

TRENDS IN EDUCATIONAL ACHIEVEMENT

The Congress of the United States
Congressional Budget Office





NOTES

Except where otherwise noted, dates used in this paper are school years rather than calendar years. For example, the results of a test administered in the fall of 1979 and the spring of 1980 are both labeled 1979. As a result, the dates used here are in some instances a year earlier than those in other published sources. This discrepancy is particularly common in the case of college admissions tests and other tests administered to high school seniors, which are often labeled in other sources in terms of the calendar year in which students would graduate.

Details in the text and tables of this report may not add to totals because of rounding.

PREFACE

At the request of the Subcommittee on Education, Arts, and Humanities of the Senate Committee on Labor and Human Resources, the Congressional Budget Office (CBO) prepared this assessment of trends in the educational achievement of elementary- and secondary-school students. This volume presents the analysis of the trends themselves; a forthcoming companion volume, *Educational Achievement: Explanations and Implications of Recent Trends*, evaluates many common explanations of the trends and discusses their implications for education policy. In accordance with CBO's mandate to provide objective and impartial analysis, neither volume contains recommendations.

Daniel Koretz of CBO's Human Resources and Community Development Division prepared the analysis under the direction of Nancy M. Gordon and Martin D. Levine. Paul L. Houts edited the report. Ronald Moore typed the many drafts and prepared the manuscript for publication.

The author thanks the following organizations for providing essential data, much of which is unpublished: the National Assessment of Educational Progress, the American College Testing Program, the Iowa Testing Programs, the College Board, CTB-McGraw Hill, Science Research Associates, the state departments of education in New York, Virginia, Texas, North Carolina, Nevada, California, and Illinois, and the school districts of Cleveland, Montgomery County (Maryland), and Houston.

Many individuals contributed in various ways to this work. Particular thanks are due H. D. Hoover of the Iowa Testing Program and Lawrence Rudner of the Office of Educational Research and Improvement, U.S. Department of Education, who provided insightful contributions at many stages of the project. Other individuals whose assistance is gratefully acknowledged include Robert Cameron and Harlan Hanson of the College Board; Douglas Coulson of the Office of Technology Assessment, U.S. Congress; Robert Forsyth of the Iowa Testing Programs; Gene Guest of CTB/McGraw-Hill; Eric Hanushek of the University of Rochester; Lyle Jones of the University of North Carolina; Jackie Woods of the American College Testing Program; and Edward M. Gramlich, Jack Rodgers, and Robertson Williams of CBO. Kenneth Rubin of CBO also provided particularly valuable comments.

1

1111

1111

CONTENTS

	SUMMARY	xiii
CHAPTER I.	INTRODUCTION	1
	The Context of the Current Controversy	2
CHAPTER II.	UNDERSTANDING MEASURES OF EDUCATIONAL ACHIEVEMENT	11
	Educational Tests Versus Educational Achievement	12
	Important Characteristics of Educational Tests	14
	Using Tests to Gauge Trends or Compare Jurisdictions	24
CHAPTER III.	AGGREGATE TRENDS IN EDUCATIONAL ACHIEVEMENT	31
	The Decline in Achievement	31
	The Recent Upturn in Achievement	46
	Differences in Trends Among Tests	48
CHAPTER IV.	GROUP DIFFERENCES IN ACHIEVEMENT TRENDS	61
	Differences in Trends Among Types of Students	61
	Race and Ethnicity	74
	Differences in Trends Among Types of Schools and Communities	79

