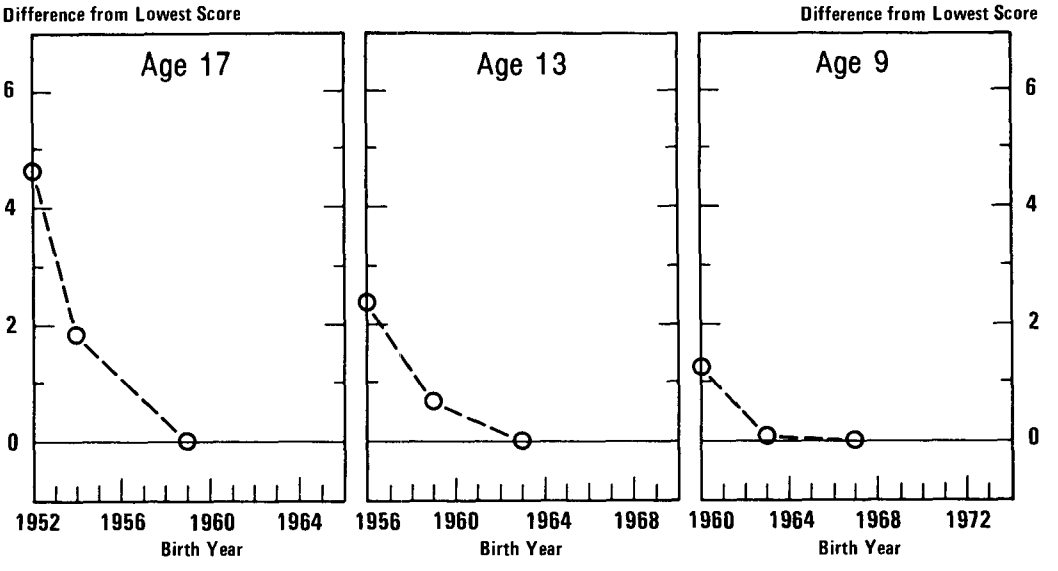
Figure B-9.

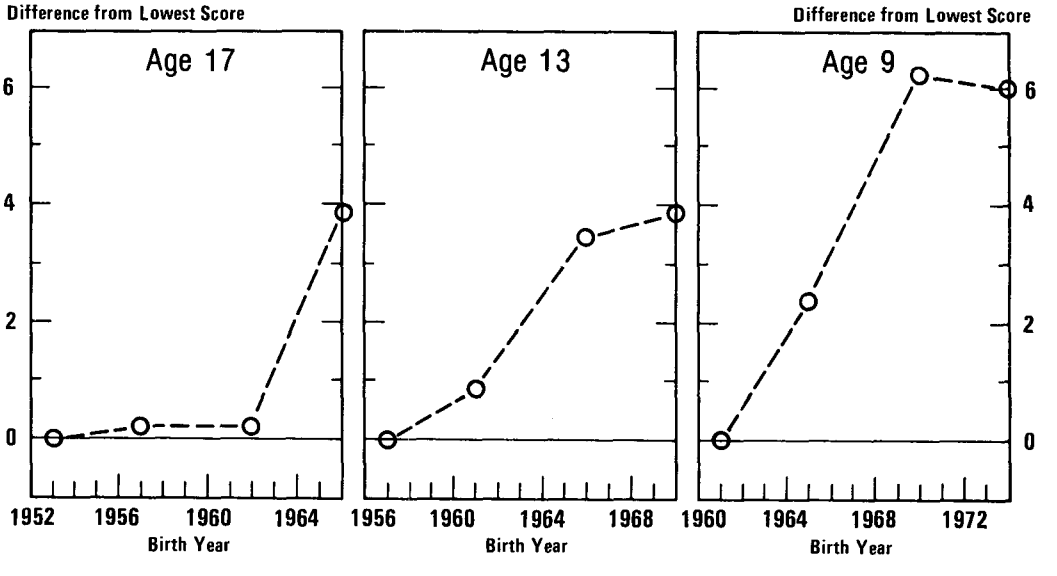
NAEP Science Scores (By birth year and age)



SOURCE: CBO calculations based on National Assessment of Educational Progress, *Three National Assessments of Science: Changes in Achievement, 1969-77* (Denver: NAEP/Education Commission of the States, 1978).

