
		Below Average	1
		Average	1
		Above Average	0
		High	0

Appendix C (page 3 of 4)

Region	District Size	SES Level	Number of Districts Available
South	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	2
		High	3
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	2
		High	0
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	3
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	4
	25,000 to 49,999	Low	2
		Below Average	5
		Average	5
		Above Average	5
		High	2
	50,000 to 99,999	Low	1
		Below Average	3
		Average	5
		Above Average	5
		High	1
	100,000 or more	Low	0
		Below Average	1
		Average	5
		Above Average	0
		High	1

Appendix C (page 4 of 4)

Region	District Size	SES Level	Number of Districts Available
West	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	2
		Below Average	2
		Average	5
		Above Average	5
		High	5
	50,000 to 99,999	Low	1
		Below Average	1
		Average	5
		Above Average	5
		High	1
	100,000 or more	Low	0
		Below Average	0
		Average	3
		Above Average	1
		High	0

Appendix D

Sample Letters, Data Collection Forms, and Questionnaires Sent to Districts

August 18, 1988

Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP

Dear Dist Phone Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the U.S. Department of Education's Office of Educational Research and Improvement (OERI). This study was stimulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you may know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. Based on his survery of states and selected school districts, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of Educational Measurement: Issues and Practice, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have technically accurate information about achievement results reported by school districts across the nation. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information from a nationally representative sample of school districts that will provide a better data base for determining not only what level of student performance is being reported, but the uses and interpretations that are being made of the results. We also are seeking information about factors that may influence test results.

Your district has been selected as part of a nationally representative sample for this study. Hence, your participation is critical to maintaining representativeness and drawing conclusions about achievement testing for the nation. Results will not be reported for individual school districts. However, participation by each sampled district is essential to ensuring an accurate picture for the nation as a whole.

We ask that you complete the enclosed questionnaire about your district's testing program. In many cases, the information that we are seeking on the forms may be provided in reports that have previously been prepared. If so, we request that you answer the general questionnaire items and send us the questionnaire along with copies of any reports that give results of districtwide assessments of student achievement or summaries of district results that have been published within the past three years. We will use those reports to obtain the requested information. Copies of press releases and newspaper articles about the test results would also be useful.

Please return the completed questionnaire in the enclosed envelope to:

Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

We also ask you to participate in a telephone interview which concerns additional questions about testing policies and practices. In order to schedule an interview, we ask that you indicate on the questionnaire dates and times which would be convenient for one of our staff members to call. The interviews consist of fifteen questions about your testing program and usually last about 30 minutes.

Thank you for your consideration. We realize that school districts receive many requests for information and that responding to such requests is a burden on your time. Your willingness to help is essential to the success of the study and to our ability to provide solid answers to the important educational questions that were raised by the Cannell report.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for Research on Evaluation, Standards, and Student Testing

August 18, 1988

Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP

Dear Dist Survey Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the U.S. Department of Education's Office of Educational Research and Improvement (OERI). This study was stimulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you may know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. Based on his survery of states and selected school districts, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of Educational Measurement: Issues and Practice, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have technically accurate information about achievement results reported by school districts across the nation. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information from a nationally representative sample of school districts that will provide a better data base for determining not only what level of student performance is being reported, but the uses and interpretations that are being made of the results. We also are seeking information about factors that may influence test results.

Your district has been selected as part of a nationally representative sample for this study. Hence, your participation is critical to maintaining representativeness and drawing conclusions about achievement testing for the nation. Results will not be reported for individual school districts.

<u>However, participation by each sampled district is essential to ensuring an accurate picture for the nation as a whole.</u>

We ask that you complete the enclosed questionnaire about your district's testing program. In many cases, the information that we are seeking on the forms may be provided in reports that have previously been prepared. If so, we request that you answer the general questionnaire items and send us the questionnaire along with copies of any reports that give results of districtwide assessments of student achievement or summaries of district results that have been published within the past three years. We will use those reports to obtain the requested information. Copies of press releases and newspaper articles about the test results would also be useful.

Please return the completed questionnaire in the enclosed envelope to:

Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Thank you for your consideration. We realize that school districts receive many requests for information and that responding to such requests is a burden on your time. Your willingness to help is essential to the success of the study and to our ability to provide solid answers to the important educational questions that were raised by the Cannell report.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for Research on Evaluation, Standards, and Student Testing

District Testing Information

District Name ______________________

State ______________________

Person Supplying Information ______________________

Address ______________________ Phone Number ______________________

Title ______________________

1 Testing Year	2 Grade	3 Test Name	4 Edition	5 Form	6 Year First Used	7 Norming Year	8 Testing Dates	9 Type of Scores
1985-1986								
1986-1987	K							
1987-1988								
1985-1986								
1986-1987	1							
1987-1988								
1985-1986								
1986-1987	2							
1987-1988								
1985-1986								
1986-1987	3							
1987-1988								
1985-1986								
1986-1987	4							
1987-1988								
1985-1986								
1986-1987	5							
1987-1988								
1985-1986								
1986-1987	6							
1987-1988								

1985-1986								
1986-1987	7							
1987-1988								
1985-1986								
1986-1987	8							
1987-1988								
1985-1986								
1986-1987	9							
1987-1988								
1985-1986								
1986-1987	10							
1987-1988								
1985-1986								
1986-1987	11							
1987-1988								
1985-1986								
1986-1987	12							
1987-1988								

Please Refer to Explanation of Information Requested - Attached

10 Number of Students Enrolled	11 Number of Students Tested	12 Number of Students' Scores Reported	13 Reading: % of Students above National 50%ile	14 Math: % of Students above National 50%ile

8 - 11. Please indicate below the name of the test used at each grade level tested, (for standardized tests, include edition and form), the number of students tested, AND THE PERCENT OF STUDENTS ABOVE THE NATIONAL 50TH PERCENTILE. (If the percent of students above the national 50th percentile is not available, please provide as much of the information on pages 4 and 5 as possible.)

		8	9	10	11
Testing Year	Grade	Test Name, Edition and Form	Number of Students Tested	Reading: % of Students above National 50%ile	Math: % of Students above National 50%ile
1985-1986					
1986-1987	K				
1987-1988					
1985-1986					
1986-1987	1				
1987-1988					
1985-1986					
1986-1987	2				
1987-1988					
1985-1986					
1986-1987	3				
1987-1988					
1985-1986					
1986-1987	4				
1987-1988					
1985-1986					
1986-1987	5				
1987-1988					
1985-1986					
1986-1987	6				
1987-1988					

Testing Year	Grade	Test Name, Edition and Form	Number of Students Tested	Reading: % of Students above National 50%ile	Math: % of Students above National 50%ile
1985-1986					
1986-1987	7				
1987-1988					
1985-1986					
1986-1987	8				
1987-1988					
1985-1986					
1986-1987	9				
1987-1988					
1985-1986					
1986-1987	10				
1987-1988					
1985-1986					
1986-1987	11				
1987-1988					
1985-1986					
1986-1987	12				
1987-1988					

12. Testing Dates (month/year) ____________________

13. Norming year of norm referenced test(s) used: ____________________

14. Year these tests were first used in your district: ____________________

If the percent of students above the national 50th percentile is provided on pages 2 and 3, pages 4 and 5 need not be completed. Skip to page 6.

If the number of students above the national 50th percentile (columns 10 and 11, pages 2-3) is <u>not</u> known, please provide as much of the following information as possible.

Testing Year	Grade	Reading		Math		Reading Score at each percentile			Math Score at each percentile		
		Mean	Standard Deviation	Mean	Standard Deviation	25	50	75	25	50	75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

Testing Year	Grade	Reading		Math		Reading Score at each percentile			Math Score at each percentile		
		Mean	Standard Deviation	Mean	Standard Deviation	25	50	75	25	50	75
1985-1986											
1986-1987	7										
1987-1988											
1985-1986											
1986-1987	8										
1987-1988											
1985-1986											
1986-1987	9										
1987-1988											
1985-1986											
1986-1987	10										
1987-1988											
1985-1986											
1986-1987	11										
1987-1988											
1985-1986											
1986-1987	12										
1987-1988											

Alternate Information Available
District Test Results

District Name: ____________________

Person Supplying Information: ____________________

State: ____________________

Address: ____________________ Phone Number: ____________________

Title: ____________________

Testing Year	Grade	15 Reading Mean	16 Reading Standard Deviation	17 Math Mean	18 Math Standard Deviation	19 Reading Score at each percentile 25	19 50	19 75	20 Math Score at each percentile 25	20 50	20 75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

1985-1986					
1986-1987	7				
1987-1988					
1985-1986					
1986-1987	8				
1987-1988					
1985-1986					
1986-1987	9				
1987-1988					
1985-1986					
1986-1987	10				
1987-1988					
1985-1986					
1986-1987	11				
1987-1988					
1985-1986					
1986-1987	12				
1987-1988					

Explanation of Information Requested

Column	Information requested
1	Testing year
2	Grade levels tested K - 12.
3	Name of test used e.g., CTBS, MAT, name of locally developed test.
4	Edition of the test used at each grade level, e.g., 1982.
5	Form of the test used at each grade level.
6	Year when test was first used.
7	Norming year of test used for reporting scores.
8	Month in which tests were administered.
9	Type of scores reported, e.g., percent correct, percentile rank, NCE. n.b. If you have more than one type of score, please provide one form of data in the preferred order as follows: Percentile Rank Grade Equivalents NCE Stanines Percent Correct ...
10	Number of students enrolled: the total number of students enrolled by grade
11	Number of students tested at each grade
12	Number of students' scores reported: If not all scores are used to compute rankings or other statewide test results, enter the number of students' scores used to compute the achievement data.
13	Reading %: The percent of students scoring above the national 50th percentile.
14	Math %: The percent of students scoring above the national 50th percentile.

n.b. If neither reading nor math data requested in 12 and 13 are available, please provide the most appropriate composite scores and indicate the nature of these on the form.

If the data requested in columns 13 or 14 (percent of students scoring above the national 50th percentile) are not available, please provide as much of the following as possible (columns 15 - 20 on the Alternate Information Sheet):

<u>Column</u>

15 Reading mean for the district.

16 Reading standard deviation.

17 Math mean.

18 Math standard deviation.

19 Reading score at each percentile: The score

- at the 25th percentile districtwide
- at the 50th percentile districtwide.
- at the 75th percentile districtwide.

20 Math score at each percentile: The math score

- at the 25th percentile districtwide.
- at the 50th percentile districtwide.
- at the 75th percentile districtwide.

Type of scores: If the type of scores reported in columns 13-20 are not the same as those indicated in column 9, please indicate the type of scores used to compute the percentiles, mean, and standard deviations.

Appendix E

District Subsample for Telephone Interviews

Appendix E

District Subsample for Telephone Interviews

The 40 cells (5 levels of SES by 8 levels of district size) within each of the 4 regions that were used to define the overall district sample were collapsed to 15 cells (3 levels of SES by 5 levels of district size) to select the subsample to be interviewed by telephone. The following levels were combined for each factor.

SES		Size	
Subsample Level	Total Sample Level	Subsample Level	Total Sample Level
1 Below Average	Low & Below Average	1 <2,500	<1,200 & 1,200-2,499
2 Average	Average	2 2,500-9,999	2,500-4,999 & 5,000-9,999
3 Above Average	Above Average & High	3 10,000-49,999	10,000-24,999 & 25,000-49,999
		4 50,000-99,999	50,000-99,999
		5 100,000 +	100,000 +

For cells of the subsample design that consisted of 2 or 4 of the cells of the total sample, one district was randomly selected. The SES = 1, size = 1 cell of the interview subsample, for example, consists of SES by size cells 11, 12, 21, and 22 in the total sample. A random number between 1 and 4 corresponding to each of those original cells was selected for each region. Following this procedure for each of the interview subsample cells that contained more than one cell from the total sample, 56 districts (4 regions x 3 SES levels x 5 size levels minus 4 void cells) for the interview subsample were selected.

Appendix E (Continued, page 2 of 2)

Using the total sample code RZS where
R = region (1 = East, 2 = North/Central, 3 = South, and 4 = West);
Z = size (1 = less than 1,200, 2 = 1,200-2,499, 3 = 2,500-4,999, 4 = 5,000-9,999, 5 = 10,000-24,999, 6 = 25,000-49,999, 7 = 50,000-99,999, and 8 = 100,000 or more); and
S = SES (1 = low, 2 = below average, 3 = average, 4 = above average, and 5 = high),
the following interview subsample was selected.

112	211	312	411
123	213	323	415
124	225	324	423
131	233	332	432
134	242	335	433
145	245	343	445
153	251	353	454
155	255	362	462
161	263	365	463
172	272	371	471
173	273	373	474
174	275(void)	374	474
181	282	382	481(void)
183(void)	283	383	483
184	285(void)	385	484

Appendix F

Grades Tested by Districts Returning Data

Appendix F (page 1 of 4)

Grades Tested by Districts Returning Data

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
1	1	1		+	+	+	+	+	+						
1	1	3		+	+	+	+	+	+	+	+				
1	1	4				+	+	+	+	+					
1	2	1		+	+	+	+	+	+	+	+	+	+		
1	2	2			+		+		+		+		+		
1	2	4			+		+		+		+		+		
1	2	5				+	+	+							
1	3	1			+	+	+	+	+	+	+				
1	3	2		+	+	+	+	+	+	+	+				
1	3	3			+	+	+	+	+	+					
1	3	4										+	+		
1	3	5	+		+	+	+	+	+	+	+				
1	4	2	+	+	+	+	+	+	+	+	+		+		
1	4	3				+		+		+			+		
1	4	4		+	+	+	+	+	+	+	+	+	+		
1	4	5			+		+	+		+		+			
1	5	3				+		+			+				
1	5	4			+		+		+						
1	5	5	Interview completed - no normed test results												
1	6	1		+	+	+	+	+	+	+	+	+	+		
1	6	2			+	+	+	+	+	+	+	+	+	+	+
1	6	4				+		+			+				
1	6	5				+		+			+			+	
1	7	1	+	+	+	+	+	+	+	+	+	+	+	+	+
1	7	2		+	+	+	+	+	+	+	+				
1	7	2		+	+	+	+	+	+	+	+	+	+	+	+
1	7	3			+				+		+	+		+	
1	7	4				+		+			+				
1	8	1		+	+	+	+	+	+	+	+	+	+	+	
1	8	2	+	+	+	+	+	+	+	+	+	+	+	+	+
1	8	2			+	+	+	+	+	+	+	+	+		
1	8	3				+		+			+				
1	8	4				+		+			+				
1	8	5				+		+			+			+	
2	1	1		+	+	+	+	+	+	+	+	+	+	+	
2	1	3	+	+	+	+	+	+	+	+	+	+	+		
2	1	4	+	+	+	+	+	+	+	+	+	+	+	+	+
2	1	5	+	+	+	+	+	+	+	+					
2	2	4		+	+	+	+	+	+	+	+				
2	2	5			+	+	+	+	+	+			+		
2	3	1				+		+		+		+	+		
2	3	2	Questionnaire completed - no usable test results												
2	3	3				+		+	+						
2	3	4				+	+	+	+	+	+				
2	4	1	+	+	+	+	+	+	+	+	+	+	+	+	+

Appendix F (page 2 of 4)

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
2	4	2		+	+	+	+	+	+		+	+	+	+	
2	4	3									+				
2	4	4				+	+	+	+			+		+	
2	5	2		+	+	+	+	+	+	+	+	+	+	+	+
2	5	3		+	+	+	+	+				+			+
2	5	4	Criterion Referenced Test results only												
2	5	5				+			+		+	+			
2	6	2		+	+	+	+	+	+	+	+	+	+	+	
2	6	3		+	+	+	+	+	+	+	+	+	+	+	
2	6	4				+			+		+		+		
2	6	5				+			+		+				
2	6	5				+		+			+			+	
2	7	1	+	+	+	+	+	+	+	+	+	+	+	+	+
2	7	2	+	+	+	+	+	+	+	+	+	+	+	+	+
2	7	2		+	+	+	+	+	+	+	+	+	+	+	+
2	7	2		+	+	+	+	+	+	+	+	+	+	+	
2	7	3	+		+			+		+			+		
2	7	3			+	+	+	+	+	+	+	+			
2	8	2	+			+		+			+			+	
2	8	3		+	+	+	+	+	+	+	+				
3	1	1	+		+	+	+	+	+	+	+	+	+	+	+
3	1	2	+	+	+	+	+	+	+	+	+				
3	1	4	+	+	+	+	+	+	+	+	+	+	+	+	+
3	1	5			+	+	+	+	+	+	+		+	+	
3	2	1	+	+	+	+	+	+	+	+	+	+	+		
3	2	2	+	+	+	+	+	+	+	+	+				
3	2	3		+	+	+	+	+	+	+	+	+		+	
3	2	4	+	+	+	+	+	+	+	+	+	+	+	+	+
3	3	1	+	+	+	+	+	+	+	+	+	+	+	+	+
3	3	2	+	+	+	+	+	+	+	+	+				
3	3	3		+	+	+	+	+	+	+	+	+	+		
3	3	4	+	+			+		+						
3	3	5		+	+		+	+		+	+		+		
3	4	1		+	+		+	+							
3	4	2		+	+	+	+	+	+	+	+	+	+	+	
3	4	3	+	+	+	+	+	+	+	+	+				
3	4	4	+	+	+	+	+	+	+	+	+	+	+	+	
3	4	5	+	+	+	+	+	+	+	+		+			+
3	5	1					+		+			+			
3	5	3			+		+			+		+		+	
3	5	4		+	+	+	+	+	+	+	+	+	+	+	
3	5	5		+	+	+	+	+	+	+	+	+	+	+	
3	6	1				+			+		+				
3	6	2				+	+	+			+			+	
3	6	3				+	+	+	+			+		+	
3	6	4					+				+			+	
3	6	5					+				+			+	

Appendix F (page 3 of 4)

			Grade												
Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
3	7	1		+	+	+	+	+	+	+	+	+	+	+	
3	7	2		+	+		+	+		+	+		+		
3	7	2	+	+	+	+	+	+	+	+	+	+	+	+	+
3	7	2		+	+	+	+	+	+	+	+	+			
3	7	3			+			+		+		+			+
3	7	3		+	+	+	+	+	+	+	+	+			
3	7	3			+	+	+	+	+	+	+	+			
3	7	3		+	+	+	+	+	+	+	+				
3	7	3		+	+		+	+		+	+		+		
3	7	4	+	+	+	+	+	+	+	+	+	+	+		
3	7	4					+				+			+	
3	7	5			+	+	+	+	+	+		+			
3	8	2	+	+	+	+	+	+	+	+	+	+	+		
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	+
3	8	3		+	+	+	+	+	+	+	+	+	+	+	+
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
3	8	5			+		+		+		+			+	
4	1	1		+	+	+	+	+	+	+	+				
4	1	2	+	+	+	+	+	+	+	+	+	+	+	+	
4	1	3		+	+	+	+	+	+	+	+				
4	2	2		+	+	+	+	+	+	+	+			+	
4	2	4			+		+		+		+		+		
4	2	4	+		+		+		+		+				
4	2	5				+	+		+	+	+		+		
4	3	1	+	+	+	+	+	+	+	+	+	+	+	+	
4	3	2		+	+	+	+	+	+	+	+		+		
4	3	3		+	+	+	+	+	+	+	+	+	+	+	
4	3	4					+				+		+		
4	3	5		+	+	+	+	+	+	+					
4	4	1		+	+	+	+	+	+	+	+				
4	4	2		+	+	+	+	+	+	+	+	+	+	+	+
4	4	4			+	+	+	+	+	+	+				
4	4	5	+	+	+	+	+	+	+	+	+				
4	5	1		+	+	+	+	+	+	+	+	+	+	+	+
4	5	2	+	+	+	+	+	+	+	+	+	+			
4	5	3	Only Chapter I test data provided												
4	5	4		+	+	+	+	+	+						
4	5	5				+	+		+	+	+	+	+		
4	6	1			+	+	+	+	+	+	+	+	+		
4	6	2			+		+		+		+				
4	6	3		+	+	+	+	+	+						
4	6	4			+	+	+	+	+	+	+	+			
4	6	5	+	+	+	+	+	+	+	+		+		+	
4	7	1			+	+	+	+	+	+	+	+	+	+	
4	7	2			+	+	+	+	+	+	+	+	+	+	
4	7	3	+	+	+	+	+	+	+	+	+	+	+	+	

Appendix F (page 4 of 4)

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
4	7	3				+		+			+				
4	7	3	+	+	+	+	+	+	+	+	+	+	+	+	+
4	7	3		+	+	+	+	+	+	+	+	+	+	+	+
4	7	3	Criterion Referenced Test results only												
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	+
4	7	4				+	+		+		+	+	+	+	
4	7	4				+			+		+	+		+	
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	
4	7	5				+		+	+		+	+		+	
4	8	3		+	+	+	+	+	+	+	+	+			
4	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
4	8	3	+	+	+	+	+	+	+	+	+	+	+	+	+
4	8	4						+		+		+		+	
Totals		153	43	40	111	123	123	123	118	104	120	82	74	66	26

Appendix G

Stem-and-Leaf Distributions of District Reports of the Percentage of Students Scoring Above the National Median in Reading and Mathematics

Appendix G

Figure G-1

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 1

Reading

Stem	Leaf	Count
9 :	6	1
9 :	01	2
8 :	9	1
8 :	013	3
7 :	588	3
7 :	34	2
6 :	55689	5
6 :	012224	6
5 :	5567789	7
5 :	001224444	9
4 :	5579	4
4 :	0134	4
3 :	56689	5
3 :	0023	4
2 :	6	1
2 :		0
1 :		0
1 :		0

P90 = 81
P75 = 66
P50 = 55
P25 = 45
P10 = 35

Mathematics

Stem	Leaf	Count
9 :	589	3
9 :	3	1
8 :	9	1
8 :	034	3
7 :	55678	5
7 :	0113	4
6 :	6899999	7
6 :	001223344444	12
5 :	5899	4
5 :	012	3
4 :	669	3
4 :	34	2
3 :	88	2
3 :	02	2
2 :	89	2
2 :	2	1
1 :		0
1 :		0

P90 = 84
P75 = 71
P50 = 64
P25 = 51
P10 = 38

Appendix G

Figure G-2

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 2

Reading

Stem	Leaf	Count
9 :		0
9 :	12	2
8 :	577	3
8 :	0012	4
7 :	5799	4
7 :	12	2
6 :	555688899	9
6 :	0012344	7
5 :	56677788999	11
5 :	0122334444	10
4 :	557778899	9
4 :	111123344	9
3 :	999	3
3 :	1	1
2 :	99	2
2 :	2	1
1 :		0
1 :		0

P90 = 80
P75 = 68
P50 = 57
P25 = 47
P10 = 41

Mathematics

Stem	Leaf	Count
9 :	559	3
9 :	013	3
8 :	67	2
8 :	001334	6
7 :	5779	4
7 :	0001112222344	13
6 :	55566788889	11
6 :	000011222	9
5 :	56677889	8
5 :	001124	6
4 :	568	3
4 :	23	2
3 :	6	1
3 :	4	1
2 :		0
2 :		0
1 :	68	2
1 :		0

P90 = 86
P75 = 74
P50 = 67
P25 = 57
P10 = 46

Appendix G

Figure G-3

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 3

Reading

Stem	Leaf	Count
9 :		0
9 :	34	2
8 :	558	3
8 :	12	2
7 :	56799	5
7 :	0122344	7
6 :	677777789	9
6 :	00111224444444	14
5 :	5566677899	10
5 :	001233344	9
4 :	556889	6
4 :	001223	6
3 :	69	2
3 :	012223344	9
2 :	89	2
2 :	14	2
1 :	5	1
1 :		0

P90 = 78
P75 = 67
P50 = 58
P25 = 45
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :	123	3
8 :	7899	4
8 :	012224	6
7 :	88	2
7 :	000112244	9
6 :	556778888899	12
6 :	000123344444	12
5 :	55567788999	11
5 :	1222333444	10
4 :	556667899	9
4 :	0224	4
3 :	69	2
3 :	334	3
2 :		0
2 :	0	1
1 :		0
1 :	1	1

P90 = 82
P75 = 70
P50 = 61
P25 = 52
P10 = 42

Appendix G

Figure G-4

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 4

Reading

Stem	Leaf	Count
9	5	1
9	00	2
8	79	2
8	001	3
7	67799	5
7	00133444	8
6	6888	4
6	000022234	9
5	5567788899	10
5	01112222244	11
4	66777899	8
4	013444	6
3	5568889	7
3	12234444	8
2	7	1
2	1	1
1		0
1	1	1

P90 = 79
P75 = 68
P50 = 55
P25 = 44
P10 = 34

Mathematics

Stem	Leaf	Count
9	9	1
9	034	3
8	69	2
8	0033	4
7	589	3
7	024	3
6	5557777888889	13
6	0000012223344	13
5	55556667778	11
5	0011222222333344	16
4	55789	5
4	011224	6
3	5579	4
3		0
2		0
2		0
1		0
1	2	1

P90 = 81
P75 = 68
P50 = 59
P25 = 52
P10 = 42

Appendix G

Figure G-5

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 5

Reading

Stem	Leaf	Count
9 :		0
9 :	03	2
8 :	5	1
8 :	00112333	8
7 :	55578	5
7 :	0011223344	10
6 :	5699	4
6 :	00112224	8
5 :	666667788	9
5 :	0001122233	10
4 :	567888999	9
4 :	11244	5
3 :	55567799	8
3 :	02334	5
2 :	679	3
2 :		0
1 :	9	1
1 :		0

P90 = 80
P75 = 72
P50 = 56
P25 = 45
P10 = 34

Mathematics

Stem	Leaf	Count
9 :	6	1
9 :	0013	4
8 :	6	1
8 :	002234	6
7 :	55777899	8
7 :	02244	5
6 :	66677778888899	14
6 :	111122344444	12
5 :	556677899	9
5 :	002222244	9
4 :	5667888899	10
4 :	1344	4
3 :	57	2
3 :	2	1
2 :		0
2 :	2	1
1 :		0
1 :		0

P90 = 82
P75 = 73
P50 = 64
P25 = 52
P10 = 45

Appendix G

Figure G-6

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 6

Reading

Stem	Leaf	Count
9 :		0
9 :	2	1
8 :	69	2
8 :	0234	4
7 :	55556	5
7 :	0001234	7
6 :	5555589	7
6 :	0144	4
5 :	66677777889	11
5 :	001223334	9
4 :	555678999	9
4 :	0122234	7
3 :	56666677889	11
3 :	00024	5
2 :	69	2
2 :		0
1 :		0
1 :	2	1

P90 = 75
P75 = 65
P50 = 54
P25 = 42
P10 = 35

Mathematics

Stem	Leaf	Count
9 :	79	2
9 :	4	1
8 :	556	3
8 :	1444	4
7 :	556778	6
7 :	123	3
6 :	5566888999	10
6 :	022222222334444	15
5 :	55556667788999	14
5 :	0011123	7
4 :	5556677889	10
4 :	22244	5
3 :	89	2
3 :		0
2 :		0
2 :	3	1
1 :		0
1 :	2	1

P90 = 84
P75 = 69
P50 = 62
P25 = 50
P10 = 44

Appendix G

Figure G-7

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 7

Reading

Stem	Leaf	Count
9 :		0
9 :	0	1
8 :	7	1
8 :	13	2
7 :	55699	5
7 :	0004	4
6 :	57789	5
6 :	001112333	9
5 :	566778	6
5 :	0011223344	10
4 :	577799	6
4 :	0334	4
3 :	7778999	7
3 :	0024	4
2 :	68899	5
2 :		0
1 :		0
1 :	0	1

P90 = 75
P75 = 64
P50 = 54
P25 = 40
P10 = 30

Mathematics

Stem	Leaf	Count
9 :		0
9 :	0333	4
8 :	6	1
8 :	00034	5
7 :	8	1
7 :	003	3
6 :	6677777789	10
6 :	0011123334	10
5 :	5667777899	10
5 :	23344	5
4 :	56778889	8
4 :	00022234	8
3 :	66788	5
3 :		0
2 :	8	1
2 :		0
1 :	9	1
1 :		0