APPENDIX A. DESCRIPTION OF MAJOR DATA SOURCES	87
The Scholastic Aptitude Test (SAT)	87
The American College Testing Program (ACT) Tests	90
The National Assessment of Educational Progress (NAEP)	91
The National Longitudinal Survey and High School and Beyond	93
Iowa Testing Programs	95
 APPENDIX B. EVIDENCE OF A COHORT EFFECT IN THE RECENT UPTURN IN ACHIEVEMENT	 99
Types of Data That Can be Used to Assess Cohort and Period Effects	100
The Fit of the Data with the Cohort and Period Models	101
An Aggregate Test of the Cohort and Period Models	118
 APPENDIX C. DIFFERENCES IN TRENDS BY SUBJECT AREA	 123
 APPENDIX D. VARIATION AMONG ACHIEVEMENT SUBGROUPS	 129
Trends in the Lowest and Highest Quartiles	131
Trends Among College-Bound Students	135
Select Students	138
Conclusion	148
 APPENDIX E. DIFFERENCES IN ACHIEVEMENT TRENDS AMONG BLACK, HISPANIC, AND NONMINORITY STUDENTS	 150
Black Students	149
Hispanic Students	159

TABLE III-1	ONSET AND END OF THE ACHIEVEMENT DECLINE, SELECTED TESTS	36
TABLE III-2	SIZE OF THE ACHIEVEMENT DECLINE, INDICATED BY SELECTED TESTS AT GRADE 6 AND ABOVE	40
TABLE III-3	SUMMARY OF NATIONAL ASSESSMENT RESULTS IN THREE SUBJECTS, AGES 9, 13, AND 17	44
TABLE III-4	RECENT TRENDS ON STANDARDIZED TESTS AMONG HIGH SCHOOL SENIORS AND JUNIORS, WITH TRENDS OVER THE SAME PERIODS ON THE SAT	50
TABLE III-5	NAEP MATHEMATICS CHANGES 1972-1977, AGE 17, BY AREA	56
TABLE III-6	NAEP READING CHANGES 1970-1979, AGE 17, BY AREA	58
TABLE IV-1	AVERAGE MATHEMATICS ACHIEVEMENT IN DISADVANTAGED URBAN COMMUNITIES AND IN THE NATION, NAEP, 1972-1981	80
TABLE IV-2	AVERAGE READING ACHIEVEMENT IN DISADVANTAGED URBAN COMMUNITIES AND IN THE NATION, NAEP, 1970-1983	81
TABLE IV-3	AVERAGE MATHEMATICS ACHIEVEMENT IN HIGH-MINORITY AND LOW-MINORITY SCHOOLS, NAEP, 1977 AND 1981	83
TABLE B-1	TIMING OF THE END OF THE ACHIEVEMENT DECLINE, BY TEST	120

TABLE C-1	MAGNITUDE OF THE ACHIEVEMENT DECLINE, BY SUBJECT	125
TABLE D-1	RECENT TRENDS IN THE NATIONAL ASSESSMENT, BY ACHIEVEMENT SUBGROUP AND ETHNICITY	132
TABLE D-2	CHANGE ON THE ILLINOIS DECADE TEST AMONG STUDENTS AT THE 25th AND 75th PERCENTILE	134
TABLE E-1	AVERAGE ACHIEVEMENT OF BLACK AND OTHER STUDENTS IN THE NLS AND HSB, BY SUBJECT	154
TABLE E-2	READING PERFORMANCE OF BLACK AND NONMINORITY STUDENTS IN THE NATIONAL ASSESSMENT	155
TABLE E-3	MATHEMATICS PERFORMANCE OF BLACK AND NONMINORITY STUDENTS IN THE NATIONAL ASSESSMENT	156
TABLE E-4	AVERAGE ACHIEVEMENT OF HISPANIC AND OTHER STUDENTS IN THE NLS AND HSB, BY SUBJECT	162
TABLE E-5	MATHEMATICS PERFORMANCE OF NONMINORITY AND HISPANIC STUDENTS IN THE NATIONAL ASSESSMENTS	164
TABLE E-6	READING PERFORMANCE OF NONMINORITY AND HISPANIC STUDENTS IN THE NATIONAL ASSESSMENTS	165

