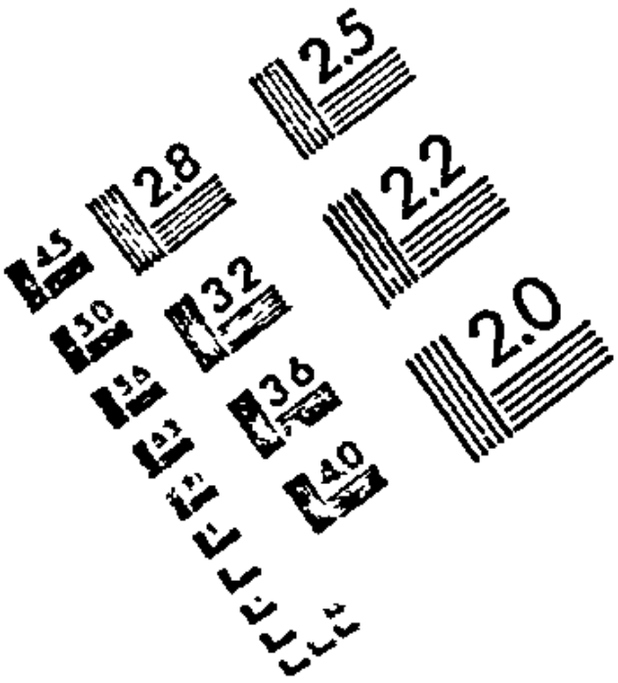

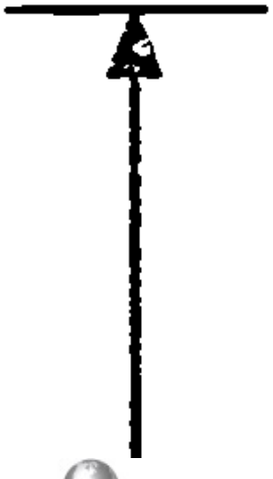

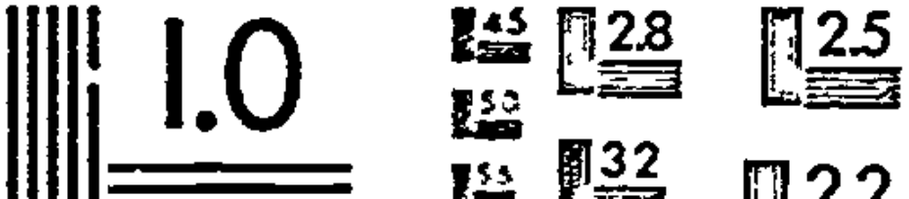

DOCUMENT RESUME

ED 320 101 CS 009 835

AUTHOR Stedman, Lawrence C.; Kaestle, Carl F.
TITLE An Investigation of Crude Literacy, Reading Performance, and Functional Literacy in the United States, 1880 to 1980. Program Report 86-2.
INSTITUTION Wisconsin Center for Education Research, Madison.
SPONS AGENCY National Science Foundation, Washington, D.C.; Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE May 86
GRANT NIE-G-84-0008
NOTE 145p.; Report prepared for the project, A Social History of the American Reading Public. Project supported by the Spencer Foundation and the National Institute of Education.
PUB TYPE Historical Materials (060) -- Information Analyses (070)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Academic Standards; *Educational History; *Educational Trends; *Literacy; *Reading Achievement; Reading Research; *Research Methodology; Standardized Tests; Test Bias; Test Reliability; *Test Score Decline; Test Validity

ABSTRACT

Focusing on the problems of validity and representativeness of samples, a study examined the history of literacy in the United States since 1880 in order to set the contemporary debate on test scores, literacy and reading performance in the longer-range perspective of the last 100 years. The quality of the data and the arguments of literacy scholars concerning crude literacy, reading performance, and functional literacy were examined. Results indicated that although the problems of concept validity, representativeness of research samples, and noncomparability across time confuse the attempt to discern trends, three historical trends were identified. These trends are that (1) in the twentieth century, self-reported outright illiteracy almost disappeared as a percentage of the whole population; (2) the big story in twentieth century literacy is the rise in school attainment, not the relative effectiveness of schools to teach children at a particular grade level; and (3) the greatest difficulty was found in making a confident statement about the reading abilities of people at different points in time with the same amount of schooling. Findings suggest that present-day literacy policy should be argued on the basis of an assessment of the current condition and on the basis of shared educational goals and not on the basis of alleged declines or rises in literacy skills. (Five footnotes are included; 15 pages of references, 20 charts, and five appendixes of data are attached.) (RS)

Program Report 86-2

AN INVESTIGATION OF CRUDE LITERACY, READING PERFORMANCE, AND FUNCTIONAL LITERACY IN THE UNITED STATES, 1880 to 1980

Lawrence C. Stedman and Carl F. Kaestle

Report from the Project on a Social History of the American Reading Public, 1880-1980

Carl F. Kaestle, Principal Investigator

Wisconsin Center for Education Research
School of Education
University of Wisconsin
Madison, Wisconsin

May 1986

NOTE

A portion of this report, Section II-B, on trends in reading performance, has been revised and expanded as an article, "The Test Score Decline Is Over: Now What?" which appeared in the Phi Delta Kappan November 1985. It is also available in slightly longer form, with more documentation, as a separate Program Report (85-8) of the Wisconsin Center for Education Research. The latter report was submitted to the ERIC system in September 1985.

A report for the project, A Social History of the American Reading Public, Carl Kaestle, Director. Supported by the Spencer Foundation and by the National Institute of Education (Grant No. NIE-G-84-0008), through the Wisconsin Center for Education Research.

Wisconsin Center for Education Research

MISSION STATEMENT

The mission of the Wisconsin Center for Education Research is to improve the quality of American education for all students. Our goal is that future generations achieve the knowledge, tolerance, and complex thinking skills necessary to ensure a productive and enlightened democratic society. We are willing to explore solutions to major educational problems, recognizing that radical change may be necessary to solve these problems.

Our approach is interdisciplinary because the problems of education go far beyond pedagogy. We therefore draw on the knowledge of scholars in psychology, sociology, history, economics, philosophy, and law as well as experts in teacher education, curriculum, and administration to arrive at a deeper understanding of schooling.

Work of the Center clusters in four broad areas:

- Learning and Development focuses on individuals, in particular on their variability in basic learning and development processes.
- Classroom Processes seeks to adapt psychological constructs to the improvement of classroom learning and instruction.
- School Processes focuses on schoolwide issues and variables, seeking to identify administrative and organizational practices that are particularly effective.
- Social Policy is directed toward delineating the conditions affecting the success of social policy, the ends it can most readily achieve, and the constraints it faces.

The Wisconsin Center for Education Research is a noninstructional unit of the University of Wisconsin-Madison School of Education. The Center is supported primarily with funds from the Office of Educational Research and Improvement/Department of Education, the National Science Foundation, and other governmental and nongovernmental sources in the U.S.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

I. INTRODUCTION

A. The Problem

On September 21, 1982, Congressman Paul Simon (D-IL) opened a congressional hearing on literacy declaring that "10 to 25 million Americans are unable to read and write." Furthermore, he declared that "an additional 35 million Americans can read only at the fifth-grade level" (Illiteracy . . ., 1984, p. 1). Concurring, Secretary of Education Terrence Bell testified that "in 1975, you have 63 million Americans that aren't proficient in meeting the educational requirements of every day adult life" (p. 5). By 1982, this had risen to 72 million, which amounted to more than a third of the adult population.

The public and many scholars have blamed the schools, arguing that test scores have been in decline for nearly 20 years, writing skills have atrophied, and permissive schooling and electronic media have short-circuited the reading abilities of our nation's youth. Supporting them, surveys in the 1970s showed that between 11 percent and 19 percent of high school graduates were functionally illiterate. After reviewing the evidence, the President's National Commission on Excellence in Education concluded in 1983 that the very security of the nation was at risk.

Findings by other experts suggest that these claims were wildly exaggerated. The U.S. Bureau of the Census, for example, estimated in 1979 that less than 1 percent of the population was illiterate. In a reexamination of functional literacy studies for the National Institute of Education, Donald Fisher (1978, p. 7), concluded that "few if any functional illiterates were actually awarded high school diplomas." Some scholars emphasize that educational attainment has risen steadily during this century in response to the rising literacy demands of a highly technological, information-laden society (Resnick & Resnick, 1977; Bormuth, 1978). Research by Roger Farr and his colleagues seems to confirm that each succeeding generation has been better educated than the last. After administering the same tests that had been given in the 1940s, they found that Indiana students in 1976 outperformed those of 1944-1945 (Farr, Fay, & Negley, 1978). Their comprehensive review of then-and-now studies also showed that students' reading skills had improved over the course of the century (Farr, Tinman, & Rowls, 1974). They concluded that "anyone who says that he knows that literacy is decreasing is . . . at best unscholarly and at worst dishonest" (p. 140).

What is an interested nonexpert to believe? Is illiteracy a serious problem or a relatively minor one? Are the trends up or down? Are they short-run or long-run? Are the people with the apocalyptic visions and the rose-colored glasses looking at the same information differently, or is each commentator mustering the data selectively? Is the measurement of literacy trends a shell game?

The purpose of this review is to set the recent debate on literacy and reading performance in the longer-range perspective of the past one hundred years. Writers engaged in policy debates rarely project literacy trends back more than twenty years. Some imply that before the test-score decline of the 1960s there was a golden age of literacy, a high plateau of impressive test scores, rigorous standards, and old-fashioned academic schooling. In general, the polemicists have focused on dubious short-run causal theories about the decline of standards in the rebellious 1960s and the deleterious effects of television.

Historians, on the other hand, have been little help in providing perspective on this issue. Although much exciting work recently has been done on the history of literacy, the story is rarely brought up into the twentieth century. The focal points for this hot specialty have been the advent of printing in sixteenth-century Europe and the expansion of literacy in late eighteenth- and nineteenth-century industrial societies, including the United States. For twentieth-century America there exist only a few summary articles about literacy trends, largely based on U.S. Census reports (Folger & Nam, 1967; Kirsch & Guthrie, 1977-1978; Selden, 1978). This article is therefore a first step toward a more detailed history of literacy in the United States since 1880, as well as a perspective on the contemporary test-score debate. For the entire period, we take a hard look at the quality of the data and the arguments of previous scholars. Much of the available data is unreliable, unrepresentative, or noncomparable over time. The attempt to determine trends is therefore perilous. Much skepticism is in order.

B. Definition of Literacy

The distinction between a literate person and an illiterate person sounds simple. It is not. Some people can read but don't; others learn how but forget later. Some can read in one context but not in another. We think of literacy as a hierarchical, measurable skill, but recent studies by linguists and anthropologists suggest otherwise (Olson, 1977; Heath, 1984). Literacy is elusive, complex. Its study requires careful definitions. Because this article reviews previous measurement efforts, we are to some degree prisoners of previous definitions. But we can clarify literacy trends by categorizing earlier measurement efforts under different concepts of literacy.

We distinguish between a vertical dimension and a horizontal dimension of literacy skills. The setting for the vertical dimension is the school. Here literacy has to do with a sequential reading curriculum, featuring authorized subject matter and expanding vocabulary recognition in standard English. Children are frequently tested as they move up through a hierarchy of graded skills and content. At the lowest level, which we shall call "crude literacy," the student learns to decode written words and to say them. Most people would not acknowledge this skill as reading unless the student also understands the meaning of the words, so we shall define crude

literacy as the ability to pronounce and understand written text using vocabulary already known to the student. As the student proceeds upward through the hierarchy, she or he learns not only more vocabulary and concepts but new logical skills and aesthetic principles. These higher level skills are all subject to testing and ranking in the hierarchy. The setting for the horizontal dimension of literacy is the world outside the school--in the family, at leisure, as a citizen, or on the job. Here literacy is more informally structured, more contextually specific, less hierarchical. Although school-based reading instruction has an obvious bearing on reading in the nonschool environments, the relationship is not predictable. Some studies of everyday reading tasks suggest that many high school students getting by in school reading tasks can't cope with various bureaucratic tasks presented to young adults, while other studies have discovered some children who fail at school but can read Sports Illustrated or an interesting cookbook. Many people label this horizontal dimension "functional literacy," a term that arose during the 1930s as a contrast to basic or crude literacy.

C. Reading Ability in America Before 1880

Studies of reading ability in early American history have focused on two purported measures of literacy: rates of signature-signing ability and rates of self-reported literacy in the U.S. Census. Despite the questionable validity of the data, it makes sense to get some estimate of crude literacy rates for a society in which large numbers of people are utterly illiterate. The consequence, unfortunately, is that historians tend to think of literacy as an either-or proposition. Historians of literacy in early America have thus been preoccupied with the question of how many people were literate at all, and how the literates differed from the illiterates in sex, race, and wealth. Until 1840, when the U.S. Census marshalls began asking people if they were literate, the only widespread and relatively systematic data on literacy consisted of signatures on public documents like deeds, wills, marriage registers, and army enlistment rosters. When people could not write their name, they marked an X. Did this mean they could not read? Conversely, could all the signers read? Some European data suggest that reading and signing ability were closely related, but other historians dispute the measure, especially in the case of women. Even if signing is accepted as an estimate of crude literacy, sampling is a problem. Different American historians have arrived at different rates of illiteracy, depending upon what sort of documents they sampled and what sort of communities they studied (Kaestle, 1985).

Historians nonetheless agree about some trends in early American literacy rates. The British colonists to New England had higher literacy rates than English people in general at the time. The New Englanders resembled the highly literate Swedes and Scots. Southern colonists of the seventeenth century were at about the English level. Male signature rates in New England, at about 60 percent, doubled female signature rates, which were at about 30 percent. Moreover, the

male signature rates made a more unambiguous climb in the eighteenth century than women's, arriving at about 90 percent by the time of the American Revolution, according to Kenneth Lockridge, while women's signature-signing rates were 40 to 50 percent (1974). Women probably were more literate than Lockridge's estimates suggest, however, because some scholars have recorded higher female signature-signing rates in certain settings, and, more important, because there is reason to believe that many women could read but not write (Auwers, 1980; Spufford, 1979; Tully, 1972). Even for males, the signature counts from deeds, wills, and marriage registers may leave out as much as one-fourth of the population who never signed such documents. Some scholars have argued that male illiteracy at the time of the Revolution was closer to 25 percent, in contrast to Lockridge's estimate of 10 percent (Soltow & Stevens, 1981; Gilmore, 1982).

Whatever the rates at the time of the Revolution, literacy had expanded by 1840, when the U.S. Census first included a literacy question. Census marshalls simply asked whether people could read and write. By the time of the 1850 Census, when sex differences were first reported in the aggregate data, 7.3 percent of white males replied no and 12.4 percent of white females said they could neither read nor write (Vinovskis & Bernard, 1978). No tests were given, so the Census data must be seen as a measure of people's willingness to state that they were illiterate. We cannot tell if census literacy figures correlate with actual reading ability, but the high white male rates and the dramatically increased female rates make sense in view of the expansion of schooling and the growing acceptance of education for girls during the early nineteenth century.

By the time of the Civil War, native-born American whites were almost all counted as literate on measures of crude literacy. While the colonial signature counts relate only to white people, the Census recorded self-reported literacy for all groups. After the Civil War, with the emancipation of enslaved black Americans and the increasing immigration of white Europeans, attention in literacy studies turned to crude literacy rates among blacks and recent immigrants. By 1880, when our analysis begins, white male and female census illiteracy rates were less than 2 percent apart, at 8.6 and 10.2 respectively. Among nonwhites, 67.3 percent of the males and 72.7 percent of the females stated that they were illiterate. About 12 percent of foreign-born whites responded that they were illiterate in any language. This was not very much higher than native-white rates in 1880, but the foreign-born rate stayed at that level until 1920, while the native-white rate dropped below 5 percent.

II. TRENDS IN U.S. LITERACY, 1880-1980

Recent publicity about an alleged decline in literacy skills among American schoolchildren has featured both the vertical and the horizontal dimensions of literacy. Students are slipping in their grade-level reading achievement, critics charge, and they are not learning functional reading skills for the practical world beyond school. An assessment of U.S. literacy trends over the past century, therefore, requires examining both types of literacy. Assertions about literacy trends, however, require accurate statistics. While historians of literacy have agonized and debated about the validity and representativeness of their measures, contemporary researchers and policy advocates have not always been so careful. Claims of test-score decline and burgeoning functional illiteracy have often ignored or de-emphasized problems of validity and sample bias. In attempting to provide a long-range perspective on reading ability, we must ask whether good estimates are even possible.

In what follows we do not deal with racial, regional, sexual, social class, or national origin differences in literacy. As might be expected, blacks, other minorities, the poor, Southerners, and the foreign born have been less literate on average than native-born, middle-class white males in the North. Although the gaps in crude literacy have narrowed substantially during the past century, some large gaps still remain on measures of the more complex skills. (See Appendix A.) We have decided to forego an analysis of literacy by groups, not because we are insensitive to these groups' special needs, but because the conceptual and measurement problems are so complicated that they demand first consideration. Because we raise serious questions about the validity of most measures, it made little sense to devote a great deal of time to analyzing findings for particular groups. Nevertheless, we must emphasize that the higher illiteracy rates for blacks and other minorities persist today even after taking SES into account, suggesting serious problems of discrimination and lack of opportunity remain. Having written that, however, we must avoid perpetuating stereotypes. While it is true that minority groups have higher illiteracy rates, most individuals in such groups are not illiterate. Indeed, most illiterates, functional illiterates, and marginally functional adults do not belong to minority groups, but are whites from a wide range of regional and social class backgrounds.

A. The Vertical Dimension: Crude Literacy

The national study of literacy began with efforts by the Census Bureau to assess the population's educational level. In each decennial census from 1840 through 1930 and in sampling surveys since, individuals have been asked whether they can read and write in any language. The Bureau classified as illiterate those who said they were "not able both to read and to write a simple message either in English or any other language" (U.S. Bureau of the Census, 1971, p. 5).

The record shows a tremendous reduction in this self-reported crude illiteracy over the past century. In 1870, 20 percent of the population was considered illiterate, while in 1979, only .6 percent was. Since the 1979 data include those who were educated many decades earlier, they do not reflect current educational conditions. Among youths aged 14 to 24 the crude illiteracy rate in 1979 was only .19 percent (U.S. Bureau of the Census, 1982, p. 17).

According to the Census, then, crude illiteracy has been largely eliminated, although it should be kept in mind that these data refer to any language, not just English, and that the absolute numbers of illiterates remain large--822,000 of those 14 and older (U.S. Bureau of the Census, 1982, p. 5).

What are we to make of the validity of this record? In more than one hundred years of inquiry, the Census Bureau never administered a literacy test. It relied upon people describing their own literacy status and never defined precisely what was meant by literacy. (See census instructions, Folger & Nam, 1967, p. 249-252; U.S. Bureau of the Census, 1948, 1960, 1971, 1982.) Because there was no test, the data reflect a myriad of personal judgments about what constitutes literacy.

This self-reporting is the fundamental weakness of the census literacy data. As a 1919 New York Times editorial stated, "Nothing could be more inexact or humorous" (February 19, 1919, p.12, column 4). Self-reporting probably produced errors in census literacy data in two major ways: some respondents misinterpreted the literacy question, and others deliberately reported themselves as literate when they were not. Misinterpretation probably resulted in an underreporting of crude illiteracy in the late nineteenth and early twentieth centuries. Historically, literacy was equated with the ability to sign one's name, so some who claimed literacy on that basis may not have been able to read and write simple messages. The Census Bureau indirectly acknowledged this. In 1930, for the first time, it warned census takers: "Do not return any person as able to read and write simply because he can write his own name" (Folger & Nam, 1967, p. 252). In a special monograph prepared for the Bureau, Folger and Nam (1967) acknowledged that the distinction between signing and being able to write simple messages was likely sometimes lost (p. 111). The illiteracy rate for simple messages in 1870, our initial data point, therefore, was likely higher than the 20 percent reported.

Because educational levels and literacy demands have increased steadily during this century, people probably came to define reading and writing in higher level terms. Thus, people who claimed literacy in recent decades were more likely to have meant they could handle simple messages. Since 1930 the Census Bureau has asked a combined question--whether the respondent could read and write--so the claim of literacy through signature-signing ability was even less likely. Over time, therefore, we conclude that census figures probably became a more accurate gauge of the Bureau's "simple message" definition of basic literacy. Because nineteenth century data would have

understated the extent of simple message illiteracy, the increase in basic literacy has been more dramatic than the record shows.

Deliberate misreporting is a more serious problem, however. Illiterates have great difficulty admitting their inability to read and write. They have developed elaborate methods of concealment and will hide their illiteracy even from other illiterates (Freeman & Kassenbaum, 1956, p. 372-3; Kozol, 1985). Although in 1870 the Census officials recognized that "great numbers of persons rather than admit their ignorance, will claim to read . . ." (Winston, 1930), by 1920 it viewed misreporting as a minor problem. "In some cases," they said, "there may be unwillingness to admit illiteracy on the part of the persons enumerated." The Bureau concluded, however, that the data were "nearly enough accurate" (U.S. Bureau of the Census, 1920, p. 1145). On the contrary, we believe that the problem probably became more serious. Given that reading levels and demands increased greatly during this century, the stigma attached to illiteracy must have also, making it much more likely that false reporting would occur. Unlike misinterpretation, therefore, misreporting probably increased, making the statistics continually less reliable. The question is, to what extent did this occur.

Although the Census Bureau never attempted to justify its reliance on self-reporting, Folger and Nam (1967), writing for the Bureau, cited two studies that purportedly demonstrated that an individual's self-reporting mirrors performance on literacy tests. They concluded that census reports of literacy are "generally accurate" (p.129, footnote 1). Both studies, however, were conducted in foreign countries, which raises questions of generalizability to the United States and, as Coles (1976) points out, in one of the two studies, test results actually demonstrated the inaccuracy of self-reporting (p. 51). The study was conducted among Iranian oil employees, and of 144 self-reported Farsi literates, nearly two-thirds could not read simple sentences! The other study, which did show the accuracy of self-reporting, was conducted in Yugoslavia in 1953 at a time when the government had launched a nationwide development program, including a major effort in adult education. It was thus beneficial for a person to admit his illiteracy and to take advantage of the reading instruction being offered. Coles argues that, by contrast, in the U.S., illiterates see no such benefits in admitting illiteracy and, in fact, view official governmental inquiries with suspicion. As he describes this process: ". . . when the 'man' comes knocking at the door asking a lot of personal questions about who can read and write, the illiterate is likely to be suspicious and uncertain of whether admission of illiteracy will help him or, in fact, harm him" (p. 51). Consequently, unlike Yugoslavs in the early 1950s, American illiterates <u>are</u> likely to hide their illiteracy. The Bureau's case for accuracy, therefore, rests on questionable evidence.