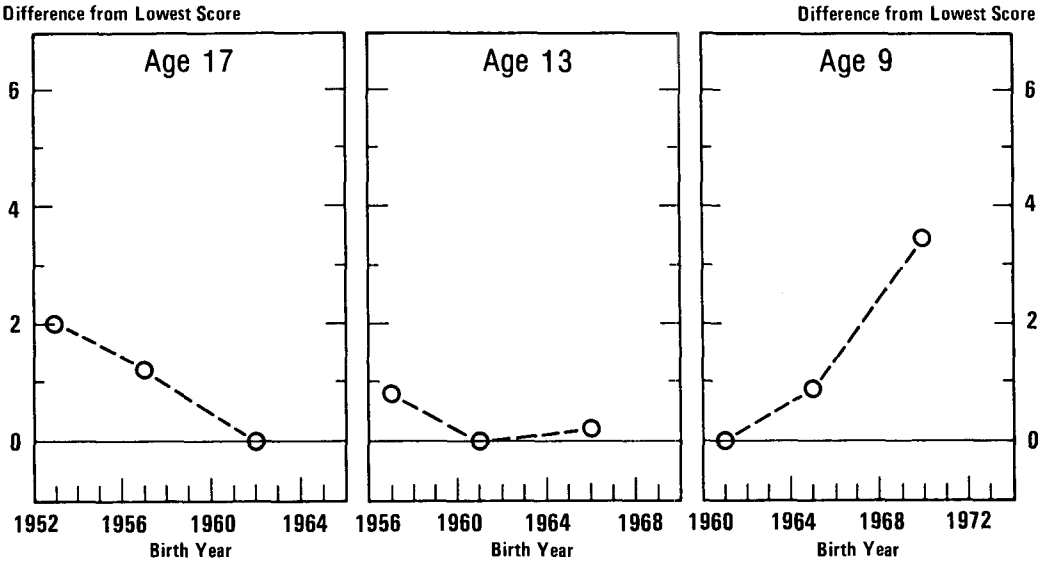
Figure B-10.

NAEP Reading Proficiency Scores (By birth year and age)



SOURCE: CBO calculations based on National Assessment of Educational Progress, *The Reading Report Card: Progress Toward Excellence in Our Schools* (Princeton: NAEP/Educational Testing Service, 1985).

Figure B-11.
NAEP Reading (Inferential Comprehension) Scores (By birth year and age)



SOURCE: CBO calculations based on National Assessment of Educational Progress, *Three National Assessments of Reading: Changes in Performance, 1970-80* (Denver: NAEP/Education Commission of the States, 1981).

Norming data offer even weaker evidence than does the NAEP for testing the cohort and period models. Like the NAEP and all other intermittent data, norming data cannot precisely pinpoint the timing of a turnaround in achievement trends. In addition, norming data usually have even longer gaps between test years than those in the National Assessment (most often, seven years). These long gaps further exacerbate the uncertainty. Norming data generally also entail testing all grade levels in all subjects at the same time. In conjunction with the long period between renorming, this factor can force the trend data to appear to be a period effect even if the true underlying pattern is a cohort effect. 9/

9. The extent of this bias depends on the time span between normings, the range of grades tested, the number of years between a given norming and the true minimum in the trend data, and the slope of the curves on both sides of the minimum. For example, suppose that grades four through six are tested in 1972 and 1979 and that the true trend is a cohort model, with grade four reaching its low point in 1972, grade five in 1973, and so on. If the declines and upturns in each grade are reasonably similar in severity, all three grades will show their lowest scores in the 1972 norming sample. If, however, the testing continues through grade 12, the older grades--beginning with grade eight or nine--would probably show their lowest scores in the 1979 norming sample.

Taken together, the ITBS norming data can be seen as consistent with either a period or a cohort model. But they do suggest--albeit weakly--that if the cohort model is correct, the low point might be a few cohorts later than in the Iowa, SAT, and New York data. In all grades from fourth through eighth, the scores of students in the norming sample reached their lowest observed levels with the norming of 1977-1978, corresponding to the birth cohorts of 1964 through 1968 (see Figure B-12). 10/ If the decline reached its end with the birth cohort of 1963, for example, one might expect fourth-and fifth-grade scores to be lowest in the prior (1970-71) norming.

The California Test of Basic Skills (CTBS) Norming Data

An equating study of the most recent (1973 and 1980) normings of the CTBS provides a somewhat stronger test of the two models, for the large span of grades tested (first through twelfth) in part compensates for the long interval between the two test dates.