SUMMARY		
FIGURE 1	IOWA AVERAGE TEST SCORES, GRADES 5, 8, AND 12, DIFFERENCES FROM POST-1964 LOW POINT	xvii
SUMMARY		
FIGURE 2	SAT-MATHEMATICS SCORES BY ETHNICITY: BLACK AND NONMINORITY STUDENTS	xviii
FIGURE I-1	SHARES OF ELEMENTARY/SECONDARY EDUCATION FUNDING BY LEVEL OF GOVERNMENT	4
FIGURE II-1	HYPOTHETICAL TEST RESULTS EXPRESSED IN STANDARD DEVIATIONS (SDs), BASED ON THE SAT-MATHEMATICS (SAT-M)	17
FIGURE III-1	HYPOTHETICAL PERIOD, COHORT, AND AGE EFFECTS, AVERAGE SCORES	34
FIGURE III-2	ITBS COMPOSITE SCORES, IOWA ONLY, BY TEST YEAR AND GRADE AT TESTING	37
FIGURE III-3	ITED COMPOSITE SCORES, IOWA ONLY, BY TEST YEAR AND GRADE AT TESTING	38
FIGURE III-4	AVERAGE SAT SCORES, BY SUBJECT, DIFFERENCES FROM LOWEST YEAR	38
FIGURE III-5	IOWA MATHEMATICS ACHIEVEMENT, DIFFERENCES FROM LOWEST YEAR	54
FIGURE IV-1	IOWA COMPOSITE, ITBS, GRADES 3-8, DIFFERENCES FROM POST-1964 LOW POINT	63
FIGURE IV-2	IOWA COMPOSITE, ITED, GRADES 9-12, DIFFERENCES FROM LOWEST YEAR	64
FIGURE IV-3	ITBS COMPOSITE, BY BIRTH YEAR	67



FIGURE IV-4	ITED COMPOSITE, BY BIRTH YEAR	68
FIGURE IV-5	TRENDS IN AVERAGE READING PROFICIENCY FOR WHITE, BLACK, AND HISPANIC STUDENTS, BY BIRTH YEAR	77
FIGURE B-1	ITBS COMPOSITE SCORES, IOWA ONLY	103
FIGURE B-2	ITED COMPOSITE SCORES, IOWA ONLY	104
FIGURE B-3	AVERAGE SAT SCORES	106
FIGURE B-4	ACT SCORES	107
FIGURE B-5	PERCENT OF NEW YORK STUDENTS SCORING ABOVE REFERENCE POINT	109
FIGURE B-6	CALIFORNIA STATE ASSESSMENT TEST SCORES	109
FIGURE B-7	VIRGINIA COMPOSITE ACHIEVEMENT	111
FIGURE B-8	NAEP MATHEMATICS SCORES	112
FIGURE B-9	NAEP SCIENCE SCORES	114
FIGURE B-10	NAEP READING PROFICIENCY SCORES	114
FIGURE B-11	NAEP READING (INFERENTIAL COMPREHENSION) SCORES	115
FIGURE B-12	ITBS NATIONAL NORMING DATA	117
FIGURE D-1	PERCENT OF SAT SCORES ABOVE 700	140
FIGURE D-2	SAT MATHEMATICS SCORES FOR SELECTED PERCENTILES	140
FIGURE D-3	SAT SCORES BY PERCENTILE OF CLASS RANK	142

FIGURE D-4	ETS/CEEB SCORES AND PERCENT OF GRADUATES TAKING TEST	145
FIGURE D-5	AVERAGE ADVANCED PLACEMENT SCORES AND PERCENT OF GRADUATES TAKING TESTS	147
FIGURE E-1	MINORITY/NONMINORITY DIFFERENCES ON THE SAT	151
FIGURE E-2	BLACK/NONBLACK DIFFERENCES ON THE ACT	153
FIGURE E-3	NORTH CAROLINA BLACK/WHITE DIFFERENCES ON CAT	159
FIGURE E-4	PERCENTAGES OF GRADE-NINE TEXAS STUDENTS PASSING MATHEMATICS AND READING TESTS, FOR THREE ETHNIC GROUPS	160

SUMMARY

Over the past several years, the educational achievement of American students has become a focus of intense public discussion and has led to a serious reexamination of schooling in America. A number of developments have contributed to this concern, including a substantial decline in test scores in the 1960s and 1970s, the weak performance of American students relative to their peers in some other countries, and the large gap in average test scores between some minority groups and nonminority students. More positive trends, though significant, have gained less notice--for example, the end of the overall achievement decline in the 1970s, a subsequent upturn in average scores, and recent gains of black and Hispanic students relative to nonminority students.