Nevertheless, there are grounds for believing the data provide a rough guide to trends in crude literacy. One possibly analogous type of data has resulted from surveys of voters' self-reported turn-out, which consistently show higher rates of reported voting than actually

occurred. One of the major sources of the discrepancy is respondent dishonesty. A nonvoter perceives the stigma of not voting and therefore claims to have voted. Wolfinger and Rosenstone (1980) reported that in the 1972 presidential elections, the actual nonvoting rate was 43.3 percent compared to a reported 33.3 percent-- a factor of 1.3. Similar results can be found in Traugott and Katosh (1979, pp. 365-7).

If we adjust literacy figures by the same factor, the 1979 illiteracy rates of .6 percent would rise to .78 percent, a small change. Even the 1870 rate of 20 percent would rise to only 26 percent. Over time, the adjusted decline in illiteracy would be just as dramatic as with raw data. But illiterates likely feel a greater stigma than nonvoters, because they are fewer and they are asked to confess to a lack of ability. We are unsure how great a factor this is, so we can only guess at the relevance of the voting studies.

Functional literary tests provide a more direct way of assessing the accuracy of recent Census figures. The results of the 1970 Survival Literacy Study, in which respondents filled out application forms, were similar to the Census data (Harris & Associates, 1970). For three of the forms--personal identification, driver's license and public assistance---those aged 16 to 24 scored above 70 percent correct, indicating they could handle simple messages. On the other two forms--loan application and Medicaid--1 percent and 6 percent, respectively, couldn't achieve this minimal level of literacy. Harris and Associates estimated that the average functional illiteracy rate in 1970 for 16- to 24-year-olds was 1 percent. The comparable figure from the 1969 census for 14- to 24-year-olds was .3 percent (U.S. Bureau of the Census, 1969).

Our confidence in the census data is further increased when we consider the major educational changes of the past 100 years. There were major gains in educational attainment; extensive basic literacy training through the CCC, the WPA, and the military; and expanded programs in adult basic education. These efforts have eliminated much of the nation's crude illiteracy, just as the record shows. We conclude, however, that illiteracy remains a serious problem. As we shall discuss, even on tests of simple reading passages and help wanted ads, over 5 percent of the population has serious trouble. They can hardly cope with the demands of contemporary U.S. society, which certainly require more than simple message literacy to survive.

B. Vertical Dimension: Reading Performance

The historical record on reading performance comes from then-and-now studies and standardized test score trends.

1. Then-and-Now Studies

The then-and-now studies aim to satisfy our curiosity about whether students today are performing better than students of yesterday. They involve giving a group of students the same test that was given to a comparable group of students years before. This is not a recent idea; one of the first then-and-now studies was performed in 1906. Riley (1908) gave all ninth-grade students in Springfield, Massachusetts, the same tests that had been given in 1846. The 1906 students performed much better on these tests, which covered spelling, arithmetic, and geography.

Previous reviewers have used then-and-now studies as evidence for their claim that reading performance steadily improved from the early part of the century through the mid-1960s (Copperman, 1978, p. 32-4; Farr, Tuinman, & Rowls, 1974.) The historical record, however, is ambiguous. Of thirteen local then-and-now studies that dealt with reading, seven did not show a clear-cut improvement. Of the seven, two showed declines, three no difference, and two had mixed results (see Chart 1). The eight state and national then-and-now studies provided more support, but three still showed no improvement. One had declines (Sligo), one mixed results (Tyler), and one lacked comparable data (the Yerkes-Gray comparison) (see Chart 2). Researchers have also used the then-and-now studies to gauge the magnitude of the test score decline of the late 1960s and 1970s. Some claimed that, in spite of the decline, 1970s students still were doing as well as those of the 1940s and 1950s, whereas others argued that the decline was so great they had fallen well behind. One study supported the more optimistic viewpoint (Farr, Fay, & Negley, 1978); two studies, the more pessimistic view (Elligett & Tocco, 1980; Eurich & Kraetsch, 1982).

On their face, therefore, then-and-now studies provide weak evidence for sweeping claims about changes in national performance. In general, their execution was so poor that their conclusions were unwarranted. There were several problems.

First, then-and-now studies are riddled with problems of comparability. Because few researchers investigated the social composition of their tested groups, we cannot rule out the possibility that the higher achievement of one group was due to a higher social-class background. This problem was particularly acute in local then-and-now studies, which were usually focused on reading achievement in only one city. With ten to twenty years between testings, the chances that a city's social composition had changed were great. Since many local studies were further limited to particular grades or schools within a city or to very small samples of students, the likelihood of uncontrolled differences in composition was large. In only two local studies did researchers ensure that the two groups had similar SES, but both of these involved such small groups that generalizations are unwarranted (Burke & Anderson; Finch & Gillenwater) (see Chart 1). Finch and Gillenwater, for example, compared only 144 sixth graders in 1931 to 198 in 1948. Farr,

Tuinman, and Rowls (1974) acknowledged that such a small size was a major limitation of the study (p. 38). Three other local studies involved even smaller groups (under 36).

Local studies with larger samples also had comparability problems (again, see Chart 1). Caldwell and Courtis compared 530 top Boston students to 12,000 low-achieving students selected nationwide. Daughtry compared two "somewhat different" sets of Florida counties. Eurich and Kraetsch (1982) found that 1928 University of Minnesota first-year students outperformed their 1978 counterparts on a standardized reading test, but the 50-year gap had produced a major difference in the type of student entering the university.

State studies, with one exception, spanned even more time than the local studies, from 20 to 32 years, yet none of the researchers compared the performance of groups matched on social class. Given the migratory patterns, demographic upheavals, and economic transformations during this century, however, this longer time span virtually ensured that a state's composition had changed greatly. Farr, Fay, and Negley (1978), for example, found that from 1944 to 1976 Indiana's population had become more urban, workers held fewer laboring and more service jobs, adults were better educated, and the proportion of blacks and Hispanics had doubled (p. 81). Such changes demonstrate the need to control social composition. Even national then-and-now studies may have involved noncomparable groups. Immigration could have increased linguistic minority students, for example, but no researchers checked the language status of the groups they compared.

Furthermore, during the long time between testings, changing educational policies altered the composition of test takers. Variations in institutionalization and mainstreaming, for example, periodically changed the number of mentally retarded and handicapped students in the schools. Raising the legal school-leaving age and emphasizing high school completion had until very recently decreased the percentage of dropouts over time. For then-and-now studies of high school reading performance, this likely meant that the original group had fewer low-achieving students than the second group, tested several decades later. On the other hand, dropping out used to be more acceptable, so many who did so were not low-achieving students. Rather, they wanted to help support their families, join the military, or get jobs and be independent. In the Indiana study, for example, researchers found that the 1944 dropout rate was more than triple that of 1976. But the legal school-leaving age in 1944 was only tenth grade, and many students who dropped out were lured by job opportunities in wartime industries and by military service (Farr, Fay, & Negley, 1978). Thus, the net effect of changing dropout rates is uncertain for the mid-twentieth century.

The second major problem with then-and-now studies is their failure to be nationally representative. Generalizing from the local studies, with their limited geographical scope and small numbers, is unwarranted. Fridan, for example, studied one parochial elementary

school in Indiana. Partlow studied schools in one Canadian city, and Bradfield studied a fifth grade in a rural California town. The state studies also were not nationally representative, coming mostly from Midwestern rural or semirural states with few minorities and no major cities. The samples in several "state" studies were not even representative of their states. Witty and Coomer (1951) described the percentages of students who passed New York State's Regents exams, but not all students take them, and the socioeconomic status of those who did probably changed between 1915 and 1947. Sligo studied "selected" Iowa high schools (Armbruster, 1977, p. 35). Tyler did the same in Ohio. Several schools declined to be in follow-up testing. No Cincinnati or Cleveland schools were included.

The national studies also were not completely representative. Bloom gave the General Educational Development (GED) test to 1943 and 1955 high school seniors during their final two months. Since the graduation rate among seventeen-year-olds in 1944 was only 43 percent, more than half the nation's students weren't included (U.S. Bureau of the Census, 1975, p. 379). Restricting the number further, Bloom did not sample technical, vocational, private, or black high schools (p. 111). His report of a slight increase in achievement from 1943 to 1955 must be seen as applying to whites who attended public schools and who, because of family income or ability, were able to remain in school through their senior year.

Tuddenham's (1948) finding that World War II draftees outperformed those of World War I on the Army Alpha test was limited to white young men. Jencks, et al. (1972) claimed that the results could be generalized to the entire male population (footnote #23, p. 113), but the World War I sample excluded foreign-language speakers as well as those with few years of schooling.

Elligett and Tocco (1980) and Gates (1961) derived national results using equating studies but tested only one school district and twelve school districts, respectively. (More on equating studies below.) We question how nationally representative the results from so few districts can be.

Despite these problems of comparability and representativeness, can we not form an estimate of trends by putting them all together? Unfortunately, for any given time period, there are only a couple of then-and-now studies, and those were of limited geographical scope. Thus, for the period between the mid-1920s and the mid to late 1930s, we have results from only two California cities, Los Angeles and Santa Monica, and only for sixth graders (Davis & Morgan; Woods; see Chart 1). It would be statistically naive to suggest that these two West Coast cities' elementary schools represented the nation. For the period of the early 1930s to the early 1950s, we have only four studies, covering Springfield, Missouri, sixth graders, Evanston, Illinois, third, fifth, and eighth graders, third- through eighth-grade students in a Canadian city, and selected Iowa high schools (Finch & Gillenwater; Partlow; Riley & Lanton; Sligo; see Chart 1). With the mixture of schooling levels and countries, and a

concentration in the Midwest, these studies can hardly be said to be a representative set of scores. They also provide little support for a steady improvement thesis: the first showed no statistical difference in achievement, the second a gain, the third mixed results, and the fourth a decline.

A third limitation of then-and-now studies, for our purposes, is that they measured skills other than reading comprehension. Although the first six local studies listed included grammar and spelling, their emphasis was on subjects such as arithmetic and geography.

The Indiana study (Farr, Fay, & Negley, 1978) involved speed reading tests; "comprehension" tests, better described as short-term memory tests, in which students answered sets of 10 questions in two minutes about reading passages they could no longer see; and sentence-meaning tests that included questions testing prior knowledge or requiring a moral judgment. Examples of these questions included: "Is treason to one's country punishable by death?" "Is it necessary for the President of the United States to be a citizen?" "Does allegiance to one's country imply loyalty?" (Farr, Fay, & Negley, 1978, p. 35-36). These unusual items make average results from that study suspect. (There were, however, paragraph-comprehension tests without these problems; we describe the results of those later.)

Tuddenham's (1948) comparisons of World War I and World War II draftees were based on the Army Alpha test, which also measured more than literacy skills. The test had eight sections, including mathematical word problems, common sense, number patterns, and general knowledge (Yoakum & Yerkes, 1920, pp. 16, 206). The test was designed to produce rankings that mirrored school grades and teachers' ratings of students' intelligence. Consequently, the test measured the vocabulary and knowledge of those with extensive formal education. The test's cultural biases are apparent even to a casual reader. (See Gould, 1981.)

Literacy data from the recruiting programs of the two wars also were unhelpful. Estimates from World War I indicated that about one-fourth of the recruits were illiterate (Yerkes, 1921, p. 744). Each camp, however, used different standards for determining literacy. In many camps, recruits were asked whether they could read newspapers and write letters home. In other camps, literacy was defined strictly in educational terms, using completion of an arbitrary grade as a presumption of literacy. A number of camps used different standards for blacks and whites; one had different criteria for northern and southern recruits (Yerkes, 1921, p. 744). The literacy data for World War I recruits, therefore, though widely cited, are a mess.

There was no comparable literacy determination for World War II. Gray's estimate of 12 to 15 percent was based on a combination of rejections, illiterates who were drafted, and those with little education (Gray, 1956, p. 39). Since the figures are based on different criteria, we conclude there was no comparable data on the reading performance of World War I and World War II soldiers.

A fourth major problem in then-and-now research is whether to present results by age or grade. We illustrate this problem by discussing one of the better known studies: Gates's (1961) renorming of his reading test comparing 1937 and 1957 students.

A renorming or equating study, as it is better described, differs from a traditional then-and-now study in that two different tests are involved. In equating, a group of present-day students is given both the old and new versions of a test. Their performance on the two different tests establishes a scale for converting scores from one test to the other. Using this scale, researchers can convert the nation's average score on the current test to its equivalent score on the earlier one and then see whether it is higher or lower than the earlier national average.

Using such a procedure, Gates converted 1957 norms for grades three through eight to their 1937 equivalents. He found that the 1937 students outperformed those in 1957 by as much as 4.5 months, with larger margins the higher the grade (see Chart 3). These results suggested that reading performance was better in the 1930s. But Gates noted that students in 1937 were older at each grade level due to stricter promotion policies. When he compared students of the same age, he found the 1957 students outperformed those of 1937 (see Chart 3 again). Gates argued that the proper comparison is by age, not grade, because a student's grade is an artifact of changing educational policies, particularly those relating to school-entering age and retention policies. We agree. By this standard, then, reading performance was better in 1957 than in 1937, at least for the upper elementary grades. But no sweeping conclusions should be reached. Even after age adjustments, first and second graders showed no difference in performance, and third graders differed by only 1 to 1.5 months. Middle of the year fifth graders in 1957 outscored 1937 students by 6.7 months, but middle-of-the year and later sixth graders in 1957 outscored their counterparts by only four months. This downward trend suggests that even smaller differences existed at the junior high school level. Furthermore, there could have been dips or peaks in the scores between the two testings, so any assertion of steady improvement is unwarranted. Finally, we have reservations about using equating studies to determine national trends, which we describe in a later section (See Section II.B.2.b.iii., "What's Wrong with These Particular Measures?" and Appendix C).

Gates's argument about comparing students of the same age is not universally accepted. Copperman and Armbruster, for example, argue that the proper comparison is between students with equal years of schooling. Nonetheless, they reached opposite conclusions about the Gates data because they disagree about students' entry age. Copperman claimed that students began school at the same age in 1937 as in 1957 and thus argued that students of the same age had received an equal number of years of schooling (p. 33). This made Gates's finding that 1957 students performed better the correct one, in Copperman's opinion. Armbruster, by contrast, claimed that 1957 students began school "much earlier" than those in 1937, and this explained why, at a

given grade, they were younger (p. 34). Consequently, he argued, the grade comparison was valid and the 1937 performance better. Neither reviewer presented data to prove his case about age of school entry.[1]

Farr, Fay, and Negley (1978) also explored the age issue. They found that in 1944 and in 1976, sixth and tenth graders scored about the same in Indiana, yet their analysis of census data showed that the 1970 sixth graders were 10 months younger, and tenth graders 14 months younger, than their 1940 counterparts. Farr, Fay, and Negley concluded that reading performance was better in 1976; they argued that, because of stricter retention policies, the students of 1944 had had more schooling for their age and thus had an advantage in the comparison (p. 53-4, 107).

Other information, however, indicated there were no differences in the amount of schooling of the two groups, or that 1976 students had more. Kindergarten was more common in later years, increasing the amount of schooling the 1976 students had had. Farr, Fay, and Negley also found that the length of the school year in 1976 averaged 9.4 months, while in 1944-1945 it was only 8.6 months (p. 89). This meant that the 1976 students had .8 month more schooling each year than those in 1944-1945 had. For sixth graders, this amounted to a 4.8-month advantage, while for tenth graders it was 8 months. Carried to its logical end, the comparison by amount of schooling should also involve the amount of reading instruction. Farr, Fay, and Negley estimated that 1976 students had 1.5 times as much daily and weekly reading instruction as those in 1944-1945 (p. 95, 97). Thus, not only were the 1976 students in school longer each year, but they also devoted more time to reading; their performance at the same level as the 1944-1945 students is an unimpressive record. It suggests that Indiana schools were teaching less effectively in 1976 than in 1944-1945. But the average 1976 student, controlling for age, still was reading better. The apparent paradox of poorer grade-level scores but higher-achieving students is the result of considering grade level first and then age level. As with the Gates' study, however, the reader should not make too much of these Farr, Fey, and Negley results.

The study had several problems. As noted, the authors failed to account for the demographic changes that occurred between testings. They also adjusted scores for age differences in a questionable manner. Because 1976 sixth graders, for example, were 10 months younger, they simply added 10 months to their scores. Such an adjustment presumed that the 1976 students would have gained one extra month for every additional month in school. This was unlikely. For one thing, it ignores ceiling effects. Sixth graders in 1976 were already near the top of the performance spectrum, so such large improvements probably were not possible. Furthermore, the sixth graders scored above the eleventh grade level, which is so far out of the tested grade range that differences in scores lose their significance.

Another problem that undermined the results was that the 1976 sixth-grade average was not based on the entire sample of students. Twenty percent of the sixth graders were excluded because they did not take all parts of the tests (Farr, Fay, & Negley, 1978, p. 39, 43). If these students were disproportionately lower-achieving, this would have artificially inflated the 1976 average. Finally, the one test which comes closest to what most educators now think of as reading comprehension, the paragraph comprehension subtest, showed that sixth graders in 1976 scored six months higher, but this was due only to the ten-month age adjustment. Had it been only a few months adjustment, they might have been outscored. Tenth graders showed no difference in paragraph comprehension performance even after Farr, Fay, & Negley made their age adjustment (p. 27). The best that can be concluded from this study, therefore, is that Indiana students in the 1970s probably were scoring about the same as they did in the 1940s.

The age issue may have implications for the findings of other studies. It suggests that students in the later groups likely would have been rated higher had their younger age been accounted for. Elligett and Tocco (1980), for example, in an equating study like Gates's, found that the reading achievement of 1979 sixth graders had declined five to ten months below that of the 1950s, but they failed to account for age differences. If the national age change was similar to Indiana's, there may have been no decline or even an improvement among students of the same age.[2] For then-and-now studies that showed no differences across time at a given grade level, age adjustments would mean that the later group actually outperformed the original group. This would resurrect some support for the claim of an achievement rise during the century, but we should not overgeneralize because local retention policies, which affect students' ages, may not have changed a great deal. Finch and Gillenwater, for example, found only a 1.56-month age difference between 1931 and 1948 sixth graders in Springfield, Missouri (Chart 1).

In addition to the four major problems--comparability, representativeness, subject matter coverage, and the age vs. grade-level criterion--researchers have left largely unexplored two testing factors that may cloud then-and-now comparisons. First, because the tests were developed at the earlier time, they favor the earlier group. Some of the vocabulary and subject matter may be foreign to students several decades later. In only one study did researchers attempt to eliminate outdated items (Cadwell and Courtis; Chart 1). Second, increased test experience, probably favoring later groups, could raise scores. Tuddenham (1948) cited this as one factor, albeit a minor one, explaining why World War II draftees outscored those in World War I. Whether these testing factors had a great enough impact to change findings, we do not know.

Given all of these difficulties, what can one conclude from then-and-now studies about reading trends? Taken at face value, especially if one considers the age factor, more of the tests showed gains than declines, while many others showed approximately equal performance rates. But few of the studies were nationally

representative. And the magnitude of the changes, up or down, was usually half a school year or less, which is well within the margin of error caused by the problems we have described.

The best national comparisons provided only limited evidence about achievement trends. The Gates (1961) study suggested that fifth and sixth graders in the late 1950s outperformed those of the late 1930s, while the Elligett and Tocco (1980) study suggested that 1950s sixth graders outperformed those of the late 1970s. But these were equating studies based on only a few school districts. Other grades in the Gates comparison showed no change, and Elligett and Tocco did not control for age differences.

There was no nationally representative then-and-now study at the high school level. Bloom's 1943-1955 GED comparison was the only national high school study but, as noted, it involved a select group of students. (It focused on high school seniors at a time when a majority did not graduate from high school; it included no black, technical, or vocational high schools.) Bloom found relatively small increases in several subjects.

The best state study compared Indiana students in 1944-1945 and 1976 (Farr, Fay, & Negley, 1978), but the age adjustments were arbitrary, social class changes were uncontrolled, and several subtests emphasized speed reading, short-term memory, and value judgments rather than reading comprehension. The results on paragraph comprehension showed little or no differences in performance between the two time periods.

Our educated guess is that school children of the same grade, age, and socioeconomic status have been performing at similar levels throughout most of the twentieth century (we consider the 1970s in detail below). But as social scientists we conclude that then-and-now studies are riddled with design and interpretation problems. The data can be used for any argument the authors support, but they can also be picked apart by critics.

2. Test Score Trends

Five major reviews of test-score trends were published in the middle to late 1970s (Armbruster, 1977; Cleary and McCandless, 1976; Copperman, 1978, 1979; Farr, Tuinman, & Rowls, 1974; Harnishfeger & Wiley, 1975). Most of these reviewers concluded that reading performance had improved steadily during the course of the century, but that since the mid-1960s, for all grades above third or fourth, it had been declining dramatically (Armbruster, 1977, p. 4; Copperman, 1978, p. 29). The declines were considered greater at higher grade levels (Armbruster, 1977, p. 40; Cleary & McCandless, 1976, p. 1; Copperman, 1978, p. 44, 49; Harnischfeger & Wiley, 1975, p. 115). There were two exceptions to these conclusions. Armbruster believed there had been drops in the early 1920s and 1940s as well as the steep decline after 1965. Farr, Tuinman, and Rowls (1974) suggested that

the post-1965 decline was slight (p. 139). (Cleary and McCandless did not comment on pre-1965 trends.)

a. The Great Test Score Decline

Knowing that a widespread concern about test scores had sparked recent educational reform proposals, we were interested in seeing how accurate the claims of a decline were. We also wanted to update these reviews and place their findings in a broader historical context. Our analysis of trends is based primarily on tests of language skills, particularly reading, because of our interest in literacy. We did not assess the evidence for mathematics or other subject areas. This is not a limitation, however, because data on language skills were a major portion of the evidence for the decline. Indeed, those who write about the test-score decline say that the drop was most steep in language skills (Cleary & McCandless,1976, p. 1; Harnischfeger & Wiley, 1975, p. 1).