P90 = 80
P75 = 67
P50 = 59
P25 = 47
P10 = 39

Appendix G

Figure G-8

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 8

Reading

Stem		Leaf	Count
9	:		0
9	:		0
8	:	56	2
8	:	233	3
7	:	67889	5
7	:	001233	6
6	:	555667889	9
6	:	0011234	7
5	:	55567777899	11
5	:	011123334	9
4	:	5667778	7
4	:	0011244	7
3	:	667789	6
3	:	11233344	8
2	:	899	3
2	:		0
1	:	9	1
1	:		0

P90 = 77
P75 = 66
P50 = 55
P25 = 41
P10 = 33

Mathematics

Stem		Leaf	Count
9	:		0
9	:	1	1
8	:	57	2
8	:	002234	6
7	:	5666788	7
7	:	023334	6
6	:	56679	5
6	:	1111222233344	13
5	:	677788999	9
5	:	12444444	8
4	:	5589999	7
4	:	0133444	7
3	:	55666789	8
3	:	0044	4
2	:		0
2	:		0
1	:		0
1	:	01	2

P90 = 79
P75 = 70
P50 = 59
P25 = 45
P10 = 36

Appendix G

Figure G-9

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 9

Reading

Stem	Leaf	Count
9 :		0
9 :	2	1
8 :		0
8 :	3	1
7 :	779	3
7 :	2	1
6 :	6889	4
6 :	1113	4
5 :	566777789	9
5 :	00111113	8
4 :	566899	6
4 :	001112344	9
3 :	55668	5
3 :	22344	5
2 :	8	1
2 :	014	3
1 :	6	1
1 :		0

P90 = 69
P75 = 58
P50 = 50
P25 = 39
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :		0
8 :	6699	4
8 :		0
7 :	559	3
7 :	1233	4
6 :	5777	4
6 :	00012234	8
5 :	589	3
5 :	00011344	8
4 :	5568999	7
4 :	12344	5
3 :	669	3
3 :	0034	4
2 :	79	2
2 :	01	2
1 :		0
1 :		0

P90 = 75
P75 = 65
P50 = 53
P25 = 44
P10 = 30

Appendix G

Figure G-10

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 10

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :		0
8 :	4	1
7 :	5	1
7 :	00334	5
6 :	568	3
6 :	00123	5
5 :	667	3
5 :	02344	5
4 :	55677889	8
4 :	0133444	7
3 :	7789	4
3 :	01344	5
2 :	578	3
2 :	0	1
1 :	5	1
1 :		0

P90 = 71
P75 = 61
P50 = 48
P25 = 38
P10 = 29

Mathematics

Stem	Leaf	Count
9 :		0
9 :	0	1
8 :	55	2
8 :	011	3
7 :	56	2
7 :	02	2
6 :	559	3
6 :	0114	4
5 :	556777789	9
5 :	134	3
4 :	689	3
4 :	1233334	7
3 :	5678888	7
3 :	04	2
2 :		0
2 :		0
1 :		0
1 :	0	1

P90 = 80
P75 = 65
P50 = 55
P25 = 43
P10 = 36

Appendix G

Figure G-11

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 11

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :	6	1
8 :	0	1
7 :	579	3
7 :	0144	4
6 :	5	1
6 :	011223	6
5 :	678	3
5 :	001123344	9
4 :	567	3
4 :	113	3
3 :	55889	5
3 :	123	3
2 :	7779	4
2 :	1	1
1 :	9	1
1 :	0	1

P90 = 75
P75 = 62
P50 = 52
P25 = 38
P10 = 27

Mathematics

Stem	Leaf	Count
9 :	6	1
9 :		0
8 :		0
8 :	023	3
7 :	599	3
7 :	22	2
6 :	67899	5
6 :	01233334	8
5 :	66889	5
5 :	00	2
4 :	578	3
4 :	244	3
3 :	5558999	7
3 :	114	3
2 :		0
2 :		0
1 :		0
1 :	0	1

P90 = 79
P75 = 68
P50 = 59
P25 = 42
P10 = 35

Appendix G

Figure G-12

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 12

Reading

Stem		Leaf	Count
9	:		0
9	:		0
8	:		0
8	:		0
7	:	79	2
7	:	24	2
6	:		0
6	:	2	1
5	:	888	3
5	:	011	3
4	:	88	2
4	:	011	3
3	:	6	1
3	:	3	1
2	:	7	1
2	:	1	1
1	:		0
1	:	3	1

P90 = 75
P75 = 58
P50 = 50
P25 = 40
P10 = 21

Mathematics

Stem		Leaf	Count
9	:	5	1
9	:		0
8	:		0
8	:		0
7	:		0
7	:	02	2
6	:	789	3
6	:	0	1
5	:	77	2
5	:	4	1
4	:	5589	4
4	:	14	2
3	:	6	1
3	:	4	1
2	:		0
2	:		0
1	:		0
1	:	0	1

P90 = 71
P75 = 67
P50 = 55
P25 = 45
P10 = 35

COMPARING STATE AND DISTRICT TEST RESULTS TO NATIONAL NORMS: INTERPRETATIONS OF SCORING "ABOVE THE NATIONAL AVERAGE"

CSE Technical Report 308

Robert L. Linn, M. Elizabeth Graue, Nancy M. Sanders
University of Colorado
UCLA Center for Research on Evaluation, Standards, and Student Testing

January, 1990

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred.

This paper was written as part of a research project sponsored by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST). It was prepared for presentation at the Annual Meeting of the American Educational Research Association, San Francisco, March, 1989.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

Acknowledgements

We thank the many state and district directors of testing and other staff who responded to our requests for data and took the time to complete the questionnaires and participate in the telephone interviews. Without their cooperation this study would not have been possible. A number of students worked part time on various aspects of collection, organization, computer entry, and analysis of the data and we are grateful for that assistance. We especially thank Frank Beaty, William Carlin, Ann Carrington, Sharon Catto, Jolene Dehning, Roy Grimm, and Kelly Strong in this regard. We also thank the members of the project advisory committee (David Bayless, Stan Bernknopf, Gordon Ensign, H.D. Hoover, Richard Jaeger, Daniel Koretz, and Lynn Winters) for their many helpful suggestions. Finally, we thank Eva Baker, Leigh Burstein, Stephen Dunbar, and Lorrie Shepard for their helpful comments on earlier drafts of this report.

Introduction

It has become commonplace for a state or district to report that its students are "scoring above the national average." Indeed, it has been suggested that all 50 states and most districts are reporting above average achievement test scores (Cannell, 1987). Is it really the case that all states claim that their students are performing above average on achievement tests? If so, how should such results be interpreted?

These are two of several questions that motivated a study of (a) norm-referenced test results that are being reported by states and school districts and (b) factors related to those scores. This report presents part of the findings of that study. Published reports and results of mail and telephone surveys of states and a nationally representative sample of school districts were used to document the degree to which "above average" achievement test results are being presented. Analyses of the possible influence of the changing meaning of norms are also presented. Subsequent reports will address a number of other factors that may have an impact on the achievement test scores of states and districts and on the proper interpretation of those results.

Background

Standardized achievement tests have long been used by schools to report student achievement to parents, policy makers, and the general public. In recent years, however, the attention given to test scores has increased dramatically. Low-stakes testing programs with results returned to teachers and reported in a low-key fashion to school boards and interested parents have given way to high-stakes testing programs that have direct and important effects on students, teachers, and school administrators. The increased emphasis on the use of test results for purposes of accountability has made questions of test quality and the trustworthiness of interpretations of major concern to educators and policy makers.

A common, albeit not the only or necessarily the best, way of providing the various audiences a means of interpreting test scores is to compare achievement test scores for a school building, a district, or a state to national norms. Slightly over half of the states and a substantial majority of the school districts rely on off-the-shelf, standardized achievement tests, for which normative comparisons provide a primary basis of interpretation. These comparisons take on a wide variety of forms, including the average grade equivalent score, the average normal curve equivalent score, the median percentile rank or percentile rank of the mean, the proportion of students scoring above the "national average," or more precisely, the national median, and the proportions of students with "below average, average, or above average" scores where the three categories correspond to stanines 1 thru 3, 4 thru 6, and 7 thru 9, respectively. In each of these examples, national norms provide the primary basis of comparison.

Norms, of course, are not the only basis of test score interpretation. Some states and districts rely on criterion-referenced interpretations of either publisher- or locally developed tests. In such cases, comparisons to past performance provide a key means of interpretation. For example, trends in the proportion of students passing a minimum-competency test, the proportion of students mastering specific objectives, or the average number of objectives mastered provide a means of comparing the current year's achievement with a benchmark. Trends may also be important in the interpretation of norm-referenced results, but the national norm still provides the major frame of reference for expressing the scores. Even states with locally developed or customized assessment programs sometimes also use comparisons to national norms to aid the interpretation of their achievement test results; these comparisons are obtained through special equating studies or item response theory links.

The pros and cons of normative comparisons have been discussed on many occasions. Discussions of appropriate and inappropriate normative interpretations are provided, for example, by Angoff (1971), Petersen, Kolen, and Hoover (1989), and in several introductory texts on educational and psychological measurement. Good discussions of appropriate and inappropriate uses and interpretations of norms may also be found in the technical manuals and interpretive guides provided by the publishers of the major standardized achievement tests.

Despite these discussions, normative interpretations continue to be misused and misinterpreted. The distinction that Angoff (1971) and others have made between the statistical meaning of "normative," which refers to "performance as it exists," and the use of the term to refer to "standards or goals of performance" (p. 533) is too often overlooked. The fact that norms for school averages or district averages differ markedly from norms for individual students is too often ignored or is given insufficient emphasis in interpretation. Because a school average is based on a range of student scores it necessarily falls somewhere in between the score of the highest scoring individual student and that of the lowest scoring student. Consequently, the distribution of school average scores is less variable than the distribution of individual student scores. The average achievement score that corresponds to the 70th percentile using school building norms, for example, may correspond to only the 60th percentile using norms for individual students.

It is widely believed that some tests have "easier" norms than others. If the norms of test A are easier or less stringent than those of test B, then a given level of achievement would be expected to appear better (e.g., result in a higher percentile rank or a larger proportion of students scoring above the national average) with test A than with test B. Note that the difficulty of norms is different than the intrinsic difficulty of test items. A test that asked easy questions could have hard norms because the norming sample was unusually able in the content area of the test. Conversely, a second test that asked relatively more difficult questions could have easier norms because the norming sample for the second test included a disproportionate number of low achieving students. The relative difficulty of norms for a particular school, school district, or state may also depend on the degree to which the test content matches the curriculum at the building or classroom levels.

The meaning of norms depends fundamentally on the definition of the reference population, and secondarily on the adequacy of sampling, the level of participation, and the motivation of the students in the norming sample, among other considerations. The year in which the norms were obtained is one of the important properties that define the reference population and it is clearly the case that norms become dated. If achievement is improving nationally, then the use of old norms will make a district or state appear to be doing better relative to the nation than would the use of current norms that provide a higher standard of comparison.

Although the above concerns about the use of norms are hardly new, questions about the meaning and trustworthiness of normative comparisons that states and districts are using to communicate test results to policy makers and the public have recently taken on increased importance. The increased importance is due, in part, to escalation in the stakes involved in testing. Concerns about normative comparisons were also exacerbated by the publication of a report by Dr. John J. Cannell (1987) titled "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average."

The Cannell report is based on a survey conducted by a community group, the Friends of Education, which found that "no state scores below the publisher's 'national norm' at the elementary level on any of the six major nationally normed, commercially available tests" (Cannell, 1987, p. 2, emphasis in original). Based on this finding, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press inflated and misleading reports on achievement levels" (p. 2).

Cannell was not the first to notice that states were reporting results that were above the national norm in greater numbers than would be expected based on past experience or common-sense notions of the likely relative standing of particular states. In 1984, the Southern Regional Education Board (SREB) reported that 9 of 11 SREB states with norm-referenced test results for elementary grades were at or above the national average (SREB, 1984). Two years later, "[i]n June, 1986, SREB first described this situation in which student achievement in nearly all states was reported to be at or above the national averages as the 'Lake Wobegon effect'—descriptive of Garrison Keillor's mythical town where all children are above average" (Korcheck, 1988, p. 3). However, it was the Cannell report that placed the issue in the national limelight.

The Cannell report attracted a good deal of attention in the press when it was released in the fall of 1987 and has been the focus of considerable debate and controversy among professional educators and measurement specialists ever since. There are undoubtedly a number of factors that helped focus attention on the findings. Dramatic statements regarding the findings such as those illustrated in the above quotes may be part of the reason. Interest in the report was probably enhanced also by the sharp criticisms of test publishers ("we believe inaccurate initial norms are the reason for high scores", p. 5, emphasis in original), of educators for the "integration of unchanging test questions into the curriculum" (p. 5, emphasis in the original), of those responsible for reporting student achievement ("no state publication honestly described norm-referenced testing," p. 6), of university and public educators serving as consultants to test publishers "who too often are mere sycophants, giving the commercial interests what they want" (p. 9), and of the U.S. Department of Education, "whose lack of knowledge of these tests constitutes nonfeasance" (p. 9, emphasis in original).

Even without the dramatic language and sharp criticism, however, the Cannell report raises serious questions and issues. The percentage of students reported to be scoring above the national 50th percentile in a number of states seems to defy common sense.

The Cannell report has been the focus of considerable discussion at national meetings and in professional journals concerned with issues of educational achievement and measurement. It was a major topic, for example, at the 1988 and 1989 Annual Assessment Conferences sponsored by the Educational Commission of the States. The report was featured along with six commentaries from test publishers and representatives of the U.S. Department of Education in the Summer 1988 issue of *Educational Measurement: Issues and Practice.* The report also led the U.S. Department of Education to arrange a meeting involving Dr. Cannell, representatives of major test publishers, and selected academics to discuss the findings and their implications in February, 1988.

Reviewers of the Cannell report (e.g., Drahozal & Frisbie, 1988; Koretz, 1988; Lenke & Keene, 1988; Phillips & Finn, 1988; Qualls-Payne, 1988; Stonehill, 1988; Williams, 1988) identified a number of factors, some of which were also suggested by Cannell, that might contribute to the seemingly anomalous finding that all states are above the national average. The fact that norms become dated was probably the most frequently mentioned potential explanation. Differences in the rules for exclusion of students from testing in norming and in operational testing programs was also proposed as a possible explanation by several reviewers (e.g., Drahozal & Frisbie; Koretz; Lenke & Keene; Phillips & Finn). Other suggested partial explanations included the possible effect of a closer match between the test and the local curriculum in operational testing programs than in norming samples (e.g., Koretz; Lenke & Keene; Phillips & Finn), and the possibilities that poor security, familiarity with the specific content of tests that are reused year after year, or teaching the test may inflate scores (e.g., Drahozal & Frisbie; Koretz; Phillips & Finn).

Reviewers (e.g., Drahozal & Frisbie, 1988; Koretz, 1988; Lenke & Keene, 1988; Phillips & Finn, 1988; Williams, 1988) also identified several shortcomings of the Cannell study and interpretations. The failure to distinguish between group and individual student norms in interpretations, aggregation bias that results when the percentage of <u>districts</u> with average scores above the national median is used to make inferences about the percentage of <u>students</u> with scores above the national median, and the treatment of the percentage of students at the 4th stanine or above as if it were an indicator of the percentage of students above the national average are among the misleading analyses and interpretations that were identified.

Despite these and other limitations, some reviewers concluded that Cannell's major findings are still probably correct. Stonehill (1988), for example, stated simply that "Cannell's evidence is compelling" (p. 23). Others were more circumspect. Koretz (1988), for example, noted that "Dr. Cannell's errors are to some extent beside the point...for they are not sufficient to call into question his basic conclusion" (p. 11), and Phillips and Finn (1988) stated that in the absence of "evidence to the contrary" they generally concurred with "the central finding of Dr. Cannell's report" (p. 10).

Procedure

The Cannell study provided part of the stimulus for the present study. Certainly the issues raised in that study are important ones that deserve to be investigated in greater detail. Of particular concern were the issues of aggregation bias, the sampling of districts to obtain estimates for states without statewide testing programs that provide normative comparisons to the nation, and the type of information obtained from districts. The Cannell study only asked districts whether their students were above or below the national average. More detailed district results would be more informative. Since the Cannell study did not include results for secondary schools, it was also important to expand the coverage to all elementary and secondary school grades.

Our interest, however, was in more than simply obtaining estimates of the number of states or the proportion of districts that report achievement test results that are above the national median or that have average achievement above the national mean. Such statistics are of interest, but are apt to raise more questions than they answer. It is evident that we also need to better understand the ways in which states and districts are using normative comparisons, the validity of those comparisons, and the factors that influence the results and the validity of test scores and their interpretation. Therefore, the present study was designed to collect data not only about the achievement scores that were reported by states and districts, but on a variety of related issues, including the way in which test results were used (e.g., public reporting, grade retention, school incentives), when and why the uses were initiated, how and when the tests were adopted, and policies regarding test administration, test security and the preparation of students for taking tests. The present report, however, is focused on the test results and the possible influence of changes in the stringency of norms over time. Other aspects of the project data are addressed elsewhere (e.g., Baker, 1989; Burstein, 1989; Shepard, 1989).

State Survey

Two national mail and telephone surveys were conducted. In the first survey, a letter and a data collection form (see Appendix A) were mailed to the directors of testing in all states. As can be seen in the sample copy in Appendix A, the state testing directors were asked to provide test results in reading and mathematics for all grades (K through 12) for the three most recent academic years (1985-86, 1986-87, and 1987-88).

States were asked to report the percentage of students scoring above the national 50th percentile statewide if the information was available. When it was not available, the states were asked to report state means and standard deviations in reading and mathematics as well as the scores corresponding to the 25th, 50th and 75th percentiles statewide. In addition to test score information, the states were asked to provide the name, edition, and form of the test used at each grade; the year the test was first used in the state; the year it was normed; the month of administration; and the way the scores were routinely reported (e.g., percentage of students above the national median). The number of students enrolled, the number tested, and the number for whom scores were reported were also requested at each grade for each of the three years in question.

Since much of the information we were seeking was already available in published reports, the state directors of testing were asked to send copies of reports containing the requested information. The reports served in place of completed data collection forms if the reports contained the necessary information. Since information about how scores are communicated to the public and how they are interpreted by the press was relevant to our interests, copies of press releases and newspaper articles about test results were requested.

Following the mailings, state directors of testing were contacted by telephone to arrange telephone interviews. Detailed results of the telephone interviews are presented in other reports of study results (see Shepard, 1989), hence only a brief description of the interview is presented here.

A copy of the telephone interview guide is shown in Appendix B. In addition to clarification questions about testing data requested on the data collection forms, testing directors were asked questions about test use, test selection, the alignment of curriculum with the test, about time spent on teaching tested objectives, about objectives given less time as a result of the test, about guidelines for test preparation, about typical and extreme practices in preparing students to take tests, and about test security practices and experience.

District Survey

A stratified random sample of districts designed to be representative of the fifty states was selected. The 1980 census data were used to stratify school districts by region, size, and socio-economic status (SES). The definitions of the levels of three stratification variables are provided in Table 1. As can be seen in Table 1, the three stratification variables, region, size, and SES, had four, eight, and five levels, respectively. Thus a total of 160 cells were defined. The SES index, which is defined in Table 1, was used to rank the school districts and then to define five strata such that approximately 15% of the students were in each of the two extreme strata (low and high), approximately 20% were in each of the adjacent strata (above and below average), and approximately 30% were in the average stratum.

Five districts were randomly selected for each cell where a sufficient number of districts was available according to the 1980 census. Five districts were available and selected for most cells; however, 15 of the cells were void and 39 of the cells had fewer than five districts. For example, there were no high SES districts with enrollments of 100,000 or more in the North/Central region and there was only one low SES district with an enrollment of 100,000 or more in the East region.

Table 1
Definitions of Stratification Variables Used to Sample School Districts

A. REGION. Region of the country was defined to have 4 strata.

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

B. SIZE. District enrollment, 1980 Census, 8 strata.

1. Less than 1,200
2. 1,200 to 2,499
3. 2,500 to 4,999
4. 5,000 to 9,999
5. 10,000 to 24,999
6. 25,000 to 49,999
7. 50,000 to 99,999
8. 100,000 or more

C. SES. Community socio-economic status index based on the 1980 census. SES equals the median family income in thousands of dollars plus 6 times the median years of education of the population 25 years old or older. SES used to define 5 strata. The labels of the strata and approximate percentage of students in each are:

1. Low (15%)
2. Below Average (20%)
3. Average (30%)
4. Above Average (20%)
5. High (15%)

The first of the randomly-ordered districts in each of the 145 non-void cells was selected for inclusion in the survey. Because achievement test results of large school districts have been the focus of considerable attention in recent years, we were particularly interested in obtaining better information about the achievement test results being reported by larger districts. Therefore, districts with enrollments of 50,000 or more were oversampled. With the oversampling of large districts, a total of 175 districts were selected for the sample. Appendix C lists the number of districts selected per cell.

After districts were selected, telephone calls were made to confirm that the district was still operating (had not, for example, been consolidated with another district since the 1980 census), to identify appropriate respondents who were responsible for the district testing program, and to obtain complete mailing addresses. Where a district no longer existed, the second listed district in the corresponding cell of the sampling design was selected as a replacement. Once addresses were obtained, letters (see Appendix D) and data collection forms were mailed.

A subsample of the districts was identified for telephone interviews, which were conducted following the mail survey (see Appendix E for a description of the procedures used to identify the interview subsample). Because telephone interviews were conducted with a subsample of the districts, two different letters requesting participation and two different data collection forms were sent to districts (see Appendix D). The same basic test data that were requested from states were also requested for all districts. Districts in the mail-survey-only subsample were also sent a brief questionnaire covering some of the interview questions about the use of test results and perceived effects of testing in the district (see Appendix D). Districts in the interview subsample did not receive a questionnaire, but were asked questions shown in the interview guide in the telephone survey (Appendix D).

Follow-up letters were sent to districts approximately three weeks and again six weeks after the initial mailing. If no response was received within three weeks after the second follow-up, attempts were made to reach respondents by telephone and urge them to respond to the survey. When district personnel declined to participate in the survey or could not be reached after repeated telephone attempts, the reason for the non-participation was recorded, and a substitute district was selected from the appropriate cell in the sampling design.

Results

States with Norm-Referenced Comparisons

A total of 35 states provided results that allowed norm-referenced comparisons for one or more grades in at least one of the three years for which data were collected (1985-86, 1986-87, and 1987-88). The remaining 15 states did not use tests with national norms. The 35 states for which norm-referenced comparisons were obtained are listed in Table 2 with an indication of the basis for the comparison and the grade levels for which test results were reported. The basis for comparisons to national norms for states that administered an off-the-shelf, norm-referenced test is obvious. However, in order to obtain estimates of the percentage of students scoring above the national median or the percentile rank of the state mean or median test score, it was sometimes necessary to convert scores from the form in which they were reported. For example, if the state reported mean grade-equivalent scores, those scores were converted to the corresponding percentile rank by reference to the test publisher's norms tables for individual pupils.

Table 2
States with Norm-Referenced Comparisons and
Grades Where at Least One Comparison is Available

State	Basis of Comparison*	Grades 1	2	3	4	5	6	7	8	9	10	11	12
Alabama	NRT	+	+		+	+		+	+		+		
Alaska	NRT	+	+	+	+	+	+	+	+	+	+	+	+
Arizona	NRT	+	+	+	+	+	+	+	+	+	+	+	+
Arkansas	NRT				+			+			+		
California	LINK			+			+		+				+
Colorado	NRT			+			+			+		+	
Delaware	NRT	+	+	+	+	+	+	+	+			+	
Georgia	NRT		+		+			+		+			
Hawaii	NRT			+			+		+		+		
Idaho	NRT						+		+			+	
Illinois	LINK			+			+		+				
Indiana	NRT	+	+	+			+		+	+		+	
Iowa	NRT	+	+	+	+	+	+	+	+				
Kentucky	NRT/LINK	+	+	+	+	+	+	+	+	+	+	+	+
Louisiana	NRT				+		+			+			
Maryland	NRT			+		+			+				
Mississippi	NRT	+			+		+						
Missouri	LINK			+			+		+		+		
Nevada	NRT			+			+						
New Hampshire	NRT				+				+		+		
New Mexico	NRT			+		+			+				
North Carolina	NRT	+	+	+			+		+				
North Dakota	NRT			+		+		+					
Oklahoma	NRT			+				+			+		
Oregon	LINK								+				
Rhode Island	NRT			+			+		+		+		
South Carolina	NRT				+	+		+		+		+	
South Dakota	NRT				+				+			+	
Tennessee	NRT		+			+		+		+			+
Texas	LINK	+		+		+		+		+			
Utah	NRT					+						+	
Virginia	NRT				+				+			+	
Washington	NRT				+				+		+		
West Virginia	NRT			+			+			+		+	
Wisconsin	NRT				+				+			+	
Number of States:	35	10	10	20	16	13	18	13	22	11	11	13	5

* NRT = Norm-Referenced Test   LINK = Equated to NRT
NRT/LINK = Some years based on NRT and others on LINK

Several of the states listed in Table 2 obtained normative comparisons indirectly by linking non-normed tests or state assessment results to a norm-referenced test through the use of special equating studies or the inclusion of norm-referenced test items with known item parameters in a customized test (see, for example, Yen, Green, & Burket, 1987, for a discussion of customized tests). States for which norm-referenced comparisons were obtained indirectly through such linkages are indicated in Table 2 by the word "LINK" in the column showing the basis of comparison.

Although comparisons to national norms either directly or through an equating link could be obtained for a total of 35 states in all, the number of comparisons varied substantially by grade level. As can be seen in Table 2, the largest number of states with results for any single grade was 22 at Grade 8. Grade 3, with 20 states, and Grade 6, with 18 states, were used for statewide testing nearly as often as Grade 8. However, there was no grade for which normative comparisons were available for a majority of the 50 states. Test results were reported by only 10 or 11 states at Grades 1, 2, 9, and 10; only 5 states reported normative test results for Grade 12.

Where possible, estimates of the percentage of students in a state who scored above the national median were obtained separately for each grade tested in reading and mathematics. Where estimates of the percentage of students above the national median could not be obtained, the state median percentile rank or the percentile rank corresponding to the statewide mean was used. Note that here, and throughout this report, it is the individual pupil norms, rather than norms for school buildings or school districts, that were used to determine percentile ranks. For some states, estimates of both the percentage of students above the national median and the median percentile rank or percentile rank of the statewide mean were available and used.

The number of states and the number of students for which estimates of the percentage of students above the national median were obtained are reported in Table 3 by year of test administration, test content, and grade. Parallel numbers are reported in Table 4 for states where estimates of the median percentile rank or the percentile rank of the statewide mean were obtained. The latter numbers were also used to obtain weighted mean percentile ranks for the states for which those results were obtained. In many cases the number of states and number of students in Tables 3 or 4 are the same for mathematics as for reading, because of the fact that both content areas were usually tested and a single number of students tested was reported for both tests. However, there are some differences (e.g., Grade 8 in Table 3), because results were available in reading but not mathematics for a given state.

Percentage of students above national median. The combined results for states of the percentage of students scoring above the national median are summarized in Figure 1. The percentages shown in Figure 1 are weighted by the number of students tested in each grade for the states reporting data for each of the three years for which data were collected. Thus each bar in the figure represents the percentage of students in the states that provided data in this form who scored above the national median for a given school year and a given grade in either reading or mathematics. For example, the first column for Grade 1, 1985-86, is based on the 281,734 first-grade students in the 7 states (see Table 3) that reported test results in this form; it shows that 54% of those students scored above the national median in reading.