With the growing concern about public education has come an increasing reliance on achievement tests as indicators of the performance of students and schools. This trend has taken many forms and is apparent from the local to the national level. Many states and localities have expanded their programs of routine testing, sometimes as a result of legislation; the additional tests are often used as minimum criteria for promotion into higher grades or for graduation. Furthermore, average test scores have become a common basis of comparisons among schools and districts, and in some communities, newspapers routinely publish test results to facilitate such comparisons. The U.S. Department of Education has begun annual publication of average college admissions test scores on a state-by-state basis, and some states have taken steps to alter their own achievement tests to make their results comparable. Test scores have in fact come to be used as a national report card, influencing decisions from the level of individual students to that of national educational policy.

In the light of this heightened reliance on achievement tests, a careful appraisal of recent trends in test scores has important ramifications for educational policy. This paper assesses test score trends among elementary and secondary school students; it also discusses the strengths and weaknesses of the information they provide. A forthcoming companion study, *Educational Achievement: Explanations and Implications of Recent Trends*, evaluates common explanations of the trends and explores implications for educational policy.

THE POLICY CONTEXT OF CURRENT CONCERNS

Although states and localities bear primary responsibility for elementary and secondary education, educational achievement is clearly a national concern. Indeed, the current debate has been national in both scope and content. It has focused in part on such national issues as the competitiveness of the American economy and national security--questions that have been recurrent themes in debate about educational policy at least since the turn of the century. Moreover, the debate has taken hold in all regions of the country, and many of the initiatives undertaken by states and localities reflect common themes and share common elements, such as increased reliance on achievement testing. As in the past, both the Congress and the Administration have been important participants in the debate through legislative proposals and the dissemination of information.

UNDERSTANDING MEASURES OF EDUCATIONAL ACHIEVEMENT

Although the use of standardized tests as indicators of educational achievement has grown sharply in recent years, scores on standardized tests are not as straightforward an indicator of achievement as they might first appear. For that reason, the strengths and weaknesses of existing tests should be kept in mind when interpreting recent trends.

The advantages of standardized tests--or, rather, the advantages that they can have if carefully constructed--are obvious and important. By imposing a uniform measure, they can avoid much of the subjectivity and extraneous variation that plagues some alternative forms of evaluation, such as grade-point averages. Standardized tests can be designed to provide valuable comparisons over time and among grade levels, tap specific types of skills, and differentiate among students at various achievement levels.

The weaknesses of standardized tests are less apparent but equally significant. In most cases, the tests are not direct and complete measures of the skills that are of concern. Rather, they are proxies for this often unobtainable ideal. Designing the proxy entails many decisions about the test's purpose, content, level of difficulty, format, the severity of time pressure, and other factors. As a result, tests vary markedly in what they measure and how well they measure it. Indeed, even apparently similar tests often produce divergent results.

Tests designed to assist in selecting students for admission to college--such as the Scholastic Aptitude Test (SAT)--provide a particularly striking example of tests as proxies for other, unobtainable measures. These

tests are intended to predict students' performance in college, which can be measured directly only long after the admissions decision must be made. Although these tests comprise multiple-choice questions, their purpose is to predict future success on some very different tasks--such as comprehending long lectures and writing fluent term papers--that help determine whether students succeed or fail in college. In the case of tests designed to measure students' current level of achievement, the contrast between the skills embodied in a test and the corresponding skills with which schools are concerned is often less striking, but it can nonetheless be substantial.