In the language of the previous reviewers, flamboyant phrases like "massive decline" (Copperman, 1978, p. 29) and "unremitting fall" (Armbruster, 1977, p. 4) recur frequently, along with trumpet-like assertions: "These are the facts . . ." and "Beyond doubt . . ." (Harnischfeger & Wiley, 1975, p. 115, 116). Two reviewers were convinced that the social movements of the 1960s had led to the deterioration of the family and the schools. Their descriptions of the times read like caricatures. Frank Armbruster, of the Hudson Institute, claimed that during the 1960s, " . . . acceptance of improper behavior, and even some types of criminal acts, was becoming commonplace. Adults, even police, could be ignored with impunity" (p. 8-9). He further claimed that "moderates in our school system lost their prominence and some apparently injudicious activist educators gained influence" (p. 7). The activists, Armbruster believed, altered curriculum and teaching methods and opened the schools to the values of the slums (p. 8). Teachers yielded to students the "responsibility of determining when, if, and within a disturbingly questionable range, even what they would study" (p. 9). When achievement declined, sympathetic media and school boards let them fix blame anywhere but on the schools. "This sympathetic attitude may have been a 'spinoff' from the Kennedy era and later emphasis on the 'War on Poverty'" (p. 8).

Paul Copperman blamed the decline to a great extent on the open education movement and a breakdown in authority relations. He criticized the "undisciplined counter-culture approach recommended by Kohl and others of his ilk" (p. 64). He described Silberman's Crisis in the Classroom, which advocated more freedom and openness in education, as "one of the most damaging pieces of educational writing to have been published in the past twenty years," and claimed that Silberman's recommendations were widely adopted (p. 68). He linked the breakdown in authority relations to the "widespread and historical antipathy to authority that permeates American society" (p. 148). His examples of opposition to authority included the Populist movement,

the labor movement, the civil rights movement, and the hostility towards the post-Vietnam and post-Watergate federal government (p. 148). He also argued that "a plethora of local organizations have set up counterculture institutions, whose primary effect is to reinforce the antisocial tendencies of many young people" (p. 170). What were these institutions? Free health clinics, runaway centers, and alternative high schools. What were their goals? "These institutions attempt to convince young people and the society at large that irresponsibility, hedonism, and laziness comprise an acceptable alternative value system" (p. 170). Such charged and dogmatic rhetoric suggests to us that the evidence on test-score trends might not have been carefully analyzed.

This proved to be the case. The reviewers' first claim, that of a steady improvement to the mid-1960s, was based on an uncritical acceptance of then-and-now studies, test renormings (equating studies) of the early 1960s, and state trends in achievement tests. This allowed Copperman (1978) to talk about the "first major skills decline in American educational history" (p. 39). As we have seen, however, the quality of the then-and-now studies was poor and cannot substantiate claims of a general rise in American achievement.

A comprehensive study of the early 1960s renorming data (Schrader, 1968), which was cited by Armbruster and covered Copperman's evidence, provided only equivocal support for a mid-1950s to 1960s rise in achievement. Schrader reviewed equating studies of both elementary and high school achievement tests. For grades five through eight, he found there had been an increase of 8 percentile points- i.e., 58 percent of the mid-1960s students had outperformed the 1950s median student. In reading, the percentage was 57 percent. Schrader described this improvement as "small, but by no means trivial" (p. 18). He noted one special factor, however, that partly explained this finding. All four of the standardized tests he studied had excluded private schools in their 1950s testing, but included them in the 1960s. Given the selectivity and higher achievement of the private school sector, higher achievement was inevitable. The question was to what extent. Schrader cited a renorming study of high school students which showed a 3- to 4-percentile impact--a similar impact at the elementary level could account for as much as half the improvement.

At the high school level, Schrader found mixed results. The School & College Ability Test (SCAT) tests, comparing 1957 and 1967, showed a 9-percentile-point improvement, but private schools had been added to the second norming group. By contrast, the renormings of the Preliminary Scholastic Aptitude Test (PSAT), 1960 and 1966, and the Iowa Tests of Educational Development (ITED), 1957 and 1962, covered the same types of schools in both years and showed scores had remained stable.

The state data was too sketchy to have produced firm conclusions about an achievement rise. Copperman, for example, citing Farr, Tuinman and Rowls (1974), mentioned three states as demonstrating a

rise from 1957 to the early 1960s. Idaho and New Hampshire, however, provided data for only one grade, and West Virginia's data was largely irrelevant or contradictory. It didn't begin for most grades until 1964 or 1965, and statewide sixth-grade data, which began in 1963, showed immediate declines. Harnischfeger and Wiley cited as support a reference to data on only two states, Minnesota and Iowa, with only one grade apiece. The evidence for skill rise from the mid-1950s to the mid-1960s, then, is shaky.

The case for a major decline beginning in the mid-1960s rested on state achievement-test trends, college entrance exams, and renorming studies of achievement tests. Contrary to the impression reviewers created, there was a substantial amount of evidence, particularly at the high school level, that test scores had remained stable or experienced only a slight decline in the 1960s.

Farr, Tuinman, and Rowls (1974) recognized that their data prevented firm conclusions, but in general, the presentation of data in this area has been both sloppy and overconfident. Reviewers generally failed to consider that a decline could be due to a state's changing social composition, and they failed to demonstrate that the states and districts they were examining were representative of the nation. Armbruster (1977), for example, claimed his conclusions were based upon data from school systems containing half the country's population and half its elementary and secondary students (p. 4, 42). Many of the 22 states, however, did not send usable data; eight provided only one year's tests. Only eight provided three or more years' results, and those typically for only a few grades.

For any given time period, data from only a few states were reported. Armbruster had 1960s data for only five states. For California, he only had third-grade data, which was irrelevant since the decline was supposed to occur in higher grades. That left four states to represent the nation: Hawaii, Iowa, New York, and South Dakota. The four comprise about one-ninth of the population. Furthermore, the data from these four states did not support the mid-1960s decline hypothesis. Hawaii's sixth-, eighth-, and tenth-grade reading scores were stable in the 1960s and didn't begin dropping until the 1970s. The same was true for Iowa's fifth graders. New York's sixth grade and Iowa's sixth- to eighth-grade reading scores dropped somewhat in the late 1960s but were level in the early 1970s. South Dakota's ninth- and eleventh-grade reading scores remained stable from 1963 through 1969. This hardly constitutes a "almost unremitting fall" beginning in the mid-1960s.

Copperman also displayed more zeal than evidence for the decline hypothesis. He claimed that the pattern of decline showed up in the "records of virtually every state office of education that makes its data available for analysis" (p. 44). As we have seen, this is quite mistaken. In addition to the stated examples, New Mexico's and Mississippi's fifth and eighth graders, and Michigan's fourth and seventh graders all had stable scores in the early 1970s (Armbruster, 1977, p. 229-36, 251). Copperman also claimed that five of the six

school districts Farr, Tuinman, and Rowls (1974) studied "reported steady declines between the mid-1960s and early 1970s" (p. 49), but two actually had stable scores in the 1960s and a third reported r, scores after 1961. The sixth district, he claimed, had "almost unusable data" and reported no change in scores, but in fact it provided five years of data for each of three grades and showed stable or rising scores.

Harnischfeger and Wiley cited Minnesota and Iowa data as proof for the decline, yet ignored data from states that showed no decline. Alabama and South Dakota, for example, also had comprehensive testing of ninth graders and juniors in the 1960s, but showed no declines.

The SAT decline is the most widely publicized piece of evidence, but compositional changes in test takers played a major role in its decline. Expanded educational opportunities in the 1960s resulted in more black ethnic minority, lower ability, and poor students taking the tests and going to college. Test taking shifted from students going primarily to selective liberal arts colleges to those attending less selective universities, two year colleges, and technical schools. The most thorough analysis of the SAT decline found that compositional changes accounted for between two-thirds and three-fourths of the 1960s decline. The Advisory Panel (1977) placed the skill decline primarily in the 1970s (p. 18).

Several other points should be made. Because the SAT is often described as having "peaked" in 1963, one might get the impression that the scores were steadily improving until then. However, SAT Verbal scores with minor fluctuations, were basically stable from 1952 to 1963.

Second, the SAT, as a general aptitude test, is not designed to measure high school students' achievement. By contrast, six major ETS Achievement Tests, including English Composition, showed increases from 1967 to 1976 at the time of the greatest decline in SATs. The students taking these tests had lower SAT scores than their predecessors but nevertheless had higher achievement scores. The Advisory Panel speculated that the SAT and the achievement tests were changing in their relevance to high school education (Advisory Panel, 1977g, p. 22-3).

Third, tests with self-selected, changing compositions, such as the SAT and ACT, are not nationally representative. As Schrader (1968) noted about the SAT, "High school seniors taking the Scholastic Aptitude Test are not representative either of high school seniors generally or of high school seniors planning to enter college" (p. 5). Trends from such testing programs, therefore, can tell us little about how the average high school student has been doing or what the average national trends in reading and literacy were.

Test renorming data, by contrast, involves representative samples, and can provide that information. Yet the high school data generally fail to support the reviewers' thesis. The Iowa Tests of Educational

Development, in national norming, showed steady increases in reading for ninth through twelfth graders from 1957 to 1962 to 1971 (Iowa Tests of Educational Development, 1971). The PSAT renormings in 1960, 1966, and 1974 showed general stability in English scores, with a slight increase in the first period and a slight decrease in the second. Mathematics scores also were stable, but with opposite trends (Breland, 1976). The SCAT tests, as mentioned previously, showed an increase of 9 percentile points for ninth through twelfth graders from 1957 to 1967 (Schrader, 1968). Project Talent data showed high school juniors had "slight gains" in reading comprehension scores from 1960 to 1970 (Flanagan, 1976, p. 9-12).

The ITED results illustrate the problems of using data from testing programs with changing, nonrepresentative samples. The American College Test (ACT), which is based directly upon the ITED (Harnischfeger and Wiley, 1975, p. 34), showed declines, whereas the ITED showed improvement. This is another case of the reviewers' selective use of evidence. Both Copperman and Harnischfeger and Wiley presented ACT and Iowa ITED results, which showed 1960s declines, but ignored the more meaningful national norms data, which showed improvement.

Renorming data from the middle grades were more supportive of the reviewers' claims of decline in the early 1970s, but there still was conflicting evidence. Stanford Achievement Tests for 1964 and 1973 showed marked declines--eighth graders dropped eight months and seventh graders five months (Copperman, 1978, p. 43). The Comprehensive Tests of Basic Skills showed larger declines between 1968 and 1973, with eighth and tenth graders falling about a year in total reading, and ninth graders dropping 3 months (Harnischfeger & Wiley, 1975, p. 58-59). The Iowa Tests of Basic Skills renormed in 1963 and 1970 showed mixed results and modest declines: seventh and eighth graders declined 1.4 and 2.1 months in reading, showed no change in language arts, and improved 1.0 and .4 months in vocabulary (Iowa Testing Program, n.d.).

Results in the opposite direction also were found. The 1962-1971 Science Research Associates comparison showed a general improvement across several subjects for grades two through eight, with reading scores remaining stable (Cleary and McCandless, 1976). The SCAT scores for grades five through eight showed a 12-percentile rise from 1957 to 1967, indicating that scores continued to improve well into the 1960s (Schrader, 1968). The Metropolitan Achievement Tests showed an increase from 1958 to 1970, with ninth graders up five months in reading, though part of the increase was due to the inclusion of private schools in the second norming (Test Department, 1971). Armbruster explained these discrepancies by arguing that tests renormed in the early 1960s and then in the early 1970s missed the peak in the late 1960s and the subsequent major decline from that peak. Although this is a possibility, the ITBS, PSAT and SAT data suggest there may have been only a slight decline. Furthermore, the studies that supposedly framed a peak indicate that the decline was not a major historical setback: 1970 and 1971 students still

outperformed those of 1958 and 1962 (MAT, ITED results), and, even as late as 1974, high school students still were doing as well as they had in 1960 (PSAT results). The evidence for a decline beginning in the 1960s--even the late 1960s--is mixed at best.

The period from 1970 to 1978, however, shows a dramatic downturn on most major tests. Two reviews of this period reached opposite conclusions (Copperman, 1979; Munday 1979a, 1979b). Munday, general manager of the Test Department of Houghton Mifflin, presented an optimistic description, uncritically describing several then-and-now studies, and downplaying the 1970-1977 ITBS decline even though it was greater than that in the 1960s. Copperman, quite rightly, faulted Munday's use of this evidence, and presented renorming data from the 1970-1978 MAT and the 1970-1977 CAT that showed declines of about a year among eighth through twelfth graders. Renorming data we collected for this time period for the SRA, CTBS, STEP, and ITED tests supported Copperman. SRA total reading scores in 1978 fell below 1971 levels, although, as the test publishers pointed out, "the decreases were relatively small in grades five, seven, and eight, and virtually zero for grade six" (Science Research Association, 1980, p. 3). Tenth graders showed a 12-percentile drop; ninth, eleventh, and twelfth graders showed 7 to 8 point drops (SRA, 1980, p. 5). The 1970-1978 STEP comparison showed eighth graders dropping 20 percentile points and twelfth graders, 13 percentile points (Educational Testing Service, n.d., p. 101). Seventh through twelfth graders on the CTBS total reading scores dropped from three months to about a year, depending on the grade. Tenth and eleventh graders showed smaller drops (CTB/McGraw-Hill, 1982a, p. 59f; CTB/McGraw-Hill, 1982b). ITED scores for tenth, eleventh and twelfth graders, after improvements from 1957 to 1971, dropped back in 1978 to around or slightly below 1962 levels, depending upon grade (Science Research Associates, 1978).

We believe the timing of the decline is an important issue. Many of the explanations that have been offered, such as social protests, challenges to authority, and curriculum upheavals, depend on a mid-1960s decline. But since the decline actually occurred in the 1970s, it is harder to blame it on the social movements of the 1960s. Even if we agreed that widespread unrest affected most high schools, student protest, such as the anti-war marches, occurred most often between 1968 and 1971. A 1977 twelfth grader, however, would have been in the third through sixth grades during these years. He or she would have been in high school from 1974 to 1977, hardly a time of protests or educational experimentation. Eighth and ninth graders would have been in their first few years of school during the time of unrest and, like the twelfth graders, would have received the bulk of their education in the 1970s. This makes the decline much harder to explain. The 1970s were a time of educational retrenchment with a renewed emphasis on the basics, the spread of statewide competency testing, and actions to end social promotions. We can hardly blame the decline on "activist educators" who, frustrated at their inability to change the schools, had effectively abandoned their efforts (See, e.g., Holt, 1976).

A careful assessment of the timing of social and educational changes and how they related to the test score decline still is needed (See Stedman and Kaestle, 1985).

b. How Bad was the Decline?

Critics presented test score data in various statistical guises, many of them quite dramatic. Copperman argued that the average (50th percentile) high school student of the late 1970s ranked at only the 39th percentile of his 1965 counterparts (p. 38). The SAT verbal drop was almost a half of a standard deviation, a big shift in the distribution of scores. Several tests showed that eleventh and twelfth graders lost a year or more in measured reading ability during the 1970s (Bode, 1981b, p. 4; CTB/McGraw-Hill, 1974b; CTB/McGraw-Hill, n.d.). If one focuses on the SAT's and on standard deviations, the test score declines do appear substantial, but there are several problems with these stark descriptions.

i. Who is Taking the Tests?

First, the figures are not unadjusted for changes in the composition of test takers. We must distinguish between decreases due to a widened or changing testing pool and those due to a general decline in students' skills. Trends on college entrance exams such as the SAT are especially difficult to interpret because they measure the performance of a self-selected group of students whose composition changes annually. Even in the 1970s, changing composition accounted for a substantial portion of test score decline. More students from characteristically lower-scoring groups continued to take college entrance tests, including minority students, those intending to pursue "career" majors as opposed to "arts and sciences" majors, women (who score lower, on average, in math), and students going to two-year community colleges and four-year public universities as opposed to highly selective liberal arts colleges. The College Board's Advisory Panel estimated that between 20 and 30 percent of the SAT decline in the 1970s was still due to such changes (Advisory Panel, 1977).

Changes in family size also had an impact, although not as great as some once claimed. First- and second-born children score higher than average, later-borns score lower. In a background report for the Advisory Panel, Breland (1977) estimated that such changes accounted for about 16 percent of the verbal SAT decline between 1964 and 1976. A recent study of the period from 1971 to 1977 produced a 4- to 9.4-percent estimate, although the effect could run higher (Zajonc and Bargh, 1980).

Combining the estimated effects of birth order and the estimated effects of changing characteristics of test takers means that at least 24 to 40 percent of the 1970s SAT decline was not due to changes in the schools' effectiveness in skill training. Many other nonschool factors also may have contributed, as we shall discuss.

Although reported trends on standardized test batteries like the CAT and the ITED are an improvement over SAT tests in that they are based on nationally representative samples, they too can be affected by nationwide changes in immigration, dropouts, or birth order. These effects may have been as substantial in the 1970s as those affecting the SAT. A falling black dropout rate and increased Asian and Hispanic immigration increased the percentage of minority students in our high schools from one-sixth to nearly one-fourth (National Center for Education Statistics, 1979, p. 17: U. S. Bureau of the Census, 1981a, p. 35). Such changes likely contributed to lowered scores (p. 35). Birth-order effects also contributed, and, for unknown reasons, are greater on standardized achievement batteries than on the SAT (Zajonc and Bargh, 1980). Another contributing factor, albeit small, was the changing age of students. Due to earlier school-entering ages and more automatic promotion policies, students coming into a given grade were increasingly younger. Researchers who have studied long-term test score trends have stressed the necessity of accounting for differential maturity of students (Gates, 1961; Farr, Fay, & Negley, 1978). Adding this factor to the birth-order and composition factors mentioned above suggests that demographic changes may account for between 30 and 50 percent of the 1970s achievement test score decline at the high school level. Critics, however, tend to assume that virtually all of the 1970s declines were due to instructional failure (Brimelow, 1983; Ravitch, 1985).

Copperman's description of the test score decline, for example, presumes that the skill decline began in the mid-1960s and dropped steadily thereafter, thus ignoring a huge compositional effect. His claim of a drop to the 39th percentile was based on an estimated 2.5 percent standard deviation (SD) drop per year from 1965 to 1978. Thirteen years of such a drop yielded an overall 32 percent drop, or 11 percentile points. Since this figure was unadjusted for compositional effects, and since the real skill decline occurred primarily in the 1970s, the overall decline was much less than Copperman claimed. Figuring 1.3 to 1.8 percent of a SD per year for seven years (the 1970s decline minus the estimated compositional effect) produces a total decline of 9.1 to 12.6 percent of a standard deviation during the 1970s. This amounts to a drop of only 4 to 6 percentile points, to the 44th or 46th percentile level. Similarly, adjusting grade-level scores for compositional effects reduces apparent declines by up to one-half. Thus, on tests in subjects that showed as much as a whole year decline, the adjusted score would be a half year.

ii. How Big was the Skill Loss?

A second problem is that critics rarely relate test score declines to actual skills. What was the difference in skills between students who scored one-half a grade level lower than another earlier group? What specific tasks could students no longer do? Standardized tests are constructed in such a way that small shifts in test performance produce large changes in percentile and grade equivalent rankings.

The decline thus sounds large when described in grade equivalent and percentile terms, even though the actual performance drop could be quite small. Oscar Buros, the late editor of the Mental Measurements Yearbook, argued against the reliance on normed scores for interpreting educational achievement. He believed that grade-level equivalents give people a vague and misleading impression of skill levels. He advocated getting closer to actual performance by using the percent of items a student knows (Buros, 1978, p. 1976). If we describe student performance in this way, we get a different sense of the magnitude of a skill decline. On many standardized tests, differences between grades amount to only a few percentage points, particularly at the high school level. On the SRA, ninth through twelfth graders' reading scores dropped a half to a full grade level between 1971 and 1978, but this corresponded to a small drop in the percent of items answered correctly. Twelfth graders, for example, had dropped a whole grade level in reading, but this was only from 72 to 68 percent correct, or a 4 percent drop. Mathematics declines were similar (Bode, 1981b, p. 4; Bode, 1981a, p, 33). Furthermore, these figures are unadjusted for compositional changes, so the actual skill decline among similar students was smaller yet. Several of the NAEP tests showed declines, but these also showed only small drops in performance. Between 1970 and 1980, for example, in inferential reading comprehension, seventeen-year-olds dropped from 64 to 62 percent, a 2 percent drop; thirteen-year-olds went from 56.1 to 55.5 percent, or only a .6 percent drop. In math, from 1973 to 1982, seventeen-year-olds dropped from 52 percent to 48 percent correct, or only 4 percentage points, while thirteen-year-olds dropped only 2 points. In science, from 1970 to 1977, seventeen-year-olds dropped only 4.7 percentage points, thirteen-year-olds 2.4 points. Other tests may show larger declines, but the point is the same: when they are expressed in terms of percent correct, the declines do not seem as great as when they are expressed in grade levels.

Another way of assessing the decline is to ask at what percentage of their former skill levels are students now performing? On the NAEP tests, for example, students were performing at 97 percent of their former levels in inferential comprehension, 92 percent in mathematics. High school students on the SRA were reading at about 95 percent of their former levels. Some may believe that even a five percent decline in skill level is worrisome. The Nation At Risk report argued that such skill declines threatened our very economic security as a nation. But what are the demonstrable educational and economic ramifications of test score declines?

In fact, the statistical links between academic success at one level and the next are relatively weak, as are those between academic performance and economic performance. The correlation between SAT scores and freshman grades, for example, is about .40 (Advisory Panel, 1977). The decline of 20 percent of a standard deviation in SAT scores in the 1970s (which accounts for compositional changes) would translate into a drop of only 8 percent of a standard deviation in freshman grades. The correlation between achievement test scores and measures of job proficiency is around .25 (Olneck, 1984, citing

Schmidt and Hunter), so the drop of 12.6 percent of a standard deviation in high school standardized achievement test scores during the entire 1970s would translate into a drop in job performance of only 3.1 percent of a standard deviation. Furthermore, as Olneck points out, new workers comprise only a small proportion of the entire work force, so recent declines in productivity can hardly be linked to recent changes in test scores. Even ten years of reduced skills among all new workers would affect only about one-fourth of the active work force. The combined effect of the modest correlations makes the critics' attempts to link declining test scores with changes in industrial productivity downright silly.