The results in Figure 1 are consistent with the general results reported by Cannell (1987) in that the overall percentage of students above the national median was greater than 50 in all of the elementary grades in both reading and mathematics for each of the three years studied. The percentage above the national median was

Table 3
Number of States and Number of Students Contributing to Estimates of Percentage of Students Above National Median by Year, Test Content, and Grade

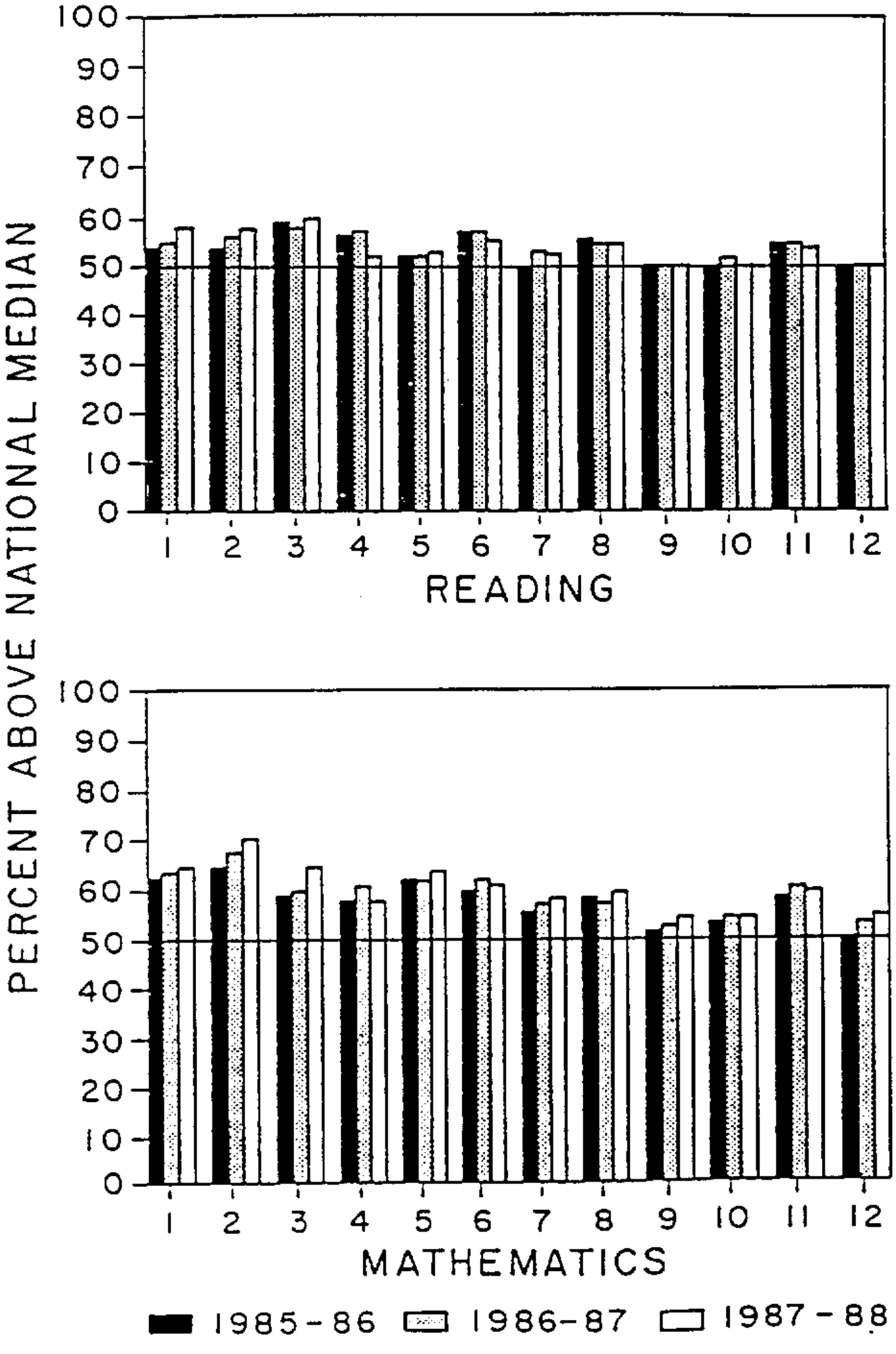
I. Reading

	1985-86		1986-87		1987-88	
Grade	Number of States	Number of Students	Number of States	Number of Students	Number of States	Number of Students
1	7	281,734	6	271,954	7	302,544
2	8	343,490	7	329,928	7	330,255
3	12	362,239	12	302,893	10	461,152
4	14	460,480	13	452,447	13	485,084
5	8	242,871	7	209,289	8	226,122
6	10	288,671	10	231,702	11	474,498
7	10	381,570	8	283,334	9	337,862
8	13	445,687	16	433,801	13	505,762
9	10	250,712	7	244,762	8	351,102
10	8	271,706	10	296,866	8	258,866
11	10	250,712	11	239,223	11	241,956
12	3	65,809	3	67,782	2	68,841
			II. Mathematics			
1	7	281,734	6	271,954	7	302,544
2	8	343,490	7	329,928	7	330,255
3	11	353,612	11	293,452	9	339,089
4	14	460,480	13	452,447	13	485,084
5	8	242,871	7	209,289	8	226,122
6	9	280,053	9	222,886	10	364,093
7	10	381,570	8	283,334	9	337,862
8	13	445,687	15	424,959	12	396,574
9	7	300,728	7	244,762	8	351,102
10	8	271,706	9	287,457	8	258,866
11	10	250,712	11	239,223	11	241,956
12	3	65,809	3	67,782	2	68,841

Table 4
Number of States and Number of Students Contributing to Estimates of Percentile Rank of State Means or Medians by Year, Test Content, and Grade

	1985-86		1986-87		1987-88	
Grade	Number of States	Number of Students	Number of States	Number of Students	Number of States	Number of Students
I. Reading						
1	5	250,628	5	264,972	6	295,840
2	6	308,342	6	323,318	7	385,391
3	11	623,579	12	336,372	12	394,641
4	11	389,954	12	446,642	13	509,839
5	7	206,325	8	250,586	11	336,191
6	8	526,312	8	245,215	11	391,526
7	8	317,994	8	281,849	11	401,015
8	11	403,406	16	471,619	14	468,180
9	6	295,903	6	239,606	8	348,617
10	6	236,868	9	291,311	8	253,699
11	9	246,555	10	234,746	10	237,583
12	3	276,030	2	65,120	2	68,841
II. Mathematics						
1	5	250,628	5	264,972	6	295,840
2	6	308,342	6	323,318	7	385,391
3	11	623,579	12	336,372	12	394,641
4	11	389,954	12	446,642	13	509,839
5	7	206,325	8	250,586	11	336,191
6	8	526,312	8	215,215	11	391,526
7	8	317,994	7	244,332	11	401,015
8	11	403,406	16	471,619	14	468,180
9	6	295,903	6	239,606	8	348,617
10	6	236,868	8	253,671	8	258,722
11	9	246,555	10	234,746	10	237,583
12	3	276,030	2	65,120	2	68,841

Figure 1
Percentage of Students Scoring Above National Median
Based on States Reporting (Weighted by Number of Students)



usually greater for mathematics than for reading. Percentages were usually higher for elementary than secondary grade levels. For Grades 1 thru 6, the percentage of students scoring above the national median in mathematics ranged from a low of 58% in Grade 4 for the 1985-86 school year to a high of 71% in Grade 2 for the 1987-88 school year, whereas the corresponding range for reading was from 52% (Grade 5, 1985-86) to 60% (Grade 3, 1987-88). For Grades 7 through 12, the percentage of students scoring above the national median ranged from 49% (Grade 12, 1985-86) to 60% (Grade 11, 1986-87) in mathematics and from 48% (Grade 9, 1986-87) to 55% (Grade 8, 1985-86) in reading.

It should be noted that while the percentages displayed in Figure 1 are generally above the naive expectation of 50%, many individual students were, in fact, receiving scores that were well below the national median. If a state reported that 55% of its students had scores at or above the national median, for example, it is obviously the case that the remaining 45% of the students in the state were receiving scores below the national median.

The results in Figure 1 provide only a very global picture since they combine the data for varying numbers of states at each grade level. They do not, for example, provide an indication of the variability from state to state. Some sense of the variability can be obtained from Figures 2 and 3, which show the distributions of the percentage of students above the national median in reading and in mathematics, respectively.

The data for the most recent year available for each state were used for the distributions in Figures 2 and 3, which for most states was the 1987-88 school year. Each point in Figures 2 and 3 represents the percentage of students in a state who scored above the national median in a particular grade.

As can be seen in Figure 2, there is considerable variability from state to state. The tendency for the percentages to be greater than 50 is quite evident for the elementary grades. However, there are some cases where the percentage is substantially below 50. It should be noted that the point in Figure 2 that is most out of line with the Cannell (1987) results is the Grade 4 reading point that corresponds to a state where only 33% of the students were reported to have scored above the national median. This state introduced a statewide test in 1987-88 and hence was not included in the results reported by Cannell.

The results shown in Figure 3 for mathematics show even greater state-to-state variability than was seen for reading. Consistent with the global results in Figure 1, the tendency for the percentages to be above 50 is more evident in mathematics than in reading. Some of the percentages in Figure 3 are extraordinarily high. Note, for example, Grade 2, where one state reported that 86% of the students scored above the national median. The only two examples of a state where the percentage is below 50 for Grades 1 through 6—the 41% at Grade 4 and the 49% at Grade 6—are both for the state that introduced statewide testing in 1987-88 and therefore was not included in Cannell's state-level data collection.

Median percentile ranks or percentile rank of state means. Since the percentage of students scoring above the national median could not be estimated for all states, the median percentile ranks or percentile ranks of state means were also analyzed. Figures 4, 5, and 6, which parallel Figures 1, 2, and 3, respectively, display the results of the latter analyses. In general, the results using these percentile rank statistics are quite similar to the results using the percentage of students scoring above the national median. This is so despite the differences in the properties of the two statistics and the fact that the two sets of analyses are based on different, albeit overlapping,subsets of states.

Figure 2
Percentage of Students Reported by States to be Scoring above the National Median in Reading (Each Point Represents a State)

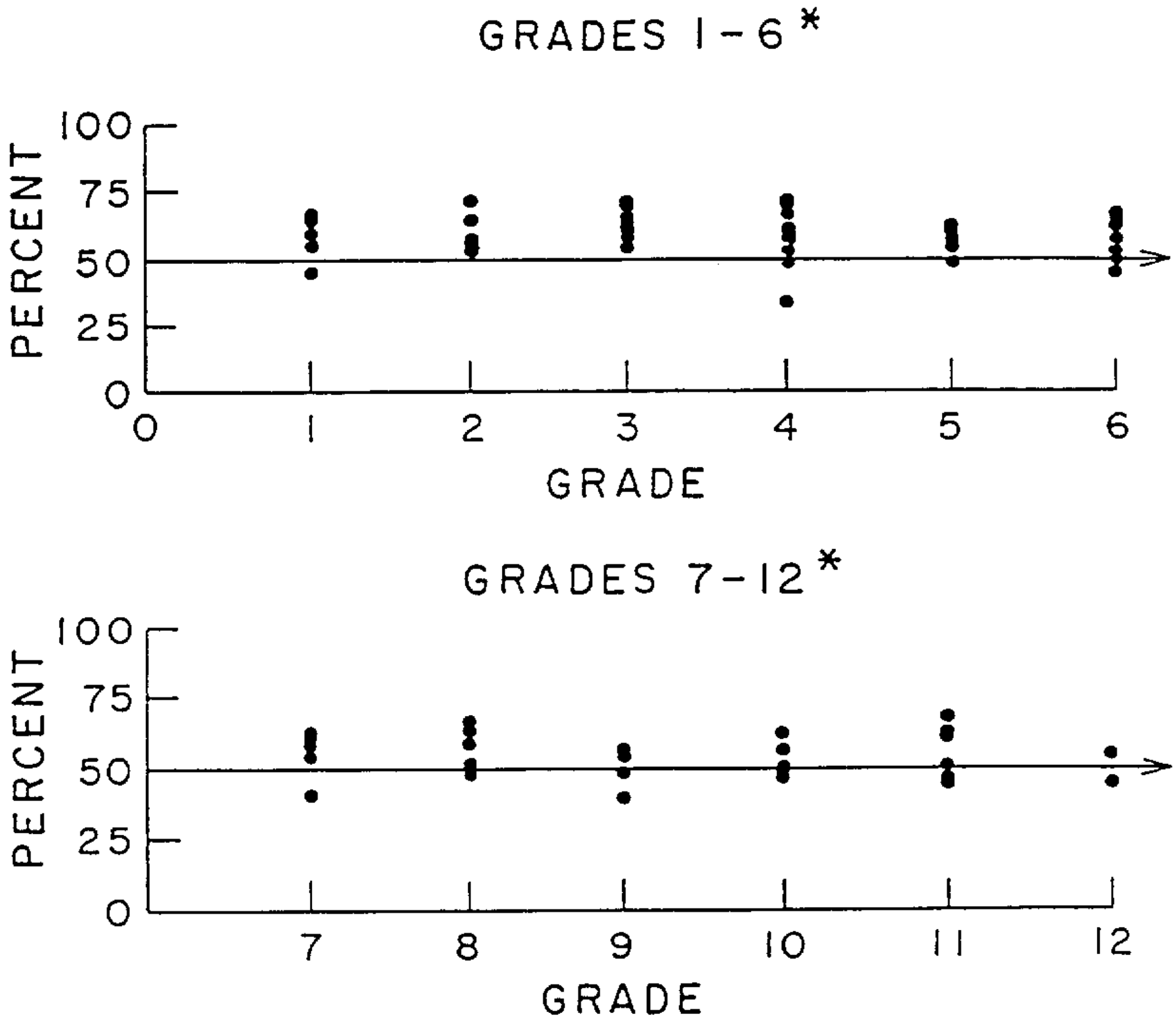
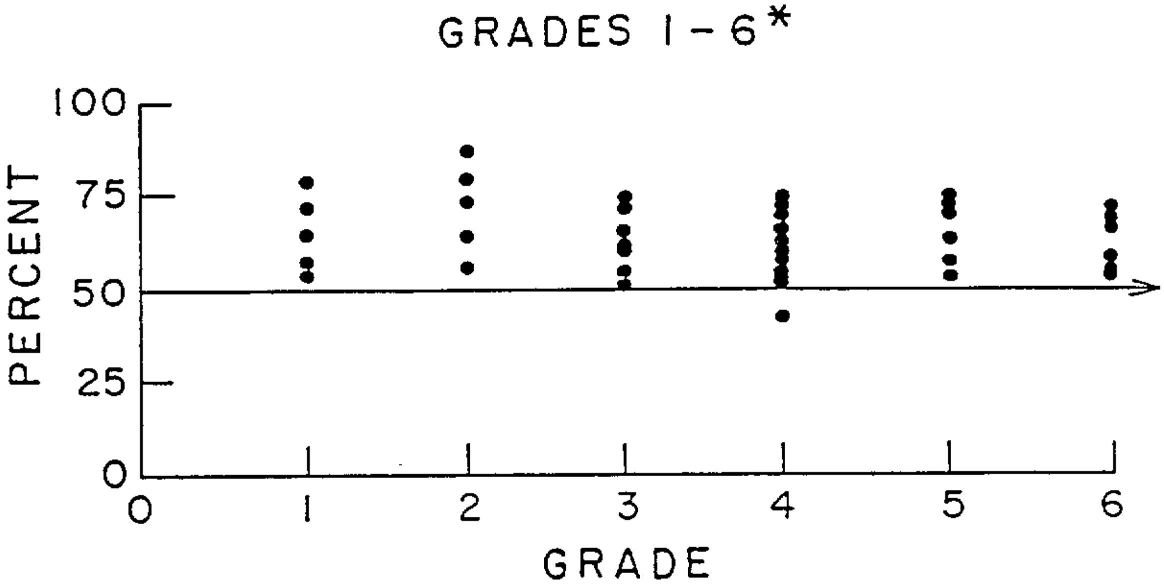


Figure 3
Percentage of Students Reported by States to be Scoring above the National Median in Mathematics (Each Point Represents a State)



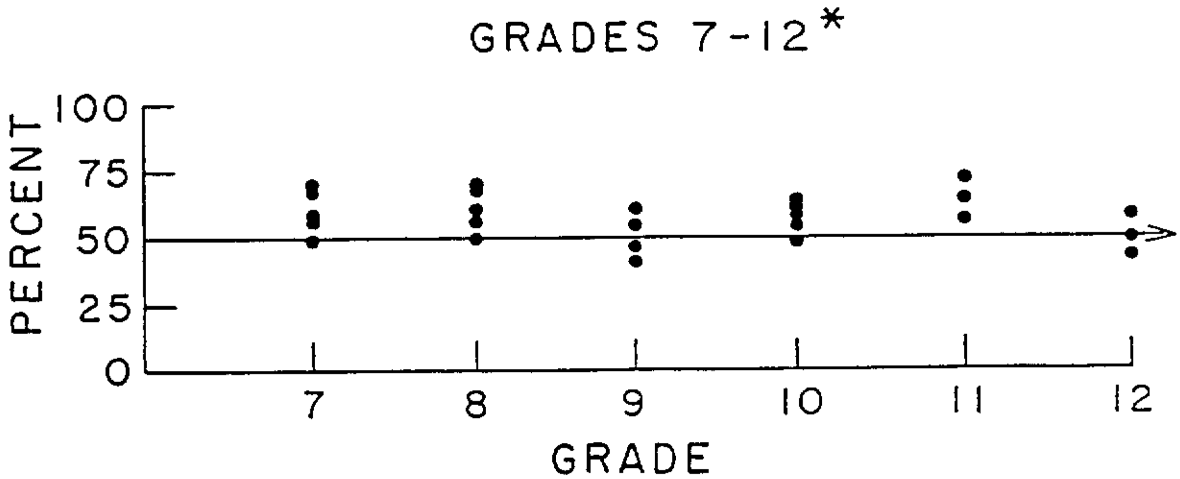


Figure 4
Weighted Mean of State Percentile Ranks

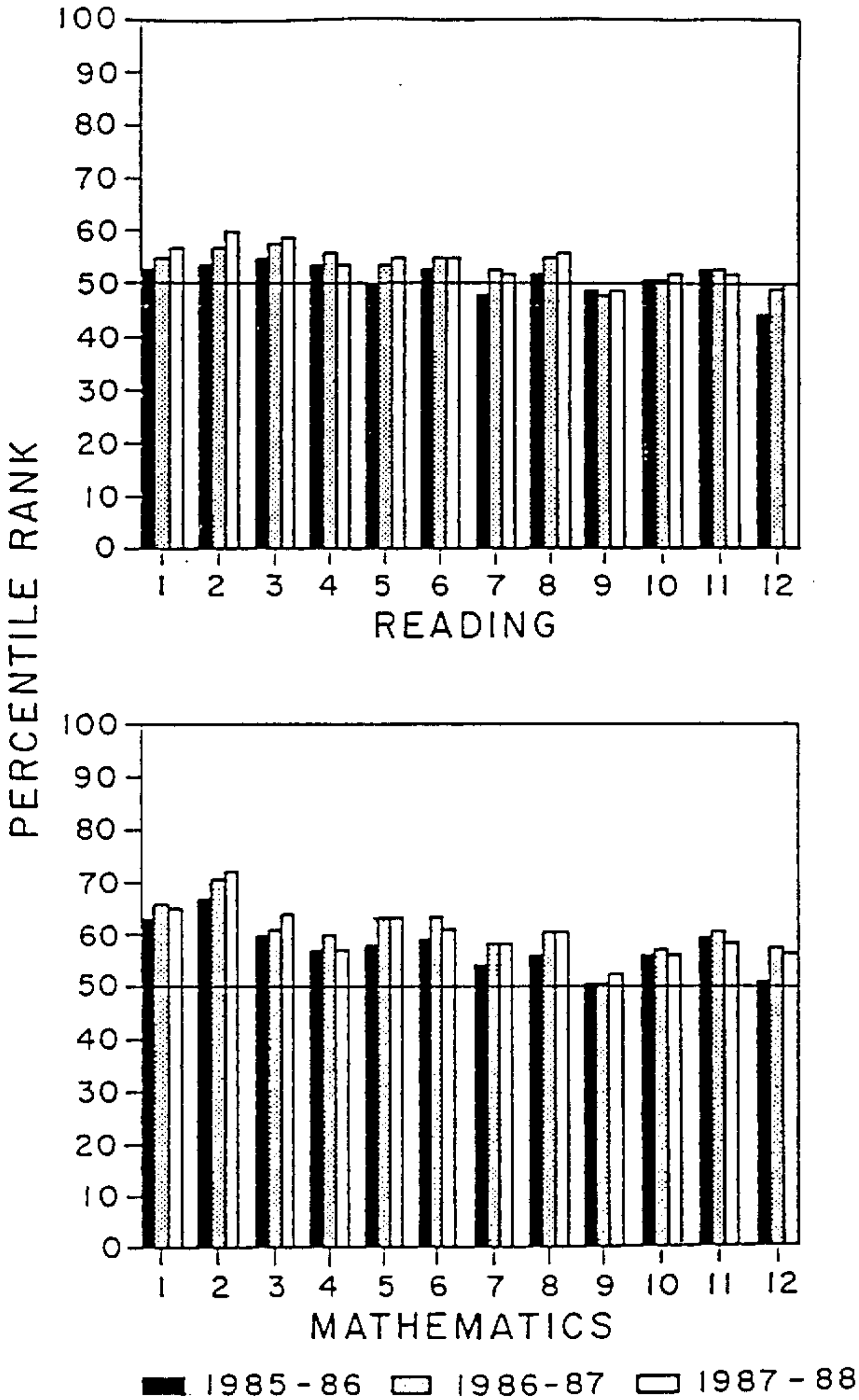
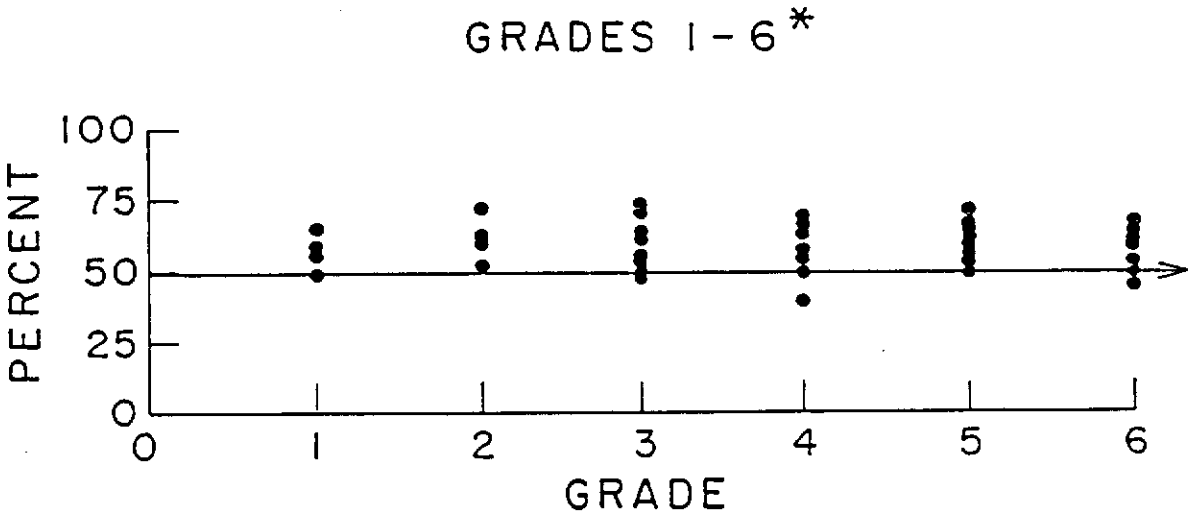


Figure 5
State Median Percentile Rank or Percentile Rank of
State Mean Test Score in Reading (Each Point Represents a State)



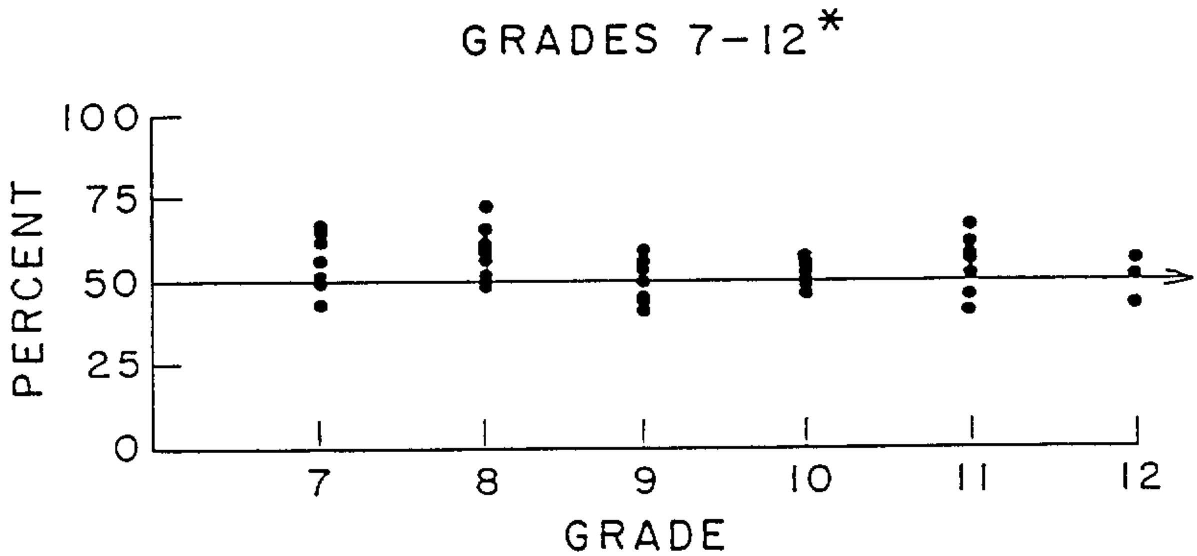
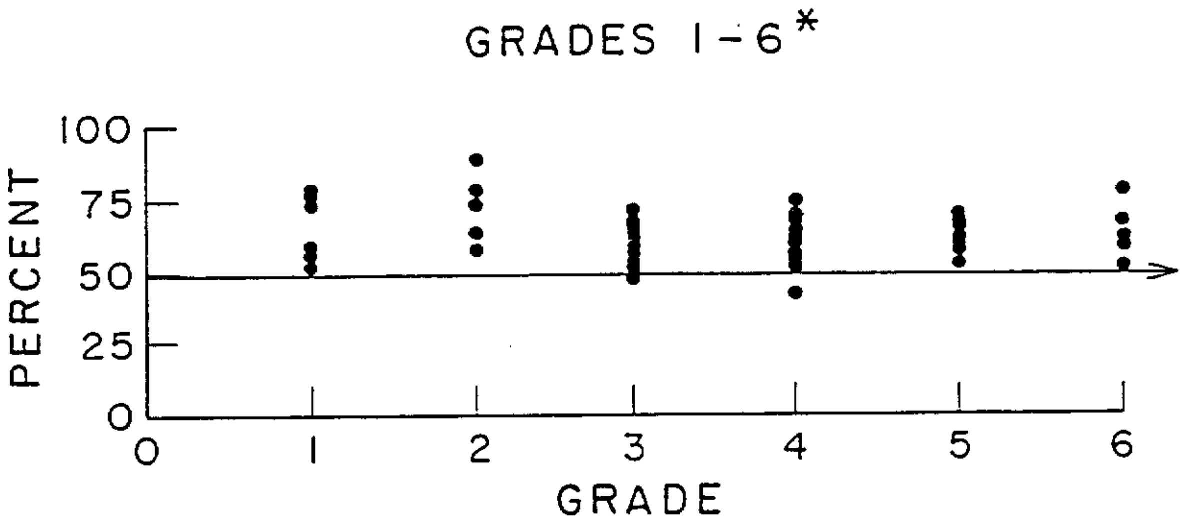
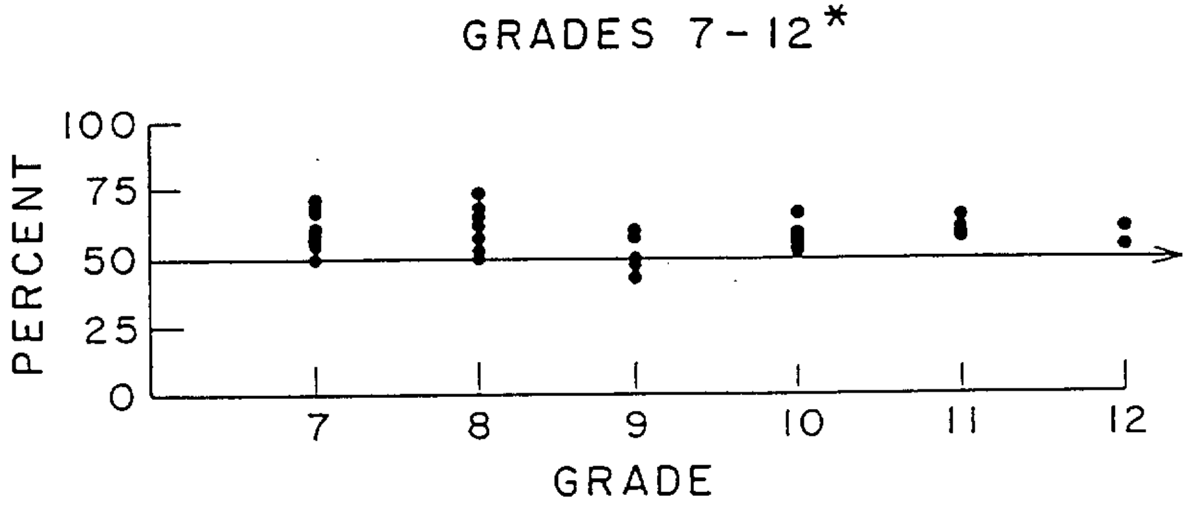


Figure 6
State Median Percentile Rank or Percentile Rank of
State Mean Test Score in Mathematics (Each Point Represents a State)





The conclusions (a) that most states are reporting results above the national average, (b) that the discrepancy is greater in mathematics than in reading, and (c) that the discrepancy is generally greater in the elementary grades than in the secondary grades do not depend on the use of a particular metric (e.g., the percentage of students above the national median). The same conclusions are supported by the use of the median percentile rank for each state or the percentile rank of the state mean.