Because of these limitations, the results of standardized tests must be interpreted cautiously. Trends should be given credence if they appear with considerable consistency in numerous tests, particularly if the tests are varied. On the other hand, trends that appear only on one test, or only among a set of very similar tests, should be considered questionable. Moreover, whether trends shown by a test are meaningful hinges on whether the characteristics of that test are appropriate for the particular issue in question. For example, if trends among students in general are at issue, college admissions tests can provide dubious information. A large number of students never take such tests, which makes the results unrepresentative of the student population as a whole. Furthermore, biases are introduced by changes in the composition of the group that does take the tests. Similarly, some minimum-competency tests provide little information about trends among high-achieving students for want of a sufficient number of difficult test items.

THE DECLINE AND SUBSEQUENT UPTURN IN ACHIEVEMENT TEST SCORES

After years of improvement, scores on achievement test scores began a sizable drop in the mid-1960s. The decline was widespread, occurring among many different types of students, on many different tests, in all subject areas, in private as well as public schools, and in all parts of the nation. 1/

Although the size of the decline varied greatly from one test to another, it was in many instances large enough to be of substantial educational concern. In general, the decline in test scores was larger in the

-
1. A few tests did not conform to this pattern. The National Assessment of Educational Progress (NAEP), for example, showed no overall drop in reading since 1970, and the American College Testing program (ACT) tests showed no decline in natural science. But these exceptions were few enough, and the conforming tests sufficiently numerous, that the generality of the decline is clear.

higher grades. Scores on tests administered in grades three and below dropped little, if at all, and tests administered in grade four showed only inconsistent and small declines. On the other hand, most tests administered in grades five and above showed declines in average scores, with the largest drops tending to occur at the high school level. Among the achievement tests assessed in this study, 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.^{2/} A different assortment of tests, however, would yield a different estimate of the decline's average magnitude.