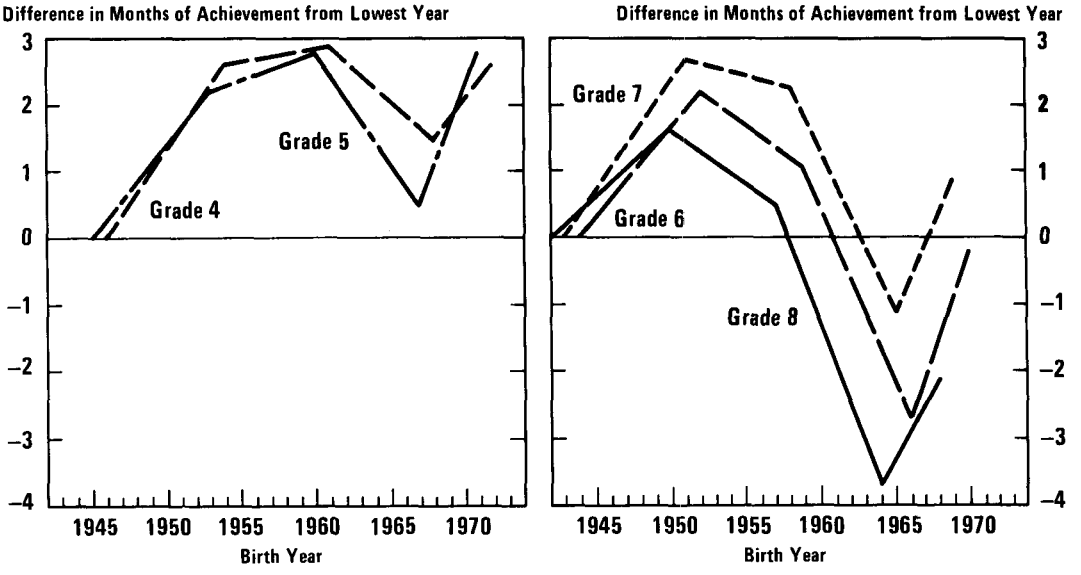
If the cohort model and the timing suggested by the Iowa, SAT, and New York data are correct, the CTBS data should show increases that are sizable in the elementary grades, gradually decrease in size in the junior-high grades, and are replaced by declines in the senior-high grades. In grade five and below, both norming samples comprise cohorts born in 1963 or later--that is, cohorts that produced increasing scores in the other data bases. In grades 6 through 11, the norming samples comprise varying mixes of post-1963 and pre-1963 birth cohorts, and the increases among the former should tend to offset the declines among the latter. Finally, both grade-12 samples were born in 1963 or earlier, so if the decline ended in 1963, the change at that grade level would reflect only years of declining achievement.

The changes in the CTBS norming samples largely conform to these predictions from the cohort model. With one exception, all comparisons at grade nine and below showed increases from 1973 to 1980, with a tendency for the largest gains to be in the lowest grades. For example, in the fall testing, the achievement of a third-grade student scoring at the 34th percentile in 1980 corresponded roughly to that of the median student in 1973, while in grade eight, a student would have had to reach the 46th percentile to score at the level of the median student of seven years earlier.

---

10. In grade three, average scores increased with every norming sample after the initial (1955) one--mirroring the negligible decline in third-grade scores in the ITBS-IA data --and are excluded from this discussion.

Figure B-12.
ITBS National Norming Data (By birth year and grade)



SOURCES: CBO calculations based on A. N. Hieronymus, E. F. Lindquist, and H. D. Hoover, *Iowa Test of Basic Skills: Manual For School Administrators* (Chicago: Riverside, 1982); and *The Development of the 1982 Norms for the Iowa Tests of Basic Skills* (Chicago; Riverside, 1983).

In contrast, students in the eleventh and twelfth grades showed a drop in achievement during that period. 11/

The California Achievement Tests (CAT) Norming Data

The 1970 and 1977 normings of the CAT were equated to each other and can be used in the same way as the 1973 and 1980 CTBS to test the cohort and period models. The two editions of the CAT, however, were more dissimilar from one another, making the procedure riskier.

Because the CAT was renormed three years earlier than the CTBS, one would expect the observed changes to switch from increases to decreases

11. California Test Bureau, McGraw Hill, unpublished tabulations. The most salient exception to this pattern occurred among ninth-grade students, who showed larger gains than any students above grade three.

three grades younger. Specifically, if the decline followed the cohort model and reached its low point with the 1963 birth cohort, one would expect grades nine and above to reflect only years of decline, while grades three through eight would reflect varying mixes of increasing and decreasing years. Only grades one and two would reflect solely increasing years, and those are grades in which the decline appears never to have occurred.