Normative Comparisons Based on District Results

Data were obtained from 153 districts, or 87%, of the target of 175 districts. Appendix F provides a listing of the region, size, and SES of each of the 153 districts that returned questionnaires, provided reports on their testing programs, or completed telephone interviews. Districtwide norm-referenced test results were available for 148 of the 153 districts. For the remaining 5 districts, districtwide normative comparisons could not be obtained for the reasons indicated in Appendix F (e.g., only criterion-referenced results were available).

Also shown in Appendix F are the grades where norm-referenced test results were reported for each district. The grades where the largest number of districts reported norm-referenced test results are Grades 3, 4, 5, 6, and 8, in which test results were obtained for between 118 and 123 districts. As was shown in Table 2, those grades, with the exception of Grade 5, were also popular choices for statewide norm-referenced testing.

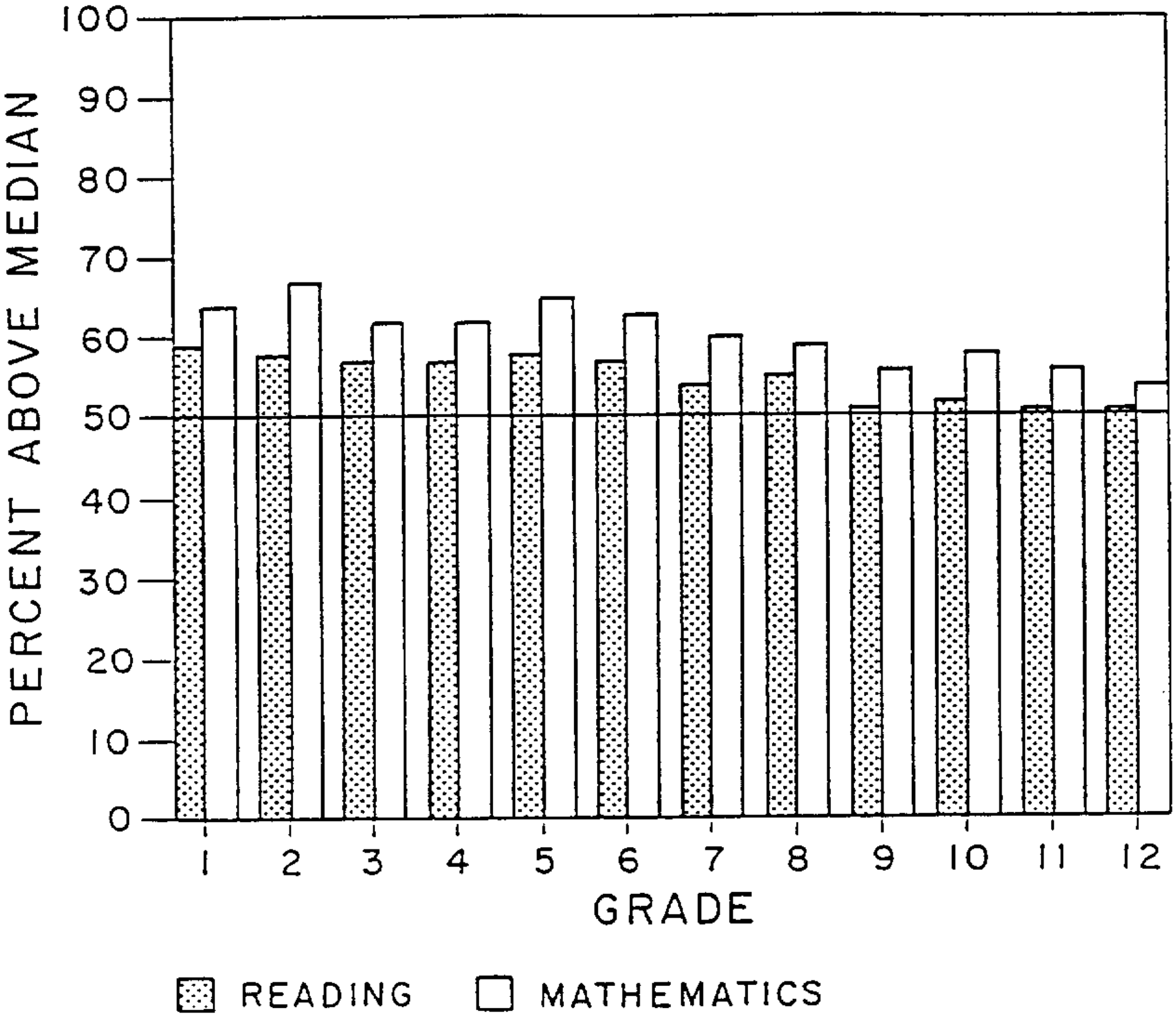
As was done for states, estimates of the percentage of students in a district who scored above the national median were obtained for each grade tested in reading and in mathematics whenever possible. Where these estimates could not be obtained, the district median percentile rank or the percentile rank corresponding to the district mean was used.

Estimates, based on the district data, of the percentage of students scoring above the national median in reading and mathematics for Grades 1 through 12 are plotted in Figure 7. The percentages plotted in Figure 7 are weighted by district size, region, and SES and thus are estimates of the percentage of students nationwide at a given grade that scored above the national median in reading or in mathematics. The number of districts on which these estimates are based varies by grade. The number of districts reporting data that could be used for the estimates in Figure 7 was 57, 77, 89, 87, 88, 85, 70, 84, 61, 52, 49, and 21 at Grades 1 through 12, respectively.

As can be seen, the estimated percentage of students scoring above the national median is consistently above 50%. For Grades 1 through 6, at least 57% of the students are estimated to have scores above the national median in reading. For mathematics, at least 62% of students are estimated to be above the national median Grades 1 through 6. In Grades 9 thru 12 the estimates of 51% or 52% for reading are closer to 50%; however, with the exception of Grade 12 with an estimate of 54%, the percentage of students estimated to have scores above the national median in mathematics is 56% or higher in every grade. Although 56% is obviously greater than 50%, it is still the case that nearly half the students (44%) received score reports below the national median when 56% scored above the median.

Figure 8 presents results that are parallel to those in Figure 7, that is, based on the data from districts where estimates of median percentile ranks or the percentile ranks of the district means were obtained. The weighted means of these percentile rank statistics are based on substantially fewer districts at each grade (the number of districts equaled 17, 27, 34, 29, 31, 27, 26, 29, 15, 16, 15, and 4 at Grades 1 through 12, respectively). Nonetheless, the results in Figure 8 lead to conclusions

Figure 7
Estimated Percentage of Students Scoring Above National Median
Based on District Results Weighted by Region, District Size and SES



that are essentially the same as those based on the estimated percentage of students above the national median. With the exception of Grade 12, where the number of districts reporting data in this form was extremely small, all of the weighted means are greater than 50. The results for the elementary grades are higher than those for the upper grades and the results for mathematics are higher than those for reading.

In addition to providing overall estimates of student performance levels, the district results provide a basis for investigating between-district variability and characteristics of districts associated with level of performance. Estimates of the percentage of students who scored above the national median in reading and mathematics were obtained for a majority of the districts that returned test results. Distributions of these percentages for districts were inspected at each grade level in both content areas. Since the complete distributions for all grades are rather voluminous, distributions for only one grade are presented and discussed in detail. Summaries of the distributions for other grades are provided and complete distributions for Grades 1 through 12 are included in Appendix G. Grade 3 was chosen for illustrative purposes since it was the earliest of the grades that were most frequently tested and reported by districts in the sample.

A total of 123 districts reported norm-referenced test results for Grade 3. Eighty-nine of those districts provided data that could be used to estimate the percentage of students scoring above the national median in reading and mathematics. The remaining districts reported data that could be used to obtain the median percentile rank or the percentile rank of the district mean, but did not provide a basis for obtaining the percentage of students scoring above the national median.

Distributions of district percentages of students scoring above the national median are illustrated by the stem-and-leaf plots in Figure 9. The "stem" corresponds to the tens digit of the percentage of students in a particular district that scored above the national median. The "leaf" reports the units digit for a district's percentage. The results for each district are depicted by a leaf (i.e., a single digit under the leaf column), that is associated with a particular stem which gives the tens digit for each leaf in that row. For example, one district reported that 93% of its students scored above the national median in reading and one district reported that 94% of its students scored above the median. Those two districts are depicted in the upper-left-hand corner of Figure 9 by the 34 under the leaf column next to a stem of 9. The lowest percentage above the median for reading that was reported by a district was 15%. The results for that district are indicated by the leaf of 5 next to a stem of 1 toward the bottom of the stem-and-leaf diagram for reading.

As can be seen in Figure 9, a majority of the districts reported that 50% or more of their students scored above the national median in reading (61 of 89 districts) and mathematics (69 of 89 districts). Only 16 of the 89 districts reported that less than 40% of their students scored above the national median in reading, but there were 12 districts that reported that three-fourths or more of their students scored above the national median. In mathematics the results show even larger numbers of districts that reported a substantial majority of their students above the median.

In order to summarize the distributions of district percentages of students reported to have scored above the national median, the 10th, 25th, 50th, 75th, and 90th percentiles of the distributions were obtained. For Grade 3, those percentiles are reported at the bottom of the two columns of Figure 9. (Parallel results for the other grades are presented in Appendix G.) These figures indicate, for example, that 10% of the districts reported that 32% or fewer of their third-grade students scored above the national median in reading. On the other hand, the 90th

Figure 8
Means of District Percentile Ranks Weighted by Region, District Size, and SES

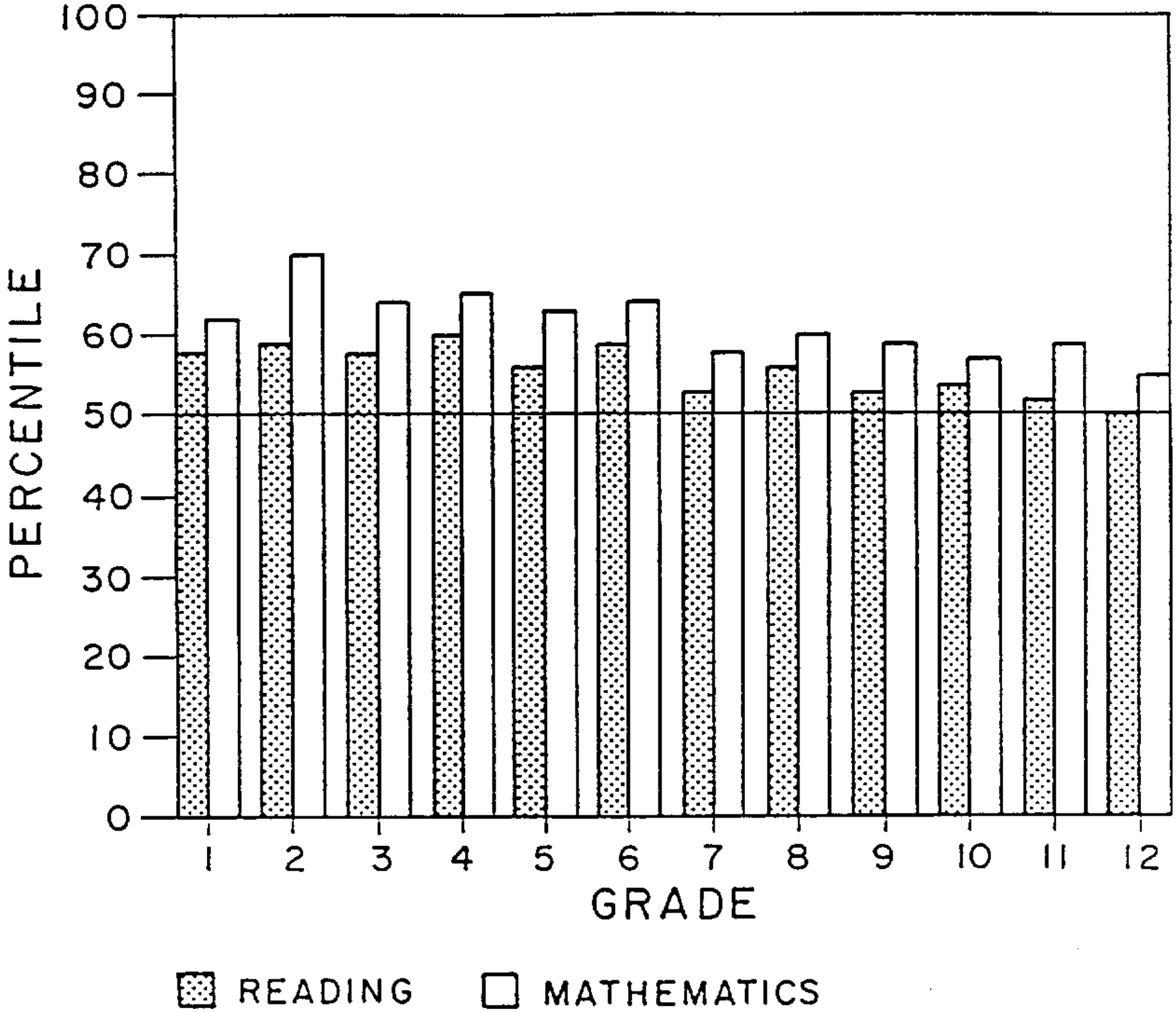


Figure 9
Stem-and-Leaf Distribution of the District Percentages of Students Scoring above the National Median at Grade 3

Reading

Stem	Leaf	Count
9 :		0
9 :	34	2
8 :	558	3
8 :	12	2
7 :	56799	5
7 :	0122344	7
6 :	677777789	9
6 :	00111224444444	14
5 :	5566677899	10
5 :	001233344	9
4 :	556889	6
4 :	001223	6
3 :	69	2
3 :	012223344	9
2 :	89	2
2 :	14	2
1 :	5	1
1 :		0

P90 = 78
P75 = 67
P50 = 58
P25 = 45
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :	123	3
8 :	7899	4
8 :	012224	6
7 :	88	2
7 :	000112244	9
6 :	556778888899	12
6 :	000123344444	12
5 :	55567788999	11
5 :	1222333444	10
4 :	556667899	9
4 :	0224	4
3 :	69	2
3 :	334	3
2 :		0
2 :	0	1
1 :		0
1 :	1	1

P90 = 82
P75 = 70
P50 = 61
P25 = 52
P10 = 42

percentile of 78 indicates that 10% of the districts reported that over three-fourths of their third-grade students scored above the national median in reading.

The five selected percentiles (10th, 25th, 50th, 75th, and 90th) of the district distributions of the percentage of students scoring above the national median were computed for all twelve grades. Those percentiles are shown in the box-and-whisker plots displayed in Figures 10 and 11 for reading and mathematics, respectively. Looking, for example, at the Grade 1 box-and-whisker plot for reading in Figure 10, it can be seen that the 10th percentile for the 57 districts reporting data at Grade 1 was 35, indicating that 1 district in 10 reported that 35% or less of its students scored above the national median. From the remaining percentiles for the Grade 1 reading results it can be seen that one district in four reported 45% or less of its students scored above the national median, half the districts reported 55% or less, three districts in four reported 66% or less, and nine districts in ten reported 81% or less.

From an inspection of Figure 10, it can be seen that districts at the 50th percentile reported that more than half (54% to 58%) of their students scored above the national median in reading in Grades 1 thru 8. Only at Grade 10 did a district at the 50th percentile report that slightly less than half (48%) of its students scored above the national median in reading. For the elementary grades, the tendency to have more than half of the students in a district scoring above the national median is much stronger in mathematics (Figure 11) than in reading (Figure 10). In Grades 1 thru 6, for example, the 25th percentile is equal to or above 50. In other words, three-quarters of the districts had more than half their students scoring above the median. Moreover, half the districts had 59% or more of their students above the national median in mathematics for Grades 1 thru 8.

The percentage of districts that had more than half of their students scoring above the national median should not be interpreted as a direct indication of the percentage of students across districts who were scoring above the median. It would be possible, for example, for a substantial majority of districts to have more than half their students above the median while less than half of all students across districts were above the median. Nonetheless, it is clear that it is more common for a district to report test results that are "above average" than ones that are "below average."

The district results provide support for the general finding that it is more common to have students scoring above the national median than it is to have them scoring below the median. However, there are more exceptions to this rule, particularly in reading, than were suggested by the Cannell study, which reported that 169 of 188 districts were "above average." Five districts refused to provide the information and only 14 districts were classified as "below average" in the Cannell study.

Cannell's results were based on a telephone survey of the largest districts in the sixteen states where statewide results were unavailable. Districts were "asked if their elementary (1-6) total battery scores were above, at, or below the national average" (Cannell, 1987, p. 22). A district was called above average if four of six grades were above the national norm, and scores on reading, language, and math were used in cases where total battery scores were unavailable.

That the frequency of districts with scores below the median suggested by Figures 10 and 11 is greater than that suggested by the Cannell results is attributable largely to the difference in definitions. For example, one district that was classified as above average based on the Cannell study reported that for Grades 2 through 6 the percentages of students scoring above the national median in reading during the 1986-87 school year were 56, 47, 35, 44, and 48, respectively. While this district would appear to be "below average" based on these reading test results, it would

Figure 10
Box-and-Whisker Plots Showing the Percentage of Students Reported to be Above the National Median in Reading by Grade for Districts at the 10th, 25th, 50th, 75th, and 90th Percentiles for the District Distributions

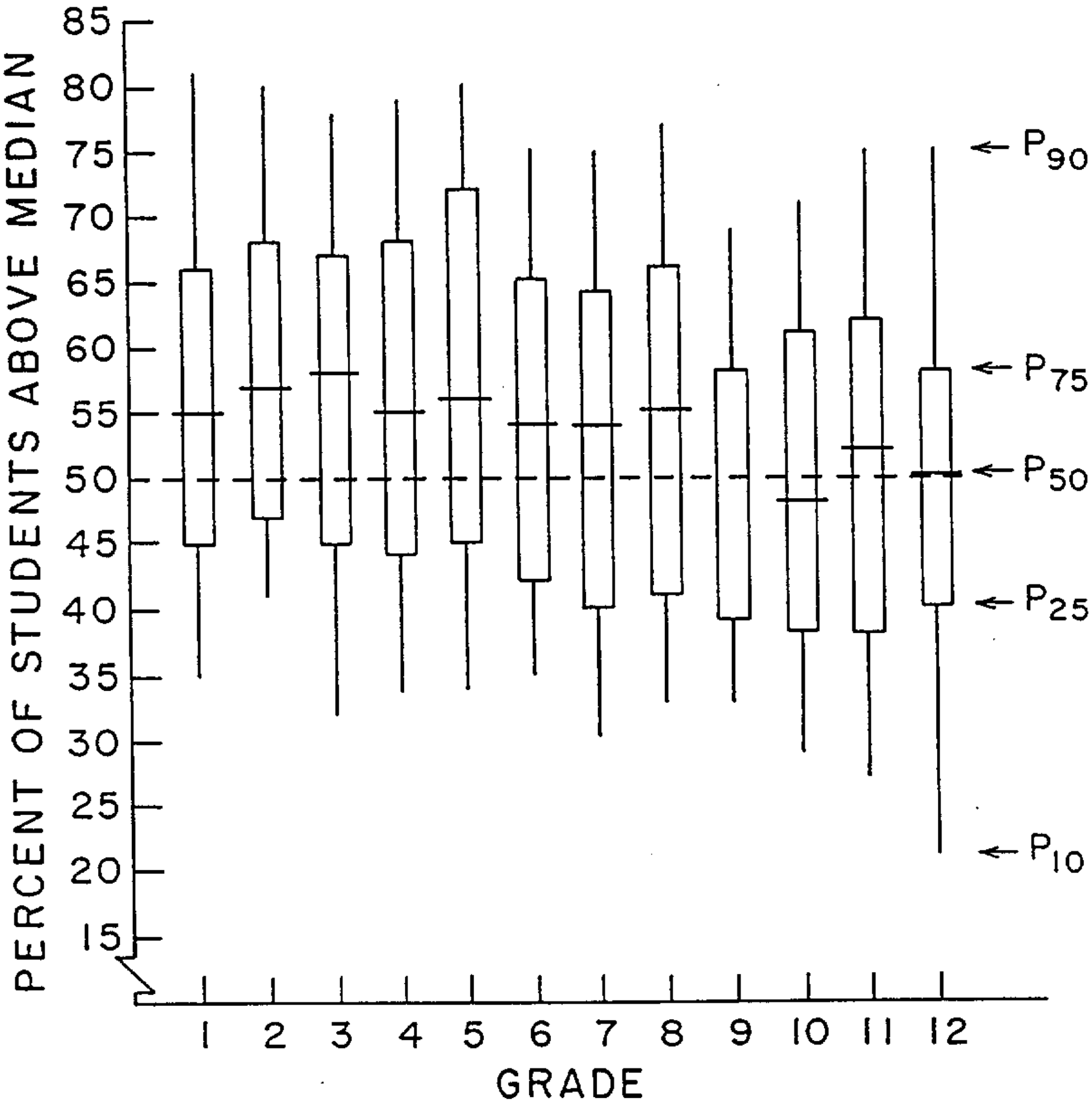
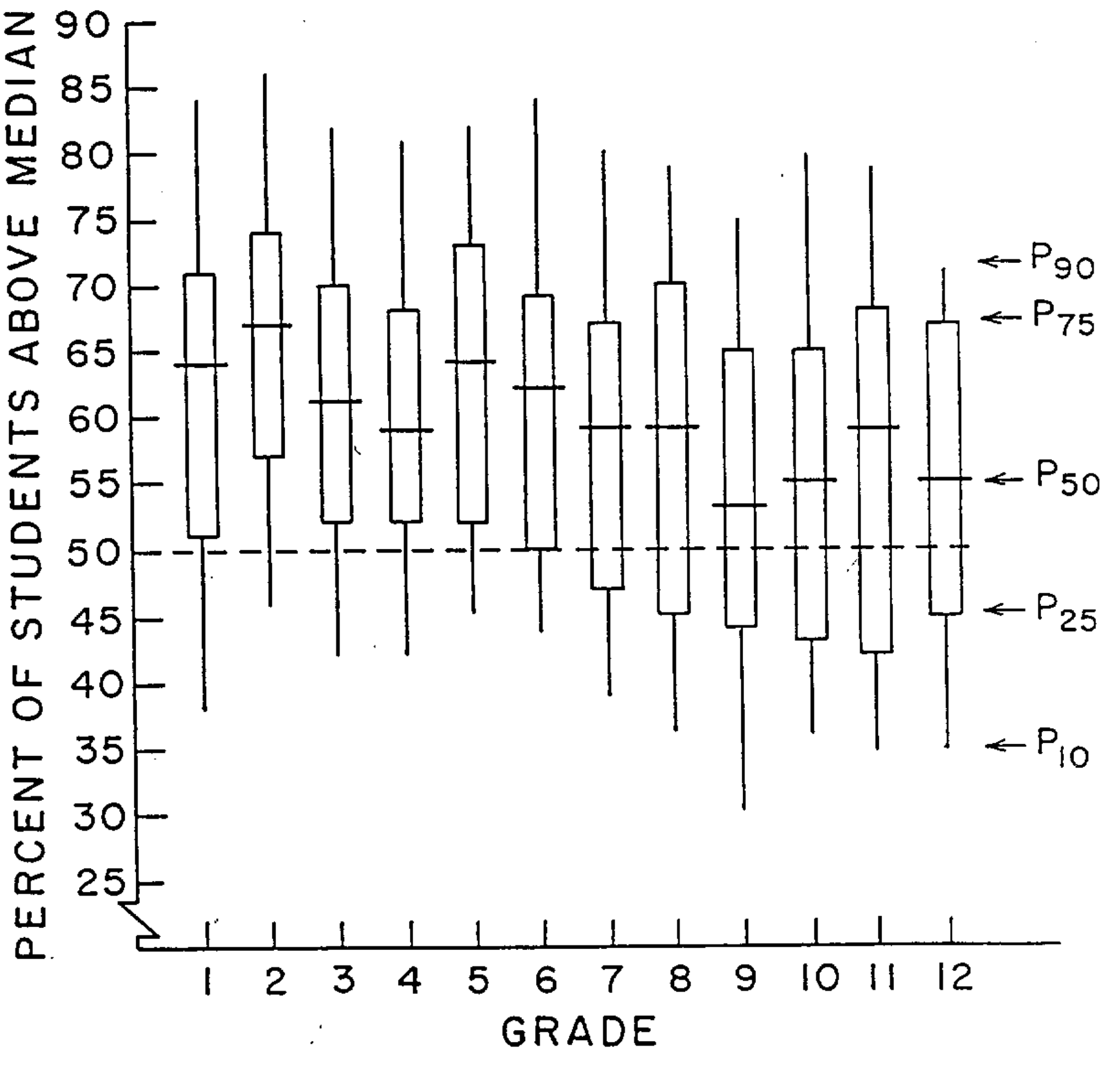


Figure 11
Box-and-Whisker Plots Showing the Percentage of Students Reported to be Above the National Median in Mathematics by Grade for Districts at the 10th, 25th, 50th, 75th, and 90th Percentiles of the District Distributions



appear to be clearly "above average" based on the corresponding percentages for mathematics (64, 64, 54, 60, and 68, for Grades 2 through 6, respectively). In general, districts reported a larger percentage of students above the national median when using total battery or mathematics scores than when using reading scores.

Summary of State and District Results

Clearly it was the exception rather than the rule for a state to report that its students, particularly its elementary school students, were performing below the national average. Although it was somewhat more common for a district than a state to report that less than half of its students were scoring above the national median, a substantial majority of districts reported that their students were performing above average (i.e., more than 50% of the students were reported to be above the national median). The tendency for students to score above the national median was especially strong in mathematics for Grades 1 thru 8. Nonetheless, it should be noted that some districts reported that substantially less than 50% of their students scored above the national median. At Grade 3, for example, 1 district in 10 reported that a third or less of its students scored above the national median in reading.

Achievement Trends and Dated Norms

Although the state and district results are generally consistent with the Cannell and earlier SREB findings which reported that achievement test results are more often above than below the national norm, they provide no real indication of the reasons that led to this result. As was discussed earlier, a wide variety of factors have been suggested as possible explanations of the apparently high test results that are being reported by states and districts. General improvement in student achievement, at least at the elementary grades, is clearly one possibility. When there are upward trends in achievement, old norms are easier (i.e., they provide a lower standard of comparison) than new norms, and thus a state or district whose students score at the current national average would score above the average defined by dated norms.

Using the aggregate results for districts, the district percentages of students scoring above the median in reading and in mathematics were related to the age of the norms used by districts at each grade (i.e., the number of years between the date of the test administration by a district and the date of the test norming by the publisher). Table 5 lists the number of districts that provided information on the year that the norms in use were obtained and the percentage of students scoring above the median for Grades 1 through 12. Also shown in Table 5 are the mean age of the norms used by districts, the mean change in the percentage of students scoring above the median for each additional year since the norms were obtained, and the estimated mean change in the percentage that resulted from the use of old norms rather than current norms.