All of the above discussion presumes that the tests are completely valid measures of academic skills. Yet performance on the tests reflects an undetermined proportion of other factors, such as motivation and test-taking skills. If these extraneous factors could be accounted for, the actual skill decline that occurred during the 1970s might have been even smaller than that described above. (See Appendix B for a more detailed discussion of the decline in terms of the tests themselves.)

iii. What's Wrong With These Particular Measures?

A third set of problems arises from the deficiencies of the particular measures most often used as evidence of the skill decline: college entrance exams, national achievement test renormings, trends in individual states' achievement test scores, and the NAEP tests.

College entrance exams provide an annual barometer of performance changes, but as noted above, the composition of the test takers changes annually, and thus it is imperative to adjust for the compositional effect. They are further limited as a national barometer because they primarily apply to the college-bound student rather than to the average student.

Standardized achievement test trends are derived from periodic renormings (five to seven years apart, generally) carried on by publishers when redesigned tests are introduced. The new tests are given to a nationally representative sample of schools. Performance on the new test, and thus current national performance, is linked to the old results through the use of "equating studies," in which samples of contemporary students are given both the new and the old tests. Problems with equating abound. Often two different contemporary groups are given the different test versions; sometimes only portions of the two test versions are administered. Also, the equating samples are usually not representative of the nation, often involving a few school districts or a small fraction of the national norming sample. Even some test publishers warn against the use of renorming data to infer national trends. Metropolitan Achievement Test publishers stated in 1978 that "these data are not appropriate for making generalizations concerning changes over time in relative achievement of American students in the basic skills areas" (The

Psychological Corporation, 1978, p. 1). More recently, these same publishers warned that there is a "popular misconception about changing norms: that a change in the norms from an old test to a new test reflects a change in the ability of the reference groups over time." On the contrary, "there are simply too many complex and confounding variables to make a sound judgment about performance over time" (Test Department, 1983, pp. 1, 2). They cite changes in the national samples of students and the changing relevance of the test content as factors that confound any generalizations. Critics who use renorming evidence to describe national trends typically do not discuss these serious limitations. (See Appendix C for a more detailed discussion of the problems with equating and norming.)

State trends on achievement tests are problematic because the data for the 1960s and 1970s were limited to a handful of states. Iowa is often used as a barometer, but because it is predominately rural and has few minority students, performance there can hardly be said to represent the nation. Even among similar states, the trends are ambiguous; for each state the critics cited as showing declines for the 1960s and early 1970s, there was a similar state that experienced no decline. Alabama and South Dakota high school scores were stable in the 1960s, for example; Mississippi eighth graders and Michigan seventh graders had stable scores in the early 1970s. (See Armbruster, 1977; Farr, Tuinman & Rowls, 1974.)

The NAEP tests probably are the best indicators of national trends. They are drawn from nationally representative samples like the standardized tests, but trends on common items are reported so there is no problem equating old and new versions of the tests; renorming studies are therefore unnecessary. Furthermore, NAEP results are reported by racial, geographical, and SES groups, so that, unlike results from recent standardized tests, trends by subgroups can be followed. Test items are also made public, so schools can independently examine what kind of skill is being measured. Like the other tests, however, NAEP tests are limited in that the test items may not reflect what is taught in schools or may relate to only a small portion of it.

iv. Did All Test Scores Decline?

A fourth flaw in the critics' argument is that they paid little attention to the contradictory evidence of the 1970s The National Assessment of Educational Progress showed that thirteen- and nine-year-olds maintained their reading scores and nine-year-olds improved theirs in 1970, 1975, and 1980 testings. Seventeen-year-olds slipped in inferential comprehension, but the drop was minor--from a 1970 level of 64 percent correct to 62 percent in 1980. Furthermore, this decline was not universal. The only region experiencing statistically significant declines in inferential skills was the Northeast; boys showed such declines, while girls did not. Blacks' scores did not fall off significantly. Some commentators argue that the NAEP reading tests are easier than the standardized high school

achievement tests and test lower-level skills (Harnischfeger & Wiley, 1975, p. 68, 70; Armbruster, 1977, p. 67). In fact, the percent of questions missed by seventeen-year-olds on NAEP is comparable to that on other achievement tests, and the proportion of the test devoted to inferential skills is also similar (See Appendix D). NAEP functional literacy results showed that seventeen-year-olds' skills remained the same between 1969 and 1979. Rhetorical skill on narrative tasks rose during the period, as did cohesion scores. A comparison of results on the Metropolitan Achievement Tests and the Stanford Achievement Tests showed a five- to six-month gain in reading and a six- to twelve-month gain in math for grades seven through ten from 1973 to 1978 (Psychological Corporation, 1978). The ACT natural science scores have remained stable over the past two decades. Flynn (1984) reviews sketchy evidence to suggest that IQ scores were stable or rising from 1972 to 1978. Finally, those who cite the SAT's as evidence for the decline rarely mention data from ETS's Achievement Tests. Scores in English composition, biology, chemistry, physics, French, and Spanish showed increases from 1967 to 1976, the time of the worst SAT decline. Thus, although the students who took these tests in 1976 had lower SAT scores than their predecessors, they outscored them on the achievement tests. The evidence about a massive, consistent skill decline, then, is much more mixed than the achievement critics' claim, and the contradictory evidence is not easily explained.[3]

We also must put the decline into its historical context. When interpreted in terms of the tremendous gains that have been achieved in educational attainment during the past several decades, the decline seems much less substantial. The median educational level of the adult population (25 years and older) rose two full years between 1960 and 1980, from 10.5 to 12.5. Between 1940 and 1980, the median level rose nearly four full years, from 8.6 to 12.5 (Grant of Eiden, 1982, Table 10, p. 16). A skill decline of at most half a year and only in certain subjects and on certain tests is minor compared to these tremendous gains. (The median also does not indicate the tremendous gains that occurred in higher education. The percentage of the population who had completed college rose from 4.6 to 17.1 percent during those same four decades. The percentage who had completed high school rose from 24.1 percent to 68.7 percent.) Since the half-year decline figure comes from standardized tests that do not directly measure the high school curriculum, and represents only a small percentage decline in skills, the decline seems even less significant when compared to the impressive gains in the number of years of schooling. Two to four years of additional schooling for the average adult represents a substantial increase in the amount of knowledge and ideas encountered and skills developed.

Finally, we must recognize that the test-score decline has ended. Most recent renormings of the major standardized test show this. On the 1982 Stanford Achievement Tests, for example, eleventh graders scored four percentile points higher in mathematics and 10 percentile points higher in reading than their 1973 counterparts. Eighth graders were up six percentile points in math and seven in reading. Students in other grades showed similar improvements, across subjects as

diverse as science and spelling. The Iowa Tests of Basic Skills, given to third through eighth graders, also showed a general improvement, with scores rising dramatically between 1977 and 1984. Preliminary analyses of the 1984 results, for example, indicate that composite scores are at an all-time high for most grades (Hioronymus, 1985). The Tests of Achievement and Proficiency also show that high school students have improved their performance in most grades and subjects in recent years.

National Assessment of Educational Progress (NAEP) results show that thirteen-year-olds' mathematics scores rose between 1978 and 1982; nine- and seventeen-year-olds' scores remained stable, ending their previous decline (NAEP, 1983). Seventeen-year-olds' reading scores rose between 1980 and 1984 (NAEP, 1985). Results for the state of Iowa show steady increases in recent years. Eighth graders, for example, improved two months in mathematics and reading during the past four years; high school students' scores also have risen, although not as much.

The score decline on college entrance tests has also bottomed out. American College Test (ACT) scores in English, social studies, and science, for example, have been stable for many years. Although ACT mathematics scores continued to drop until 1983, they have since risen. Scholastic Aptitude Tests (SAT) math scores have increased a few points in recent years. SAT Verbal scores have been fluctuating up and down a point or two for several years. The test score decline is over.

Although we dispute the critics' claim that there was a massive score decline and that the schools were to blame, we are not letting the schools off the hook. Historically, schools have had trouble educating a substantial portion of their students, whether this is judged by essay writing, mathematical computation, or foreign language skills, or by high school graduation, employment, or yearly retention rates. Minorities and the poor have never done as well as they should have. Students' higher-order skills, such as problem solving, typically have been underdeveloped. A serious reassessment of the purposes, organization, and control of schooling is in order.

We are particularly concerned that the recent increase in test scores has been brought about, in part, by an excessive focus on testing. Many schools are spending too much time "teaching the test." This narrows the focus of the curriculum and can undermine the development of good reading skills and critical thinking. So while trends in reading scores may be in doubt, the need for improved reading and writing skills is not. (For an elaboration of this argument see Stedman & Kaestle, 1985.)

C. Horizontal Dimension: Functional Literacy

Functional literacy has been measured in four ways: by educational attainment, by tests of applied reading skills, by

comparing the reading grade level of a population to that of frequently encountered reading materials, and by job literacy measures. In describing these four approaches, we pay attention to how functional literacy was defined and how accurately it has been measured. We are interested in two questions: How extensive is functional illiteracy in the United States today and how has this changed over time?

1. Educational Attainment

When the Civilian Conservation Corps coined the term "functional literacy," in the 1930s, they defined it as the reading ability of someone with three or more years of schooling. The level of education necessary to be considered functionally literate has been steadily rising. During World War II, the Army used the term to refer to a fourth-grade educational level and, until manpower demands became overwhelming, rejected recruits who had less schooling (Ginzberg & Bray, 1953; U.S. Census, 1948). In 1947, the Census Bureau applied the term "functional illiterates" to those with fewer than five years of schooling and ceased asking their questions about crude literacy to those with more schooling (U.S. Census, 1948, P-20, No. 20, p. 3). In 1952, the Bureau raised the functional literacy level to the sixth grade (P-20, No. 45). By 1960, the U.S. Office of Education was using eighth grade as the standard (Fisher, 1978, p. 38; Harmon, 1970, p. 226-243). Finally, by the late 1970s, some noted authorities were describing functional literacy in terms of high school completion (Hunter & Harmon, 1979, p. 27; Carroll & Chall, 1975, p. 8).

In each case of successively steeper criteria, educational attainment was considered to be a proxy measure of an individual's ability to function in society. The connection rests on the assumption that people who reach a certain grade have learned to read at a certain minimal level. This is a shaky generalization when applied to individual cases, but it seems reasonable to assume that a substantial increase in school attainment would raise the average reading ability of a population. Data on educational attainment comes from the Census Bureau, which since 1940 has routinely determined the years of schooling of different age groups in the U.S. population. We have retrojected this data to 1910; the resulting data show that functional illiteracy as measured by educational attainment has greatly diminished. In 1910, 23.8 percent of the population had completed fewer than five years of schooling, while in 1980, only 3.3 percent had. These rates were for those 25 years and older. Among those aged 25 to 29 in 1980, the rate was only .7 percent, indicating we have virtually eliminated functional illiteracy by this standard among the younger generation. For the eighth grade standard, the drop has been from about 45 percent of the population in 1910 to 9.7 percent in 1979. Only 2.8 percent of the 25 to 29-year-olds in 1979 had not completed eighth grade (Folger & Nam, 1967, p. 133; March 1979, P-20, No. 356, p. 11). For the high school completion standard, progress has been less marked. In 1910, 86.5 percent of the population had not completed high school, while in 1980, 31.3 percent

had not. Among 25- to 29-year-olds, the figure was 14.2 percent (Digest of Educational Statistics, 1982, Table 10, p. 16).

Like the crude literacy data, this data was self-reported, but it is probably more accurate. There is less stigma to revealing the number of years of schooling one has completed than to confessing outright illiteracy. Furthermore, in 1960, the Census Bureau did follow-up interviews with a sample of census respondents and found that only 2 percent had overreported their grade levels (Folger & Nam, 1967, p. 212). In another investigation there were various discrepancies, but the self-reported Census data was found to parallel the enrollment statistics gathered from schools by the U.S. Office of Education (Folger & Nam, 1967, pp. 223, 226).

Although the data may be accurate, the trends as presented are misleading. Applying the same standard over the course of the century fails to account for changes in the content of schooling over time or for increases in societal literacy demands. An eighth-grade education in 1980 may not mean the same as it did in 1910; furthermore, the literacy skills demanded by jobs and modern living are likely more complex in 1980 than in 1910. If we accept the proposition that literacy demands have risen continually, one might portray twentieth-century literacy development as a situation in which school attainment levels were rising but functional literacy was falling. This would be the case, for example, if we applied to the data on educational attainment the previously described judgments about what educational level constituted functional literacy in different time periods. In 1930, about 88 percent of the population had a third-grade education or more; in 1950, 88.9 percent had a fifth-grade education or more; in 1960, 78 percent had at least an eighth-grade education; and in 1980, 68.7 percent had completed high school (Folger & Nam, 1967, 133; Digest of Educational Statistics, 1982, Table 10, p. 16). If we use a more conservative, and perhaps more realistic, standard for 1980--an eighth-grade education--functional literacy has increased slightly in the past 50 years--from 88 percent (completing third grade) in 1930 to 90.3 percent (completing eighth grade) in 1980.

There are three limitations to this approach to estimating functional literacy. First, the line between functional literacy and illiteracy is somewhat arbitrary, and authorities often do not justify their particular cutoffs. Drawing the line at a different point can have a huge effect on the percentage of the population considered functionally illiterate. Furthermore, drawing a line means that functional literacy is conceptualized in dichotomous terms: a person is either functionally literate or functionally illiterate. This makes little sense. A person who has completed eighth grade does not suddenly become able to function effectively in society while the person with only seven years of schooling is unable to cope. There are gradations in the ability to function, and a person's performance varies by setting and task. Arbitrary and dichotomous definitions of functional literacy also are problems with more direct tests of reading ability, as we shall see. Using school attainment as a

measure of functional literacy has the further problem, however, of equating schooling with learning. Many children, obviously, do not perform at grade level, so the number of people who are functionally illiterate may be considerably greater than the traditional school attainment figures suggest. A proper assessment of functional literacy requires testing the population on functional literacy tasks.

2. Tests of Functional Literacy

The second approach to functional literacy is to identify the skills necessary for functioning in society, develop a test that measures such skills, and then administer the test to a representative sample of the population.

The pioneering effort in assessing functional literacy was made by Guy Buswell at the University of Chicago in 1937. He administered a test to 897 Chicago-area adult residents across occupational, educational, and age distributions. The test consisted of five sections: identifying product prices from a mail-order catalog; finding phone numbers in a directory; matching movies and theaters from newspaper ads; answering questions about a series of reading passages; and a multiple-choice vocabulary test. He found that the average adult answered correctly about 51 of the 93 items. These results varied by education. Those with six or fewer years of schooling answered only 25 correctly on average; those with some college answered 65 correctly. The results, as might be expected, also varied with reading habits. The 100 highest-scoring adults regularly read magazines and books, while the 100 lowest read only occasionally.

After Buswell's effort, no similar work was done until the 1970s, when five major studies were conducted (Harris and Associates, 1970, 1971; Gadway and Wilson, 1976; Murphy, 1975; Adult Performance Level Project, 1977). (The National Assessment of Educational Progress is currently investigating the functional literacy of young adults. Results should be available this spring. See Kirsch, 1985.) Investigators tested skills such as map reading, dictionary use, deciphering help-wanted ads, using train schedules, and reading product labels. The functional illiteracy rates from these studies ranged from 3 percent to 54 percent.

The variations were due to differences in the tasks tested and the level at which functional literacy was set. The lowest rate, 3 percent, which corresponded to 4.3 million functionally illiterate adults, was derived from a test in which people filled out forms such as driver's licenses applications (Harris, 1970). A relaxed standard which allowed individuals to make several mistakes before they were classified as functionally illiterate, was set. The highest rate, 54 percent, was derived from a test of functional "competency," not just literacy, and the 54 percent included "marginally competent" adults, as well as those deemed functionally "incompetent" (Adult Performance Level, 1977). It led Secretary of Education Terrence Bell to testify

to Congress in 1982 that as many as 72 million adults were not literate enough to function effectively in society (Bell, 1982).

The seriousness of the functional illiteracy problem obviously depends upon which estimate one accepts. Part of our purpose in reviewing these studies is to determine the validity of these estimates. Critics have argued that the tests did a poor job of measuring functional literacy and that the estimates were greatly exaggerated. Some critics recalculated the percentage of functional illiterates and revised the estimates downwards. Other critics, however, argue that functional illiteracy rates were underestimated. In what follows, we first describe the tests and their findings, then explore the major criticisms, and finally review evidence concerning historical trends.

a. The Tests

The first study was conducted by Louis Harris and Associates for the National Reading Council, a group of 40 appointed by President Richard Nixon in 1970, was called the Survival Literacy study. It tested the ability of those sixteen years and older to read, understand, and fill out application forms. Of the five forms, one requested personal identification information, while the others were applications for a bank loan, a social security number, a driver's license and Medicaid. The researchers established three criterion levels: 70 percent correct, 80 percent correct and 90 percent correct. A correct answer required respondents to provide information appropriate to the request (Chart 4 shows Form III, Application for Driver's License). Those who answered correctly less than 70 percent of the time, the "low survival" group, were considered functionally illiterate. Those who answered correctly 90 percent or more of the time, the "likely survival" group, were considered functionally literate. In between were the "questionable survival" (70 to 80 percent correct) and "marginal survival" (80 to 90 percent correct) groups. The intermediate categories softened the problem of arbitrary and dichotomous definition.

Harris and Associates found that, on average, 3 percent of the population were functionally illiterate on a given form, that is, scored below 70 percent correct. The percentages ranged from less than .5 percent on the public assistance form to 9 percent on the Medicaid form (see Chart 5). In absolute numbers, this meant that 4.3 million people sixteen years and older were functionally illiterate. As can be seen from the chart, many more people didn't reach the functional literacy level of 90 percent correct. On average, 13 percent fell short of this level, for a total of 18.5 million people who were in the low, questionable, or marginal survival categories.

The second study, also conducted by Harris and Associates for the National Reading Council, was called the 1971 National Reading Difficulty Index and was similar to the Survival Literacy Study. Researchers asked a national sample of people sixteen years and older

to fill out an application form derived from various official forms such as passport, driver's license, and credit card applications. Researchers also tested the population's ability to read three types of materials: telephone dialing and rate information, classified housing ads, and classified employment ads (see Chart 6 for examples).

For each portion of the test, researchers reported the percentage of people who answered correctly a given number of items (see Chart 7).

After weighing the items for difficulty, Harris and Associates found that 4 percent of the sample answered correctly less than 80 percent of test questions. They concluded that these people "suffer from serious deficiencies in functional reading ability" and that their ability to survive in practical reading situations was in "serious doubt." This referred to 5.7 million people aged 16 years and over. Another 11 percent (15.5 million) people scored below 90 percent correct, and the researchers concluded that these individuals would need to make a "serious effort" to handle real-life reading situations.

The third and fourth studies examined everyday reading tasks with particular attention to the different formats of such tasks. The third study was conducted by the National Assessment of Educational Progress for the U.S. Office of Education's National Right to Read Effort (Gadway & Wilson, 1976). NAEP's test, the Mini-Assessment of Functional Literacy (MAFL), assessed the ability of 17-year-old students in 1971, 1974 and 1975 to read such everyday formats as word passages, reference materials, and graphic materials, including charts, maps, pictures, coupons, and forms. The Adult Functional Reading Study (AFRS), organized by the Educational Testing Service, tested the ability of the population aged 16 and older to deal with such formats as advertisements; legal documents; instructions, as in recipes and manuals; and listings, such as telephone directories and train schedules (Murphy, 1975).

In spite of their similarities, the tests had different emphases. The MAFL exercises were designed to assess a range of reading skills, including the ability to glean significant facts, comprehend main ideas, and draw inferences. By contrast, the AFRS test was more concerned with covering major areas of functioning in society, such as citizenship (e.g. reading an election ballot), health (understanding an anti-rabies form), and recreation (following directions on a seed packet). The MAFL provides an assessment of functional literacy up and down the hierarchy of skills, the vertical dimension of literacy, whereas the AFRS provides a more comprehensive assessment of functional literacy in the horizontal or applied direction. (Later in the article we present examples from these tests.) MAFL researchers chose the 75-percent level as the functional literacy threshold. They found that 87.4 percent of the nation's seventeen-year-old students in 1975 scored 75 percent or better and thus considered them functionally literate. Twelve and six-tenths percent, or about 13 percent, did not reach this level and thus were considered functionally illiterate. In

the spring of 1983, the President's National Commission on Excellence in Education used this 13 percent teenage-illiteracy rate as one of its 13 indicators that the future of the nation was at risk.

AFRS researchers did not report a functional literacy rate. They were leery of imposing a specific criterion level and also recognized the widespread disagreement over what constitutes functional literacy (Murphy, 1975a, 1975b). Instead, they reported the average percent correct on each item. Since this was the only way the data was reported, we do not have a functional literacy rate from this study. Kirsh and Guthrie (1977-8, p. 501), however, reanalyzed the AFRS test data for two groups of tasks. They found that the average "maintenance" item was answered correctly by 82 percent of the population, or, as they put it, "almost one out of five respondents could not complete reading tasks involving a table of contents, common signs, and train schedules" (p. 501). An average of one out of four could not handle occupation items, which dealt with sick leave, discrimination information, and employment applications.

The fifth study, the Adult Performance Level Project, (Adult Performance Level, 1977) was conducted by researchers at the University of Texas under sponsorship of the U.S. Office of Education. It differed from the other studies in three major ways. First, it was a study of functional competency rather than functional literacy. Thus, the skills it assessed were not confined to reading, but included writing, computation, and problem-solving. Second, the test designers conceived of competency partly in terms of knowledge and thus tested information as well as skills. Third, the test was deliberately designed to distinguish between those who were successful in the society, i.e., who had completed high school and were in white-collar or professional jobs, and those who were unsuccessful, that is, had fewer than eight years of schooling, were unskilled or unemployed, and lived in poverty. The researchers did this to develop "competency profiles" associated with different levels of adult success.