The results of the CAT renorming study largely conform to these predictions based on the cohort model. Grades one and two showed gains of over 0.6 standard deviation. These increases rapidly tapered off with increasing age, so that grades five and six showed essentially no change. Grades seven and eight showed declines of less than 0.2 standard deviation, while the higher grades all showed drops larger than 0.3 standard deviation. 12/

AN AGGREGATE TEST OF THE COHORT AND PERIOD MODELS

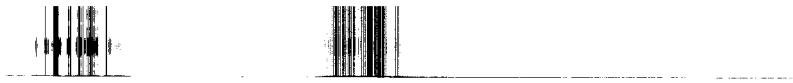
Another method of testing the cohort and period models is to assess which model yields the least variable estimates of the timing of the end of the decline, considering only those continuous data bases that show a clear low point. That is, the timing of the decline's end can be estimated in terms of both test years and birth cohorts, and the relative variation in those estimates indicates which of the models fits the data more closely. This approach, however, suffers from the relatively small number of data bases that can be applied.

Among the data bases meeting these criteria, the cohort model fits the data more closely than does the period model (Table B-1). The end of the decline, expressed in test years, showed a mean of 1976 and a 12-year range (from 1970 to 1982). When expressed in terms of birth cohorts, the decline's end showed a mean of 1962 and a range of only seven years (from 1958 to 1965). The standard deviation of the estimate is roughly 60 percent larger when test years are used. 13/

12. California Test Bureau, McGraw Hill, unpublished tabulations.

13. There are several ambiguities, noted in the text above, in specifying single years as the end of the decline in each data series, and these uncertainties apply to the patterns shown in Table B-1 as well. The most striking ambiguity entails the ACT mathematics assessment, which continued to decline, though slightly and inconsistently, for several years after the substantial decline ended. Table B-1 uses the year that the decline ended entirely. Substituting the year that the major decline in mathematics scores ended (1976), however, would not alter the conclusions. While it would make the relative fit of the cohort and period models more similar, the cohort model would still fit appreciably better.

The closer fit of the cohort model is much more striking if the ACT is excluded. The ACT is anomalous in two respects among continuous data bases showing an achievement decline--the early end of its decline and the lack of a subsequent upturn. Because these anomalies are unexplained, retesting the cohort model without the ACT seems warranted. When the ACT is excluded, the test years marking the end of the decline shows a nine-year range (from 1970 to 1979), while the birth cohorts show only a three-year range (from 1961 to 1964). Similarly, the difference between the test-year and cohort-year standard deviations is much larger--the former is nearly 3.5 times as large as the latter.



APPENDIX C

DIFFERENCES IN TRENDS BY SUBJECT AREA

As discussed in Chapter III, among all of the tests considered in this paper, no single subject area consistently showed the most severe decline in average scores. Nor was the decline consistently more substantial in either "directly" or "indirectly" taught subjects. This appendix provides the information on which those conclusions are based.

Not all of the data sources discussed in this paper could be used for making comparisons among subject areas. Only those tests that included more than one subject area and that could be converted to standard deviations (SDs) could be used, since only in those instances could the relative size of the decline among subject areas be ascertained. The most serious omission for this reason is the National Assessment of Educational Progress; the NAEP staff did not retain sufficient information on SDs to convert published raw scores.

This appendix includes data from tests administered both annually and less frequently, but comparisons among subject areas often have a somewhat different meaning in the two cases. When annual data are available, the beginning and end of the decline in each subject can be ascertained, and the tabulations in this appendix represent the total amount of each decline, regardless of its duration. In those instances, the largest decline need not be the most rapid. A subject showing a slower decline than others, for example, can drop more in total if its decline is sufficiently long in duration.

In the case of tests administered less often than annually, however, the beginning and end of the decline cannot be pinpointed. In those instances, the tabulations in this appendix represent the amount scores dropped during a fixed period for all subjects in one test battery--for example, the period between two normings, or between the National Longitudinal Study (1971) and the High School and Beyond study (1979).[1/] If the period used does not include years of rising scores, these comparisons indicate the relative rate of decline among subject areas, as well as the

1. The NLS and the HSB tests were administered in the springs of 1972 and 1980, respectively--that is, in the 1971 and 1979 school years.

amount of the decrease over that period. The comparisons need not, however, indicate the relative total decline among different subjects, since they cannot take into account differences in the duration of the decline. Moreover, because the time span used can encompass varying periods of rising scores, these comparisons are less reliable than those based on annual data. 2/