As can be seen in Table 5, the average district that returned data was using norms that were four or five years old. Although most districts were using the most recent norms available from the publisher for the test being used, there was still an average of four or five years between the date of test administration by the district and the date of norming because publishers typically have collected norms only about every seven years. With a single exception, the percentage of students scoring above the median increased in both reading and mathematics with each additional year since the norms were obtained. The exception was for reading at Grade 10. By using norms that were four or five years old rather than current norms, assuming the latter were available, the percentage of students scoring above the median was estimated to be higher in all but Grade 10 in reading and in every grade for mathematics. For Grades 1 through 8 the expected increase ranges from

Table 5
Changes in District Percentages of Students
Above the National Median with Increasing Age of Norms

Grade	Number of Districts	Mean Age of Norms (Years)*	Mean Change in Percentage Above Median per Year		Estimated Mean Change (Old Minus Current Norms)	
			Reading	Math	Reading	Math
1	46	4.7	1.3	1.7	6	8
2	63	4.8	1.0	1.9	5	9
3	73	5.1	1.2	1.7	6	9
4	70	4.3	1.3	1.4	6	6
5	73	5.2	1.4	1.9	7	10
6	69	4.5	1.0	2.3	5	10
7	61	4.8	0.5	2.2	2	11
8	70	5.1	1.7	2.2	9	11
9	49	4.7	0.5	2.3	2	11
10	42	4.7	-0.3	1.1	-1	5
11	42	5.0	1.1	2.3	6	12
12	14	5.4	0.2	1.2	1	6

* Mean age of norms is the average number of years between the date of test administration and the date that the norms used to report district results were collected by the publisher.

2% to 9% in reading and from 6% to 11% in mathematics. Taking differences of the latter magnitude into account would largely eliminate the tendency for these districts to report results that are above the national median.

Trends over Several Years for Selected States

The district results in Table 5 show that there is a relationship between the age of norms used and the level of achievement test scores for the districts in this sample. These results are cross-sectional, and there may be a variety of other district characteristics associated with the age of norms for the test used as well as the level of student achievement. Therefore, these results do not provide a sufficient basis for concluding either that older norms are easier than newer norms or that achievement has been going up.

Figures 1 and 4, which were considered earlier, present achievement test results for three years. Neither of these figures provides a very clear indication that achievement scores went up or down during the three years for which data were collected. There is some suggestion from both of these figures that scores went up in Grades 1, 2, and 3. However, the direction of change is not only unclear at most other grades, but would be difficult to interpret in any event because the subset of states for which data were obtained changed somewhat from year to year. Furthermore, three years is too short a time interval to assess long-term trends.

Though not a specific part of the data collection design, results included in the state assessment reports for some of the states made it possible to look at trends for longer time intervals. Achievement trends for four states are summarized in Figure 12.

The upper-left-hand quadrant of Figure 12 shows a plot of the percentage of students in one state (State A) scoring above the national median in reading and mathematics at Grade 4 for each of the past six school years. During this interval a single test form of a single edition of a test was administered each year and results were based on comparisons to the 1980-81 national norms provided by the test publisher. As can be seen, the first year the test was administered, 1982-83, the percentage of students scoring above the national median was well below 50 for both reading (41%) and mathematics (44%). During each of the following five years these percentages increased, most notably in mathematics. In 1987-88, 57% of the students scored above the national median in reading and 68% scored above the national median in mathematics.

Similar results using the alternative statistic of the percentile rank in the individual pupil norms corresponding to the statewide mean test score are shown for another state (State B) in the upper-right-hand quadrant of Figure 12. As in the previous example, the results are shown for a six-year period during which a single form of a single edition of a test was administered each year. Comparisons were to norms obtained in 1978 in this case. Although the trend for State B was less steep than the one for State A and was based on a different metric, there was a clear upward trend during the six years in both reading and mathematics.

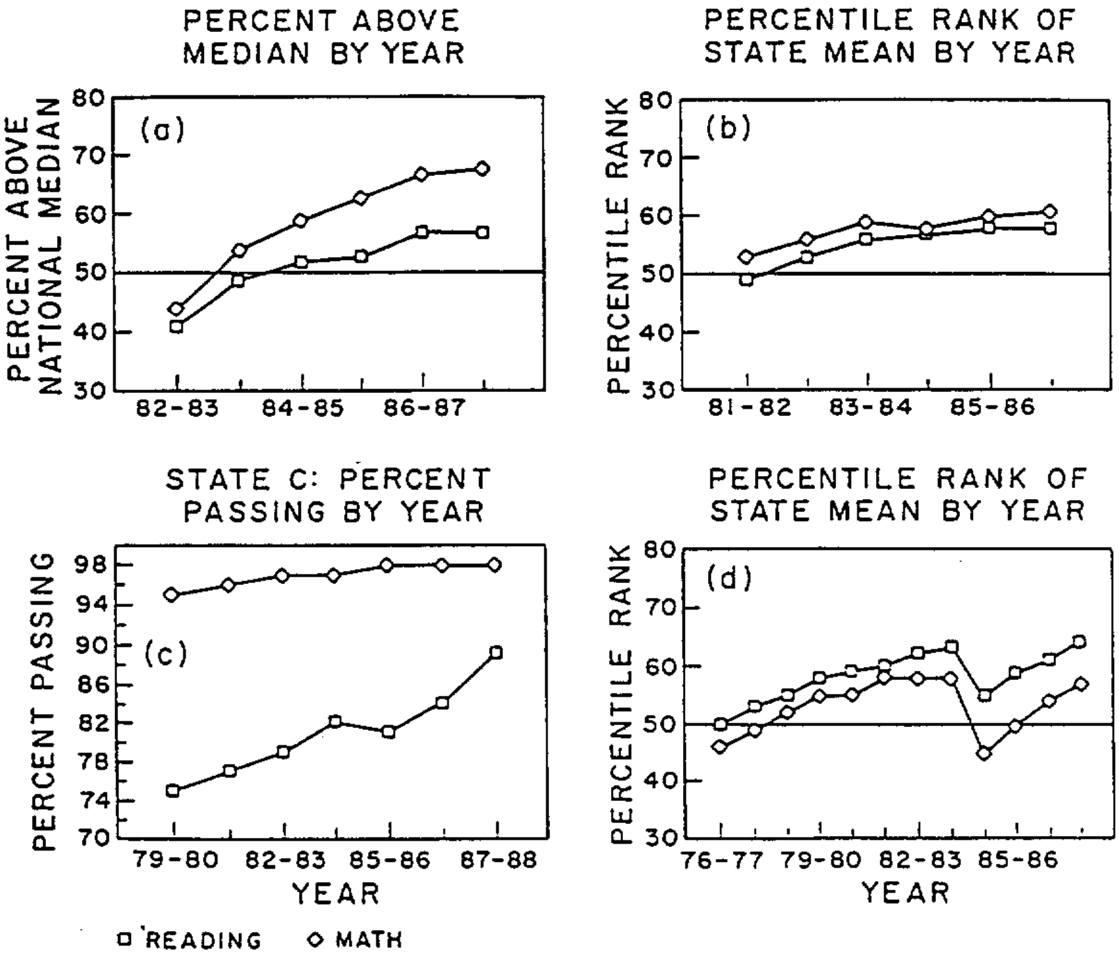
The third example, State C, shown in the lower-left-hand quadrant of Figure 12, uses an entirely different metric than has been considered so far. The plots for State C show the percentage of students passing statewide minimum-competency tests in reading and mathematics for each of seven years. In mathematics the percentage passing was 95 in the first year and gradually increased to 98% over time. For reading, where there was more room for movement, the increases between the first and most recent years of test administration were more substantial.

The final plot shown in the lower-right-hand quadrant of Figure 12 displays the percentile ranks of the state means in reading and mathematics based on individual pupil norms for Grade 3 in State D. The State D results not only span the longest time interval, twelve school years, but include a change in test editions within the period of time that was covered. A single form of a single edition of a test was used for the eight years starting in 1976-77 and running through 1983-84. The pattern for those first eight years was reasonably similar to the ones shown for the other three states in Figure 10. There was a consistent upward trend during those years.

The feature of the plot for State D that most clearly sets it apart from the plots for the other three states in Figure 12 is the sharp decline shown in percentile rank between the 1983-84 and 1984-85 school years, followed by increases over the next three years to bring 1987-88 results back to approximately where they were in 1983-84. As was previously indicated, during the 1984-85 school year the new edition of the test was introduced and the same form of that edition was administered in each of the last four years covered in the plot of results for State D. Thus the sharp decline corresponds to the introduction of the new test edition.

The sharp decline in performance relative to national norms that State D experienced when the new edition of the test was introduced is not unique. Figures 13 and 14, for example, show the results for two large school districts that introduced new editions during the 1987-88 school year. As can be seen, both districts experienced large declines in the percentage of students scoring above the national median between 1986-87 and 1987-88.

Figure 12
Trends in Reading and Mathematics Achievement for Four States



There are several possible interpretations of the trend results shown in Figures 12, 13, and 14. The most straightforward interpretation of the trends in Figure 12 is that achievement in reading and mathematics for the grades in question improved rather steadily in all four states. The dip when a new edition was introduced in State D could simply reflect general increases in student performance across the nation, which made the more recent norms associated with the newer edition more stringent than the norms associated with the older edition of the test. This same interpretation could also explain the dips in performance levels associated with a new test edition for the two districts shown in Figures 13 and 14.

An alternative interpretation of these results, however, is that increases in test scores simply reflect increasing familiarity with a given test form and more focused instruction on the content of that specific form. By administering the same form of a test for several years teachers are apt to become increasingly familiar with the specifics of the test content and alter instructional emphases to better match the content of the test. As indicated by Mehrens and Kaminski (1988) and by Shepard (1989), test familiarity might influence instruction in a wide variety of ways, ranging from practices that would generally be considered sound uses of test results (e.g., identifying and working on objectives where students show weaknesses) to those that most educators consider unethical (e.g., teaching the specific items on a test just prior to test administration).

It is not possible to distinguish whether the trends in Figures 12, 13, and 14 were due to improvements in achievement, to increased familiarity with the tests, or to some alternative explanation, solely from the results presented in those figures. However, other data can be brought to bear on the issue. In particular, the questionnaire and interview results which are discussed in other reports based on this project (e.g., Shepard, 1989) speak to some of these issues. Only the question of whether norms are changing in difficulty with time as a result of increases in student achievement nationally will be considered here.

Achievement Trends and Changes in the Difficulty of Norms

National changes in achievement levels obviously lead to differences in the meaning of norms. During a period of declining performance such as the nation experienced in the 1960s and the first part of the 1970s (Harnischfeger & Wiley, 1975; Koretz, 1986; 1987), newer norms provide a less stringent standard of comparison than older norms. Koretz (1987), for example, estimated that during the period of the much publicized test score decline (roughly the early or mid 1960s to the mid 1970s) "the average decline in grades six and above was large enough that the typical (median) student at the end of the decline exhibited the same level of achievement as was shown before the decline by students at the 38th percentile" (p. 2). Thus a state or district using old norms in the mid 1970s could have appeared to be well below the national average when in fact their students were scoring at the then current national average. On the other hand, when performance on achievement tests is increasing, newer norms become harder and the use of old norms can make a state or district that would have only average or below average scores in terms of current national norms appear to be above average. Clearly, national trends in achievement tests scores have importance for understanding normative comparisons.

Although increases in test performance have not received as much attention as the decline of the 1960s and 1970s, several sources of evidence suggest that achievement test scores have been going up. National Assessment of Educational Progress (NAEP) reports (e.g., Dossey, Mullis, Lindquist, & Chambers, 1988; NAEP, 1985) indicated that there were some increases in reading and mathematics between the early or mid 1970s to the mid 1980s. Based on his review of NAEP and data from several other tests, Koretz (1987) concluded that the decline in test scores ended

Figure 13
Percentage of Students Above National Median for District A
Before and After a Change of Test Editions (New Edition in 1987-88)

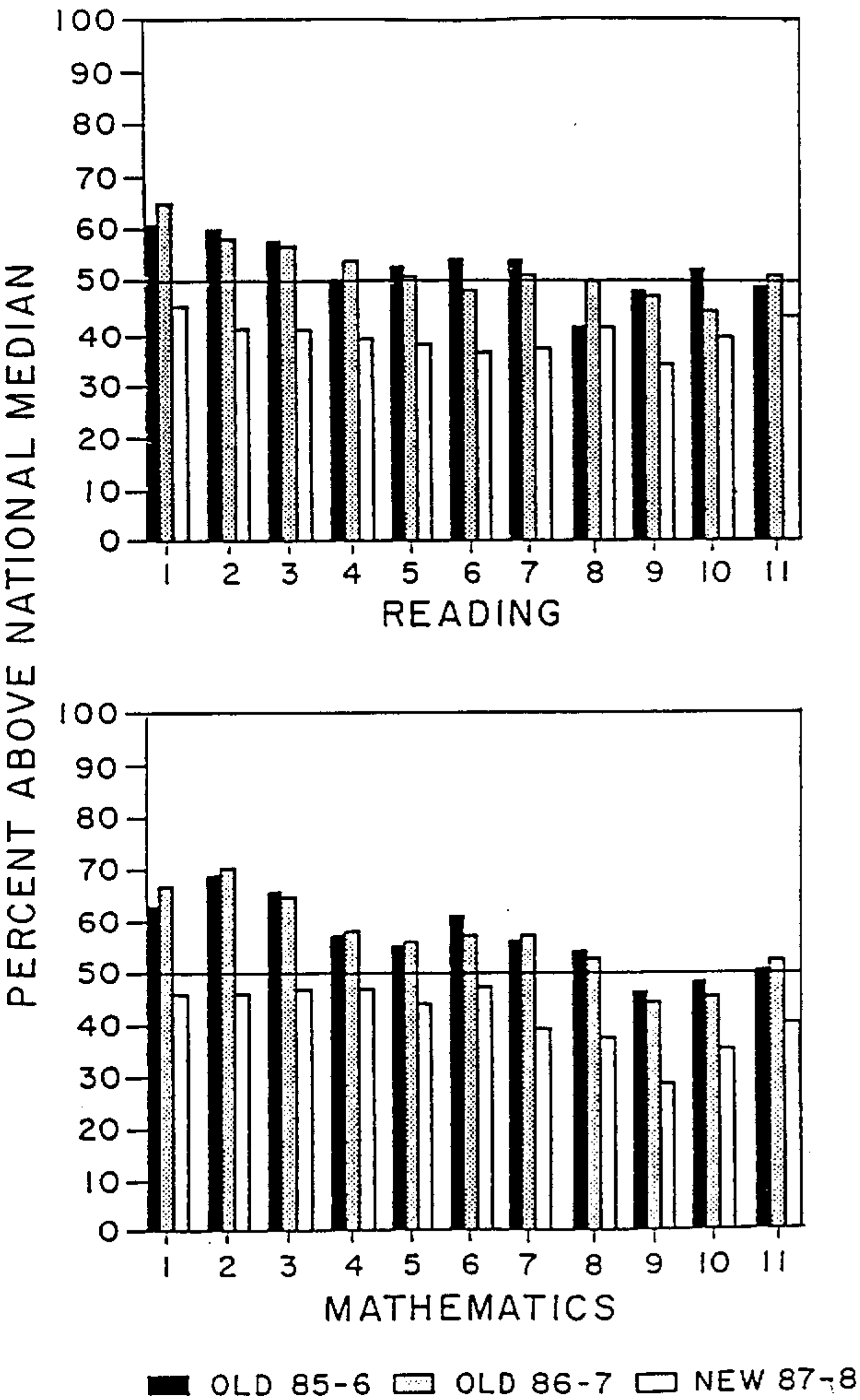
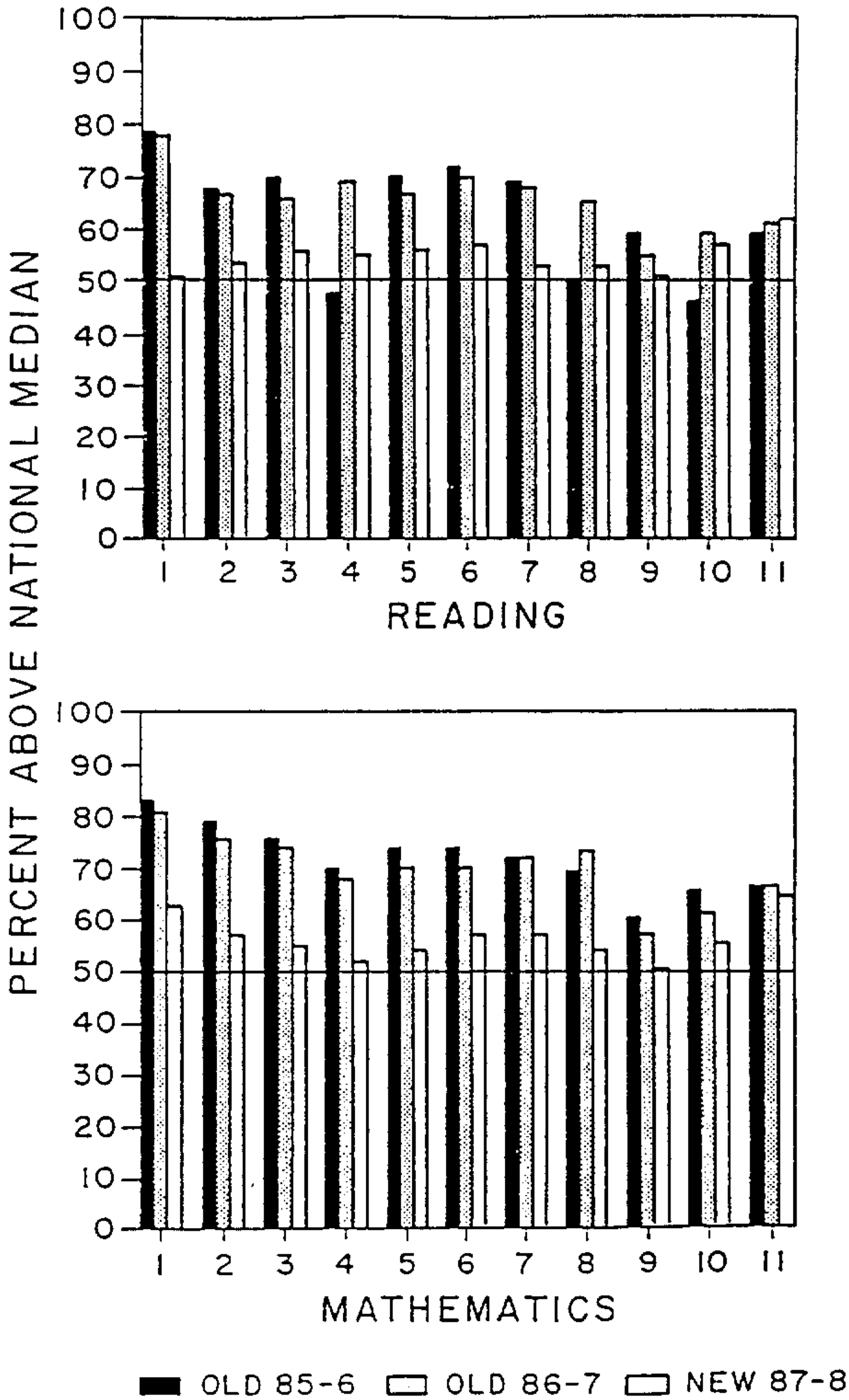


Figure 14
Percentage of Students Above National Median for District B
Before and After a Change of Test Editions (New Edition in 1987-88)



with cohorts of students that entered school in the late 1960s and that subsequent cohorts of students "produced a sharp rise in scores on most, but not all, tests. In the majority of instances in which scores increased, the rise has been steady—with each cohort tending to outscore the preceding one—and often roughly as fast as the decline" (p. 2).

Norming studies conducted periodically for standardized tests also provide evidence regarding trends in national achievement. When a new edition of a standardized test is introduced, it is customary not only to collect new normative data for the new edition but also to equate the old and new editions of the test. The equatings make it possible to estimate the extent to which achievement has increased or decreased over the years between the norming of the two editions. In some cases, new norms are collected for a previously normed edition of a test, which again provides a means of comparing national performance on the test at two points in time.

Several test publishers reported increases in achievement based on the results of their norming studies. CTB/McGraw-Hill (1987), for example, noted when the norms for Form E of the California Achievement Tests (CAT) were reported and compared to the norms for the CAT Form C to which Form E was equated that "the CAT E norms are more difficult than the CAT C norms. This seems to indicate that students in 1984-85 were achieving at a higher level than in 1977, when CAT C was normed" (p. 3-4). Increases in performance were reported when Form G of the Iowa Tests of Basic Skills (ITBS) was published. "Between 1977-78 and 1984-85, the improvement in ITBS test performance more than made up for previous losses in most test areas. Composite achievement in 1984-85 was at an all time high in nearly all test areas" (Hieronymus & Hoover, 1986, p. 148). Increases in performance have also been reported for the Stanford Achievement Test (SAT7) (Wiser & Lenke, 1987) and the Comprehensive Tests of Basic Skills (CTBS) (Rothman, 1988) and increases can be inferred from comparisons of the norms for the Metropolitan Achievement Tests (MAT6) (Psychological Corporation, 1988) and norms for equivalent scores on the previous edition of the MAT (Prescott, Balow, Hogan, & Farr, 1978; 1986).

Table 6 provides a summary of the changes in the percentile rank of achievement test scores that were at the national median at one of the two times that norms were obtained for the six most used standardized achievement tests. The numbers are estimates of the changes in national percentile rank in reading and mathematics between the two norming years indicated at the head of each column of the table. Also shown for comparative purposes are estimated changes in national percentile ranks based on NAEP.

As is indicated in the footnotes to Table 6, the numbers in each column of Table 6 are derived from different sources and involve different types of comparisons. In the case of the CTBS, the comparison is between 1981 norms and estimates of 1987 norms for the same test form based upon a weighting of user data. The Stanford results are based on 1981-82 and 1986 norming studies for the same test form. The other published test comparisons involve norming studies for successive editions of the test battery. However, the numbers in Table 6 all have a similar interpretation. A positive number indicates that performance was higher when measured at the more recent of the two norming years indicated at the top of each column. For example, the number 14 shown for reading achievement on the California Achievement Tests (CAT) in Grade 2 indicates that an equated Form C or Form E score that would have placed a student at the national 50th percentile using the 1977 Form C norms would lead to a national percentile rank of only 36 using the 1984-85 Form E norms. The 14 shown in Table 6 is the difference between the percentile ranks of 50 in 1977 and 36 in 1984-85.

Table 6
Estimated Changes in National Percentile Rank of
Achievement Scores at the National Median at One Point in Time

I. Reading Achievement

Source/Years Being Compared*

Grade	CAT[1] 77 to 84-5	CTBS[2] 81 to 87	ITBS[3] 77-8 to 84-5	MAT[4] 77-8 to 84-5	SRA[5] 78 to 83-4	Stanford[6] 81-2 to 86	NAEP[7] 74-5 to 83-4
1	28	7	9	20	-3	11	
2	14	10	12	5	1	4	
3	12	2	11	13	1	6	3
4	11	8	12	5	-1	2	
5	14	5	11	7	2	2	
6	11	8	12	6	-3	2	
7	16	6	11	9	-2	2	0
8	11	5	10	7	-4	1	
9	15			9	2	3	
10	8			-5	2	0	
11	4			-3	-2	4	2
12	1			-5	-7	3	

II. Mathematics Achievement

Grade	77 to 84-5	81 to 87	77-8 to 84-5	77-8 to 84-5	78 to 83-4	81-2 to 86	77-8 to 85-6
1	16	18	3	12	10	15	
2	14	22	5	9	3	10	
3	13	13	5	15	-6	9	4
4	11	14	9	7	-2	8	
5	13	17	8	11	3	8	
6	13	17	8	10	0	7	
7	15	15	10	2	1	6	5
8	18	11	10	5	0	7	
9	14			0	1	4	
10	8			4	4	4	
11	5			7	-2	4	0
12	2			6	-4	5	

*Footnotes on following page

Footnotes for Table 6

1 Differences in California Achievement Tests (CAT), Form E (1984-85 norms) percentile ranks and corresponding CAT, Form C (1977 norms) percentile ranks of 50 (CTB/McGraw-Hill, 1987, Table 38, p. 3-35).

2 Differences in Comprehensive Tests of Basic Skills (CTBS), Form U percentile ranks in 1981 and those required to have a percentile rank of 50 on the CTBS in 1987 (based on November, 1988, CTB-McGraw-Hill press release, "CTB/McGraw-Hill Studies Show Students Achieving at Higher Levels in Basic Skills", see also, Rothman, 1988, p. 20). The 1987 norms are estimated from weighted user data.

3 Differences in Iowa Tests of Basic Skills (ITBS), Form G (1984- 85 norms) percentile ranks and corresponding ITBS, Form 7 (1977-78 norms) percentile ranks of 50 (Hieronymus & Hoover, 1986, Table 6.31, p. 153).

4 Differences in Metropolitan Achievement Tests (MAT6), Survey Forms L and M (1984-5 norms) and corresponding MAT, Forms J and K (1977-78 norms) percentile ranks of 50 (Psychological Corporation, 1988; Prescott, Balow, Hogan, & Farr, 1978; 1986).

5 Differences in SRA Achievement Series, Forms 1 and 2 (1983-84 norms) percentile ranks and corresponding SRA Achievement Series Forms 1 and 2 (1978 norms) percentile ranks of 50 (Science Research Associates, 1979; 1986).

6 Differences in Stanford 7 Plus (1986 norms) percentile ranks and corresponding Stanford Early School Achievement Test, 2nd edition; Stanford Achievement Test, 7th edition, and Stanford Test of Academic Skills (TASK), 2nd edition (1981-82 norms) percentile ranks of 50 (Gardner, Madden, Rudner, Karlsen, Merwin, Callis, & Collins, 1983; 1987).

7 Differences for the National Assessment of Educational Progress (NAEP) are based on age (9, 13, and 17) rather than grade (3, 7, and 11) cohorts. For reading, the differences are between the 1983-84 assessment percentile ranks and the corresponding 1974-74 assessment percentile rank of 50 (NAEP, 1985). For math, the differences are between the 1985-86 assessment percentile ranks and the corresponding 1977-78 percentile rank of 50 (NAEP, 1988; frequency distributions provided by Beaton).

With the exception of the SRA Achievement Series, the differences for Grades 1 thru 8 are all positive, indicating that more recent norms are more stringent than older norms for five of the six tests. For Grades 10 through 12 the differences are generally smaller than those shown for the earlier grades and two of the four tests with results for the high school grades have some differences that are negative, indicating a decline in performance and therefore easier recent norms in those instances.

The changes in percentile ranks shown in Table 6 are based on various time intervals between norming studies. More direct comparison can be made by dividing the changes in percentile ranks in Table 5 by the number of years between the norming studies to obtain estimates of yearly changes in percentile ranks. Such yearly changes in percentile ranks for Grades 1 thru 8 are presented graphically in Figures 15 and 16 for reading and mathematics, respectively.