The Adult Performance Level Project used three competency groupings to report its scores. Adult Performance Level 1 was the group of adults who were "by and large, 'functionally incompetent'"; those in Adult Performance Level 2 were described as "marginally competent"; and those in Adult Performance Level 3 were "most competent" (Adult Performance Level, 1977, 17). Nowhere have the criterion levels, that is, the percent correct associated with each category, been described. Researchers found that 19.1 percent of the adult population, ages 18 to 65, were in the Adult Performance Level 1 category, and 33.9 percent were in the Adult Performance Level 2 category. Researchers concluded: "approximately one-fifth of the U.S. adults are 'functionally incompetent'" (Adult Performance Level, 1977). This included about 23 million people (Northcutt 1975, 48). If one considered as well those who were marginally incompetent, one would estimate that 53.6 percent of the adult population, as Secretary Bell estimated in 1982, 72 million adults have difficulty functioning. (Illiteracy . . ., 1984, p. 5).

Adult Performance Level also provided breakdowns for the tested skills: reading, writing, computation, and problem-solving. The rates for Adult Performance Level 1 ranged from 16.4 percent for writing to 32.9 percent for computation. This meant that one-third of the population was computationally incompetent. For reading, 21.7 percent of those tested fell into Adult Performance Level 1; 32.2 percent into Adult Performance Level 2.

The results from these five studies are summarized in Chart 8.

b. The Criticisms

The functional illiteracy rates shown in Chart 8 are much higher than self-reported crude illiteracy rates or functional illiteracy measured by years of educational attainment. At first glance, they indicate that functional illiteracy is a major educational problem. However, serious criticisms have been levied against these tests.

i. Validity

Creators of functional literacy tests assumed two things: that it is possible to identify a set of skills needed for daily living and that, once identified, it is possible to create a set of test items that can mirror the set of real-life tasks. Acland (1976) noted that: "any test of competency assumes we can judge what it takes to get by and having judged it, we can measure it" (p. 25). Critics of functional literacy tests have found fault both with the definition of the domain of skills and with researchers' selection of items. Several critics questioned whether it was even possible to identify a set of tasks that could be called functional literacy or functional competency. As Fisher (1978) writes: " . . . one must question the adequacy of a general assessment instrument. By most accounts, functional literacy is a concept relative to the country in which one lives. It seems reasonable to assume that it is also relative to a given subpopulation. The literacy demands on one subpopulation may include only some of the demands on other subpopulations. A subpopulation could be functionally literate with respect to itself, but not with respect to other subpopulations" (p. 57).

Acland (1976) makes much the same point, arguing that people do not face the same problems, that the skills it takes to function successfully vary with a person's social group, and that it is nearly impossible to identify a single set of tasks that can be used to test functional literacy.

Griffith and Cervero (1977) take this argument one step further, claiming that whatever tasks are included on the test reflect an unrecognized value position of the test designers. They write: "Functional competence can only be defined from a specific value perspective . . . It is accurate to say that the content of any definition of functional competence flows from the value orientation

of the test developers, an orientation which the test developers have chosen not to discuss" (p. 218). They linked the Adult Performance Level to previous efforts in life-adjustment education, arguing that it simply was one more attempt to force the individual to adjust to the majority society.

Although we are sympathetic to this line of argument, it is not altogether successful. Acland (1976), for example, presented three examples from the Adult Performance Level to show how hard it is to find universal problems and to suggest that the functional illiteracy rates were not that alarming. Fourteen percent of the adult population has trouble reading highway maps, but, as he points out, 16 percent do not own cars. Thirty percent have trouble with airline schedules, but 45 percent have never flown. Similar results were found for filling out checks and for maintaining checking accounts. Acland argued that those who answered incorrectly probably had little experience with the specific tasks being tested, so it was little wonder that so many had trouble on the test. He believed people develop the skills necessary for the tasks they face. As Acland noted, however, his argument would have been stronger had the data on car ownership and flying been been gathered from the same people who took the test. Only then could we argue that those who have trouble reading highway maps don't own cars.

There are several reasons to believe that there is an actual skill problem. Related items show greater percentages of people having trouble than those Acland cited. Thirty percent, for example, could not identify the meaning of "right-of-way." Fifty-nine percent could not read a parking ticket and determine the date by which payment was due. Both of these rates are much larger than the 16 percent who don't own cars. The percentage of test subjects who filled out deposit slips improperly exceeded the percentage who didn't have checking accounts. Furthermore, we question whether the proper comparisons were made in these examples. Map reading, for example, is not exclusively needed by those who own cars. Finally, many problems tested a general skill rather than the specific task presented. The airline schedule problem is a case in point. It was designed to be representative of all schedule-reading tasks, including bus, train and airline timetables. Although Acland acknowledged this, he believed that airline schedules have peculiarities that only those who have flown can decipher. Results from the AFRS, however, suggest otherwise. On that test, 33.4 percent of the adult population couldn't read a train schedule. The problem, we suspect, is not a lack of travel experience but problems with schedule reading.

The danger in Acland's critique is that it can lead us away from recognizing generally needed functional skills. It led Acland, for example, to talk of "rich people's problems" and "poor people's problems," as if map reading, schedule reading, and check writing are the province of the well-to-do. Such stratification of literacy needs makes us uncomfortable. Although we should be sensitive to the issues Acland raises about some people's unfamiliarity with specific tasks, this should not lead us to the conclusion that the tests did not cover

skills needed eventually by most people. Filling out application forms, understanding earnings statements, reading product labels and coupons, as well as the skills mentioned above, are tasks almost universally faced. Such tasks formed substantial portions of the five tests discussed above.

But even if we can identify a set of generally needed functional tasks, what about Griffith and Cervero's point that this is necessarily a value-laden process? They argued that test objectives are dependent upon the group designing the test. They claimed that different groups would inevitably identify a different "universe of behaviors" as fundamental (213). As evidence, they noted that although consumer economics was a key knowledge area on the Adult Performance Level test, it had ranked only 168th in importance out of 170 curriculum areas in a recent survey of adult basic education teachers. One disagreement, however, does not prove their case. We would expect that in most cases different but largely overlapping sets of skills would be identified. Furthermore, it should be noted that the Adult Performance Level objectives and test items resulted from the input of diverse groups. Test designers consulted with representatives of state and federal agencies, held regional conferences on adult needs, and interviewed undereducated and underemployed people (Adult Performance Level, 1977, p. 6-7). The Adult Performance Level was also field tested on 3,500 undereducated and underemployed people in 30 states. Nevertheless, of the five tests, the Adult Performance Level was particularly subject to the problem of bias because it was designed to test knowledge as well as skills. Several of the objectives used to define functional competency were culturally and politically loaded. For the health area, one objective read, "to understand the importance of family planning, its physical, psychological, financial and religious implications" (A5). A consumer economics objective read, "to understand the implication of consumption vis-a-vis finite world resources and to recognize that each individual's pattern of consumption influences the general welfare" (A5). With such value-laden objectives, it was inevitable that some culturally biased questions would be included on the test (See examples).

Example #1:

"Concerning the right to peaceful assembly, 12 percent of the sample felt that permission to have peaceful meetings should not be given to certain kinds of groups; e.g., 'radicals' and 'troublemakers'" (Emphasis added; from Adult Performance Level, 1977, p. 25).

Example #2: (From the American College Testing program's version of the APL). The city garbage truck has not picked up Esther Maxey's garbage for three weeks. Esther is having trouble keeping the flies and mice away. What should she do?

a. Take the garbage down the street to an empty lot.
b. Call the hospital to complain about the mice.
c. Call the sanitation department about the problem.
d. Cover the garbage with a sheet.

(From Fischer, Haney & David, 1980, p. 69)

As Heller et al. noted (cited in Fischer, Haney & David, 1980, p. 69) example #1 asks for an opinion rather than a recognition of the constitutional guarantee. As for example #2, the notorious quality of many cities' public services makes more than one answer plausible. As Fischer, Haney & David write, "If this test item measures anything at all, surely it is only the test-takers' ability to ferret out the item-writer's sense of social propriety" (p. 69).

The issue, of course, is the extent of the bias problem. None of the critics systematically analyzed all of the objectives for the test. Although some of the test's 65 stated objectives were vague and some were of questionable relevance, few were as culturally or politically loaded as those the critics cited. Most of the actual test items did not show any particular bias; their most striking quality was their mundaneness. Later, we describe each of the published items. For now, though, we present three representative examples. Individuals were asked to address an envelope (with or without a zip code), to follow directions on a medicine-bottle label that stated "take two pills twice a day," and to determine change from a twenty-dollar bill given a sales receipt (Adult Performance Level, 1977, p. 28, 22 and Thompson, 1983, p. 480). It is difficult to see what "value-orientation" was being advanced by these tasks. The finding that 13 percent, 21 percent, and 28 percent of the population, respectively, could not perform these tasks suggests that there is a functional literacy problem.

The other four tests also involved various groups in their design, thus lessening the possibility of bias. In their surveys, Harris and Associates were directed by the National Reading Council, a group of 40 men and women including businessmen, teachers, Congressmen, civic leaders, and entertainers. The MAFL test was designed by a group of reading specialists on the National Right to Read staff. For the AFRS test, ETS conducted a national survey of adult reading activities, covering what people read, and for how long, and asking them to rate the activities' importance. The results of the survey, along with a set of test items designed by ETS, were submitted to a panel of representatives from industry, education, journalism, and consumer groups. On the panel's advice, the items were simplified and multiple-choice items were eliminated. The field test was conducted with 2,100 New York and New Jersey adults, most of whom were in Manpower Training Centers. Over one-third of the items were eliminated as a result of the field testing (Murphy, 1975). These four tests have not received the criticism on cultural grounds that the Adult Performance Level has. Bias was probably less a problem

since these tests were focused on functional skills rather than knowledge. Tests that involve telephone dialing instructions, applications forms, and everyday reading such as train schedules, store coupons, and report cards are less prone to cultural prejudice. These five tests, for the most part, did not promote values from a particular social milieu. However, the tests were all in English. Only bilingual testing could erase this bias.

ii. Test Quality and Test Construction

Critics also questioned how well the tests measured functional literacy or competency. Here again, they singled out the Adult Performance Level. Fischer, Haney and David (1980), for example, concluded that, on the basis of its design, conduct, and reporting, the Adult Performance Level findings were "altogether untenable" (p. 69). They acknowledged, though, that several of the criticisms had only partial merit. In a review of adult competency tests, for example, Nafziger et al. (1975) rate the Adult Performance Level poor in terms of administrative usability and technical excellence, fair in measurement validity, and good only in examinee appropriateness. Their evaluation method, however, was described by a noted testing expert as "incredibly subjective" (Anastasi quoted in Fischer, Haney, & David, 1980, p. 65).

APL's construct validity has also come under fire. Adult Performance Level test designers conceptualized functional competence as a matrix of five general skills applied across five general knowledge areas. After performing a factor analysis on the Adult Performance Level, Cervero (1980) found that it measured only three independent dimensions: verbal ability, writing and computation. Problem-solving, purported by the Adult Performance Level designers to be one of the five essential functional skills, did not emerge as a separate dimension. As a result, Cervero questioned the APL's construct validity, arguing that it measured an individual's ability in the three R's rather than their functional competence. However, reading, writing, and computation are still essential ingredients of functional competence, and finding that the test items largely measured them is exactly what we would expect.

With more merit, critics have questioned the basis for including certain items on the Adult Performance Level test (Cervero, 1980; Heller et al, cited in Fischer, Haney, & David, 1980; Fisher, 1978). As noted, Adult Performance Level designers wished to profile the skills associated with advanced education and job status and to distinguish these from the skills possessed by those in unskilled jobs, with little education, living in poverty. As a consequence, an important factor in determining whether an item was included on the test was whether it discriminated between the haves and the have-nots. In the course of field testing and revisions, items were eliminated if they failed to correlate positively with occupation, education, and income levels. As Heller et al. pointed out, however, this approach obscured the distinction between survival and success. Identifying

the skills and knowledge associated with success in contemporary society is quite different from identifying the minimal essential skills necessary for survival (See Fischer, Haney & David, 1980, p. 63). If the Adult Performance Level was supposed to measure basic functional skills, as has been widely assumed, items should have been included because they measured such skills and not because particular socioeconomic groups performed well on them. As Donald Fisher (1978) argued: "There is no a priori reason for discarding items that fail to correlate positively with the Adult Performance Level levels. If an item tests an area of knowledge or a set of skills which are logically, though not empirically, important, the item should be included" (p. 53).

We largely agree with these critics. APL's method of item selection is reminiscent of the World War I psychological testing program. The Army Alpha tests were field tested on graduate students and officer training school candidates, among others. Items that did not distinguish between them were discarded. Later, the psychologists "proved" their test's validity by showing that scores on it were highly correlated with education and income--the very factors it had been designed around. In the same way, in their final report, Adult Performance Level researchers emphasized the fact that functional incompetence varied with level of education and income. This fact, however, was not meaningful because the test was designed to produce these very results.

Two caveats to these criticisms are in order. First, Adult Performance Level test designers used a two-stage process to select items. They created a large pool of items that measured the basic competencies, and then they excluded items on the basis of their correlation with education, income, and job status. The test items, therefore, still represented a set of basic functional skills. Second, the correlations between Adult Performance Level scores and the socioeconomic indices were not very high (education, .56; income, .40; and occupational status, .31), so the selection of items was neither as systematic nor as successful as the critics implied. The weak correlations indicate that the items measured more than just the socioeconomic status of the individual or the skills associated with a given social class. A judicious examination of the Adult Performance Level, therefore, still can provide important insights into the extent and nature of functional competency.

Critics have raised concerns about test quality that applied more generally to all five tests. Several criticisms were persuasive, although for each one there are strong counter-arguments.

According to one information processing perspective, educators have widely misinterpreted the functional literacy tests. The poor performance of many people was due not to deficits in their skills or knowledge but simple errors in their processing and answering of test questions. Donald Fisher (1981), for example, presented a model of functional literacy question-answering designed to show how people process typical test material. Analyzing data from the AFRS, he

showed that most errors fit the model, that they were the result of breakdowns at one stage or another in information processing caused by the complexity of the question and the material. Fisher suggested that if his findings were experimentally confirmed, it would mean that functional literacy was not as serious a problem as many believed, and it would indicate a different approach to remediation. While we agree that some errors occurred because people had difficulties processing the questions, we do not agree with Fisher's suggestion that the typical error was "more or less mechanical" (p. 443). His model describes a dynamic information processing system that cannot be clearly distinguished from functional literacy. His various stages are important processes involved in functional reading, namely "identify target and locator propositions," "derive search clues," "encode passage," "identify untagged propositions," "modify status of propositions," etc. People who repeatedly make errors in such processing stages have functional reading problems. His model describes the problem in more detail; it does not make the problem disappear.

In a related analysis, Murphy (1975) administered the AFRS to 100 people in adult learning centers and then interviewed them concerning the errors they made. He found "two very simple causes of incorrect responses" (p. 14): unfamiliarity with everyday words and difficulties with everyday formats. People had trouble understanding words and phrases such as "recipe," "to call up," "lives," "locker," "circle," "fourth," "severe," and "mild." Even when they had the vocabulary, they did not know how to handle the words in varying formats such as doctor's bills, train schedules, applications forms, election ballots, etc. His finding reinforces the view that the errors represent serious deficits rather than lapses in routine, mechanical processing.

Poor performance also was attributed to fatigue and to poorly constructed test items. Fisher (1978) was interested in determining what effect these factors had on the performance of high school graduates. He wanted to test reports that the high schools were graduating large percentages of functional illiterates. He estimated the impact of the two factors by assuming they had caused much of the "functional illiteracy" found among the college educated. Using the two Harris surveys and the Adult Performance Level study, he recalculated the functional illiteracy rates for high school graduates. He found that the rates on the three tests dropped respectively from 1 percent to .7 percent, from 2 percent to .6 percent, and from 9 percent to 7 percent when controlling for fatigue and poorly constructed items. He thus concluded reports of extensive functional illiteracy among high school graduates were exaggerated.

There are three problems with Fischer's argument. First, several of the tests were too short or varied to have produced much fatigue. The AFRS, for example, involved only 17 questions per person and took less than thirty minutes to complete (Murphy, 1975, pp. 5, 14). The Adult Performance Level test took 60 minutes and involved a wide range of formats and tasks. Although the Survival Literacy Study involved

five application forms, four had twelve or fewer items. The second Harris survey involved only one application form. The rest of the test was a series of short oral questions--four about telephone dialing and three each about three short housing and employment ads. Second, Fisher assumed that poorly constructed test items accounted for a substantial portion of the functional illiteracy among the college educated. But college-educated individuals, having had extensive test experience, likely recognized most test construction errors and realized what the tester "had in mind." The errors they made on the test were more likely to have been real functional errors. Third, the original functional illiteracy rates already accounted for fatigue and errors due to poor items. Each of the tests set functional illiteracy criteria that were far short of perfect performance. In other words, a respondent could miss a substantial portion of the test and still not be considered functionally illiterate. On the first Harris study, for example, the threshold was set at 70 percent, so an individual could miss up to 30 percent of the test and still not be labelled functionally illiterate. On the second Harris survey, a person had to answer incorrectly more than 20 percent of the weighted test items before being considered functionally illiterate; on the MAFL, it was 25 percent. Although no thresholds were reported for the Adult Performance Level, an estimate can be made from information found in the appendix of the final report (Adult Performance Level, 1977, Appendix B, page B11). The bottom quartile, which encompassed the Adult Performance Level 1 group, scored below 71 percent. So an individual could miss 29 percent of the test and still not be considered functionally incompetent. It seems to us that such large margins of error were adequate to cover problems due to fatigue and poor item construction.

Several other researchers also raised questions about the test items, but few assessed how extensive this problem was on any given test. Fisher, David, and Haney (1980), for example, faulted the Adult Performance Level for having factually incorrect items, but cited only one example of this. Caughran and Lindlof (1972) questioned the inclusion of the Medicaid form in the Survival Literacy study because it was five to 13 times as difficult as the other forms (see Chart 5). The issue, though, isn't whether it was harder than the other forms, but whether it adequately measured functional literacy. Although the form is a bureaucratic labyrinth that should be changed, the ability to find one's way through it is a necessary reading task for many people today. Cervero (1980, 168) questioned whether paper-and-pencil test items can measure functional competence, but gave no examples. Although paper-and-pencil test items cannot always mirror real-life tasks, they often do. The ability to fill out application forms, for example, is a functional skill and is readily measured by a written functional literacy test. Including everyday reading materials such as report cards, parking tickets, and earning statements on the tests also ensured that they approximated real-life problems. Furthermore, it is not widely recognized that questions and tasks were posed orally on three of the tests, either for the entire test (AFRS) or for important sections of it (Harris II, Adult Performance Level). This approach eliminated problems caused by an inability to read the

questions and made the items closer to real-life tasks, in which an individual often knows what he must do and then applies his functional skills.

Acland (1976) made perhaps the most significant criticism of the items when he observed that, in real-life settings, individuals solve many of the problems posed by the tests by relying upon environmental cues and help from others. He presented several examples from the Adult Performance Level to illustrate his point. Although 26 percent of the sample could not determine which of the three cereal products was the cheapest per unit weight, Acland noted that some supermarkets now provide unit pricing labels that make this skill unnecessary. Twenty-seven percent of the sample did not know the normal human body temperature, but as Acland pointed out, every thermometer clearly marks this point, making it unnecessary to remember this particular piece of information. Sixty-one percent had trouble determining the right tax from a tax table, but as Acland discovered from the IRS, only 6 percent of actual tax filers made any kind of arithmetic error in their returns. The reason for this is that in real-life situations, people get assistance. IRS surveys revealed that at least 60 percent of the population gets help filling out their tax forms. Sticht and his colleagues make a similar point about the employment findings (Sticht, 1975, p. 149). Even though many people have trouble reading the ads, most jobs are found not from newspapers but from friends, employment agencies, or direct contact with employees.

Another source of assistance that Acland didn't mention was feedback. In real-life settings, unlike test situations, if an individual makes a mistake, such as not paying a sufficient amount of money for a purchase or misreading a menu, observers will point out his error and often suggest the proper response. In other instances, what seems like a problem on the test proves inconsequential in a real-life setting. An example from the Adult Performance Level illustrates both points. People were asked to write a note to their child's teacher excusing the child for missing a day of school due to a sore throat. Twenty-two percent did not include a salutation, which seemed to indicate a functional literacy problem; in fact, the salutation is conventional and unnecessary. Seven percent did not identify the child, but because the child is bringing the note to the teacher, this is also unnecessary. Twenty-nine percent failed to sign the note, but in most schools, teachers would give the note back to the child and ask them to have their parents sign it. The feedback provided would ensure a functional response. Other errors on this task, however, suggested there are more serious functional competency problems. Seven percent did not produce a comprehensible message; another three percent wrote notes that were so illegible their content couldn't be judged.

In spite of our sympathy to Acland's line of reasoning, it downplays functional literacy problems. At some level all functional tasks could be solved if you got someone else to help you or to do them for you. Being functionally literate or competent, however, implies a large degree of self-reliance. Are you able to handle that

task when the environmental cues are not there? Many supermarkets do not have unit pricing. Filling out forms incorrectly often causes delays and frustration. Seeking help may work, but a person should not have to depend upon others to solve the basic tasks of daily living. Furthermore, help may not always be available. As Sticht and his colleagues note, filling out application forms is a solitary enterprise for which many are unprepared. In short, although the tests ignore the alternative strategies of real-life settings, they still measure a person's self-reliant functional literacy.

Fisher (1978) raised yet another argument about test quality. He questioned the validity of the functional literacy measures because a surprising number of professionals and managers were categorized as functionally illiterate. The 1971 Harris study, Fisher said, showed a 5 percent rate, the Adult Performance Level 11 percent, and the AFRS, 14 percent. Since individuals in professional and managerial jobs have clearly succeeded in society's terms, they can hardly be considered "functionally illiterate." According to Fisher, the tests must be mislabelling these individuals. Furthermore, since the rates among high school graduates parallel those for professional and managers (4 percent, 11 percent and 19 percent), Fisher inferred that few, if any, high school graduates were functional illiterates.