The majority of the tests considered here showed the largest declines on language-related subtests, but the exceptions were frequent enough to suggest that this ranking is more a reflection of the attributes of individual tests than an underlying consistency in the achievement trends (see Table C-1). In addition to the SAT, test batteries that showed the greatest decline in language-related tests include the NLS and HSB comparison, the grade 12 Iowa state data (ITED-Iowa), the Illinois Decade Study, and, for the most part, the Project TALENT 15-year comparison (1960 and 1975). In contrast, Iowa state elementary school data (ITBS-Iowa) show the opposite pattern: the decline in mathematics was much more severe than that in any of the language-related subjects. Senior high school norming data for the California Achievement Test (CAT-US) also show a greater decline in mathematics than in other areas. Other test batteries--such as the national norming data for the elementary-level Iowa test battery (ITBS-US)--show a more complex pattern, with the various language-related tests bracketing the mathematics test in terms of the magnitude of the decline. The ACT showed a slightly larger decline in English than in mathematics. It also showed its largest decline in social studies, however, and no decline at all in science.

The various tests are also inconsistent in terms of the relative declines in "directly taught" and "indirectly taught" subjects. Some of the language-related tests that showed particularly steep declines--such as the vocabulary tests in the Project TALENT data and the NLS-to-HSB comparison--might be viewed as being largely indirectly taught subjects. Other language-related tests that declined markedly, however, presumably are much more reliant on formal instruction--such as the language test in the national ITBS data and the expression test in the national ITED data, both of which are tests of language usage. In addition, mathematics, which has been used as an example of a directly taught subject, showed the steepest decline in several test batteries.

---

2. In the case of tests administered less often than annually, the tabulations used here are based on a single interval during which all subjects evidenced declines. If an adjacent interval showed declines in some subjects but not others--as was the case, for example, with the grade-eight ITBS norming data--that adjacent period was ignored.

TABLE C-1. MAGNITUDE OF THE ACHIEVEMENT DECLINE, BY SUBJECT

Test	Grade	Subject	Total Decline (Standard Deviations)
SAT	12	Verbal	0.48
	12	Mathematics	0.28
NLS to HSB	12	Vocabulary	0.22
	12	Reading	0.21
	12	Mathematics	0.14
ITED-US	12	Expression	0.28
	12	Mathematics	0.26
	12	Vocabulary	0.23
ITED-US	10	Mathematics	0.32
	10	Expression	0.29
	10	Vocabulary	0.22
ITED-Iowa	12	Reading a/	0.40
	12	Social Studies	0.36
	12	Expression	0.32
	12	Vocabulary	0.30
	12	Science	0.28
	12	Mathematics	0.27
ITED-Iowa	10	Reading a/	0.32
	10	Mathematics	0.31
	10	Expression	0.29
	10	Social Studies	0.27
	10	Vocabulary	0.25
	10	Science	0.25
ITBS-Iowa	8	Mathematics	0.47
	8	Language	0.37
	8	Reading	0.35
	8	Vocabulary	0.26

(Continued)

TABLE C-1. (Continued)

Test	Grade	Subject	Total Decline (Standard Deviations)
ITBS-Iowa	6	Mathematics	0.38
	6	Language	0.25
	6	Reading	0.17
	6	Vocabulary	0.10
ITBS-US	8	Language	0.32
	8	Mathematics	0.28
	8	Vocabulary	0.23
	8	Reading	0.20
ITBS-US	6	Language	0.32
	6	Mathematics	0.28
	6	Vocabulary	0.19
	6	Reading	0.17
CAT-US	12	Mathematics	0.34
	12	Reading Comprehension	0.24
	12	Vocabulary	0.23
	12	Language	0.18
CAT-US	9	Mathematics	0.30
	9	Language	0.28
	9	Vocabulary	0.21
	9	Reading Comprehension	0.05
ACT	12	Social Studies	0.55
	12	Mathematics	0.42
	12	English	0.37
	12	Science	0.06
Illinois Decade	11	English 2	0.49
	11	English 1	0.38
	11	Social Studies	0.35
	11	Math 2	0.26
	11	Science	0.19
	11	Math 1	0.05

(Continued)

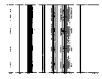
TABLE C-1. (Continued)