In general, the results in Figures 15 and 16 are fairly consistent with those based on the analyses of the district data that were reported in Table 5. The estimates of yearly changes derived from the district data are greater than those

Figure 15
Estimated Yearly Changes in Reading Percentile Rank:
Publisher Results at the Median

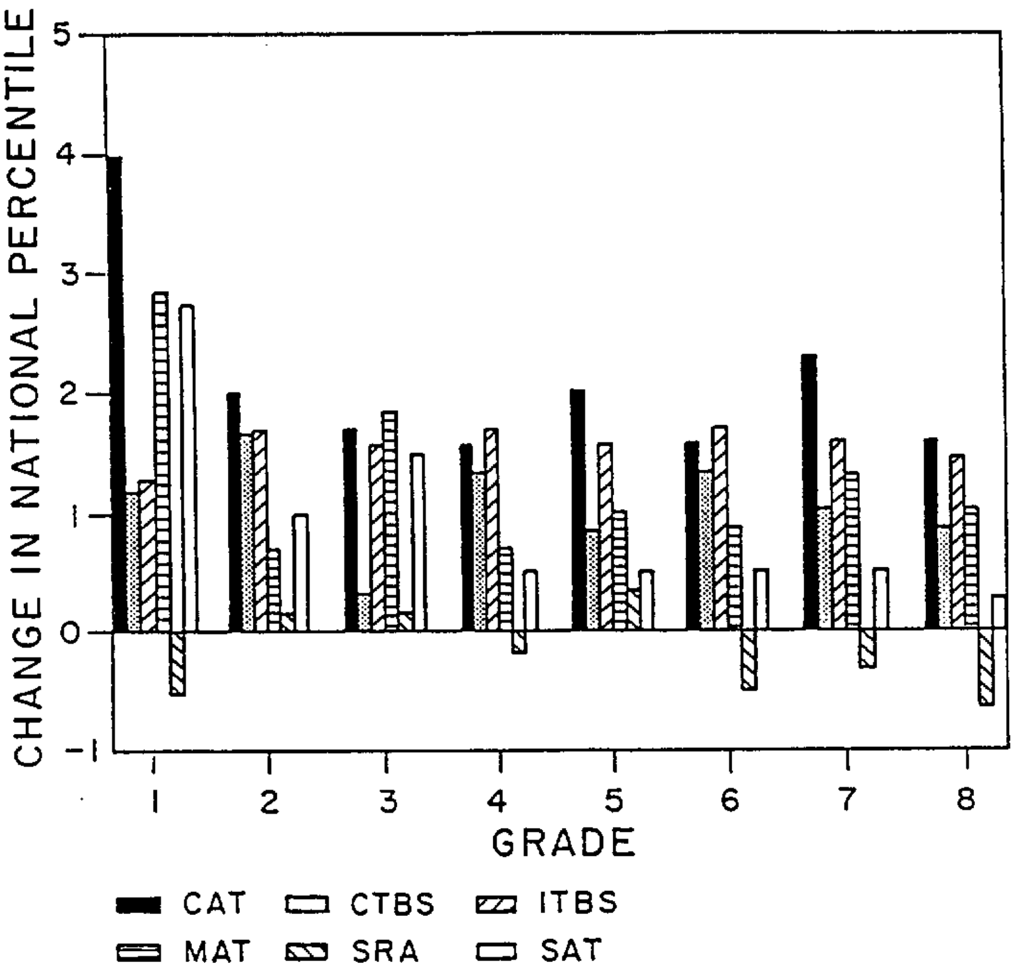
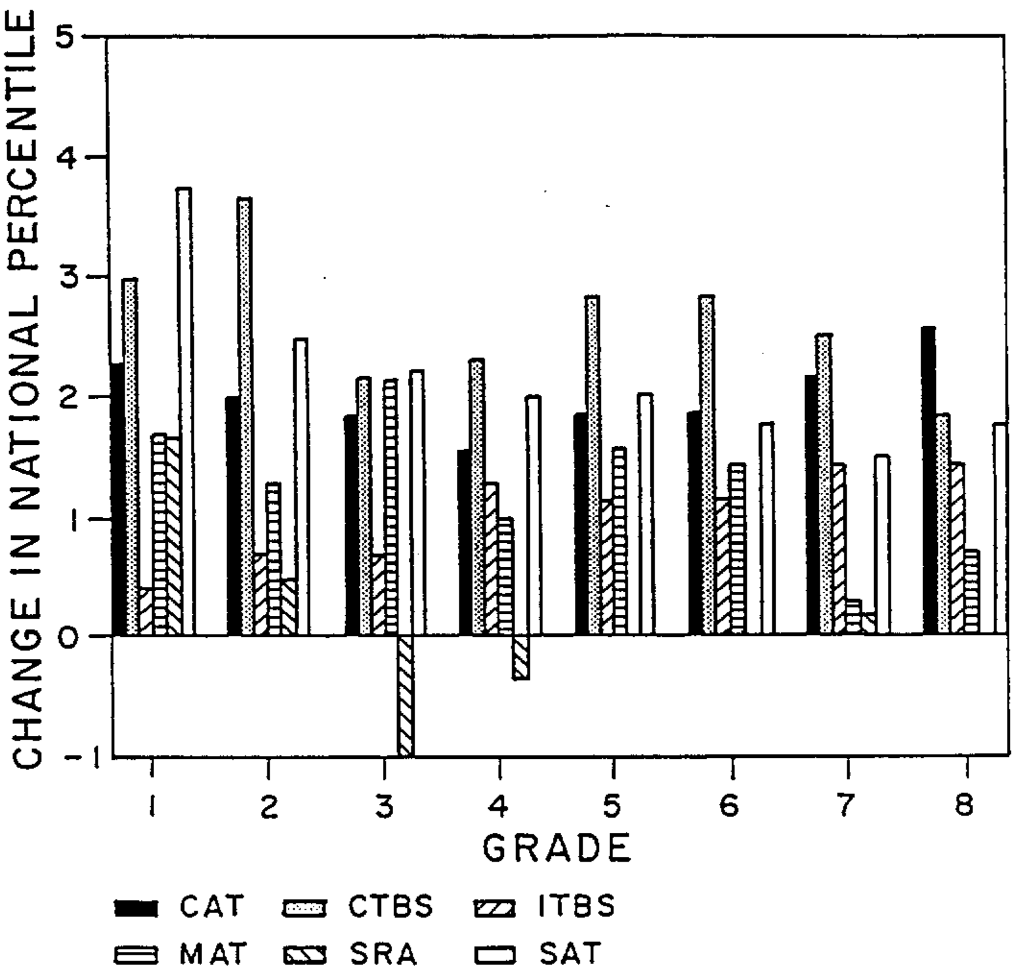


Figure 16
Estimated Yearly Changes in Mathematics Percentile Rank:
Publisher Results at the Median



shown in Figures 15 and 16 for some tests but smaller than those for other tests. The Table 5 estimates of changes in norm-referenced performance that would be expected as a result of a change in the date of the norms, however, are of the same order of magnitude as those shown in Figures 15 and 16.

Although the NAEP trend results are based on age cohorts rather than grade cohorts, the NAEP results represent the best available independent means of estimating national changes in achievement. Changes in percentile ranks estimated from NAEP results between the 1974-75 and 1983-84 assessments for reading and between 1977-78 and 1985-86 for mathematics are plotted in Figures 17, 18, and 19 for 9-, 13-, and 17-year-olds, respectively. Also shown in these figures are the changes for the six norm-referenced tests at the modal grades for 9-, 13-, and 17-year-olds, that is, Grades 3, 7, and 11.

As can be seen in these figures, the different data sources vary a good deal in the magnitude of change in performance. The NAEP results suggest either some increase in performance (ages 9 and 17 in reading and ages 9 and 13 in mathematics) or no change during the interval in question. The increases indicated by NAEP are smaller than those shown by some, but not all, of the standardized tests. Comparing the publisher Grade 3 results with NAEP age 9 results (Figure 17), it can be seen that four of the six standardized tests show larger gains in reading and five of the six show larger gains in mathematics than would be estimated by NAEP. At age 13 (Figure 18) NAEP shows no change in reading and two of the standardized tests (SRA and Stanford) indicate only small changes at Grade 7, but the remaining four tests suggest more substantial increases in performance. In mathematics, two standardized tests suggest smaller changes at Grade 7 than NAEP obtained for 13-year-olds, one standardized test shows a change similar to the one obtained by NAEP, and the remaining three standardized tests show larger gains in performance. At Grade 11 or age 17 (Figure 19), relatively little change is indicated by any of the data sources for reading and relatively small and inconsistent changes are indicated for mathematics.

Of course, the dates of the first and second normings are not the same for all the tests and the tests differ in content coverage and in the specifics of the samples on which the norms were based. Nonetheless, the different data sources give rather different answers in some cases to the question of the degree to which test performance has increased during the past decade. The discrepancy between increases suggested by NAEP and most of the standardized tests raises questions about the possibility that artifacts may inflate the norm-referenced test results.

One possible artifact is that the norms obtained for a standardized test may be biased because of differential participation rates in norming studies by school districts according to whether the districts were already using the standardized test being normed (Baglin, 1981). If school districts that are already using a standardized test are more likely to participate in the norming of a new edition of the test than districts using another publishers test, and if districts that are using a given test generally have curricula that match more closely the objectives of both the new and old editions of that test or emphasize those objectives because the test is used, then the norms could be more difficult. In other words, such an influence would run counter to the observed tendency for states and districts to report that more than 50% of their students score above the national median.

To investigate the latter possibility, Wiser and Lenke (1987) compared the performance of user and non-user groups when the 1986 norms for the Stanford were obtained. They found that "users performed as well or better than non-users in all subject areas through Grade 6." For Grades 7 through 12 the results were more mixed, with users performing better in some subject areas at some grades but non-users performing better for other combinations.

Figure 17
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 3 (NAEP, Age 9)

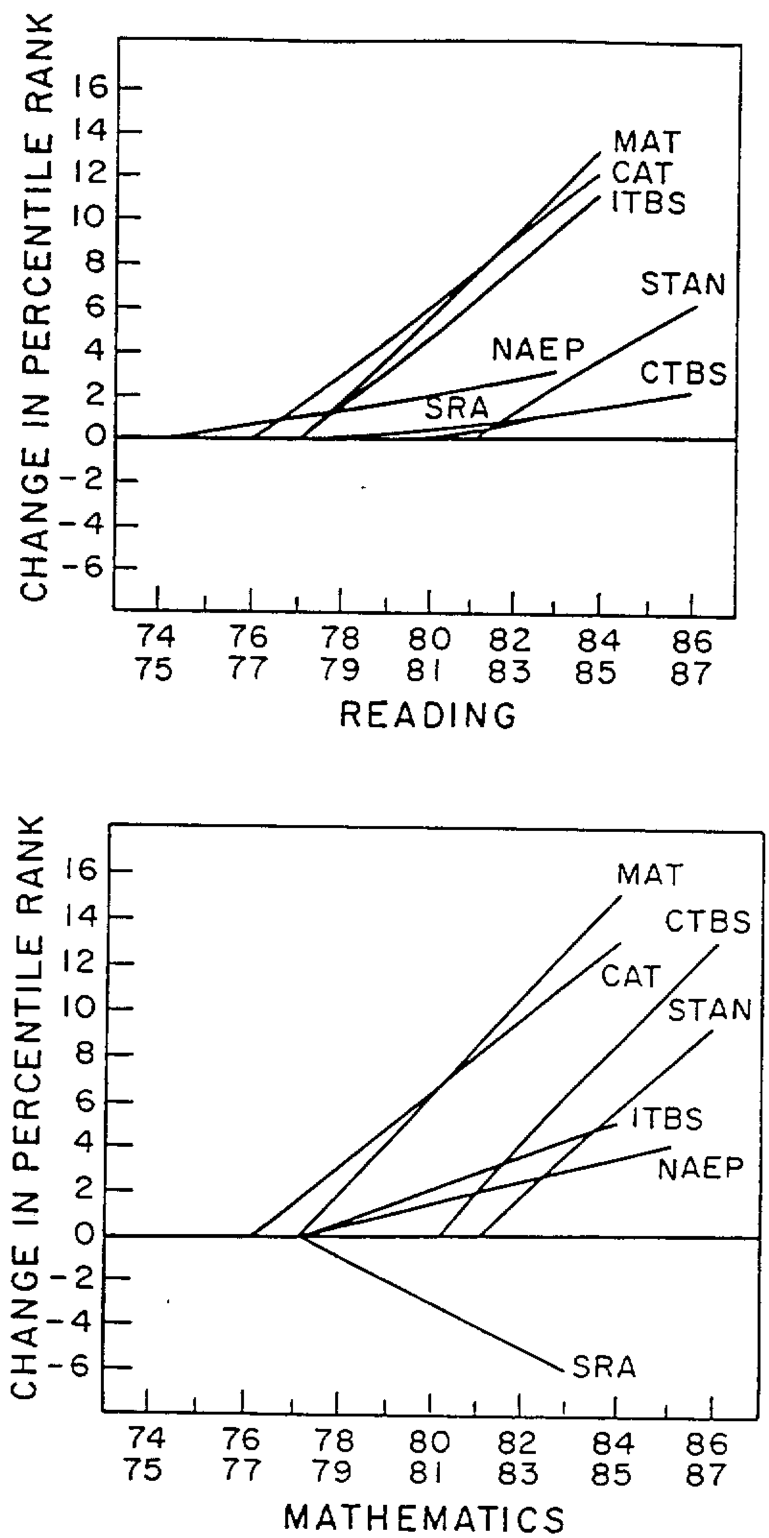
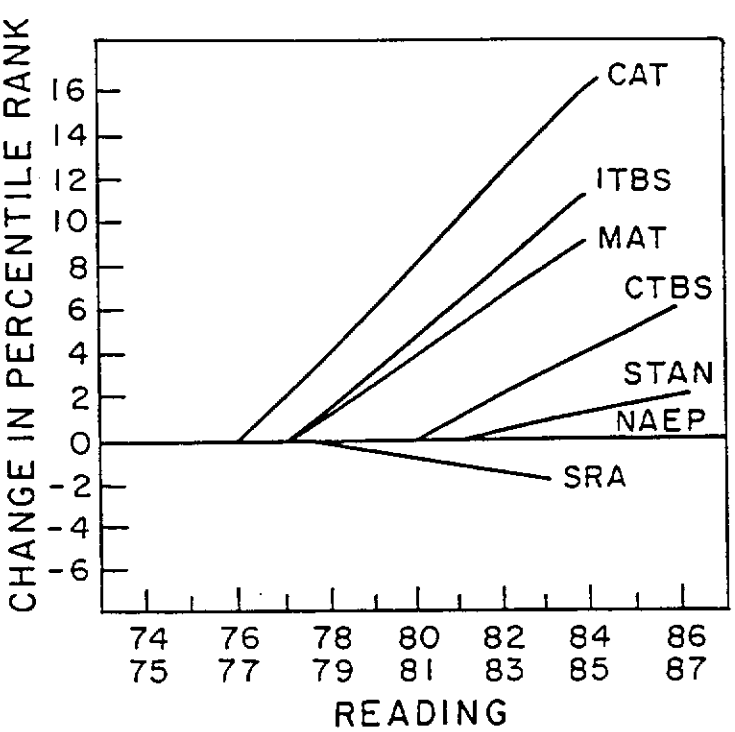


Figure 18
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 7 (NAEP, Age 13)



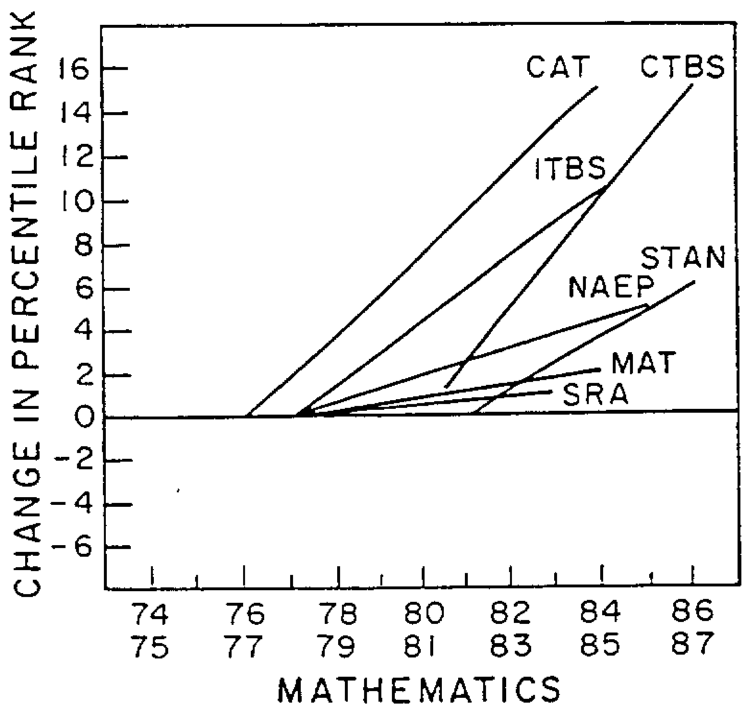
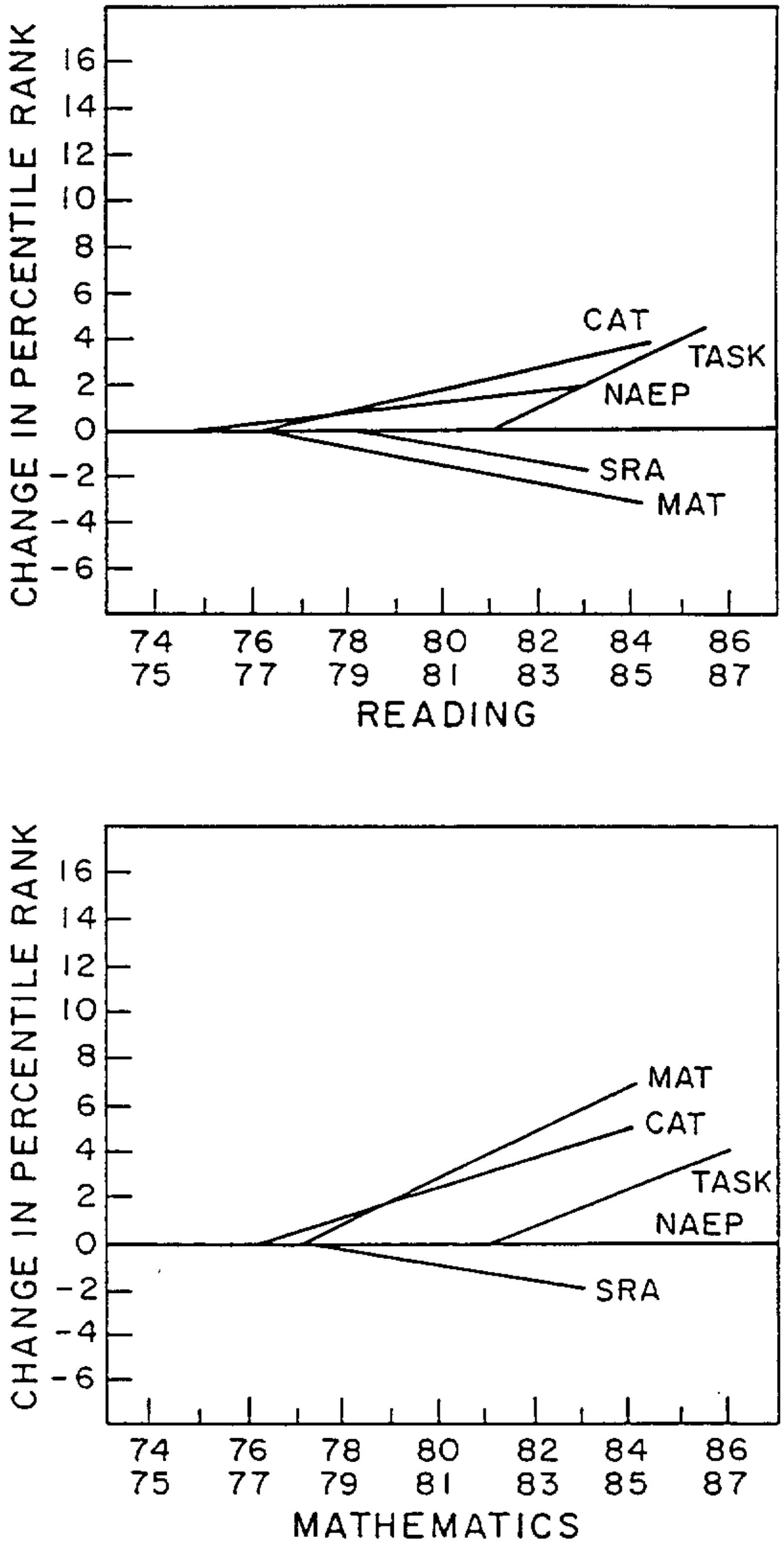


Figure 19
Estimated Change at the Median in National Percentile Ranks of Achievement Test Scores at Grade 11 (NAEP, Age 17)



Wiser and Lenke noted that the comparison of particular interest in their results was between the 1986 non-users and the 1982 norming sample. Since the Stanford 7 was a new edition at the time of the 1982 norming, the participants in the norming sample had not previously used the edition and were comparable in that sense to the 1986 non-user sample. The 1982 sample and the 1986 non-user samples were also matched on school ability as measured by the Otis-Lennon School Ability Test. Thus, a comparison of the 1982 and 1986 non-user results provides an estimate of the change in achievement that is uncontaminated by the familiarity that users have with the particular edition of the test.

We used the scaled score means and standard deviations reported by Wiser and Lenke (1987) to calculate two estimates of the changes in average test scores in terms of 1982 standard deviation units for total reading and total mathematics. The first estimate is simply the mean for the full 1986 norming sample (users and non-users) minus the 1982 mean, all divided by the 1982 standard deviation. The second estimate is the 1986 mean for non-users only minus the 1982 mean, all divided by the 1982 standard deviation. The two sets of standardized differences are summarized in Table 7.

Table 7
Estimated Standardized Average Changes in Achievement Test Scores on the Stanford from 1982 to 1986 (Based on Wiser & Lenke, 1987)

	Reading		Mathematics	
Grade	Total Group[a]	1986 Non-users[b]	Total Group	1986 Non-users
1	.17	.10	.34	c
2	.13	.04	.18	.10
3	.13	.12	.15	.12
4	.03	-.01	.12	.12
5	.03	-.02	.17	.16
6	.03	-.02	.10	.06
7	.03[d]	.03[d]	.08	.06
8	.00[d]	-.08[d]	.10	.11
9	.08[d]	.03[d]	.05	.07
10	.05	.05	.04	.03
11	.10	.11	.03	.05
12	.13	.14	.05	.08

[a]The mean for the full 1986 norming sample (users and non-users) minus the 1982 mean all divided by the 1982 standard deviation.

[b]The mean for the 1986 non-users only minus the 1982 mean all divided by the 1982 standard deviation.

[c]Not available.

[d]Reading Comprehension.

For Grades 1 and 2 the non-user group data results in estimates of the gain in achievement in reading between 1982 and 1986 that are substantially smaller than the estimates based on the total norming sample. The gain in reading achievement appears to be about 40% smaller (i.e., 100x(.17-.10)/.17) at Grade 1 and about 70% smaller at Grade 2 with non-user data than with the data from the total norming sample. This difference is consistent with the premise that familiarity with a test form leads to inflated estimates of achievement gains. However, large differences in estimates based on non-user and total norming sample data such as those for reading in Grades 1 and 2 are not found consistently.

The non-user estimates of standardized gains in reading achievement are smaller for the total-norming-group estimates in Grades 1 through 6 and Grades 8 and 9, albeit by only a trivial amount at Grade 3. The two sets of estimates are the same to two decimal places in Grades 7 and 10, and the non-user estimates are actually larger than those based on the total norming sample at Grades 11 and 12. For mathematics, non-user estimates of achievement gains are 20% or more lower than total group estimates only at Grades 2, 3, 6, and 7, while they are larger by an equal percentage or more at Grades 9, 11, and 12.

Overall, the Wiser and Lenke results suggest that increasing familiarity with a particular test form may explain part of the apparent growth in norm-referenced test performance. The generally higher scores obtained by non-users in 1986 than were obtained in the 1982 norming of the then new edition of the test, however, suggest that there also has been some more generalized improvement in performance, particularly in mathematics.

Results recently reported by Hoover (1989) for the Iowa Tests of Basic Skills (ITBS) suggest that much of the increase in performance on a test form may occur on the first operational administration of the form. From user data weighted to estimate national performance, Hoover estimated that approximately 55% of the students scored above the 1984-85 national median across Grades 3 through 8 on the Battery Composite when Forms G and H were first administered operationally in 1985-86. In the second and third years of operational administration the average percentage of students across Grades 3 thru 8 who scored above the 1984-85 national median increased to 59% (1986-87) and then to 60% (1987-88).

The gains from the first year to the second and third years of operational use reported by Hoover may be attributable to a combination of real gains in achievement and increasing familiarity with a test form. The relatively large gain in the first year that the test was used operationally, however, may be due to a combination of several additional factors such as (a) the selection of a test that was most closely aligned with the state or district curriculum, (b) greater emphasis on the importance of good test performance when the test was used operationally than when it was normed, and (c) the exclusion of a larger fraction of less able students in operational test administrations than in norming studies. Indirect support for the latter explanation comes from Hoover's finding that only about 6%, rather than the expected 10%, of the students scored below the 10th percentile during the first year of operational administration of Forms G and H of the ITBS. High scores (at or above the 90th percentile), on the other hand, occurred at the expected rate of 10% in the first year of operational test use.

Discussion

Weighted estimates from the district sample suggest that at least 57% of the students in Grades 1 through 6 are obtaining scores above the national median on norm-referenced reading tests. The corresponding figure for mathematics is 62%. The comparable figures for Grades 7 through 12 are lower, but still somewhat greater than 50%. The state results are quite consistent with the district estimates. Thus, the results of the present study provide additional support for the general finding by Cannell and by the SREB that for the elementary grades almost all states and the majority of districts are reporting norm-referenced achievement test results that are above the national median.

While supporting Cannell's general finding that it is more common for a state or district to obtain test results that are "above the national average," our analyses lead us to conclusions that are different, and certainly less sensational, than the ones he reached. To begin with, it is important to put the "above average" findings in context. Many students are receiving scores that are "below average" even in districts or states that are reporting that substantially more than 50% of their students are "scoring above the national average." When a district reports that 57% of its students obtained reading scores that are at or above the national median, for example, the other 43% of the students obviously scored below the median. It should also be emphasized that although most districts report results that are "above the national average," there are still many districts throughout the nation that are reporting results that are below average. One out of 10 districts in our sample, for example, reported that a only about a third of its students at a given grade scored above the national median in reading.

Cannell (1987) concluded that norm-referenced achievement tests are producing inflated reports from states and districts on the achievement of their students. But the finding that more than half the students are scoring above the national median that was obtained when the norms were established does not necessarily imply that the results are inflated. There are many factors that may lead to the general finding, but it seems clear that the use of "old" norms is one of the major factors that contributes to the abundance of "above average" scores.

The evidence reviewed provides strong support for the conclusion that norms obtained for Grades 1 through 8 during the late 1970s or early 1980s are easier on most tests than are more recent norms. Consequently, a state or district where the average student scores at the <u>current</u> national average will be accurately reported to be above a national average that is defined by norms that are several years old. It appears that a substantial fraction of the "Lake Wobegon" phenomenon may be attributable to the use of old norms. It should be noted that the use of "old" norms is not purposeful on the part of school districts or states; they generally use the most recent norms available. Since standardized tests are usually normed every seven years, the most recent norms available will be, on average, 3.5 years old in most school years.

Concerns about dated norms have led to suggestions that publishers should produce current annual norms (e.g, Cannell, 1988; Phillips and Finn, 1988) and publishers are now attempting to do this by obtaining weighted estimates of national results from user data (e.g., Rothman, 1988). As Shepard (1989) has pointed out, however, annual norms based on user data potentially have several serious defects. If users differ from nonusers in ways other than those reflected by the demographic variables used for weighting, then user-based annual norms may be worse than dated norms where there is at least an understood frame of reference. In particular, if test familiarity leads to higher test performance, a state or district that changes publishers and administers a several-years-old test form for the first time will be at a disadvantage when results are compared to user norms (Shepard, 1989).

The alternative of conducting special national norming studies every year, or even every other year, is not a realistic or desirable possibility. Norming is not only expensive, but the quality of the results is very dependent on voluntary participation of schools and well-motivated students. Current participation rates in norming studies conducted roughly every six or seven years by a publisher are already far lower than would be desired. More frequent attempts to norm tests would surely lower the participation rates still further and thereby degrade the quality of the norms. Finally, it should be noted that although more recent norms provide a more stringent standard of comparison when scores are going up as they have been during the last decade, they would provide a less stringent standard during periods of decline in scores such as that experienced between the mid 1960s and the mid 1970s. Thus, we do not believe that the use of annual norms is an appropriate or effective way to deal with problems caused by dated norms.

In any reporting of test scores emphasis needs to be given to the changing meaning of norms and the age of the norms that are used. It obviously is not sufficient to report that "students in state X are scoring above the national average" without clearly indicating the year in which the norms were obtained. Simply noting the year of the norms is not enough, however. An explanation of the implications of shifting norms also needs to be provided along with an indication of what is known about recent trends in the stringency of national norms.