There are two problems with his argument. First, the functional illiteracy rates he cited for professionals and managers are questionable in two cases. The 1971 Harris study actually showed only 2 percent being functionally illiterate, (Harris and Associates, 1971, p. 52, 56). And the AFRS rate must also be questioned. The AFRS study never produced functional illiteracy rates. Results were published on an item-by-item basis. For each item, we know what percentage of professionals and managers answered incorrectly. Fisher took these figures and averaged them. The 14 percent rate represents, therefore, the percentage of professionals and managers who missed the items on average. This is quite different, however, from the usual way functional illiteracy is determined. For the AFRS, we do not know what percentage of professionals and managers missed what percentage of items. If a 75 percent correct threshold indicates functional literacy, for example, we do not know whether 14 percent, 5 percent, or only 1 percent fell below that point. Second, and more important, Fisher assumed that any person who had achieved professional or managerial status had functional literacy skills. Job literacy tasks, however, can be quite different from those practical daily tasks that were tested. Many people might be able to handle their jobs but not be able to negotiate airline schedules, social security applications, miles per gallon calculations, etc. Indeed, they may be relying upon subordinates to accomplish these tasks. We must also break down a monolithic impression of professionals and managers. They are not all lawyers, doctors and corporate executives. The professional and managerial job classification includes small business proprietors, owners of minor construction and manufacturing concerns, professional athletes such as prizefighters and football players, and fine-arts professionals such as dancers and musicians. Many of these individuals may be quite talented but still have trouble handling

everyday reading tasks. We also must realize that many professionals and managers assumed positions in earlier decades and consequently may have less educatic than their modern counterparts. In 1974, for example, 1 percent of white male professionals and 5.7 percent of managers had not attended school beyond eighth grade. Three percent of the white male professionals had not graduated from high school, while 14 percent of the managers had not. The percentages for women were comparable, while those for black males were greater at 3.8 percent and 19.7 percent (U.S. Department of Labor, 1974).

These facts suggest that a significant portion of professionals and managers could have difficulties with functional tasks, and that the functional illiteracy rates were not as anomalous as Fisher believed.

iii. Criterion Levels

Functional literacy rates are crucially dependent upon the criterion level. Is an individual who scores below 60 percent on a test functionally illiterate? below 70 percent? below 90 percent? Researchers can control functional illiteracy rates by raising or lowering the cutoff point. Yet, without exception, researchers for the five studies did not explain or justify their particular cutoffs.

The MAFL test provides the most striking illustration of how crucial the cutoff is. MAFL researchers chose the 75 percent level to define functional illiteracy and, as a consequence, found that 12.6 percent of the nation's seventeen-year-old students were functional illiterates. Yet had they chosen the 60 percent level, only 2.9 percent would have been so labelled. Some explanation of the criterion level seems crucial, but it simply wasn't offered. Without an explanation, we must speculate. A 60 percent criterion seems inappropriately low, because the MAFL was designed so that all seventeen-year-old students could actually answer all items correctly (Gadway and Wilson, 1976, p. vii). At the 90 percent level, on the other hand, a whopping 45.4 percent of the seventeen-year-old students would have been classified functionally illiterate. The 75 percent cutoff probably is more reasonable in that it allowed a larger margin of error for fatigue and poor item construction. The cutoffs on the other tests also seem reasonable, given that simple functional tasks were being tested and that they provided a large margin of error.

The variations across criterion levels illustrate that many so-called "functional illiterates" are not completely incapacitated but do have some basic functional skills. On the MAFL, for example, most of the "functional illiterates" scored above 60 percent, thereby demonstrating that they can handle a majority of the functional tasks thrown at them. The danger in a loaded, dichotomous term like "functional illiterate" is that the individuals falling in that category are stigmatized as completely incompetent when, in fact, many are not. This is not to say that such individuals are not having serious difficulties functioning--after all, the tests measure simple

daily tasks and they missed large portions of them--but they may be coping satisfactorily in many situations. There is, of course, a subset of the group labelled "functionally illiterate" who can handle hardly any functional tasks at all and who truly have serious problems coping. We call the first group "significantly dysfunctional" and the second "absolutely dysfunctional." We conceptualize the functional spectrum, therefore, as running through four groups that shade into each other--namely, from functional to marginally functional to significantly dysfunctional to absolutely dysfunctional. The second Harris study, like the MAFL, confirms the existence of the two dysfunctional groups. On many test sections, one set of people missed several items while another set could hardly answer any questions at all (see Chart 7). Recognizing the existence of these two sets of dysfunctional individuals should alter the ways people conceptualize and respond to the functional illiteracy problem. The problems of the two groups likely differ, as does the remedial help that should be provided.

What are the relative size of the two groups? Although the MAFL showed that most teenage "functional illiterates" are significantly dysfunctional rather than absolutely dysfunctional, the ratio is probably different in the general population. Since fewer people received schooling in the past, we would expect to see higher proportions of absolutely dysfunctional individuals among the elderly and the general population. The findings of the two Harris studies and the Adult Performance Level seem to bear this out.[4]

iv. Calculation of Rates

Beyond the problem of how criterion levels were set, critics have also questioned how the functional illiteracy rates were calculated for the particular criterion. Caughran and Lindlof (1972) faulted Harris and Associates for simply averaging the results across the five forms on the Survival Literacy Study (see Chart 5). Contrary to their own definition of functional illiteracy, Harris and Associates did not combine all five forms and determine what percentage of individuals scored below 70 percent across all forms. Averaging the way they did can produce an inflated illiteracy rate, but this is not, as Caughran and Lindlof argue, necessarily the case. Averaging may indeed underestimate the illiteracy rate, depending upon how one defines functional illiteracy. If we think of each application as a separate test, the separate rates are meaningful. They can be interpreted in light of the demands of each form. Averaging the results on the five forms tells us how well the population does on an average application form, so the Harris findings are still useful if seen in this light.

Fisher (1978) faulted MAFL researchers for not weighting their items as to importance when they calculated functional illiteracy rates. Although this sounds like a good idea, it really would have introduced the subjectivity the critics decried. In many cases, it is not clear how it could be done. Is it more important to know when a

parking ticket is due or to be able to read a train schedule? Of course, they could have presented both weighted and unweighted results. Harris and Associates (1971) did this, judging an item's importance by how hard it was for the population to answer. Official items were assigned a greater weight. The logic of this weighting is questionable, however, since it may be that the items few individuals can answer are the least important, while those that most can handle are the most important to know. Their weighting scheme seems to have inflated their rates. With items weighted, they found that 15 percent of the population had functional reading problems (Scored below 90 percent). Looking at the unweighted results, however, we find only 5 percent of the population scored below 90 percent, a much less alarming result (see Chart 9). We estimate that only 2 percent of the population scored below 80 percent correct, which was Harris and Associates[1] functional literacy cutoff.

v. Minimizing the Seriousness of the Findings

Critics of functional literacy tests have attempted to minimize the seriousness of the findings in two ways. Citing the distribution of intelligence as propounded by Terman and Merrill, Caughran and Lindlof (1972) held that 2 percent of the noninstitutionalized population would always be illiterate and that, therefore, functional illiteracy rates must be considered in that light. In other words, like the poor, the illiterates ye shall always have with you. We are skeptical, however, when one of the early pioneers of psychological testing is cited as an authority on the capability of humankind. As has been well documented, many of these early testers, including Terman, held racist and distorted views of human intelligence and seriously underestimated the impact the environment can have on developing an individual's latent abilities. Furthermore, their criticism was misplaced since mentally retarded individuals were excluded from the testing. Harris and Associates, for example, excluded from their testing anyone who had trouble communicating with the interviewers, while the MAFL researchers excluded any emotionally, physically, or mentally handicapped individuals deemed incapable of taking the test.

At several points, Fisher (1978) predicted future illiteracy rates using present-day distributions of rates by age. Since it is true that functional illiteracy rates are much higher among the elderly, it is likely that in a few decades the population's overall rate will be lower. Does this mean we can afford to be complacent about the problem? Certainly not. The illiteracy problems of younger cohorts are substantial; we can take scant comfort in the death of aged illiterates.

vi. Raising the Estimates

Several arguments can be made that the reported rates actually underestimated the extent of functional illiteracy in American society

in the 1970s. The first concerns sampling. The MAFL involved only seventeen-year-old students and did not test dropouts. Given that almost 20 percent of high school students drop out, and that their average performance is poorer than those who stay in school, we estimate that the MAFL's 12.6 percent overall functional illiteracy rate needs to be raised by around 4 percent (20 percent of the students x 20 percent estimated rate = 4 percent). Furthermore, non-English-speaking seventeen-year-olds were excluded, so the percentage may be higher yet, perhaps another percentage point or two (1.4 percent of persons aged 5 to 17 speak English poorly or not at all. U.S. Bureau of the Census, 1982b). This would raise the MAFL rate to about 18 percent. For the Survival Literacy Study, Harris and Associates (1970, p. 29) were unable to test 12 percent of those sampled--8 percent because of a communication barrier such as language or deafness, 3 percent who refused to fill out the forms, and 1 percent who were unable to complete the test because of blindness or visual defects (1970, p. 29). Figuring that 2.2 percent of the population age 18 and older speaks English poorly or not at all and that a portion of those who refused to take the tests did so to hide their illiteracy, one can estimate that between 2.5 percent and 3 percent should be added to the Harris rates. This would raise their figure from 3 percent to 6 percent. In their second study, Harris and Associates (1971, p. 25) were unable to test 7.8 percent of the population--6.8 percent due to a language barrier, .8 percent due to deafness, and .2 percent due to blindness or visual defect. Again, we would raise figures by 3 percent, meaning that 7 percent have serious functional deficiencies. Our unweighted estimates of 2 percent would rise to 5 percent.

The second reason the rates may underestimate the extent of illiteracy is that, in certain cases, the tasks used on the tests were easier than those required in real-life situations. Caughran and Lindlof (1972), for example, compared the Survival Literacy's Application for Public Assistance, form IV, with the original government form upon which it was based. The Harris form was easier in three important respects. The format had been changed, it had proportionately fewer "difficult" words, and its readability level was grade five to six compared to grades seven to eight of the original. Caughran and Lindlof determined that its readability was similar to the forms III and II Harris used (driver's license and personal bank loan). This would raise the functional illiteracy rate on form IV from less than .5 percent to about 1.5 percent. (The rate of persons scoring below 90 percent would rise from 3 percent to about 9.5 percent.) If the other forms were similarly simplified, a percentage point or more would have to be added to the estimate of functional illiteracy. Coupled with corrections for the language barrier and refusals, the Survival Literacy results increase from 3 percent to about 7 percent. Sticht noted that on Harris II (1971) the questions were asked by the interviewers, which makes the employment ad reading easier than in real life where the individual must figure out his own questions. Also, Harris simplified both the housing and employment ads by removing the abbreviations. Thus the Harris II rates also are likely underestimated.

The third point relates to interpreting the results from studies limited to younger, better-educated age groups. This applies specifically to the MAFL findings. Since only seventeen-year-olds were tested, the results underestimate the functional illiteracy rate of the general population. We can extrapolate them, however, based upon the results of other surveys. On the first Harris study, (1970, p. 20), 1 percent of the 16 to 24-year-olds were functionally illiterate compared to 3 percent in the general population; on the second Harris study (1971, p. 51), it was 1 percent compared to 4 percent. On the Adult Performance Level (p. 19), 16 percent were functionally illiterate compared to 19.1 percent in the general population. In each case, the general population had a rate about 2 to 3 percent higher. Thus, for MAFL type items, we would estimate that about 21 percent of the population would be functionally illiterate (again, adjusted for dropouts and foreign language speakers). This is remarkably similar to the 21.7 percent Adult Performance Level rate for functional incompetence in reading.

The same age warning applies to the forthcoming results from the NAEP functional literacy study of young adults, ages 21 to 25 (See Kirsch, 1985). Many observers probably will compare the new results to those of the Adult Performance Level and mistakenly conclude there has been a substantial drop in illiteracy, or that the Adult Performance Level results were without merit, without taking into account the youth and education of the target group.

The fourth point relates to the tests' limited task domains. This applies to the two Harris studies, particularly to the Survival Literacy study. Fisher (1978) makes much of the low illiteracy levels on these narrow tests. After making his readjustments for fatigue and item construction for the two Harris studies, he concluded that the rates among high school graduates were "negligible" (p. 7). The two Harris studies, however, primarily involved filling out application forms and thus did not sample a wide range of functional tasks. The Adult Performance Level, by contrast, did, and even the adjusted rate for high school graduates showed 7 percent. This was hardly "negligible" since it amounted to over 6.5 million high school graduates (Grant and Eidon, 1982, p. 17). Thus, we must keep the task domain in mind when discussing results. After our adjustments, for example, the Harris rates were 7 percent and 5 percent--substantially lower than the MAFL and Adult Performance Level rates of 21 percent and 21.7 percent. In all this, of course, we must keep in mind that many of those labelled "functionally illiterate" are not absolutely dysfunctional.

c. Understanding the Estimates

Individual items provide an effective way of understanding the dimensions and nature of the functional illiteracy problem. On Chart 10, we briefly describe the published items and their associated illiteracy rates. We have arranged them by test and ordered them then from the most difficult to the least difficult items. The reader

should keep in mind that the rates are unadjusted for sampling or difficulty. They may, therefore, underestimate the actual extent of illiteracy. Using these charts, as well as other information from the literacy tests, we can arrange functional tasks on a scale from the easiest to the most difficult. What are the functional tasks that most U.S. adults (16 and older) can handle? Chart 11, an example from the AFRS, shows that only .1 percent have difficulty recognizing the milk bottle. This item tests lowest-level survival reading skill. The results suggest that virtually all adults can recognize simple words that are frequently encountered. Only 2 percent of the test subjects made one or more errors in telephone dialing instructions. Three percent of the population were unable to read and understand housing ads (allowing two errors). If we allow only one error on these very simple passages, 5 percent had trouble. These results suggest that nearly everyone, 95 percent to 97 percent, can read simple advertisements and find rate and area code information in a table. Although the percentages who cannot read accurately at this level are not great, it should be remembered that three to five percent corresponds to 4 to 7 million people aged 16 and older.

The next set of tasks concerns application forms. The Harris II study showed 5 percent of the subjects performed below 90 percent correct on a test which was substantially (two-thirds) composed of such items. Chart 12, another example from the AFRS, shows the depth of the problem. Seven and two-tenths percent were unable to mark the correct spot where the name of their emergency contact should be entered. On the driver's license form of the Harris Survival study (previously displayed, Chart 4) 8 percent scored below 90 percent correct, yet the test asked only for simple information such as height, weight, name, color of eyes, number of times previously examined for license, and day and hour it would be convenient to take the test. (Only 1 percent scored below 70 percent, meaning that most people filled out much of the application. These items suggest that 5 percent to 8 percent of the population have trouble completing job applications, applying for loans, and filling out auto insurance forms.

The next set of items were those that about one-seventh (or 14 percent) of adults could not handle. These included map reading, addressing an envelope, writing a check properly, and verifying the figures on an earnings statement.

Chart 10 then moves to those items on which about one-fifth, or 20 percent, of the population had trouble. These include understanding a check-cashing-policy sign, reading dosage information on medicine bottles, determining price-per-unit weight, finding the monthly repayment on a loan repayment chart, determining what size product a coupon was good for, and understanding a housing inspection notice. Substantial percentages of subjects could not handle items that seem straightforward to a highly literate person. Twenty-one percent, for example, could not determine how often houses should be inspected for termites when presented with a government brochure that read "periodic

inspections should be made at least every six months if you live where termites are common" (Thompson, 1983, p. 481).

About one-third of the subjects failed at reading airline and train schedules, figuring out how much change should come from a purchase, and determining which subjects had improved on a report card (Train example in Chart 13). Forty percent could not read two short passages describing a blood donation program and identify the person they should contact (Chart 14). Forty percent could not determine how much they would save when given a certain percentage off a retail price. Finally, those items on which a majority of adults had trouble included determining when a parking ticket fine was due, locating the tax for a given income on a tax table, determining total purchase price for a mail order, reading and filling out a W-4 form, calculating miles per gallon, and knowing how to put a return address on a business letter.

Given the data previously presented and that on the charts, we find it reasonable to estimate that about 20 percent of the adult population, or about 35 million people, have serious difficulties with common reading tasks. Another 10 percent or so are probably marginal in their functional literacy skills.

What are the consequences of low functional literacy skills for day-to-day living? Many people, perhaps between 10 and 20 percent, have trouble using the phone book because their sorting skills are poor. They may have difficulty using the Yellow Pages. Some may have trouble finding emergency phone numbers for police and fire. Among the lowest fifth in functional literacy skills are many who are unable to read product labels and must depend upon brand-name logos for item selection in a grocery store. Many are unable to determine whether they are receiving correct change. Many cannot read recipes very well and cannot follow the directions on frozen-food packages. Their ability simply to navigate city streets is hampered by their difficulties with traffic signs, street names, and bus or subway schedules. They may have trouble filling out job applications. On the job, they may have trouble completing forms requesting sick leave or holidays. Parents' roles as guardians of their children and providers for their welfare can be seriously undermined by a lack of functional skills. Many cannot read letters from school or report cards. Many cannot communicate effectively with the teachers or help their children with homework. They may threaten their own health and that of their families because they cannot understand prescriptions or warnings on medicine bottles, or read notices aimed at preventive medicine. Even in their leisure time, they may be frustrated by reading problems. If they choose to go out for dinner, they may not be able to read the menu. They may not be able to read the newspaper movie ads well enough to determine where and at what time a given film is showing. Even in their own homes, they may have trouble deciphering the local TV schedule.

Poor literacy skills foster passivity and increase the possibility of exploitation. Most of the people in this lowest quintile of

functional literacy skills are not outright illiterates. Most can read to some degree, and they devise many ways to cope with the reading tasks on which they have difficulty. Moreover, the forces of poverty and discrimination, and the lack of broadening experiences and opportunities in many people's lives, render many literacy tasks irrelevant. These problems would not disappear if we could suddenly improve everyone's functional literacy. Nonetheless, low literacy skills can be crippling, and a target group of 20 percent of our population is alarming. Is functional illiteracy increasing, or have we just rediscovered a long-standing problem?

d. Historical Trends in Functional Literacy

The five functional literacy tests discussed above all were administered in the 1970s, so we do not have a historical record to judge long-term trends in functional literacy. Nevertheless, there are three ways of gauging short-run historical trends: comparing scores on the MAFL of 1975 to scores on the NAEP reading test of 1971, comparing the performance of various age groups on the functional literacy tests, and analyzing test-score trends on the work-study skills sections of standardized tests.

Since sixty-four of the MAFL items were taken from the 1971 NAEP reading test, NAEP researchers compared scores on those items (The MAFL was given in 1974 as well as 1975). They found that the average student's functional literacy performance rose from 83.7 percent correct in 1971 to 85.9 percent in 1975. We thus know that during the early 1970s, seventeen-year-old students somewhat improved their functional literacy.

Comparing the literacy of different age groups is a widely used method of determining changing educational conditions. On the crude literacy and educational attainment measures of the Census, for example, each successive younger age group has a lower illiteracy rate. This indicates that each successive 10-year cohort is becoming more literate and attaining more schooling than the last. On the functional literacy tests, however, we do not see this consistent pattern. As Copperman (1978, p. 47) pointed out for the Adult Performance Level, given in 1974, the highest competency rate was for the 30 to 39-year-olds, while the 18 to 29 age group had nearly the same rate as the 40 to 49-year-olds. If we trace which years these groups were in school, we see the better educational conditions prevailed (see Chart 15). Since the 30 to 39-year-olds were in school during the 1950s, this period may have been a hey day for American schooling and the training of functional competency skills. Since the 18 to 29 year olds performed similarly to the 40 to 49-year-olds, who would have graduated at the latest in 1952, Copperman concluded that students' performance in the 1970s had deteriorated to the levels of the early 1950s (p. 48). In fact, the 18 to 29-year-olds performed better than the 40 to 49-year-olds, so Copperman's conclusion isn't fully supported. Nevertheless, the performance of those graduating in the mid-1960s and through the early 1970s appears to have worsened in

comparison to the 1950s graduates. The 1973 AFRS shows a similar pattern. The 16 to 19-year-olds, who would have been in junior high and high school in the late 1960s and early 1970s, missed 28 percent of the items. This is more than either the 20 to 29-year-olds (23 percent) or the 30 to 59-year-old group (26 percent).

The designers of the Adult Performance Level found a similar pattern, but they reached a different conclusion than Copperman. They argued that the youngest adults scored more poorly simply because they had had less experience with functional literacy tasks (Adult Performance Level, 1977, p. 37). In an analysis of the Adult Performance Level and AFRS results, Fisher (1978, p. 9) argued that the younger groups scored more poorly because they had less education. On the AFRS, for example, the 16 to 19-year-olds had 11.3 years of schooling compared to 12.6 years for those 20 to 29. Many of the teenagers would graduate from high school and attend college. Fisher goes on to show that when the differences in educational attainment are considered, the younger cohorts are actually doing better than would be expected. According to Fisher, schooling has become more efficient, at least as far as functional literacy skills are concerned.

Copperman's argument also is contradicted by the results of the other three functional literacy tests. The two Harris studies showed that the youngest cohort, ages 16 to 24, was the most literate (See Chart 15). Furthermore, the MAFL results showed that seventeen-year-olds improved their skills in the early 1970s. These results suggest that the population's functional literacy skills have improved during the past few decades.

Other evidence about functional literacy trends comes from the portions of standardized tests labelled "work-study" skills. These skills are similar to many of those used on the functional literacy tests: using indexes, alphabetizing, deciphering maps and graphs, and so forth. They cover a narrower range and are more academic in nature, but they are conceptually similar. . . . The longest time frame on functional literacy skills comes from a then-and-now study performed by Farr, Fay, and Negley (1978) for Indiana, comparing 1944-1945 performance with 1976. Using the Iowa Tests of Silent Reading, Farr and his colleagues found that sixth graders in 1976 were eight months ahead of 1944 students in alphabetizing and four months ahead in the use of indexes (p. 58). Tenth graders performed 7 percentile points higher on index usage and 2 percentile points higher on the selection of key words, a reference task (p. 69). The margins were even greater after the researchers took age into account. These results also suggest that functional literacy has improved greatly although, as we have discussed earlier, there are serious difficulties interpreting the results from then-and-now studies.