Test	Grade	Subject	Total Decline (Standard Deviations)
Talent 15-Year Follow-Up	9,10,11	Vocabulary	0.40
	9,10,11	English	0.30
	9,10,11	Quantitative Reasoning	0.22
	9,10,11	Reading Comprehension	0.06
	9,10,11	Computation	0.23
	9,10,11	Mathematics	-0.07
	9,10,11	Abstract Reasoning	-0.24
	9,10,11	Creativity	-0.34

SOURCES: CBO calculations based on Hunter M. Breland, *The SAT Score Decline: A Summary of Related Research* (New York: The College Board, 1976); The College Entrance Examination Board, *National College-Bound Seniors, 1978 and 1985* (New York: The College Board, 1985); Donald A. Rock, Ruth B. Ekstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated with Decline of Test Scores of High School Seniors, 1972 to 1980* (Washington: Center for Statistics, U.S. Department of Education, 1985); Robert Forsyth, Iowa Testing Programs, personal communications, April, 1984; "Mean ITED Test Scores by Grade and Subtest for the State of Iowa" (Iowa City: Iowa Testing Programs, undated and unpublished tabulations); "Iowa Basic Skills Testing Program, Achievement Trends in Iowa: 1955-1985" (Iowa City: Iowa Testing Programs, undated and unpublished tabulations); A. N. Hieronymus, E. F. Lindquist, and H. D. Hoover, *Iowa Tests of Basic Skills: Manual For School Administrators* (Chicago: Riverside, 1982); *The Development of the 1982 Norms for the Iowa Tests of Basic Skills* (Chicago: Riverside, 1983); CTB/McGraw-Hill, unpublished tabulations, December 1977; L. A. Munday, *Declining Admissions Test Scores* (Iowa City: American College Testing Program, 1976); American College Testing Program, *National Trend Data for Students Who Take the Act Assessment* (Iowa City: ACT, undated); *Student Achievement in Illinois, 1970 and 1981* (Springfield: Illinois State Board of Eduction, 1983); John C. Flanagan, "Analyzing Changes in School Levels of Achievement Using Project TALENT Ten- and Fifteen-Year Retests," in G. R. Austin and H. Garber (eds.), *The Rise and Fall of National Test Scores* (New York: Academic Press, 1982), pp. 35-49.

NOTE: This table is limited to data that span a sizable portion of the decline and that permit exclusion of the subsequent upturn. Only selected grade levels are presented for the sake of simplicity.

a. This is the "Interpretation of Literary Materials" test. Reading skills are also measured by the ITED social studies and science tests.




APPENDIX D

VARIATION AMONG ACHIEVEMENT SUBGROUPS

As discussed in Chapter IV, there is inconsistent evidence about the relative trends in test scores among different achievement subgroups--that is, among groups of students categorized by their differing levels of achievement. Because this issue has received considerable public attention, and because the conclusions presented in the paper are not entirely in keeping with those presented by some other writers, this appendix provides additional detail about the evidence that underlies the following five generalizations, presented in Chapter IV:

- The achievement decline and the subsequent upturn occurred among both low- and high-achieving students.

- During the mid- and late 1970s--that is, during the end of the achievement decline and the beginning of the subsequent upturn--students in the top achievement quartile on the National Assessment of Educational Progress (the top fourth of all students, when ranked by achievement) lost ground relative to those in the bottom quartile.

- Other data, however, do not consistently suggest a narrowing gap between the top and bottom achievement quartiles. The narrowing evident in the NAEP data might be limited to the short time period of that particular assessment (roughly half of the 1970s), or it might be limited to certain types of tests. Alternatively, more detailed analyses than those now available might show the narrowing to be a more general pattern.

- Test scores of students taking college-admissions tests--currently, about half of all high-school graduates--declined more than those of high school seniors in general, but this difference primarily reflects the changing composition of the group taking those tests rather than a greater decline in achievement among high-achieving students.

- Select students--those scoring highest on tests, taking the most advanced courses, and so on--experienced both the decline and

the subsequent upturn in achievement. Select students did not show a consistently greater decline than the average student. Indeed, by some measures, select students appear to have gained relative to the average, particularly in the area of mathematics. The sketchiness and inconsistency of data on select students, however, cloud these conclusions.

As noted in Chapter IV, however, both differences and similarities among trends in achievement subgroups must often be taken with a grain of salt. They can be simple artifacts of technical aspects of the tests used--specifically, the scaling of the test, its content, and the measure of change that is reported. For example, if both the top and bottom achievement quartiles show a decline of 5 percentage points in the average number of test items answered correctly, these seemingly equivalent changes could in fact reflect very different real changes in skills. The change would be proportionately larger in the bottom quartile. Moreover, the typical students in each quartile answer very different questions correctly, and only detailed information about the content and difficulty of the additional items answered incorrectly by each quartile would indicate whether the loss of skills in each group are qualitatively or quantitatively similar. 1/ Technical solutions of this ambiguity are complex and have rarely been applied to the specific question of relative trends among different achievement subgroups.