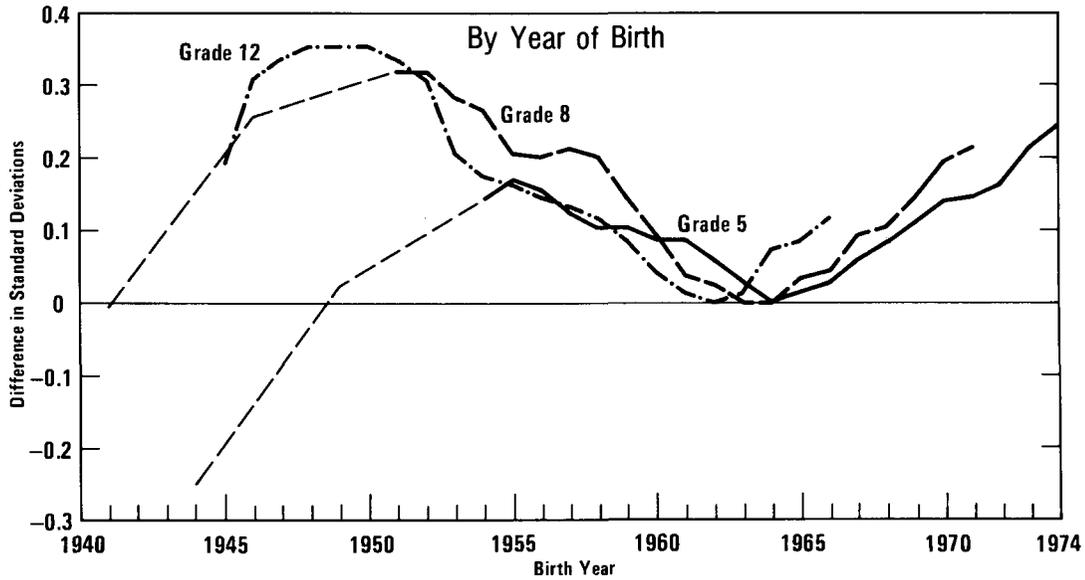
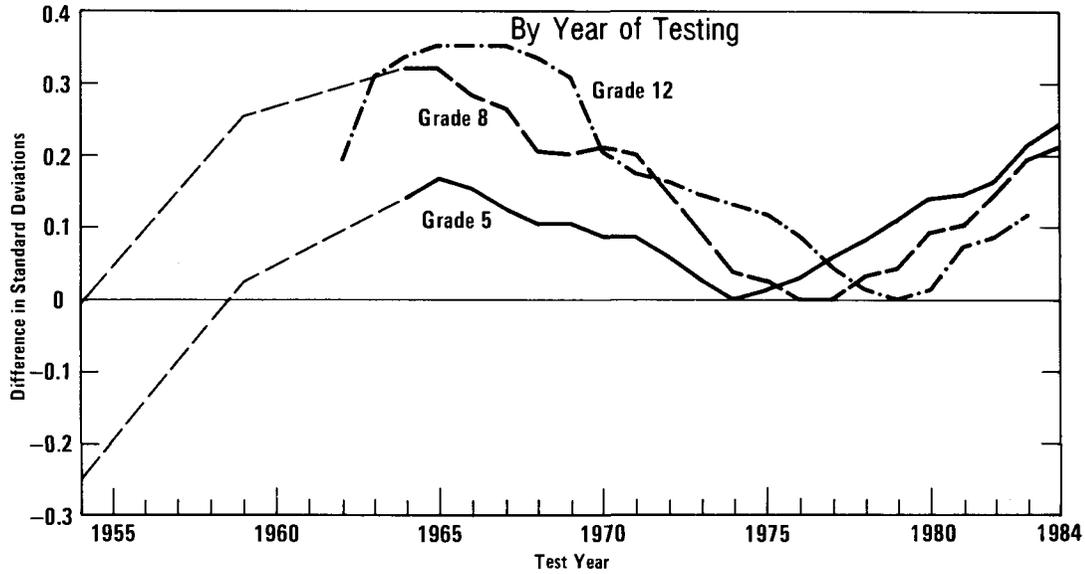
Although not all skills commonly considered "basic" escaped serious deterioration, the score decline appears to have been greater in areas involving higher-order skills. For example, between 1972 and 1977, the National Assessment of Educational Progress in mathematics showed no change in the performance of 17-year-olds in the simple recall of facts and definitions, but substantial declines took place on test items tapping deeper understanding and problem-solving skills. Items testing arithmetic computation showed a mixed pattern; in general, the more complex items evidenced the sharpest drops in success rates. This larger drop in higher-level skills might be one cause of the greater test score decline in the higher test grades.

The overall decline in test scores generally ended with the cohorts of children born around 1962 and 1963--that is, with the cohorts that entered school in the late 1960s. Thus, the decline's end first appeared in tests administered in the upper elementary grades in the mid-1970s. Thereafter, it moved into the higher grades at a rate of roughly a grade per year as those birth cohorts aged, reaching the senior high school grades in the late 1970s (see Summary Figure 1). This pattern, however, has gained relatively little attention. Perhaps because of the greater notice accorded to tests at the senior high school level, there has been a widespread misconception that the decline ended only within the past few years.

In fact, subsequent cohorts of children--those entering school in the late 1960s and throughout the 1970s--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. As a result, achievement in the elementary grades is now by some measures at its highest level in three decades. At the other extreme, scores on tests administered to high school students, such as the Scholastic Aptitude Test