Almost as long a time frame comes from the national renorming of various standardized tests. The general pattern parallels that which we found for standardized test scores in reading and math--improvements up to 1970, a decline from 1970 to 1977 or 1978,

and recent improvements. Renorming of the Iowa Tests of Basic Skills in 1955, 1963, and 1970 demonstrated improvements in work-study skills for each time period. The average student improved two to five months, depending upon grade (Hieronymus, Lindquist, & Hoover, 1982, p. 109-111; Riverside Publishing Company, 1983, p. 6). (Also see Armbruster, 1977, p. 36-7.) The Iowa Tests of Educational Development were renormed in 1957, 1962, and 1971, and eleventh and twelfth graders greatly improved their scores on sources of information. Ninth and tenth graders improved their scores during the first period, but dropped somewhat in the second period (The Iowa Tests of Educational Development, 1971; Science Research Associates, 1978). From 1970 to 1977, scores substantially deteriorated. On the ITBS, fourth through eighth graders dropped one to four months; on the ITED, tenth through twelfth graders dropped below 1950s levels. Recently, work-study skills have improved--up two months between 1978 and 1982 on the ITBS for third through eighth graders and up one to two months on the Tests of Achievement and Proficiency for ninth through twelfth graders (Riverside Publishing Company, 1983, p. 5, p. 10).

The discrepancy between these results and those previously cited was likely due to the fact that the standardized tests measure a narrower and more academic range of skills. But, as previously discussed, using renorming results to infer literacy trends is highly questionable (See Appendix C on renorming for further discussion).

The longitudinal evidence, though of dubious quality, thus suggests that from the 1940s into the 1970s, the population's general functional literacy skills remained stable or improved somewhat, and that while students' academic functional literacy skills weakened in the 1970s, they recently improved.

3. Reading Grade Levels

Researchers also have attempted to measure the extent of functional literacy by assessing the population's reading grade level and then comparing these estimates to the reading grade level required of frequently encountered printed materials. Thus they can gauge how great a deficiency exists between reading-grade levels and various reading materials. Three major research efforts will be described.

a. The Research Techniques

The Brief Test of Literacy was developed to accompany the 1966-1970 Health Examination Survey conducted by the National Center for Health Statistics (Vogt, 1973). The test was administered to a representative sample of the nation's 12- to 17-year-olds and tested their ability to understand simple reading passages (see Chart 16).

Twenty-one questions were asked, three for each of seven passages. The literacy cutoff was established in a field test with a group of fourth graders. Although the research report says that literacy was

defined at beginning fourth-grade reading performance, the cutoff score was, in fact, set at 10.75, which corresponded to the bottom 20 percent of the fourth graders.[5] Of the nation's youth, ages 12 to 17, 4.8 percent were found to be illiterate on this 1973 test, that is, they performed at a level equivalent to that of the lowest 20 percent of fourth graders.

The second estimate comes from the Defense Department's Armed Services Vocational Aptitude Battery, which was administered to a representative sample of 9,000 18- to 23-year-olds in 1980 (Kirsch, 1985). The tests were developed for the military so they could identify qualified recruits and make assignments to military occupations and training programs. Schools can use the test to counsel students about career opportunities. The ASVAB consists of 10 subtests "designed to measure general cognitive abilities and acquired information in specific areas" (U.S. Department of Defense, 1984, p. 3). The sections include: general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension, and electronics information. Composite academic and occupational scores were produced. The three academic scores were: academic ability, verbal, and math. Interestingly, the verbal score comes from the word knowledge, paragraph comprehension, and general science subsections. The median reading level for 18- to 23-year-olds was 9.6, with 18 percent falling below the seventh grade level (Kirsch, 1985).

Corder produced the third set of estimates (Cited in Fisher, 1978, p. 36). After reviewing data from several standardized test publishers, he determined the reading levels of students in various grade levels. Thirteen percent of twelfth graders, for example, read below an eighth-grade level, while 14 percent of those in eighth grade read below a fifth-grade level. Extrapolating these figures to the general population on the basis of educational attainment, Fisher (1978, p. 36) reported that, of those fourteen years and older, 12.25 million people were reading below the fifth-grade level, and 45 million people were below the eighth-grade level. These numbers corresponded to about 7 percent of the population below a fifth-grade reading level and about 30 percent below an eighth-grade level.

These three estimates are roughly consistent, giving them face validity. When compared to the reading levels of everyday home and work materials, they suggest a serious problem. The reading level of 18 percent of young adults and 30 percent of the entire adult population lags behind common materials by several years.

Sticht (1975), for example, reports that lead articles in such well-known magazines as Reader's Digest, Saturday Evening Post, Popular Mechanics, Ladies Home Journal, and Harper's average around twelfth- to thirteenth-grade level. This has been true for the past forty years. Another study found the typical magazine article averages around eleventh-grade level (Monteith, 1980). Newspaper articles vary between ninth- and twelfth-grade level. Wire service

stories tend to be around eleventh grade, non-wire service, ninth to tenth grade. Stories on crime, local news, weather, and national political reporting is at the ninth- to tenth-grade level, while news on the economy, peace, international affairs, and nonpolitical state and national news is at the eleventh- to twelfth-grade reading level. Monteith (1980) reports that newspaper election coverage tends to be at the college level. Only best sellers have a reading level that matches the reading ability of the best of the lower 30 percent of the population. Ranging from sixth-grade level to ninth, best sellers' reading levels have averaged 7.3 for the past fifty years. As noted, though, 18 percent of those aged 18 to 23 read below the seventh-grade level. Bureaucratic forms average between eighth and sixteenth grade level. Surprisingly, one study found that the Scholastic Aptitude Test averages only around the ninth-grade level, but this still is above the abilities of many.

Many job materials also appear to be beyond the ability of the lowest 30 percent. More than one-half of the materials in seven U.S. army careers were at the eleventh-grade to college level in difficulty. Several of these careers have civilian counterparts. The reading level of materials for cooks, for example, was estimated at the ninth-grade reading level, those for repairmen at 14.5, and those for supply clerks at sixteenth grade (Sticht, 1975. p. 51). Most of 400 military technical manuals were above the ninth-grade reading level of the average soldier. Training manuals for recruits were at the eleventh-to twelfth-grade level, above average 10.7 level for recruits and well above the reading level of the lowest 30 percent.

Recent studies show that many social tasks involve materials of great reading difficulty (Wellborn, 1982). Apartment leases, for example, are written at the college level, insurance policies are at the twelfth-grade level, and an aspirin bottle labels are at the tenth-grade level. Directions for preparing a TV dinner and filling out a tax form were at the eighth-grade level. Only a driver's license manual with a reading grade level of sixth grade fell within the ready grasp of the bottom 30 percent. Ironically, food stamp notices are written at the twelfth-grade level (Kozol, citing Chall, 1985, p. 228). The list of everyday materials that many find difficult to comprehend is extensive. We end with a particularly striking one: the antidote instructions on a bottle of corrosive kitchen lye are at the ninth-grade level (Kozol, 1985, p. 10).

However, as with previous functional literacy measures, validity questions about these measures bring the findings into doubt. The Brief Test of Literacy, for example, was field tested on only 180 fourth graders. Their performance can hardly be considered nationally representative, so any claim that the cutoff corresponds to fourth-grade level is misleading. Nevertheless, the passages were simple, so the 4.8 percent who scored below 10.75 can be considered illiterate in terms of simple reading materials. As Fisher (1978) noted, the reading grade level estimates of the population are fraught with problems. The level depends upon the particular reading test considered, and it varies according to the sampling used, the types of

skills included, the difficulty of the items, and floor effects. Nevertheless, Fisher concluded that "grade level equivalents provide some of the best general measures of literacy and functional literacy skills currently available" (p. 37). We must be careful, though, in interpreting the relationship between the population's reading level and the reading level of everyday printed materials. Individuals can read and understand many materials that are rated above their reading grade level. There is no rigid dividing line between understanding and confusion, functional literacy is a continuum. On the major standardized tests, the average sixth grader comprehends about 80 percent of the material understood by the average eighth grader. This suggests that people with sixth-grade reading levels still will be able to understand much of the material written at an eighth-grade level. Keeping this is mind makes many of the previously noted less alarming.

On the other hand, we must not think that just because materials are assigned an eighth grade reading skill level that they are fully understood by a person with an eighth-grade reading level. The reading level of materials has a technical definition--that grade at which the average student can understand 75 percent of what is presented. Thus, even individuals whose grade level matches that of the material will fail to comprehend about one-fourth of it. Thus, a person with a sixth-grade reading level would understand only 60 percent of eighth-grade material (80 percent of 75 percent). Thus, a gap of two grades or more between everyday reading materials and the population's average reading level does suggest large gaps in understanding. Furthermore, for certain materials, such as antidote instructions and warning labels, 100 percent comprehension is essential; for other materials, complete understanding is not as crucial.

We must be careful, however, in using any reading grade level findings. The process by which researchers assign reading grade levels to materials has serious problems.

Researchers used readability formulas to determine a passage's reading level. Although different formulas are based on different characteristics of the passage, they are applied to passages in the same way. By counting such things as the number of one- and two-syllable words, and the number of common words, and by determining the average sentence length, researchers were able to determine whether a passage was easy or difficult and, more precisely, whether its reading grade level was low or high. The FORCAST formula, for example, which has been used to measure the reading level of job materials, derives the reading grade level of a passage from the number of one-syllable words in a 150-word section.

$$RGL = 20 - \frac{\#\text{ of one-syllable words}}{10}$$

The Flesch formulas (Flesch, 1948), developed by one of the early promoters of readability formulas, determined readability by the following formula:

readability = (1.599 x the number of one-syllable words per 100 words) - (1.015 x the average number of words in the sentences) - 31.517

(for 100-word samples)

Several questions arise immediately about the estimates of materials' difficulty. Which formula was used? The impression was created that a hard and fast RGL could be assigned to a passage, but different readability formulas will yield different results. How were passages selected for analysis? The passages may not have been representative of the text or document in question. Furthermore, assigning only one grade level to a document is misleading because different parts of it usually differ in their complexity (See e.g., Sticht, 1975, 31). But our major concern is with the readability formulas themselves. Their precision is illusory. They were derived in a questionable manner.

The first step was to determine the RGL's of a diverse set of passages. Researchers created multiple-choice tests based on each passage and administered them to a group of students. The grade level at which the average student reached an established comprehension level--usually 75 percent correct--was considered the reading grade level of the passage. If adults were tested, a standardized reading test also was administered to determine their reading grade levels. If adults with a tenth-grade reading level answered 75 percent of the questions about a passage correctly, the passage was assigned a tenth-grade reading level.

After examining many passages, an empirical relationship was established between the characteristics of the passages and their reading grade levels (see Chart 17). In this case, the number of one-syllable words in each passage is plotted against its reading grade level. The resulting line yields a formula relating RGL to the number of one-syllable words. Once the formula was derived, it was used as a shortcut in determining the reading level of new passages. Researchers simply had to count the number of words, syllables, etc., in the passages, rather than create multiple-choice tests for each passage and administer them to large numbers of people.

One problem with using readability formulas is that the relationship between the characteristics of the passages and their RGLs is imprecise. The FORCAST formula, for example, was derived from 12 passages, and the correlation between RGL and number of one-syllable words was .87 (Sticht, 1975, p. 28). Although such correlations are often labelled as "high," they in fact lead to RGL estimates from the formula that are quite different from those derived from the comprehension analysis. The FORCAST formula, for example, imprecisely estimated the reading level of the original 12 passages.

The average discrepancy was 1.2 grade levels, with one being off 2.6 grade levels (Sticht, 1975, p. 23). Passages with the same reading grade level were rated differently by the formula. Two passages, for example, had reading grade levels of 12.0 but were rated 12.2 and 13.2. A third at 12.1 was rated 11.3 (Sticht, 1975, p. 23).

These discrepancies can be even greater when applied to new passages. When the FORCAST formula, for example, was validated on a new set of 12 passages, the correlation dropped from .87 to .77. Even this is a somewhat better empirical relationship than that found with other widely used formulas. The original Flesch formula and its modified form correlated only .74 and .70 with reading grade level (Flesch, 1948).

Another imprecision enters because of the way multiple-choice tests were used to determine the passage's reading level. The RGL derivations were not done on nationally representative samples. They simply chose a single community for their test population. If a different local sample had been given the tests, the passages might have been given different reading grade levels. If different multiple-choice tests had been constructed for these same passages, different reading grade levels might have been assigned. Different standardized reading tests produce different results, so had different tests been used to assess the adults' reading grade level, passage RGL's would have been different. Finally, given the unreliability of standardized tests, if the same tests had been readministered, the adults' reading grade level might have been different. All of these problems suggest that the empirical foundation of readability formulas is shaky.

If this weren't bad enough, some researchers assigned reading grade levels to the passages without creating and administering multiple-choice tests. Researchers who developed the FORCAST formula, for example, determined the reading grade level of their twelve passages by using the cloze procedure.

In the cloze procedure, every fifth word of a passage is deleted and the individual is instructed to fill in the missing words. Completion rates of 40 percent have been found to correspond roughly to 75 percent comprehension levels. Sticht and his colleagues used 35 percent completion rates for a slightly lower 70 percent comprehension level. The cloze procedure saves the researcher the trouble of constructing a multiple-choice test, and avoids the possible pitfalls of test construction into which an inexperienced test developer might fall. In Chart 18, we invite the reader to attempt a cloze exercise based on an article about Rosalynn Carter's girlhood (Answers and discussion in Appendix E).

The cloze procedure is a substitute and, as such, produces results only roughly similar to the results of a comprehension test. Empirical results have been mixed but suggest the cloze procedure probably is an inadequate substitute. Bormuth (1967) did find a correlation of .946 between cloze and comprehension scores, but this

was an average over nine passages, so we do not know the relationship for particular passages. Only fourth and fifth graders were involved, and the range of readability levels for the materials was only from 4.5 to 6.5, which limits the generalizability of the results. In another study, Bormuth (1969) gave results by passage and sub-tests. The strongest relationship was between cloze scores and making inferences, ranging from .78 to .84. The weakest was with main ideas, .35 to .46. So for the strongest inferences the cloze test was unable to explain 30 percent to 40 percent of the comprehension variation. For the main idea, 79 percent to 88 percent couldn't be explained. These are weak supports to justify substituting cloze results for comprehension test results. Findings by Rankin and Culhane (1969) also showed little support for the substitution. These researchers presented findings by each passage and the cloze-comprehension relationship ranged from .54 to .77. The five passages averaged .68. This leaves a majority of the variation in comprehension unrelated to cloze scores and again suggests cloze results are a weak substitution. These findings also were of limited generalizability. The difficulty level of passages ranged from fifth through eighth grades, and only fifth graders were tested.

We also should note that some researchers of job literacy have used the cloze test rather than standardized reading tests to assess the reading level of adults (Mikulecky, 1981; Mikulecky & Winchester, 1983). This, too, is an unreliable substitution. As Bormuch (1969) reports, cloze tests correlate .70 to .85 with standardized reading tests. Sticht and his colleagues found .78 and .87 for two different sets of passages (Sticht, 1975, p. 23). If researchers were trying to "explain" or account for the variation in the independent variable, these results would be strong. In these cases, however, they are trying to determine whether substituting the cloze test for the comprehension or reading test is valid. There is no question that it is easier to administer the cloze test than to construct the comprehension test, but the unrelated variation is so great, the substitution has been injudicious. At the least, any results generated by cloze tests should have standard errors presented.

Furthermore, as Sticht and his colleagues point out, the cloze procedure can produce different RGLs depending upon which word the deletion process starts with (Sticht, 1975, p. 25). Although the differences were in part attributable to differences among the men taking each variation of the test, one passage ranged from 17 percent to 40 percent average completion rate, another from 35 percent to 57 percent average completion rate, depending upon which words were deleted.

Applying the formulas to assess job literacy has its own peculiarities and also produces questionable results. The problem is particularly acute when used to assess a job's demands. Sticht (1975) described two initial problems: identifying the domain of job reading materials and properly sampling them (p. 87). Because the purpose of the research is an examination of the impact of literacy, there may be a tendency to identify materials that require more reading (Sticht &

McFann, 1975, p. 69, 70). Sticht and his colleagues found that formally prescribed materials differed from those used on the job. The materials supervisors identified for the cook's job, for example, were at the eleventh-grade reading level, while those identified by the cooks were at the ninth-grade level (Sticht, 1975, p. 87).

The real issue, of course, is whether these materials were necessary to doing the job competency. Jobs are much more than their reading components. Workers are not dependent upon their reading materials. They can take shortcuts; partial understandings can suffice; asking others about a particular task can preclude the need for detailed reading; and repeated referrals to particular materials can make them understandable even if they are at "higher" levels. Sticht and his colleagues found that workers "more frequently learn and perform job tasks by watching and talking with others" (Sticht, 1975, p. 59). Diehl and Mikulecky (1980), in an in-depth, on-the-job study of workers in a wide range of occupations found that "almost 80 percent of the reading tasks cited were felt not to be necessary to completing job tasks" (p. 224). More importantly, Diehl and Mikulecky report other studies that show workers can "successfully read and apply information from job materials up to two grade levels above their assessed reading levels" (p. 225). Similarly, Sticht and his colleagues found that the reading level assigned to a job on the basis of its materials were much higher--often three to four grades higher--than the reading level workers needed to perform their jobs satisfactorily (Sticht, 1975, p. 85, 86). (We discuss these findings in a later section.)

b. Historical Trends in Reading Grade Levels

Readability formulas weren't popularized until the 1940s. A 1963 review by Klare described applied studies of readability formulas, but the studies that had been completed at that point in time were limited in scope. In the work world, they covered materials such as employee magazines and handbooks, management letters, corporate reports, union contracts, industry newsletters, and financial reports--in short, nearly everything but reading materials used on the job. A vast number of studies have been performed on school textbooks; although studies of historical changes in textbooks are interesting, they do not provide information about the gap between adult reading ability and reading materials in the work world. Few studies followed the difficulty of similar materials over time. The only information that approximated this was reviewed by Sticht (1975, p. 170). He found that various military materials had remained at roughly the same level of reading difficulty over the past three decades. Unfortunately, these were training manuals rather than job materials, and Sticht did not trace the average military personnel's grade-level reading ability over time. Those studies that related to magazines and best sellers over time have already been mentioned, but to determine truly the changes in functional literacy over time would require data on the reading grade level of the population over time.

In sum, we found no studies that tried to assess the reading grade level of the population at different points in time and to compare these to the reading grade level of everyday reading materials. Data on the educational attainment of the population at various points in time could be used to estimate changes in its reading ability over time, but this approach would be fraught with problems. There would be no way to account for changes in the quality of instruction or to determine whether the reading level of the average seventh grader or tenth grader of yesterday differs from his modern counterpart. As we have seen, then-and-now studies are complicated and provide shaky comparisons. Furthermor , we still would have the problems inherent to using readability formulas to estimate the reading difficulty of materials. An additional problem concerns the age of the formulas. An eighth-grade passage in the 1940s might differ from what would now constitute an eighth-grade passage. Therefore, we cannot say anything definitive about long-range trends in functional literacy as measured by reading grade levels.

4. Literacy and Job Performance

Many observers have worried that there is a growing discrepancy between the population's literacy level and the skills required to perform successfully most of society's jobs. In this section we discuss this argument, the research upon which it is based, and the validity of the estimates of the job-literacy gap.

Research on job literacy has a long history, but most of it has been of little worth. Efforts at synthesis suffer because existing research is fragmented, local, descriptive, and noncomparable. The most useful work is of recent vintage, which limits historical perspective. Only a few jobs have been analyzed, which limits any generalizations about the extent of the literacy gap. Of these jobs, most have been military, but this is not in itself a serious limitation because the jobs that have been examined, such as cooking and repair work, have civilian counterparts (See, e.g. Sticht, 1975).

The recent research exemplifies two general approaches, sometimes combined in the same study. The first approach is designed to yield specific quantitative estimates about the literacy abilities of workers and the literacy demands of their jobs. As we described above, researchers have determined the reading difficulty of job materials and compared these to the reading levels of workers or of the population as a whole. They also have compared workers' reading levels with their job proficiency, measured in several ways. Researchers also have attempted to determine what literacy level is associated with satisfactory performance at a given job. As we shall discuss, this line of research is fraught with problems, yet categorical conclusions have been reached about the gaps between jobs demands and worker preparation.

The second approach is more qualitative, involving interviews with workers and on-the-job observations to determine their reading

strategies, the purposes for which they read, and the content and format of the reading materials they encounter at work (Diehl and Mikulecky, 1980; Kirsch and Guthrie, 1984; Mikulecky, 1981, 1982; Mikulecky and Winchester, 1983). Unlike the first approach, the literacy demands of the job and the literacy ability of the worker are not reduced to single numbers that can be compared, but are conceptualized as a rich set of diverse skills and purposes. Researchers focus on the degree of congruence between the skills the job requires and those the workers possess, and they seek to understand ways in which the worker could acquire better skills. One advantage of this approach is that findings are based upon in-depth analysis of workers on the job; a disadvantage is that researchers have been able to study only a limited number of workers--for example, 99 in Kirsch and Guthrie's case, 107 in Diehl and Mikulecky's, 27 in Mikulecky and Winchester's. Given their complex purposes and the fact that only a few such studies have been performed, this research does not lend itself to easy generalization. It does suggest, however, the limitations of a reductionist numerical approach.

We turn now to the first line of research, which has produced the generalizations about the growing job-literacy gap. The military research is the best example.

a. The Research

In response to manpower problems during the Vietnam War, military officials embarked on a functional literacy research program. They were concerned that many draftees were not sufficiently literate to handle military jobs and wanted to determine how best to close the gap between personnel skills and tasks. Although researchers did make the traditional estimates of the reading level of job materials, they also directly compared the reading levels of personnel to their job performance. This is a more accurate method of assessing job literacy than relying upon readability formulas. Sticht and his colleagues performed much of this monumental research, and their report forms the basis of our discussion (Sticht, 1975). Much of our critique is derived directly from their own self-criticisms.

Sticht and his colleagues developed four measures of job proficiency. The first, called the job-reading-task test, involved reading tasks of the same type and format that workers encountered on the job. Cooks, for example, were given cookbooks and asked questions that required them to look up recipes in order to answer correctly. The second measure was a paper-and-pencil job-knowledge test keyed to the worker's particular job. The third measure was the job-sample test, in which the workers were graded on typical tasks from their jobs, including reading and nonreading aspects. Repairmen, for example, fixed vehicles with the usual manuals available. The fourth measure was supervisor ratings, derived from a standard army evaluation report and a questionnaire designed to assess the worker's competence (Sticht, 1975, p. 61-2).