The test results cited in this section differ in the certainty of their conclusions about achievement subgroups. At one extreme, the results of the Illinois Decade study are very ambiguous, because two available measures of change lead to different conclusions about achievement subgroup differences. At the other extreme, some--but not all--of the relevant tabulations from the National Assessment are clear-cut, because some show increases in the lowest quartile concurrently with decreases in the top quartile. Use of different scaling or reporting conventions would generally not alter the conclusion of a narrowing achievement gap in those cases.

---

1. This ambiguity also arises with other common measures of change, such as scaled-score or standardized-score changes.

   Technically, the problem has several aspects. One is that the metrics commonly used are not ratio scales; indeed, they are arguably not even interval scales. The construction of the tests poses additional problems, for a single test is unlikely to be a comparably comprehensive measure of mastery at two very different levels of achievement and therefore may understate the relative change of students at one level. The tabulation and reporting of results further complicates comparisons, since information on the additional items correctly or incorrectly answered is rarely reported, particularly for achievement subgroups.

TRENDS IN THE LOWEST AND HIGHEST QUARTILES

The most extensive and best-known information on the relative trends among students in the top and bottom achievement quartiles is from the NAEP. Relevant information is also available, however, from the SAT, the ACT, the ITBS, and the Illinois Decade Study.

The National Assessment of Educational Progress

In general, the currently available NAEP tabulations show a narrowing of the gap between the top and bottom quartiles in all three age groups (9, 13, and 17) and subjects (reading, science, and mathematics) for which the analysis was conducted. The comparative data, however, span only four or five years during the 1970s. Comparable tabulations of the NAEP are unavailable for the remaining middle half of the student population.

These particular NAEP trends show great variation--changes ranged from sizable improvements to large declines--which complicates comparison of achievement subgroups. This variation probably results in part from the period over which changes were measured--beginning between 1972 and 1974 and ending between 1976 and 1979, depending on the subject tested. Given the cohort pattern shown by the end of the decline, it is likely that these particular assessments of trends among nine-year-olds began about the time that their brief and small decline ended. The trend for 13-year-olds probably spanned the last years of the decline and the first years of the upturn, while the trend among 17-year-olds corresponds roughly to the last years of the decline. Consistent with this cohort pattern, the NAEP data described here show few declines among 9-year-olds, few gains among 17-year-olds, and a more mixed pattern among 13-year-olds. Comparisons are thus clearest if made within any one age group.

In the lowest quartile, nine-year-olds showed improvement in two of three subject areas and no change in the other. This held true for both black and white students (see Table D-1). In the top quartile, black students also showed improvement. White students did not, however; they showed sizable declines in two subjects and no change in a third.

TABLE D-1. RECENT TRENDS IN THE NATIONAL ASSESSMENT, BY ACHIEVEMENT SUBGROUPS AND ETHNICITY

Group	Subject Area		
	Reading	Science	Mathematics
9-Year-Olds in the 4th Grade			
Lowest Quartile			
Black students	Improvement: gain of 8.4 percentage points	No significant change in performance	Improvement: gain of 2.9 percentage points
White students	Improvement: gain of 4.6 percentage points	Improvement -gain of 1.7 percentage points	No significant change in performance
Highest Quartile			
Black students	Improvement: gain of 3.0 percentage points	No significant change in performance	Improvement: gain of 2.6 percentage points
White students	No significant change in performance	Significant decline: 2.4 percentage points	Significant decline: 3.3 percentage points
13-Year-Olds in the 8th Grade			
Lowest Quartile			
Black students	Improvement: gain of 3.5 percentage points	No significant change in performance	Improvement: gain of 2.6 percentage points
White students	Improvement: gain of 1.5 percentage points	Improvement: gain of 2.0 percentage points	No significant change in performance
Highest Quartile			
Black students	Improvement: gain of 2.5 percentage points	No significant change in performance	Significant decline: 2.5 percentage points
White students	No significant change in performance	Significant decline: 4.1 percentage points	Significant decline: 3.2 percentage points
17-Year-Olds in the 11th Grade			
Lowest Quartile			
Black students	No significant change in performance	No significant change in performance	Improvement: gain of 1.6 percentage points
White students	Significant decline: 1.7 percentage points	No significant change in performance	Significant decline: 1.8 percentage points
Highest Quartile			
Black students	No significant change in performance	Significant decline: 3.9 percentage points	Significant decline: 5.5 percentage points
White students	No significant change in performance	Significant decline: 4.2 percentage points	Significant decline: 4.3 percentage points

SOURCE: National Assessment of Educational Progress, "Educational Winners and Losers, the Whos and Possible Why," (press release, February 8, 1983).