There is ample evidence that scores on norm-referenced tests have been going up in Grades 1 through 8 in recent years. But the more important question is: Has student achievement improved in recent years? Unfortunately, the answer to the latter question is equivocal.

Achievement test scores are of interest to the degree that they enable valid inferences to be made about broader achievement domains. But little attention has been given to the issue of the degree to which valid generalizations about broad achievement domains can be made from state or district test results.

Comparisons of the changes in norms of standardized tests with estimates of changes in achievement based on NAEP results suggest that test norms may be changing more rapidly than is student achievement as measured by NAEP. The Wiser and Lenke (1987) findings that apparent increases are generally smaller for non-users than for users of a given test series suggest that part of the apparent growth in achievement based on norm-referenced test results may be due to increased familiarity with a particular form of a test. Only part of the apparent gain can be explained in this way, however.

The differences between the gains in performance indicated by NAEP and by norm-referenced tests, and between Wiser and Lenke's total norming sample and their non-users suggest at the very least that caution is needed in interpreting gains in norm-referenced test scores as reflections of the amount of improvement that has taken place in achievement, more broadly defined. More direct assessments of the degree of generalizability of results to other tests and to other indicators of student achievement are greatly needed.

Hoover's (1989) finding that only about 6% of the students scored below the 10th percentile in the first year of operational administration of Forms G and H of the ITBS suggests that roughly a third to a half of the difference between the percentage of students scoring above the national median and the naive expectation of 50% may occur in the first year of use and may be due to what happens with the least able students. This suggests that greater emphasis in reporting needs to be given to the lower end of the score distribution and to the students who are excluded from testing when results are reported by states or districts. It may be quite appropriate, indeed desirable, to exclude students with limited English proficiency or students receiving particular types of special education services from a norm-referenced test administration. Such students should not be ignored, however, when district or state achievement results are reported: At minimum, the number of such students and the reasons for exclusion from testing should be reported.

The practice of using a single form of a test year after year poses a logical threat to making inferences about the larger domain of achievement. Scores may be raised by focusing narrowly on the test objectives without improving achievement across the broader domain that the test objectives are intended to represent. Worse still, practice on nearly identical or even the actual items that appear on a test may be given. As Dyer aptly noted some years ago, "if you use the test exercises as an instrument of teaching you destroy the usefulness of the test as an instrument for measuring the effects of teaching" (1973, p. 89).

Current accountability pressures place great emphasis on test scores. It is unlikely that any single test, no matter how well constructed, normed, and validated, can withstand the pressures to serve both as an instrument of instruction and as an instrument for measuring the effects of instruction. Making valid inferences about broad achievement domains from test scores has always been a challenging and difficult undertaking, but it is made all the harder by current demands for accountability and the use of standardized test results as primary indicators of accountability.

References

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Baglin, R.F. (1981). Does "nationally" normed really mean nationally? *Journal of Educational Measurement, 18*, 97-107.

Baker, E.L. (1989). *What's the use? Standardized tests and educational policy* (Report to OERI, Grant No. G-86-0003). Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Burstein, L. (1989). *Looking behind the "average": How are states reporting test results?* (Report to OERI, Grant No. G-86-0003). Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Cannell, J.J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average* (2nd ed.). Daniels, WV: Friends for Education.

Cannell, J.J. (1988a). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice, 7*(2), 5-9.

Cannell, J.J. (1988b). The Lake Wobegon effect revisited. *Educational Measurement: Issues and Practice, 7*(4), 12-15.

CTB/McGraw-Hill. (1987). *Technical Report: California Achievement Tests, Forms E and F, Levels 10-20.* Monterey, CA: Author.

CTB/McGraw-Hill. (1988). *CTB/McGraw-Hill studies show students achieving at higher levels in basic skills* (CTB/McGraw-Hill press release). Monterey, CA: Author.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., & Chambers, D.L. (1988). *The mathematics report card. Are we measuring up? Trends and achievement based on the 1986 national assessment.* Princeton, NJ: Educational Testing Service.

Drahozal, E.C., & Frisbie, D.A. (1988). Riverside comments on the Friends of Education report. *Educational Measurement: Issues and Practice, 7*(2), 12-16.

Dyer, H.S. (1973). *Recycling the problems in testing. Proceedings of the 1972 Invitational Conference on Testing Problems.* Princeton, NJ: Educational Testing Service.

Gardner, E.F., Madeen, R., Rudman, H.C., Karlsen, B., Merwin, J.C., Callis, R., & Collins. (1983). *Stanford Achievement Test Series: Multilevel norms booklet: National.* San Antonio: The Psychological Corporation.

Gardner, E.F., Madeen, R., Rudman, H.C., Karlsen, B., Merwin, J.C., Callis, R., & Collins. (1987). *Stanford 7 Plus: Multilevel norms booklet: National.* San Antonio: The Psychological Corporation.

Harnischfeger, A., & Wiley, D.E. (1975). *Achievement test score decline: Do we need to worry?* Chicago: ML-Group for Policy Studies in Education.

Hieronymus, A.N., & Hoover, H.D. (1986). *Iowa Tests of Basic Skills Forms G/H: Manual for school administrators, Levels 5-14.* Chicago: The Riverside Publishing Co.

Hoover, H.D. (1989, June). *Reactions to "Lake Wobegon" one year later: Results from a replication of the Cannell study.* Symposium presented at the 1989 Assessment Conference of the Educational Commission of the States, Boulder, CO.

Korcheck, S.A. (1988). *Measuring student learning: Statewide student assessment programs in the SREB states.* Atlanta: Southern Regional Education Board.

Koretz, D. (1986). *Trends in educational achievement.* Washington, DC: Congressional Budget Office.

Koretz, D. (1987). *Educational achievement: Explanations and implications of recent trends.* Washington, DC: Congressional Budget Office.

Koretz, D. (1988). Arriving at Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8-15, 46-52.

Lenke, J.M., & Keene, J.M. (1988). A response to John J. Cannell. *Educational Measurement: Issues and Practice, 7*(2), 16-18.

Mehrens, W.A. & Kaminski, J. (1988, April). *Using commercial test preparation materials for improving standardized test scores: Fruitful, fruitless or fraudulent?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

National Assessment of Educational Progress. (1985). *The reading report card: Progress toward excellence in our schools: Trends in reading over four national assessments, 1971-1984.* Princeton, NJ: Educational Testing Service.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming, and equating. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.

Phillips, G.W., & Finn, C.E. Jr. (1988). The Lake Wobegon effect: A skeleton in the testing closet? *Educational Measurement: Issues and Practice, 7*(2), 10-12.

Prescott, G.A., Balow, I.H., Hogan, T.P., & Farr, R.C. (1978). *Metropolitan Achievement Tests Forms JS and KS, Complete Survey Battery: Teacher's manual for administering and interpreting.* San Antonio: The Psychological Corporation.

Prescott, G.A., Balow, I.H., Hogan, T.P., & Farr, R.C. (1986). *MAT6, Metropolitan Achievement Tests Forms L and M: Survey battery multilevel national norms booklet.* San Antonio: The Psychological Corporation.

The Psychological Corporation. (1988). *MAT5 to MAT6 Survey: Corresponding scaled scores of the fifth and sixth editions of the Metropolitan Achievement Test Survey Battery.* San Antonio: The Psychological Corporation.

Qualls-Payne, A.L. (1988). SRA Response to Cannell's article. *Educational Measurement: Issues and Practice, 7*(2), 21-22.

Rothman, R. (1988, November, 16). Pupil gains not just mythical, test publisher says study shows. *Education Week*, p. 20.

Science Research Associates. (1979). *SRA Achievement Series Forms 1 and 2: Norms and conversion tables.* Chicago: Science Research Associates, Inc.

Science Research Associates. (1986). *SRA Achievement Series Forms 1 and 2: Answer keys, norms, and conversion tables.* Chicago: Science Research Associates, Inc.

Shepard, L.A. (1989). *"Inflated test score gains": Is it old norms, or teaching the test?* (CSE Technical Report 307). Los Angeles: UCLA Center for the Study of Evaluation.

Southern Regional Education Board. (1984). *Measuring educational progress in the South: Student assessment.* Atlanta: Southern Regional Education Board.

Stonehill, R.M. (1988). Norm-referenced test gains may be real: A response to John Jacob Cannell. *Educational Measurement: Issues and Practice, 7*(2), 23-24.

Williams, P.L. (1988). The time-bound nature of norms: Understandings and misunderstandings. *Educational Measurement: Issues and Practice, 7*(2), 18-21.

Wiser, B., & Lenke, J.M. (1987, April). *The stability of achievement test norms over time.* Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.

Yen, W.M., Green, D.R., & Burket, G.R. (1987). Valid normative information from customized tests. *Educational Measurement: Issues and Practice, 6*(1), 7-13.

Appendix A

Sample Letter and Data Collection Form for Directors of State Testing Programs

July 22, 1988

States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP
States Data:NOT ON DESKTOP

Dear States Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the Office of Educational Research and Improvement (OERI). This study was sitmulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. The interpretation of normative comparisons was called into question by Cannell's finding that "no state scores below the publisher's 'national norm' at the elementary level on any of the six major nationally normed, commercially available tests" (p. 2 of second edition of Cannell Report). The value of assessment results was further challenged by Cannell's conclusion that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of <u>Educational Measurement: Issues and Practice</u>, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have a better understanding of the magnitude and prevalence of the apparently high achievement results reported by Cannell. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information that will provide a better data base for determining not only what proportion of students score above determining not only what proportion of students score above the 50th percentile according to national norms, but other important characteristics of the test results such as changes in means over time and the variability in scores. We also need to obtain information on the way in which test results are currently used (e.g., public reporting, grade retention, school incentives, etc.), when these uses were instituted, and planned changes in the use of test results. Finally, we are seeking information about policies regarding test security and guidelines on preparation of students for taking tests.

A CRESST staff member will be contacting you by phone to seek your assistance and to arrange for a time for a phone interview with an appropriate person on your staff. The information that will be requested is outlined on the enclosure. We will send you more detailed worksheets between now and the time of the telephone interview to help organize the requested information.

In many cases, the information that we are seeking may be provided in reports that have previously been prepared. Thus we request that you send us copies of any reports that give summaries of district results that have been published within the past three years. Copies of press releases and newspaper articles about the test results would also be useful. If you send us reports and press releases as quickly as possible, we will use the reports to extract as much of the requested information as possible. We will call you to ask questions after we have "done our homework".

Please send reports to: Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Thank you for your consideration. We will phone you within the next two weeks to answer questions and to try to arrange a time for a telephone interview. A return postcard is enclosed so that you can indicate the name, phone number, and best times for us to try to contact the appropriate person for the telephone interview.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for the Study of Research on Evaluation Standards, and Student Testing

Explanation of Information Requested

Column	Information requested
1	Testing year
2	Grade levels tested K - 12.
3	Name of test used for statewide assessment e.g., CTBS, MAT, name of locally developed test.
4	Edition of the test used at each grade level, e.g., 1982.
5	Form of the test used at each grade level.
6	Year when test was first used.
7	Norming year of test used for reporting scores.
8	Month in which tests were administered.
9	Type of scores reported, e.g., percent correct, percentile rank, NCE. n.b. If you have more than one type of score, please provide one form of data in the preferred order as follows: Percentile Rank Grade Equivalents NCE Stanines Percent Correct ...
10	Number of students enrolled: the total number of students by grade statewide.
11	Number of students tested.
12	Number of students' scores reported: If not all scores are used to compute rankings or other statewide test results, enter the number of students' scores used to compute the achievement data.
13	Reading %: The percent of students scoring above the national 50th percentile statewide.
14	Math %: The percent of students scoring above the national 50th percentile statewide.

n.b. If neither reading nor math data requested in 12 and 13 are available, please provide the most appropriate composite scores and indicate the nature of these on the form.

If the data requested in columns 13 or 14 (percent of students scoring above the national 50th percentile) are not available, please provide as much of the following as possible (columns 15 - 20 on the Alternate Information Sheet):

<u>Column</u>

15 Reading statewide mean.

16 Reading statewide standard deviation.

17 Math statewide mean.

18 Math statewide standard deviation.

19 Reading score at each percentile: The score at the 25th percentile statewide.

- at the 50th percentile statewide.
- at the 75th percentile statewide.

20 Math score at each percentile: The math score at the 25th percentile statewide.

- at the 50th percentile statewide.
- at the 75th percentile statewide.

Type of scores: If the type of scores reported in columns 13-20 are not the same as those indicated in column 9, please indicate the type of scores used to compute the percentiles, mean, and standard deviations.

Statewide Testing Information

____
State Name

____
Person Supplying Information

____
Title

1 Testing Year	2 Grade	3 Test Name	4 Edition	5 Form	6 Year First Used	7 Norming Year	8 Testing Dates	9 Type of Scores
1985-1986								
1986-1987	K							
1987-1988								
1985-1986								
1986-1987	1							
1987-1988								
1985-1986								
1986-1987	2							
1987-1988								
1985-1986								
1986-1987	3							
1987-1988								
1985-1986								
1986-1987	4							
1987-1988								
1985-1986								
1986-1987	5							
1987-1988								
1985-1986								
1986-1987	6							
1987-1988								

1985-1986								
1986-1987	7							
1987-1988								
1985-1986								
1986-1987	8							
1987-1988								
1985-1986								
1986-1987	9							
1987-1988								
1985-1986								
1986-1987	10							
1987-1988								
1985-1986								
1986-1987	11							
1987-1988								
1985-1986								
1986-1987	12							
1987-1988								

Please Refer to Explanation of Information Requested - Attached

10 Number of Students Enrolled	11 Number of Students Tested	12 Number of Students' Scores Reported	13 Reading: % of Students above National 50%ile	14 Math: % of Students above National 50%ile

Alternate Information Available

State Name

Person Supplying Information

Title

		15	16	17	18	19			20		
Testing Year	Grade	Reading		Math		Reading Score at each percentile			Math Score at each percentile		
		Mean	Standard Deviation	Mean	Standard Deviation	25	50	75	25	50	75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

1985-1986											
1986-1987	7										
1987-1988											
1985-1986											
1986-1987	8										
1987-1988											
1985-1986											
1986-1987	9										
1987-1988											
1985-1986											
1986-1987	10										
1987-1988											
1985-1986											
1986-1987	11										
1987-1988											
1985-1986											
1986-1987	12										
1987-1988											

Appendix B

Interview Guide

code

District

State

Interviewer

date

Person(s) Interviewed

name | name

title | title

Background information: Number of schools in district

Size (range)

Center for the Study of Research on Evaluation, Standards, and Student Testing, Robert L. Linn, School of Education, University of Colorado at Boulder

Part I: District Testing Data (to be recorded on the forms provided)

YEARS TESTED 1. Are districtwide test results available for:

_____1987-88

_____1986-87

_____1985-86 If none, then the most recent year: _____

If there is no districtwide testing, ask only 12, 13, 19 - 22, and 26 for large districts.

ENROLLMENT BASIS 2. What is the basis for the enrollment figures used to give the number of students in each grade? (e.g., ADA= Average Daily Attendance)

ENROLLMENT SOURCE 3. What office provides the enrollment figures? [name of person and phone number if easily available]

TESTED = REPORTED 4. Is the number of students tested the same as the number of students that are included in the reported test results?

Yes_____ No_____

If no, how does the number included in the reported test results differ from the number tested?

probe: special education

SAMPLING PLAN

5. Were all eligible students in the grade tested or is a sampling plan used?

___ universal testing by grade ___ sampling plan

Please describe any sampling procedures used.

TESTING EXCLUSIONS

6. What rules are used to determine students who are excluded from testing?

request: copies of any written policies that describe these rules

% EXCLUDED

7. How many students (or what percent of the students) are excluded using these rules?

MAKE-UP TESTING

8. What are the policies for make-up testing (for students who are absent)?

request: if in writing

[Ask the following only if needed:]

LOCALLY CONSTRUCTED TEST

9. If a specially constructed test is used, is it linked to a norm-referenced test? If so, what is the name and edition of the norm referenced test?

REPORTING NATIONAL COMPARISONS

10. If the percent of students above the 50th percentile is unknown, please describe the way in which scores are reported and comparisons are made to the national norm.

LOCAL FACTORS IN TEST SCORES

11. Are any factors of schools or the characteristics of their students taken into account in reporting test scores?

(e.g., percent minority, percent eligible for free lunch, Chapter I)

[BEGIN TAPE RECORDING]

Part II: Testing Policies and Perceptions

USES AND IMPORTANCE

12. What are the uses of test results?

___ -local district and school instructional and evaluation decisions

___ -reporting to parents about individual student progress or school programs

___ -School Board attention (And if so, how have Board members used test results-- to increase testing programs or other forms of accountability?

___ -state or local politician use of scores in campaigning or proposing legislation

___ -changing general funding levels for schools

___ -targeted funds or mandating programs such as remediation

___ -superintendent, principal, or teacher performance rating or jobs

___ -media coverage and community awareness

*** How important are test scores in your district?

/______________/______________/______________/______________/

extremely very moderately slightly not important

REFORMS

13. Have major educational reforms been introduced in your district in the past five years?

request: Would you briefly describe these or send us written descriptions that are available?

TEST SELECTION

14. Who selected the standardized test(s) being used? (If locally developed, how was the content selected?)

probe: committee composition, e.g., teachers, parents,...

CURRICULUM ALIGNMENT

15. Have there been efforts to assure that the curriculum and the test are aligned?

If so, please describe those efforts.

TIME ON SPECIFIC OBJECTIVES

16. Do you think that teachers spend more time teaching the specific objectives on the test(s) than they would if the tests were not required?

How much more time?

IMPORTANT OBJECTIVES GIVEN LESS TIME

17. To what extent do you think important objectives are given less time or emphasis because they are not included on the test?

What kinds of objectives are neglected?

INFORMAL GUIDELINES ABOUT TEST PREPARATION

18. Do you or members of your staff provide informal guidelines about test preparation? What kind of advice do you give schools about how to prepare students to take tests?

probes:

- length of time to practice
- minimum and maximum recommended time for practice
- whether to use items in a specific format for practice

TECHNICAL ASSISTANCE ABOUT TEST PREPARATION

19. What kind of technical assistance or materials do you provide to schools about test preparation?

request: Would you send us copies of the materials or descriptions of the assistance?

probes:

- practice tests
- testwiseness packages
- curriculum domain materials but not specific test items
- amount of these activities

TYPICAL PRACTICES OF TEST PREPARATION

20. Can you describe typical practices of test preparation?

probes:

- If they say, one school does X, ask how common this is, or how many other schools do the same.
- Do schools use the materials and assistance you provide?
- What else do they do beyond what you recommend?

EXTREME PRACTICES OF TEST PREPARATION

21. Can you describe <u>extreme</u> cases of test preparation?

probes:

If they describe a worst case, ask what they would think of as a best case. (as well as what is more typical, above)
Examples of cases which violate your recommendations?

TEST ADMINISTRATION AND SECURITY POLICIES

22. Do you have written policies regarding test administration and security procedures?
If not, do you have informal guidelines?

<u>request</u>: written policies

WHO ADMINISTERS OR HAS TESTS OR KNOWS TESTS

23. Who administers the tests?
Do teachers in some schools have copies of the tests prior to test administration?

How familiar are teachers with the specific items on the tests?

probes:
teachers administering same test over years
principals or teachers having test files

DETECT ANOMALIES

24. Do you have any formal procedures for detecting anomalies in the data?

request copies

probes:
check for missing test booklets
computer detection of significant numbers of erasures
" " of extraordinary gains from one year to the next
check numbers of students tested against enrollment

TYPICAL AND EXTREME PRACTICES

25. Can you give examples of both typical and extreme testing practices?

Have you withheld score reports because of suspected cheating?

probes:

good practices: consistent, successful make-up testing

examples of cheating-
- teachers filling in answers
- extending time limits for tests
- teaching specific items on the test
- discrepancies in numbers of students tested

[Ask the following only in districts designated as 7's or 8's- large districts]

REACTIONS TO CANNELL REPORT

26. What are your reactions to the Cannell report and its conclusions?

FACTORS IN ACHIEVEMENT TRENDS

27. What do you think are the primary factors that contribute to the recent trends in achievement test scores in your district?

probes:
- educational reforms
- norms (unrepresentative or old)
- pressure on teachers to have high scores

Closing:

When finishing and thanking them for their time, review the things which you may have requested in writing.

Checklist of Requested Written Information

___ testing data on years not yet received (e.g., all three years 1985-1988)

___ testing data such as distribution measures

___ #3- name and phone of office or person with enrollment figures

___ #6- Rules for testing exclusions

___ #8- Policies for make-up testing

___ #13- Educational reforms in the state

___ #19- Technical assistance or materials for test preparation

___ #22- Test administration and security policies

___ #24- Procedures for detecting anomalies

The address for mailing is:

Dr. Robert Linn 303-492-8280 (Bob)
University of Colorado or -2124
(Nancy)
School of Education or -3108
(Lorrie)
Campus Box 249
Boulder, CO 80309

If you have missing answers and have to schedule another call, please indicate that in the telephone log.

Appendix C

Districts Available by Cells of Sampling Design

Appendix C

Number of Districts Available by Cells in Sampling Design

Region	District Size	SES Level	Number of Districts Available
East	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	2
		Below Average	4
		Average	0
		Above Average	1
		High	1
	50,000 to 99,999	Low	1
		Below Average	2
		Average	1
		Above Average	1
		High	0
	100,000 or more	Low	1
		Below Average	2
		Average	0
		Above Average	2
		High	1

Appendix C (page 2 of 4)

Region	District Size	SES Level	Number of Districts Available
North/ Central	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	1
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	0
		Below Average	2
		Average	5
		Above Average	5
		High	4
	50,000 to 99,999	Low	1
		Below Average	3
		Average	2
		Above Average	0
		High	0
	100,000 or more	Low	0
		Below Average	1
		Average	1
		Above Average	0
		High	0

Appendix C (page 3 of 4)

Region	District Size	SES Level	Number of Districts Available
South	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	2
		High	3
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	2
		High	0
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	3
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	4
	25,000 to 49,999	Low	2
		Below Average	5
		Average	5
		Above Average	5
		High	2
	50,000 to 99,999	Low	1
		Below Average	3
		Average	5
		Above Average	5
		High	1
	100,000 or more	Low	0
		Below Average	1
		Average	5
		Above Average	0
		High	1

Appendix C (page 4 of 4)

Region	District Size	SES Level	Number of Districts Available
West	Less than 1,200	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	1,200 to 2,499	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	2,500 to 4,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	5,000 to 9,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	10,000 to 24,999	Low	5
		Below Average	5
		Average	5
		Above Average	5
		High	5
	25,000 to 49,999	Low	2
		Below Average	2
		Average	5
		Above Average	5
		High	5
	50,000 to 99,999	Low	1
		Below Average	1
		Average	5
		Above Average	5
		High	1
	100,000 or more	Low	0
		Below Average	0
		Average	3
		Above Average	1
		High	0

Appendix D

Sample Letters, Data Collection Forms, and Questionnaires Sent to Districts

August 18, 1988

Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP
Dist Phone Data:NOT ON DESKTOP

Dear Dist Phone Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the U.S. Department of Education's Office of Educational Research and Improvement (OERI). This study was stimulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you may know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. Based on his survery of states and selected school districts, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of Educational Measurement: Issues and Practice, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have technically accurate information about achievement results reported by school districts across the nation. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information from a nationally representative sample of school districts that will provide a better data base for determining not only what level of student performance is being reported, but the uses and interpretations that are being made of the results. We also are seeking information about factors that may influence test results.

Your district has been selected as part of a nationally representative sample for this study. Hence, your participation is critical to maintaining representativeness and drawing conclusions about achievement testing for the nation. Results will not be reported for individual school districts. However, participation by each sampled district is essential to ensuring an accurate picture for the nation as a whole.

We ask that you complete the enclosed questionnaire about your district's testing program. In many cases, the information that we are seeking on the forms may be provided in reports that have previously been prepared. If so, we request that you answer the general questionnaire items and send us the questionnaire along with copies of any reports that give results of districtwide assessments of student achievement or summaries of district results that have been published within the past three years. We will use those reports to obtain the requested information. Copies of press releases and newspaper articles about the test results would also be useful.

Please return the completed questionnaire in the enclosed envelope to:

Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

We also ask you to participate in a telephone interview which concerns additional questions about testing policies and practices. In order to schedule an interview, we ask that you indicate on the questionnaire dates and times which would be convenient for one of our staff members to call. The interviews consist of fifteen questions about your testing program and usually last about 30 minutes.

Thank you for your consideration. We realize that school districts receive many requests for information and that responding to such requests is a burden on your time. Your willingness to help is essential to the success of the study and to our ability to provide solid answers to the important educational questions that were raised by the Cannell report.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for Research on Evaluation, Standards, and Student Testing

August 18, 1988

Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP
Dist Survey Data:NOT ON DESKTOP

Dear Dist Survey Data:NOT ON DESKTOP:

We seek your assistance in a study that is being conducted by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) on behalf of the U.S. Department of Education's Office of Educational Research and Improvement (OERI). This study was stimulated by the report "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above Average" by Dr. John J. Cannell. As you may know, this report attracted considerable attention in the press and has been of great interest at OERI and among those concerned about the assessment of educational achievement.

Cannell's findings and conclusions are both provocative and controversial. Based on his survery of states and selected school districts, Cannell concluded that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels" (p. 6 of special issue of Educational Measurement: Issues and Practice, 1988, Vol. 7, No. 2).

Given the importance that is attached to student achievement and the widespread use of normative comparisons, Cannell's findings and conclusions deserve close scrutiny. We need to have technically accurate information about achievement results reported by school districts across the nation. We also need to have a better understanding of the factors which may contribute to and explain the findings.

To achieve these goals, we need your help in collecting information from a nationally representative sample of school districts that will provide a better data base for determining not only what level of student performance is being reported, but the uses and interpretations that are being made of the results. We also are seeking information about factors that may influence test results.

Your district has been selected as part of a nationally representative sample for this study. Hence, your participation is critical to maintaining representativeness and drawing conclusions about achievement testing for the nation. Results will not be reported for individual school districts.

However, participation by each sampled district is essential to ensuring an accurate picture for the nation as a whole.

We ask that you complete the enclosed questionnaire about your district's testing program. In many cases, the information that we are seeking on the forms may be provided in reports that have previously been prepared. If so, we request that you answer the general questionnaire items and send us the questionnaire along with copies of any reports that give results of districtwide assessments of student achievement or summaries of district results that have been published within the past three years. We will use those reports to obtain the requested information. Copies of press releases and newspaper articles about the test results would also be useful.

Please return the completed questionnaire in the enclosed envelope to:

Robert L. Linn
School of Education
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Thank you for your consideration. We realize that school districts receive many requests for information and that responding to such requests is a burden on your time. Your willingness to help is essential to the success of the study and to our ability to provide solid answers to the important educational questions that were raised by the Cannell report.

Sincerely,

Eva L. Baker
UCLA

Robert L. Linn
University of Colorado-Boulder

Co-Directors, Center for Research on Evaluation, Standards, and Student Testing

District Testing Information

District Name

Person Supplying Information

State

Address

Phone Number

Title

1 Testing Year	2 Grade	3 Test Name	4 Edition	5 Form	6 Year First Used	7 Norming Year	8 Testing Dates	9 Type of Scores
1985-1986								
1986-1987	K							
1987-1988								
1985-1986								
1986-1987	1							
1987-1988								
1985-1986								
1986-1987	2							
1987-1988								
1985-1986								
1986-1987	3							
1987-1988								
1985-1986								
1986-1987	4							
1987-1988								
1985-1986								
1986-1987	5							
1987-1988								
1985-1986								
1986-1987	6							
1987-1988								

1985-1986								
1986-1987	7							
1987-1988								
1985-1986								
1986-1987	8							
1987-1988								
1985-1986								
1986-1987	9							
1987-1988								
1985-1986								
1986-1987	10							
1987-1988								
1985-1986								
1986-1987	11							
1987-1988								
1985-1986								
1986-1987	12							
1987-1988								

Please Refer to Explanation of Information Requested - Attached

10 Number of Students Enrolled	11 Number of Students Tested	12 Number of Students' Scores Reported	13 Reading: % of Students above National 50%ile	14 Math: % of Students above National 50%ile

8 - 11. Please indicate below the name of the test used at each grade level tested, (for standardized tests, include edition and form), the number of students tested, AND THE PERCENT OF STUDENTS ABOVE THE NATIONAL 50TH PERCENTILE. (If the percent of students above the national 50th percentile is not available, please provide as much of the information on pages 4 and 5 as possible.)