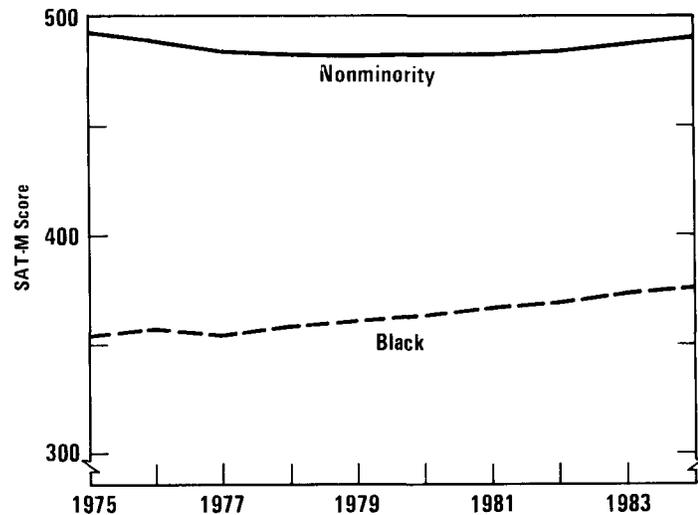
2. The average decline on these tests was roughly 0.3 standard deviation.

Summary Figure 1.
Iowa Average Test Scores, Grades 5, 8, and 12,
Differences from Post-1964 Low Point



SOURCES: CBO calculations based on "Iowa Basic Skills Testing Programs, Achievement Trends in Iowa: 1955-1985" (Iowa Testing Programs, unpublished and undated material); A. N. Hieronymus, E. F. Lindquist, and H. D. Hoover, *Iowa Tests of Basic Skills: Manual For School Administrators* (Chicago: Riverside, 1982); "Mean ITED Test Scores by Grade and Subtest for the State of Iowa, 1962 to Present" (Iowa Testing Programs, unpublished and undated tabulations); and Robert Forsyth, Iowa Testing Programs, personal communication, August 1984.

Summary Figure 2.
 SAT-Mathematics
 Scores by Ethnicity:
 Black and
 Nonminority Students



SOURCE: The College Entrance Examination Board, "College Board Data Show Class of '85 Doing Better on SAT, Other Measures of Educational Attainment" (press release, The College Board, September 1985).

(SAT), still remain relatively close to their low points of the late 1970s, probably because of the shorter interval since scores began to rise again in those age groups. While it appears that these improvements are occurring at many skill levels, the data raise disturbing questions of whether the improvements on some tests have been larger in the more basic skills than in areas requiring deeper understanding.

Another important issue in the debate over educational achievement is the performance of minority students on standardized tests. Over the past 10 to 15 years--a period that encompassed both declining and improving test scores--the average scores of some minority students rose compared with those of nonminority students. The relative gains of black students appear on every test for which separate trend data for black students are available. Although the gap in average scores between black and nonminority students remains large, it has narrowed appreciably (see Summary Figure 2).^{3/} Some

3. On the SAT, for example, the rate at which the scores of black and nonminority scores have converged over the past nine years is comparable to that of the total decline in scores among all students taking the test--a trend that few observers have labeled insignificant.

test results suggest that the scores of black students showed lesser decreases than did those of nonminority students during the final years of the achievement decline, stopped declining earlier, and showed greater improvement during the first years of the overall upturn in scores.

In addition, Hispanic students, who also typically have average scores well below those of nonminority students, showed relative gains over the past decade. The improvement appears to have been greater among Mexican-American students than among other Hispanics. These patterns are less clear-cut, however, because of more limited data, ambiguities in the classification of diverse Hispanic students, and the relatively small number of Hispanics in the test data.

The period since 1970 also included relative improvement of average test scores in certain characteristically low-achieving types of schools and communities. Between 1977 and 1981, mathematics scores on the National Assessment of Educational Progress rose much more sharply in high-minority schools (those with minority enrollments of 40 percent or more) than in other schools. This upturn suggests that the gains of minority students cannot be attributed entirely to those attending schools with low concentrations of minority students. Students in disadvantaged urban schools also showed relative gains in the National Assessments of mathematics and reading. In mathematics, for example, average scores of 9- and 13-year-old students in disadvantaged urban communities rose markedly after 1972, while those of students in other localities rose little or not at all. These relative gains were sizable; by 1981, a fourth to a third of the gap in test scores between disadvantaged urban communities and the rest of the nation had been overcome.