Among 13-year-olds as well, the lowest quartile showed mostly improvements in performance, albeit typically smaller than among the younger children. (This, too, is expected in light of the cohort pattern.) White students in the highest quartile again showed declines in two of three subjects; among blacks in this quartile, gains and losses were approximately balanced.

A similar discrepancy between the highest and lowest quartiles also appeared among the 17-year-olds, although overall--as expected--declines predominated over gains. Blacks in the lowest quartile showed no change in two subjects and a small gain in a third. Their white counterparts showed slight declines in two of three subjects. In contrast, in the top quartile, both races showed large declines in two subject areas.

Other Data

Data from other sources, however, are partially inconsistent with the NAEP data and call into question whether there was a general closing of the gap between high- and low-achieving students on a variety of tests and over the entire period of the achievement decline.

Tabulations of SAT candidates categorized by self-reported class rank show a similar narrowing of the gap between high- and low-achieving students since 1975. Moreover, this pattern occurred over most of the range of achievement; each group declined relative to all others ranking lower, bringing the scores of high-ranking and low-ranking students closer to each other. (These data unfortunately do not include reliable information about the bottom 20 percent.)

Ambiguous evidence on the relative trends among students in the top and bottom quartiles is found in the "Illinois Decade Study," a comparison of scores on a fairly high-level achievement test administered to Illinois high school juniors in the 1970 and 1981 school years. Declines in raw scores were consistently larger among students at the 75th percentile, albeit sometimes by a very small margin (see Table D-2). 2/ On the other hand,

---

2. *Student Achievement in Illinois, 1970 and 1981* (Springfield: Illinois State Board of Education, September 1983). Note that these data are not entirely comparable to the NAEP achievement subgroups analysis. Rather than reporting the average scores of all students above the 75th percentile--as in the NAEP reports--the Illinois Decade study reports results for the students at the 75th percentile. The same distinction applies to the data on scores at the 25th percentile. Thus, the NAEP analyses incorporate students who are further apart in their levels of achievement.

TABLE D-2. CHANGE ON THE ILLINOIS DECADE TEST AMONG STUDENTS AT THE 25th and 75th PERCENTILE

	75th Percentile		25th Percentile	
	Raw Change	Percent Change	Raw Change	Percent Change
Mathematics 1	-0.7	-4.6	-0.2	-2.7
Mathematics 2	-1.5	-12.5	-0.8	-13.1
English 1	-3.4	-16.0	-1.8	-13.8
English 2	-3.1	-15.2	-2.9	-22.1
Social Studies	-2.6	-15.0	-1.0	-11.0
Natural Science	-0.8	-6.6	-0.6	-8.8

SOURCE: CBO calculations based on Illinois State Board of Education, *Student Achievement in Illinois, 1970 and 1981*, Exhibit A-5; and J. Fyans, personal communication.

when the changes are expressed in proportional terms, this pattern disappears. The percent change in scores at the 25th percentile were sometimes smaller but sometimes larger than those at the 75th percentile.

Data from other tests, however, and from the SAT earlier in the period of decline (before 1975), cast doubt on the NAEP results. A tabulation of changes in SAT scores among groups of students divided by their percentile rankings on the SAT itself showed no comparable narrowing of the gap in the years before 1975. Indeed, in mathematics, the gap appears to have widened slightly (see the section below on "select students"). In addition, if the gap between the top and bottom quartiles were narrowing, one would expect a shrinking standard deviation--that is, a narrower distribution of scores. 3/ Since the beginning of the 1970s,

3. The standard deviation would shrink unless there were other, offsetting changes in the distribution of scores--such as a change in the distribution of scores in the middle two quartiles. Moreover, without such other distributional shifts, changes in the composition of the test-taking group would not alter this link between the standard deviation and the gap between the top and bottom quartiles. Any change in the standard deviation attributable to compositional changes (such as an increase resulting from lower dropout rates) would also be reflected in the gap between high- and low-achieving students.