		8	9	10	11
Testing Year	Grade	Test Name, Edition and Form	Number of Students Tested	Reading: % of Students above National 50%ile	Math: % of Students above National 50%ile
1985-1986					
1986-1987	K				
1987-1988					
1985-1986					
1986-1987	1				
1987-1988					
1985-1986					
1986-1987	2				
1987-1988					
1985-1986					
1986-1987	3				
1987-1988					
1985-1986					
1986-1987	4				
1987-1988					
1985-1986					
1986-1987	5				
1987-1988					
1985-1986					
1986-1987	6				
1987-1988					

Testing Year	Grade	Test Name, Edition and Form	Number of Students Tested	Reading: % of Students above National 50%ile	Math: % of Students above National 50%ile
1985-1986					
1986-1987	7				
1987-1988					
1985-1986					
1986-1987	8				
1987-1988					
1985-1986					
1986-1987	9				
1987-1988					
1985-1986					
1986-1987	10				
1987-1988					
1985-1986					
1986-1987	11				
1987-1988					
1985-1986					
1986-1987	12				
1987-1988					

12. Testing Dates (month/year) ______________________

13. Norming year of norm referenced test(s) used: ______________________

14. Year these tests were first used in your district: ______________________

If the percent of students above the national 50th percentile is provided on pages 2 and 3, pages 4 and 5 need not be completed. Skip to page 6.

If the number of students above the national 50th percentile (columns 10 and 11, pages 2-3) is <u>not</u> known, please provide as much of the following information as possible.

Testing Year	Grade	Reading		Math		Reading Score at each percentile			Math Score at each percentile		
		Mean	Standard Deviation	Mean	Standard Deviation	25	50	75	25	50	75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

Testing Year	Grade	Reading		Math		Reading Score at each percentile			Math Score at each percentile		
		Mean	Standard Deviation	Mean	Standard Deviation	25	50	75	25	50	75
1985-1986											
1986-1987	7										
1987-1988											
1985-1986											
1986-1987	8										
1987-1988											
1985-1986											
1986-1987	9										
1987-1988											
1985-1986											
1986-1987	10										
1987-1988											
1985-1986											
1986-1987	11										
1987-1988											
1985-1986											
1986-1987	12										
1987-1988											

Alternate Information Available
District Test Results

District Name ____________________

Person Supplying Information ____________________

State ____________________

Address ____________________ Phone Number ____________________

Title ____________________

		15	16	17	18	19			20		
Testing Year	Grade	Reading Mean	Reading Standard Deviation	Math Mean	Math Standard Deviation	Reading Score at each percentile 25	50	75	Math Score at each percentile 25	50	75
1985-1986											
1986-1987	K										
1987-1988											
1985-1986											
1986-1987	1										
1987-1988											
1985-1986											
1986-1987	2										
1987-1988											
1985-1986											
1986-1987	3										
1987-1988											
1985-1986											
1986-1987	4										
1987-1988											
1985-1986											
1986-1987	5										
1987-1988											
1985-1986											
1986-1987	6										
1987-1988											

1985-1986					
1986-1987	7				
1987-1988					
1985-1986					
1986-1987	8				
1987-1988					
1985-1986					
1986-1987	9				
1987-1988					
1985-1986					
1986-1987	10				
1987-1988					
1985-1986					
1986-1987	11				
1987-1988					
1985-1986					
1986-1987	12				
1987-1988					

Explanation of Information Requested

Column	Information requested
1	Testing year
2	Grade levels tested K - 12.
3	Name of test used e.g., CTBS, MAT, name of locally developed test.
4	Edition of the test used at each grade level, e.g., 1982.
5	Form of the test used at each grade level.
6	Year when test was first used.
7	Norming year of test used for reporting scores.
8	Month in which tests were administered.
9	Type of scores reported, e.g., percent correct, percentile rank, NCE. n.b. If you have more than one type of score, please provide one form of data in the preferred order as follows: Percentile Rank Grade Equivalents NCE Stanines Percent Correct ...
10	Number of students enrolled: the total number of students enrolled by grade
11	Number of students tested at each grade
12	Number of students' scores reported: If not all scores are used to compute rankings or other statewide test results, enter the number of students' scores used to compute the achievement data.
13	Reading %: The percent of students scoring above the national 50th percentile.
14	Math %: The percent of students scoring above the national 50th percentile.

n.b. If neither reading nor math data requested in 12 and 13 are available, please provide the most appropriate composite scores and indicate the nature of these on the form.

If the data requested in columns 13 or 14 (percent of students scoring above the national 50th percentile) are not available, please provide as much of the following as possible (columns 15 - 20 on the Alternate Information Sheet):

Column

15 Reading mean for the district.

16 Reading standard deviation.

17 Math mean.

18 Math standard deviation.

19 Reading score at each percentile: The score

- at the 25th percentile districtwide
- at the 50th percentile districtwide.
- at the 75th percentile districtwide.

20 Math score at each percentile: The math score

- at the 25th percentile districtwide.
- at the 50th percentile districtwide.
- at the 75th percentile districtwide.

Type of scores: If the type of scores reported in columns 13-20 are not the same as those indicated in column 9, please indicate the type of scores used to compute the percentiles, mean, and standard deviations.

Appendix E

District Subsample for Telephone Interviews

Appendix E

District Subsample for Telephone Interviews

The 40 cells (5 levels of SES by 8 levels of district size) within each of the 4 regions that were used to define the overall district sample were collapsed to 15 cells (3 levels of SES by 5 levels of district size) to select the subsample to be interviewed by telephone. The following levels were combined for each factor.

SES		Size	
Subsample Level	Total Sample Level	Subsample Level	Total Sample Level
1 Below Average	Low & Below Average	1 <2,500	<1,200 & 1,200-2,499
2 Average	Average	2 2,500-9,999	2,500-4,999 & 5,000-9,999
3 Above Average	Above Average & High	3 10,000-49,999	10,000-24,999 & 25,000-49,999
		4 50,000-99,999	50,000-99,999
		5 100,000 +	100,000 +

For cells of the subsample design that consisted of 2 or 4 of the cells of the total sample, one district was randomly selected. The SES = 1, size = 1 cell of the interview subsample, for example, consists of SES by size cells 11, 12, 21, and 22 in the total sample. A random number between 1 and 4 corresponding to each of those original cells was selected for each region. Following this procedure for each of the interview subsample cells that contained more than one cell from the total sample, 56 districts (4 regions x 3 SES levels x 5 size levels minus 4 void cells) for the interview subsample were selected.

Appendix E (Continued, page 2 of 2)

Using the total sample code RZS where
R = region (1 = East, 2 = North/Central, 3 = South, and 4 = West);
Z = size (1 = less than 1,200, 2 = 1,200-2,499, 3 = 2,500-4,999, 4 = 5,000-9,999, 5 = 10,000-24,999, 6 = 25,000-49,999, 7 = 50,000-99,999, and 8 = 100,000 or more); and
S = SES (1 = low, 2 = below average, 3 = average, 4 = above average, and 5 = high),
the following interview subsample was selected.

112	211	312	411
123	213	323	415
124	225	324	423
131	233	332	432
134	242	335	433
145	245	343	445
153	251	353	454
155	255	362	462
161	263	365	463
172	272	371	471
173	273	373	474
174	275(void)	374	474
181	282	382	481(void)
183(void)	283	383	483
184	285(void)	385	484

Appendix F

Grades Tested by Districts Returning Data

Appendix F (page 1 of 4)

Grades Tested by Districts Returning Data

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
1	1	1		+	+	+	+	+	+						
1	1	3		+	+	+	+	+	+	+	+				
1	1	4				+	+	+	+	+					
1	2	1		+	+	+	+	+	+	+	+	+	+		
1	2	2			+		+		+		+		+		
1	2	4			+		+		+		+		+		
1	2	5				+	+	+							
1	3	1			+	+	+	+	+	+	+				
1	3	2		+	+	+	+	+	+	+	+				
1	3	3			+	+	+	+	+	+					
1	3	4										+	+		
1	3	5	+		+	+	+	+	+	+	+				
1	4	2	+	+	+	+	+	+	+	+	+		+		
1	4	3				+		+		+			+		
1	4	4		+	+	+	+	+	+	+	+	+	+		
1	4	5			+		+	+		+		+			
1	5	3				+		+			+				
1	5	4			+		+		+						
1	5	5	Interview completed - no normed test results												
1	6	1		+	+	+	+	+	+	+	+	+	+		
1	6	2			+	+	+	+	+	+	+	+	+	+	+
1	6	4				+		+			+				
1	6	5				+		+			+			+	
1	7	1	+	+	+	+	+	+	+	+	+	+	+	+	+
1	7	2		+	+	+	+	+	+	+	+				
1	7	2		+	+	+	+	+	+	+	+	+	+	+	+
1	7	3			+				+		+	+		+	
1	7	4				+		+			+				
1	8	1		+	+	+	+	+	+	+	+	+	+	+	
1	8	2	+	+	+	+	+	+	+	+	+	+	+	+	+
1	8	2			+	+	+	+	+	+	+	+	+		
1	8	3				+		+			+				
1	8	4				+		+			+				
1	8	5				+		+			+			+	
2	1	1		+	+	+	+	+	+	+	+	+	+	+	
2	1	3	+	+	+	+	+	+	+	+	+	+	+		
2	1	4	+	+	+	+	+	+	+	+	+	+	+	+	+
2	1	5	+	+	+	+	+	+	+	+					
2	2	4		+	+	+	+	+	+	+	+				
2	2	5			+	+	+	+	+	+			+		
2	3	1				+		+		+		+	+		
2	3	2	Questionnaire completed - no usable test results												
2	3	3				+		+	+						
2	3	4				+	+	+	+	+	+				
2	4	1	+	+	+	+	+	+	+	+	+	+	+	+	+

Appendix F (page 2 of 4)

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
2	4	2		+	+	+	+	+	+		+	+	+	+	
2	4	3									+				
2	4	4				+	+	+	+			+		+	
2	5	2		+	+	+	+	+	+	+	+	+	+	+	+
2	5	3		+	+	+	+	+				+			+
2	5	4	Criterion Referenced Test results only												
2	5	5				+			+		+	+			
2	6	2		+	+	+	+	+	+	+	+	+	+	+	
2	6	3		+	+	+	+	+	+	+	+	+	+	+	
2	6	4				+			+		+		+		
2	6	5				+			+		+				
2	6	5				+		+			+			+	
2	7	1	+	+	+	+	+	+	+	+	+	+	+	+	+
2	7	2	+	+	+	+	+	+	+	+	+	+	+	+	+
2	7	2		+	+	+	+	+	+	+	+	+	+	+	+
2	7	2		+	+	+	+	+	+	+	+	+	+	+	
2	7	3	+		+			+		+			+		
2	7	3			+	+	+	+	+	+	+	+			
2	8	2	+			+		+			+			+	
2	8	3		+	+	+	+	+	+	+	+				
3	1	1	+		+	+	+	+	+	+	+	+	+	+	+
3	1	2	+	+	+	+	+	+	+	+	+				
3	1	4	+	+	+	+	+	+	+	+	+	+	+	+	+
3	1	5			+	+	+	+	+	+	+		+	+	
3	2	1	+	+	+	+	+	+	+	+	+	+	+		
3	2	2	+	+	+	+	+	+	+	+	+				
3	2	3		+	+	+	+	+	+	+	+	+		+	
3	2	4	+	+	+	+	+	+	+	+	+	+	+	+	+
3	3	1	+	+	+	+	+	+	+	+	+	+	+	+	+
3	3	2	+	+	+	+	+	+	+	+	+				
3	3	3		+	+	+	+	+	+	+	+	+	+		
3	3	4	+	+			+		+						
3	3	5		+	+		+	+		+	+		+		
3	4	1		+	+		+	+							
3	4	2		+	+	+	+	+	+	+	+	+	+	+	
3	4	3	+	+	+	+	+	+	+	+	+				
3	4	4	+	+	+	+	+	+	+	+	+	+	+	+	
3	4	5	+	+	+	+	+	+	+	+		+			+
3	5	1					+		+			+			
3	5	3			+		+			+		+		+	
3	5	4		+	+	+	+	+	+	+	+	+	+	+	
3	5	5		+	+	+	+	+	+	+	+	+	+	+	
3	6	1				+			+		+				
3	6	2				+	+	+			+			+	
3	6	3				+	+	+	+			+		+	
3	6	4					+				+			+	
3	6	5					+				+			+	

Appendix F (page 3 of 4)

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
3	7	1		+	+	+	+	+	+	+	+	+	+	+	
3	7	2		+	+		+	+		+	+		+		
3	7	2	+	+	+	+	+	+	+	+	+	+	+	+	+
3	7	2		+	+	+	+	+	+	+	+	+			
3	7	3			+			+		+		+			+
3	7	3		+	+	+	+	+	+	+	+	+			
3	7	3			+	+	+	+	+	+	+	+			
3	7	3		+	+	+	+	+	+	+	+				
3	7	3		+	+		+	+		+	+		+		
3	7	4	+	+	+	+	+	+	+	+	+	+	+		
3	7	4					+				+			+	
3	7	5			+	+	+	+	+	+		+			
3	8	2	+	+	+	+	+	+	+	+	+	+	+		
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	+
3	8	3		+	+	+	+	+	+	+	+	+	+	+	+
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
3	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
3	8	5			+		+		+		+			+	
4	1	1		+	+	+	+	+	+	+	+				
4	1	2	+	+	+	+	+	+	+	+	+	+	+	+	
4	1	3		+	+	+	+	+	+	+	+				
4	2	2		+	+	+	+	+	+	+	+			+	
4	2	4			+		+		+		+		+		
4	2	4	+		+		+		+		+				
4	2	5				+	+		+	+	+		+		
4	3	1	+	+	+	+	+	+	+	+	+	+	+	+	
4	3	2		+	+	+	+	+	+	+	+		+		
4	3	3		+	+	+	+	+	+	+	+	+	+	+	
4	3	4					+				+		+		
4	3	5		+	+	+	+	+	+	+					
4	4	1		+	+	+	+	+	+	+	+				
4	4	2		+	+	+	+	+	+	+	+	+	+	+	+
4	4	4			+	+	+	+	+	+	+				
4	4	5	+	+	+	+	+	+	+	+	+				
4	5	1		+	+	+	+	+	+	+	+	+	+	+	+
4	5	2	+	+	+	+	+	+	+	+	+	+			
4	5	3	Only Chapter I test data provided												
4	5	4		+	+	+	+	+	+						
4	5	5				+	+		+	+	+	+	+		
4	6	1			+	+	+	+	+	+	+	+	+		
4	6	2			+		+		+		+				
4	6	3		+	+	+	+	+	+						
4	6	4			+	+	+	+	+	+	+	+			
4	6	5	+	+	+	+	+	+	+	+		+		+	
4	7	1			+	+	+	+	+	+	+	+	+	+	
4	7	2			+	+	+	+	+	+	+	+	+	+	
4	7	3	+	+	+	+	+	+	+	+	+	+	+	+	

Appendix F (page 4 of 4)

Region	Size	SES	K	1	2	3	4	5	6	7	8	9	10	11	12
			Grade												
4	7	3				+		+			+				
4	7	3	+	+	+	+	+	+	+	+	+	+	+	+	+
4	7	3		+	+	+	+	+	+	+	+	+	+	+	+
4	7	3	Criterion Referenced Test results only												
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	+
4	7	4				+	+		+		+	+	+	+	
4	7	4				+			+		+	+		+	
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	
4	7	4	+	+	+	+	+	+	+	+	+	+	+	+	
4	7	5				+		+	+		+	+		+	
4	8	3		+	+	+	+	+	+	+	+	+			
4	8	3	+	+	+	+	+	+	+	+	+	+	+	+	
4	8	3	+	+	+	+	+	+	+	+	+	+	+	+	+
4	8	4						+		+		+		+	
Totals		153	43	40	111	123	123	123	118	104	120	82	74	66	26

Appendix G

Stem-and-Leaf Distributions of District Reports of the Percentage of Students Scoring Above the National Median in Reading and Mathematics

Appendix G

Figure G-1

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 1

Reading

Stem	Leaf	Count
9 :	6	1
9 :	01	2
8 :	9	1
8 :	013	3
7 :	588	3
7 :	34	2
6 :	55689	5
6 :	012224	6
5 :	5567789	7
5 :	001224444	9
4 :	5579	4
4 :	0134	4
3 :	56689	5
3 :	0023	4
2 :	6	1
2 :		0
1 :		0
1 :		0

P90 = 81
P75 = 66
P50 = 55
P25 = 45
P10 = 35

Mathematics

Stem	Leaf	Count
9 :	589	3
9 :	3	1
8 :	9	1
8 :	034	3
7 :	55678	5
7 :	0113	4
6 :	6899999	7
6 :	001223344444	12
5 :	5899	4
5 :	012	3
4 :	669	3
4 :	34	2
3 :	88	2
3 :	02	2
2 :	89	2
2 :	2	1
1 :		0
1 :		0

P90 = 84
P75 = 71
P50 = 64
P25 = 51
P10 = 38

Appendix G

Figure G-2

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 2

Reading

Stem	Leaf	Count
9 :		0
9 :	12	2
8 :	577	3
8 :	0012	4
7 :	5799	4
7 :	12	2
6 :	555688899	9
6 :	0012344	7
5 :	56677788999	11
5 :	0122334444	10
4 :	557778899	9
4 :	111123344	9
3 :	999	3
3 :	1	1
2 :	99	2
2 :	2	1
1 :		0
1 :		0

P90 = 80
P75 = 68
P50 = 57
P25 = 47
P10 = 41

Mathematics

Stem	Leaf	Count
9 :	559	3
9 :	013	3
8 :	67	2
8 :	001334	6
7 :	5779	4
7 :	0001112222344	13
6 :	55566788889	11
6 :	000011222	9
5 :	56677889	8
5 :	001124	6
4 :	568	3
4 :	23	2
3 :	6	1
3 :	4	1
2 :		0
2 :		0
1 :	68	2
1 :		0

P90 = 86
P75 = 74
P50 = 67
P25 = 57
P10 = 46

Appendix G

Figure G-3

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 3

Reading

Stem	Leaf	Count
9 :		0
9 :	34	2
8 :	558	3
8 :	12	2
7 :	56799	5
7 :	0122344	7
6 :	677777789	9
6 :	00111224444444	14
5 :	5566677899	10
5 :	001233344	9
4 :	556889	6
4 :	001223	6
3 :	69	2
3 :	012223344	9
2 :	89	2
2 :	14	2
1 :	5	1
1 :		0

P90 = 78
P75 = 67
P50 = 58
P25 = 45
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :	123	3
8 :	7899	4
8 :	012224	6
7 :	88	2
7 :	000112244	9
6 :	556778888899	12
6 :	000123344444	12
5 :	55567788999	11
5 :	1222333444	10
4 :	556667899	9
4 :	0224	4
3 :	69	2
3 :	334	3
2 :		0
2 :	0	1
1 :		0
1 :	1	1

P90 = 82
P75 = 70
P50 = 61
P25 = 52
P10 = 42

Appendix G

Figure G-4

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 4

Reading

Stem	Leaf	Count
9 :	5	1
9 :	00	2
8 :	79	2
8 :	001	3
7 :	67799	5
7 :	00133444	8
6 :	6888	4
6 :	000022234	9
5 :	5567788899	10
5 :	01112222244	11
4 :	66777899	8
4 :	013444	6
3 :	5568889	7
3 :	12234444	8
2 :	7	1
2 :	1	1
1 :		0
1 :	1	1

P90 = 79
P75 = 68
P50 = 55
P25 = 44
P10 = 34

Mathematics

Stem	Leaf	Count
9 :	9	1
9 :	034	3
8 :	69	2
8 :	0033	4
7 :	589	3
7 :	024	3
6 :	5557777888889	13
6 :	0000012223344	13
5 :	55556667778	11
5 :	0011222222333344	16
4 :	55789	5
4 :	011224	6
3 :	5579	4
3 :		0
2 :		0
2 :		0
1 :		0
1 :	2	1

P90 = 81
P75 = 68
P50 = 59
P25 = 52
P10 = 42

Appendix G

Figure G-5

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 5

Reading

Stem	Leaf	Count
9 :		0
9 :	03	2
8 :	5	1
8 :	00112333	8
7 :	55578	5
7 :	0011223344	10
6 :	5699	4
6 :	00112224	8
5 :	666667788	9
5 :	0001122233	10
4 :	567888999	9
4 :	11244	5
3 :	55567799	8
3 :	02334	5
2 :	679	3
2 :		0
1 :	9	1
1 :		0

P90 = 80
P75 = 72
P50 = 56
P25 = 45
P10 = 34

Mathematics

Stem	Leaf	Count
9 :	6	1
9 :	0013	4
8 :	6	1
8 :	002234	6
7 :	55777899	8
7 :	02244	5
6 :	66677778888899	14
6 :	111122344444	12
5 :	556677899	9
5 :	002222244	9
4 :	5667888899	10
4 :	1344	4
3 :	57	2
3 :	2	1
2 :		0
2 :	2	1
1 :		0
1 :		0

P90 = 82
P75 = 73
P50 = 64
P25 = 52
P10 = 45

Appendix G

Figure G-6

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 6

Reading

Stem	Leaf	Count
9 :		0
9 :	2	1
8 :	69	2
8 :	0234	4
7 :	55556	5
7 :	0001234	7
6 :	5555589	7
6 :	0144	4
5 :	66677777889	11
5 :	001223334	9
4 :	555678999	9
4 :	0122234	7
3 :	56666677889	11
3 :	00024	5
2 :	69	2
2 :		0
1 :		0
1 :	2	1

P90 = 75
P75 = 65
P50 = 54
P25 = 42
P10 = 35

Mathematics

Stem	Leaf	Count
9 :	79	2
9 :	4	1
8 :	556	3
8 :	1444	4
7 :	556778	6
7 :	123	3
6 :	5566888999	10
6 :	022222222334444	15
5 :	55556667788999	14
5 :	0011123	7
4 :	5556677889	10
4 :	22244	5
3 :	89	2
3 :		0
2 :		0
2 :	3	1
1 :		0
1 :	2	1

P90 = 84
P75 = 69
P50 = 62
P25 = 50
P10 = 44

Appendix G

Figure G-7

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 7

Reading

Stem	Leaf	Count
9 :		0
9 :	0	1
8 :	7	1
8 :	13	2
7 :	55699	5
7 :	0004	4
6 :	57789	5
6 :	001112333	9
5 :	566778	6
5 :	0011223344	10
4 :	577799	6
4 :	0334	4
3 :	7778999	7
3 :	0024	4
2 :	68899	5
2 :		0
1 :		0
1 :	0	1

P90 = 75
P75 = 64
P50 = 54
P25 = 40
P10 = 30

Mathematics

Stem	Leaf	Count
9 :		0
9 :	0333	4
8 :	6	1
8 :	00034	5
7 :	8	1
7 :	003	3
6 :	6677777789	10
6 :	0011123334	10
5 :	5667777899	10
5 :	23344	5
4 :	56778889	8
4 :	00022234	8
3 :	66788	5
3 :		0
2 :	8	1
2 :		0
1 :	9	1
1 :		0

P90 = 80
P75 = 67
P50 = 59
P25 = 47
P10 = 39

Appendix G

Figure G-8

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 8

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :	56	2
8 :	233	3
7 :	67889	5
7 :	001233	6
6 :	555667889	9
6 :	0011234	7
5 :	55567777899	11
5 :	011123334	9
4 :	5667778	7
4 :	0011244	7
3 :	667789	6
3 :	11233344	8
2 :	899	3
2 :		0
1 :	9	1
1 :		0

P90 = 77
P75 = 66
P50 = 55
P25 = 41
P10 = 33

Mathematics

Stem	Leaf	Count
9 :		0
9 :	1	1
8 :	57	2
8 :	002234	6
7 :	5666788	7
7 :	023334	6
6 :	56679	5
6 :	1111222233344	13
5 :	677788999	9
5 :	12444444	8
4 :	5589999	7
4 :	0133444	7
3 :	55666789	8
3 :	0044	4
2 :		0
2 :		0
1 :		0
1 :	01	2

P90 = 79
P75 = 70
P50 = 59
P25 = 45
P10 = 36

Appendix G

Figure G-9

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 9

Reading

Stem	Leaf	Count
9 :		0
9 :	2	1
8 :		0
8 :	3	1
7 :	779	3
7 :	2	1
6 :	6889	4
6 :	1113	4
5 :	566777789	9
5 :	00111113	8
4 :	566899	6
4 :	001112344	9
3 :	55668	5
3 :	22344	5
2 :	8	1
2 :	014	3
1 :	6	1
1 :		0

P90 = 69
P75 = 58
P50 = 50
P25 = 39
P10 = 32

Mathematics

Stem	Leaf	Count
9 :		0
9 :		0
8 :	6699	4
8 :		0
7 :	559	3
7 :	1233	4
6 :	5777	4
6 :	00012234	8
5 :	589	3
5 :	00011344	8
4 :	5568999	7
4 :	12344	5
3 :	669	3
3 :	0034	4
2 :	79	2
2 :	01	2
1 :		0
1 :		0

P90 = 75
P75 = 65
P50 = 53
P25 = 44
P10 = 30

Appendix G

Figure G-10

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 10

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :		0
8 :	4	1
7 :	5	1
7 :	00334	5
6 :	568	3
6 :	00123	5
5 :	667	3
5 :	02344	5
4 :	55677889	8
4 :	0133444	7
3 :	7789	4
3 :	01344	5
2 :	578	3
2 :	0	1
1 :	5	1
1 :		0

P90 = 71
P75 = 61
P50 = 48
P25 = 38
P10 = 29

Mathematics

Stem	Leaf	Count
9 :		0
9 :	0	1
8 :	55	2
8 :	011	3
7 :	56	2
7 :	02	2
6 :	559	3
6 :	0114	4
5 :	556777789	9
5 :	134	3
4 :	689	3
4 :	1233334	7
3 :	5678888	7
3 :	04	2
2 :		0
2 :		0
1 :		0
1 :	0	1

P90 = 80
P75 = 65
P50 = 55
P25 = 43
P10 = 36

Appendix G

Figure G-11

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 11

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :	6	1
8 :	0	1
7 :	579	3
7 :	0144	4
6 :	5	1
6 :	011223	6
5 :	678	3
5 :	001123344	9
4 :	567	3
4 :	113	3
3 :	55889	5
3 :	123	3
2 :	7779	4
2 :	1	1
1 :	9	1
1 :	0	1

P90 = 75
P75 = 62
P50 = 52
P25 = 38
P10 = 27

Mathematics

Stem	Leaf	Count
9 :	6	1
9 :		0
8 :		0
8 :	023	3
7 :	599	3
7 :	22	2
6 :	67899	5
6 :	01233334	8
5 :	66889	5
5 :	00	2
4 :	578	3
4 :	244	3
3 :	5558999	7
3 :	114	3
2 :		0
2 :		0
1 :		0
1 :	0	1

P90 = 79
P75 = 68
P50 = 59
P25 = 42
P10 = 35

Appendix G

Figure G-12

Stem-and-Leaf Distribution of the District Percents of Students Scoring Above the National Median at Grade 12

Reading

Stem	Leaf	Count
9 :		0
9 :		0
8 :		0
8 :		0
7 :	79	2
7 :	24	2
6 :		0
6 :	2	1
5 :	888	3
5 :	011	3
4 :	88	2
4 :	011	3
3 :	6	1
3 :	3	1
2 :	7	1
2 :	1	1
1 :		0
1 :	3	1

P90 = 75
P75 = 58
P50 = 50
P25 = 40
P10 = 21

Mathematics

Stem	Leaf	Count
9 :	5	1
9 :		0
8 :		0
8 :		0
7 :		0
7 :	02	2
6 :	789	3
6 :	0	1
5 :	77	2
5 :	4	1
4 :	5589	4
4 :	14	2
3 :	6	1
3 :	4	1
2 :		0
2 :		0
1 :		0
1 :	0	1

P90 = 71
P75 = 67
P50 = 55
P25 = 45
P10 = 35