The researchers examined four jobs: cook, repairman, supply clerk, and armor crewman. They found a range of reading levels for each, depending upon the proficiency measure and the level of job performance considered satisfactory. The reading grade levels were seventh through ninth grades for cooks, eighth through twelfth grades for repairmen, ninth through thirteenth grades for supply clerks, and eighth grade for armor crewmen.

Some of these ratings are above the ninth-grade average reading level of Army personnel, suggesting that many people in the Army would have trouble at reading tasks on their jobs. Furthermore, in 1973, around the time of the research, 12 percent of the new recruits read below the sixth-grade level. In the early 1970s, 20 percent of recruits were classified in Mental Category IV, many of whom read below the seventh-grade level (p. 169-1970). As noted above, in 1980, the Armed Services Vocational Aptitude Battery showed that 18 percent of the nation's eighteen- through twenty-three-year-olds read below the seventh-grade level. The research results led Sticht and his fellow researchers to conclude: ". . . the evidence is overwhelming in indicating that many adults, young or older, cannot read and use with facility much of the written materials needed to function well in our society" (p. 184).

The research, however, has serious deficiencies. The relationship between a serviceman's reading level and his job proficiency was so weak as to suggest that literacy had little to do with job performance. The correlations between personnel's reading grade level and three of the job-proficiency measures were quite low. Depending upon the job, reading grade level correlated from only .40 to .57 with the job-knowledge test, .26 to .40 with the scores on the job-sample test, and were unrelated to supervisor ratings. Even the relatively higher correlation with the job-reading-task test can be explained because the test was, by necessity, a reading test. Furthermore, it was given only to recruits in training and thus does not tell us about job performance. The absence of any relationship with job-supervisor ratings cannot be explained away by attributing it to the imprecision or bias of the typical supervisor evaluations. In this case, two detailed evaluation questionnaires were completed. Furthermore, any bias that reduced the actual correlation would have penalized those with greater literacy, which is unlikely.

As Sticht noted (1975, p. 68), however, the relationship between grade level and the three proficiency measures was likely attenuated since they were dealing with workers already on the job. Presumably, many workers with poor reading skills had already demonstrated their poor performance and had been transferred. It is true that few of the servicemen they studied read below the fifth-grade level. Still, 44.5 percent of them read below the eighth-grade level and nonetheless performed their jobs satisfactorily (p. 63-64). Contrary to the "literacy" gap perspective, low reading-achievement scores (fifth through eighth) were not <u>per se</u> a barrier to job competence among these subjects.

Other studies also have shown a weak relationship between reading level and job performance. Mikulecky and Winchester (1983), for example, in a study of nurses, found no relationship between general reading ability, job reading ability, and job performance rating. The range of reading performance was wide--for those judged competent, job reading level ranged from 8.3 to 15.8 grade level. In a study of workers in diverse occupations, Mikulecky (1981) found that some workers (5 percent) were severely limited in their ability to read a ninth-grade passage. Mikulecky commented: "It is possible, it seems, to hold a job if one can barely read" (p. 179). Obviously, severe reading deficiencies would interfere with the ability to acquire and hold many jobs, but above a certain th shold, reading level has little to do with job performance. The real issue in the "gap" debate is what percentage of jobs have high reading demands and what percentage of individuals lack the necessary skills to meet those demands. We address this further on.

In spite of such weak correlations, Sticht and his colleagues proceeded to determine the reading grade level associated with "job competence." This introduced the familiar criterion problem into their analysis. At one point, Sticht and his colleagues defined functional literacy as "that level of reading ability that is minimally sufficient for satisfactory job performance" (Sticht, 1975, p. 75). But what constitutes "satisfactory?" There usually are two components to the criterion level--a particular percentage of correct answers and the percentage of individuals who reach that level of correct answers. Thus, on the job-reading-task test, Sticht and his colleagues chose an 80/70 criterion--that is, the job literacy level of a given job would be defined as that reading grade level at which 80 percent of the individuals could answer 70 percent of the job reading items correctly. Varying either the percentage of individuals or the percentage correct would have altered the reading grade level assigned to a job. With the 80/70 criterion, they rated the repairman's job at a ninth to tenth-grade reading level. Elsewhere, Stich et al. noted that the military usually uses a 70/70 rule (70 percent getting 70 percent correct). This would lower the assigned reading grade level of repairmen to eighth grade, 1.5 grades lower (p. 113). If, instead of lowering the percentage of individuals, we lowered the percentage correct (for example, by arguing that the tasks on the test were too heavily reading-laden to be properly representative of the job itself), and if we used an 80/60 criterion (i.e. 80 percent getting 60 percent correct), the reading grade level of repairmen drops to about seventh grade--2.5 grades less than the 80/70 results. For the supply clerk, the results when moving from 80/70 to 80/60 are even more dramatic: the reading grade level drops from thirteenth to tenth grade, or three full grades. What criterion level is the proper one? If the reading grade level of jobs is used to exclude individuals from consideration for a job or as evidence of a literacy problem, an adequate justification of the criterion is necessary.

In response to the criterion problem, Sticht and his colleagues properly argued that "there is, then, no single unitary skill to be

designated as the job reading level requirement. Rather, there are as many levels of reading requirement as there are levels of job proficiency . . ." (Sticht, 1975, p. 83). Choosing a criterion, they argue, is a "judgmental decision" (p. 76). Furthermore, they point out that factors such as the supply of labor and the availability of literacy training may affect how low the criterion level will be set (p. 57). Thus, Sticht et al. are aware of the problems involved in selecting a criterion. Yet they conclude that determining a sufficient reading proficiency level for a job is not the proper province of the researcher, but rather a decision to be made by responsible managers (p. 83). This contention is disingenous since they not only establish criteria that are different from those of the management, but they also explicitly criticize at least one criterion level established by management (p. 82). If their dictum were followed, management would have the power to establish arbitrary, inflated, or discriminatory requirements for jobs. Unless the researchers explore what constitutes satisfactory reading performance on the job, workers may be left at the mercy of employers. Unfortunately, in this most crucial of matters, Sticht and his colleagues failed to apply their typical ingenuity. They might have used interviews with supervisors and workers about relative levels of competence and about what elements of workplace behavior comprise unsatisfactory performance. Tests to measure these skills could have been developed. Instead, they developed their own criterion levels without justification and entirely independent of the management's 70/70 standard. The reading level assigned to a job was that at which individuals were not over-represented in the bottom quarter of job proficiency. For example, 33 percent of armor crewsman at the seventh- to eighth-grade level fell in the bottom fourth of job proficiency while only 22 percent of those in the eighth- to ninth-grade level did, so an eighth-grade reading level was assigned. This procedure defines no performance level as satisfactory; the criterion is entirely relative. Let us say, for example, that everyone in the bottom quarter should be considered unsatisfactory performers. If so, 22 percent of those with eighth-grade reading ability are not performing well, while 67 percent of those with seventh-grade reading level are! Which is the appropriate reading level? No research findings on job literacy demands can be accepted without a specification of the criterion and a justification of its selection.

Even if these research designs were satisfactory, too few jobs have been examined to assess the workplace gap of the society. There are more than 13,000 different jobs in the society, yet only a few have been examined in these studies. Sticht and his colleagues, for example, studied seven for readability and only four for job proficiency. Mikulecky (1981) examined jobs in four broad categories: professional/technical, retail/clerical, service, and blue collar, but he ignored distinctions between and within subcategories.

Even if we show that the reading demands of many jobs are greater than the average reader's reading grade level skill, we can conclude little. Only if the majority of jobs required more skill than the

average person possessed could we justifiably argue that there was a general literacy "gap" in the population. The only proper way of determining whether there is a literacy gap is to compare the distribution of reading ability among the population to the distribution of reading demands in different jobs. Lerner (1981), attempting to do this, argued that unskilled people considerably outnumbered unskilled jobs (p. 1060). She based her claim on the Mini-Assessment of Functional Literacy study, which showed that 12.6 percent of seventeen-year-olds were functionally illiterate, twice the Department of Labor's estimate that 6.1 percent of jobs were unskilled in 1970. The comparison is misleading. Those who were labelled functionally illiterate by the MAFL were not "unskilled." The functional illiterate label was applied to those who scored below the 75-percent criterion. Only 2.9 percent, however, scored below the 60-percent criterion. Thus, most had some literacy skills. Furthermore, the MAFL was focused on general functional skills like reading maps, using the dictionary, reading a telephone directory, understanding tax instructions, interpreting traffic signs, and reading labels. Failure to do well did not necessarily mean that an individual lacked minimal job-reading skills. Also, as we have seen, an individual's job-reading levels are typically higher than general reading levels on standardized tests. Lerner presented no evidence of the reading skills necessary on these jobs. The skills of most of these teens probably were greater than those required in unskilled jobs. Even semi-skilled jobs appear to require little reading. In a study of a black working-class community, Heath (1980) examined closely the reading demands of residents at home, play, school, and work. Observing the lives of 90 semi-skilled workers, she found that "on the job, community members were not often called on to read" (p. 129). Their job applications were filled out by a personnel officer. Employees were instructed orally about their new jobs. Insurance information and new regulations were posted, but since these were usually explained orally, employees "did not find it necessary to read the bulletin board notices" (p. 130). Foremen were not sent memos describing new production strategies but were briefed and in turn explained the changes orally to the workers. Time charts and safety records did require some reading and writing, but they were routinely completed without difficulty. Heath's findings are by no means definitive. Kirsch and Guthrie found that reading was a regular part of the job for semi-skilled workers. A major difference between the studies, though, was that Kirsch and Guthrie focused on the Analytical Instrumentation Division of a Fortune 500 company. Only 33 workers were interviewed, and these were service as well as semi-skilled workers. If, indeed, most semi-skilled workers were required to read only rarely, a very large percentage of low-literacy jobs would be available. The literacy gap would vanish.

An alternative, and perhaps better, barometer of job readiness is the Brief Test of Literacy, which showed that 4.8 percent of those aged twelve to seventeen were reading below the fourth-grade level. For several reasons, it makes sense to accept this level as the proportion of "unskilled" people. The military results, for example, suggested that those with a fifth- to eighth-grade reading level could

hold their jobs, and the Mikulecky study found that 5 percent of those holding jobs were poor readers. This proportion (about 5 percent) is lower than the percent of unskilled jobs (about 6 percent). It is lower still than the percentage of unskilled and semi-skilled jobs combined. From this perspective, there is no job-literacy gap. The figures suggest that at least for the lowest literacy levels, the dimensions of the alleged gap have been exaggerated.

A more comprehensive index of job reading requirements is available through the Department of Labor's Dictionary of Occupational Titles, which identifies 13,800 U.S. jobs and assigns to them among other things, an index of General Educational Development (GED). The GED is a measure of the verbal, reasoning, and mathematical skills necessary to perform the job. This classification provides a national data base of jobs and their requirements; it has been used by researchers to judge the gap between workplace demands and worker preparation. In spite of its national scope, however, this data has several limitations for an analysis of workplace literacy demands.

The GED scores are based upon formal job descriptions rather than analysis of the jobs' reading demands. Two concerns emerge here. Formal descriptions may not match actual job tasks. GED levels are assigned subjectively. Sticht and McFann (1975), for example, criticized the lack of specificity in the rules analysts use to assign GED scores (p. 73). (See also Fine, 1968, p. 367.) Other critics have suggested that analysts were rating the social standing of the occupation rather than the skills the job required (See, e.g., Spenner, 1980, p. 247). Second, the verbal, reasoning, and mathematical demands are estimated separately for each job on a six-point scale (seven points in the 1950s). The published GED score is the highest of the three ratings. This means that the GED does not necessarily indicate the reading demands of the job. A job that seems to have high mathematical but low reading demands, for example, will receive a high GED score. Furthermore, the verbal component is not limited to reading skills, but also refers to speaking and listening (U.S. Department of Labor, 1965, p. 652). GED scores often are translated into educational levels, yet the Department of Labor deliberately rejected such an approach because of variations in the quality of schooling and the possibility of learning from experience (Berg & Gorelick, 1971, p. 43). Years of schooling is not much help anyway because, as we noted, the GED is keyed to the highest of the three skill areas rather than to reading. Furthermore, there is no fixed relationship between GED scores and years of schooling. Berg's analysis provided five different models based upon five different assumptions of the relationship between the two. Depending upon which assumptions were used, a worker preparation gap emerged or disappeared. There is no definitive way of determining which of the five models is the most reasonable. As Berg and Gorelick admitted, "In each case, some assumptions are attractive and some are unacceptable" (p. 58). Finally, since GED scores are derived from job descriptions rather than job performance, we do not know what GED levels are associated with what levels of job proficiency. For these

reasons we have not pursued further the Dictionary of Occupational Titles for use as a data base.

A final weakness of the reductionist quantitative approaches is their traditional concept of reading. The reading grade levels of workers are measured by standardized tests (or by cloze tests substituting for them). Similarly, the difficulty level of the job materials are determined from formulas derived from standardized tests. Unlike many errors in the research, this one serves to understate the gap between real job reading demands and workers' reading proficiency. In-depth research has demonstrated that schools develop a set of literacy skills unlike those typically needed on the job (Sticht, 1975, p. 183-6; Diehl and Mikulecky, 1980, p. 224-5; Kirsch and Guthrie, 1984, p. 231). As Sticht noted, high school English teachers focus on the reading and interpretation of literature and on general composition. They usually ignore technical writing and editing, or job materials like technical manuals, advertising copy, flow charts, memoranda, government pamphlets, and guidebooks. Mikulecky has noted that general reading strategies differed considerably from reading to accomplish a task, and evaluating the usefulness of material. He found that job-reading tasks are more integrated with other tasks, more immediately applied, and more repetitious than school-reading tasks. He and Diehl (1980) distinguish "reading-to-do" and "reading-to-learn" and hypothesize major differences in information processing between the two types of reading (p. 225).

This analysis suggests that the standardized tests that measure school learning are inappropriate to analyzing workplace literacy demands. Kirsch and Guthrie also have questioned the construct validity of tests on these grounds and further argued that merely sampling job materials is insufficient since job-reading processes do not inhere in the materials. Redesigned tests, focused on job-reading skills, might nonetheless show a relationship between literacy and job performance. Indeed, although Mikulecky and Winchester (1983) found that general reading ability and the reading level of job materials did not relate to job competence, "reading to assess while performing a task" did. Sticht and his colleagues could have used this approach. Their job-reading-task tests more closely captured job-reading skills than standardized tests. They could have been used to relate worker's reading skill to the job proficiency measures. Instead, the authors used general reading ability as measured by standardized tests, and they found only weak or insignificant relationships.

b. Trends Over Time

Given how poor our knowledge of reading in the contemporary world is, it is not surprising that our knowledge of historical trends in job literacy demands also is very limited. The evidence is thin, and the earlier work suffers from the same defects as the research of the 1970s. Two broad types of evidence can be described: changes in the

occupational structure and empirical studies of workers' skills in different jobs.

Chart 19 shows some changes in the occupational structure during the century. As can be seen, between 1900 and 1970, the occupations that expanded most were in the white-collar categories, particularly professional-and-technical workers and clerical workers. The decreases were for nonfarm laborers, farmworkers, and private household workers. If we assume that white-collar jobs require a higher level of literacy, high-literacy jobs increased from 17.6 percent to 48.3 percent and low-literacy jobs decreased from 55.4 percent to 10 percent. This is a major transformation, but it is not a very helpful finding. Even from a simple quantitative perspective, we do not know what reading grade level is required in these different jobs or how it has changed over time. The low-literacy job may have required much more reading in 1970 than in 1900. Farmworkers include managers, farmwork has become more technical, and household workers face greater reading demands. From a qualitative perspective, we do not know what kinds of literacy skills have been needed or how these have changed. Finally, we do not know how literacy demands vary by job within these categories. All we can say is that over the course of the century the literacy requirements of work in general probably increased. The limited usefulness of this finding is clear when one considers changes in the educational level of the population. During the same time, those with little education (0 to 4 years) dropped from 23.8 percent (1910) to 5.3 percent (1970), while those with 0 to 8 years dropped from about 45 percent (1910) to 9.7 percent (1980). High school graduates increased from 13.5 percent (1900) to 55.2 percent (1969). Thus, while literacy demands have increased, so have educational levels. We cannot determine on the basis of these data whether one has outstripped the other. The issue is further complicated because we are not interested simply in years of school attended but in actual reading ability.

Since its inception, empirical research on job skill levels has been primarily designed for job counselling purposes and for personnel selection by large organizations. For both purposes, employers began by giving workers skill tests, usually written, although sometimes mechanical tasks, covering one or more of the following: general intelligence, clerical skills, manual dexterity, verbal ability, finger dexterity, mechanical assembly, spatial relations, numerical ability, vocabulary, form perception, motor coordination, self-sufficiency, and social dominance (Dodge, 1935; Dvorak, 1935; U.S. Department of Labor, 1979). Researchers have then created Occupational Ability Patterns (OAPs) that portray the levels of particular skills supposedly required for jobs. By comparing an individual's skill profile to the OAP, counsellors and personnel managers can determine the best assignment of the individual. Only recently have researchers gone on to determine the general population's possession of these particular skills. In principle, this makes possible a workplace literacy gap assessment. Unfortunately, it has yet to be done. Our review of the secondary literature did not reveal any such study.

Empirical approaches to job skill assessment were pioneered for the Army during World War I (See Yoakum and Yerkes, 1920). In one approach, the job was simply assigned a reading level equal to the median level of test scores of workers in that job; in a second approach, the level assigned depended upon the relationship between test scores and job performance measures. The first approach is not very useful for estimating trends in workplace literacy. In many studies there was no test of reading, and in others it was included only as part of a composite score (Yoakum & Yerkes, 1920; see review of early research in Dodge, 1935; Dvorak, 1935). Consequently, most of this research cannot help us determine literacy skills among workers, or historical trends. The results also were inconclusive. The data could not clearly differentiate jobs by reading demands because workers' skills in different jobs overlapped greatly. Berg and Gorelick (1971) noted that variations "within occupational groups have been found to be as great as variations among these groups" (p. 41). Dodge (1935) found the same and concluded that OAPs based on median or average scores were of "little value for guidance purposes" (p. 76). They may be overstating the case. Tests of clerical skills do seem to distinguish clerical workers from lower-literacy workers (Dvorak, 1935), and a substantial performance appears to exist between workers in very high and very low prestige jobs (Youkum & Yerkes, 1920, p. 198-9). Still, the vast overlap among individuals' skill levels for the bulk of occupations undermines this approach to rating job literacy levels. Furthermore, there was no observation of the actual job situation in this research. The skills the tests measured might have been different from those essential to the job, or the measured skill levels could be artificially high or low. Higher levels might have been the result of job entrance requirements not directly relevant to the work at hand (credentialling effects), or they might have reflected learning on the job, and thus workers with lower scores could enter and successfully perform the job. Finally, we do not know how the skills levels related to job proficiency. Consequently, studies such as Stewart's (1947) comparing the relative standing of World War I and World War II Army occupations are of little value for determining historical trends (Cited in Baer & Roebar, 1977).

The second approach also has met with little success. Sometimes the empirical relationship between test scores and job proficiency has been too low to assign a job level; sometimes the results have been mixed, making assignment ambiguous. World War I results, for example, demonstrate this mixture of high and low correlations. Job performance rankings of 765 men by infantry company commanders were nearly identical to those from the Army test of job skills (Yerkes & Youkum, 1920, p. 32). But officer ratings of 374 men in 12 other companies correlated only .536 with test scores (p. 30), and Army rank, excluding medical officers, showed no relationship to test scores (p. 40). Extremely high or low test performance was correlated with extremely high or low officer ratings, but middle performance was a poor discriminator (p. 33). Prediction of job success varied by job type. Test scores correlated positively with clerical workers' job performance, but were negatively correlated with machine operators'

performance (p. 201). Given the nature of the Army test, and the select sample being studied--only white young men in a draft characterized by high rejection rates--these results are of little use for historical comparisons, save to illustrate the futility of the method.

The business world's experience with such testing also reveals its limited usefulness. In the 1920s and 1930s, industry abandoned most employee testing because it had failed to be a good predictor of on-the-job success (Hale, 1982, p. 18-20). Primarily, clerical testing was retained. After World War II, corporations reembraced testing, using personality tests in its quest for the organization man (Hale, pp. 29-30). Clerical testing, however, still predominated.

In the post-war period, the most comprehensive occupational-aptitude testing program has been that of the U.S. Employment Service. Its General Aptitude Test Battery has been validated in more than 550 studies, covering a representative sample from 12,000 jobs. Depending upon the job, however, the correlations between GATB scores and civilian job performance ranged from only .23 to .58 (Department of Defense, 1984, p. 19). The Armed Services Vocational Aptitude Battery has produced similar outcomes, with correlations of .36 to .52 for jobs within the communications area, .39 to .77 for data processing specialties, and .53 to .73 for clerical and supply specialties (U.S. Department of Defense, 1984). The success measure, however, was not job performance, but training performance.

Contemporary sociological research also has found very weak relationships between test scores and job success (r=.3) (Jencks, et al., 1972, p. 186-7). These results demonstrate a number of things about predicting job success and hence assigning a test score to a job. First, the ability to predict success varies greatly by type of occupation. Second, within a given occupational category, it varies greatly by particular job. Third, the ability to predict success is very poor for many job specialties. For many jobs, test scores were such that no level, or very low levels, would be assigned. These facts make a national assessment of the job-literacy gap difficult.

We also are back to the problem of what constitutes an acceptable success criterion. We need to know what percentage of those who score below the test score criterion are performing satisfactorily. Many of the criticisms raised more than 50 years ago by Dvorak, one of the earliest and most optimistic practitioners, still apply (Dvorak, 1935). More jobs need to be studied, more skills need to be tested, and tests need to better assess the demands of particular occupations.

A national analysis using carefully reviewed GATB or ASVAB job levels, which determined the proportion of jobs at particular levels and the proportion of adults at these levels, might provide some measure of the workplace literacy gap. Unfortunately, the verbal measure of the GATB is a vocabulary test, while that of the ASVAB is a combination of word knowledge, paragraph comprehension, and, rather

arbitrarily, science knowledge. Better predictors would be needed for a national assessment. Alternatively, if GED job ratings were extensively validated through observations of workers, and a national profile of adults' language skills was assembled, similar findings could be developed. Although the GED analysis would be more complicated, it would be more compelling because it covers a range of language skills--not just vocabulary, as on the GATB. Before embarking on such an ambitious project, however, researchers should recognize that the work likely would have limited value because of the wide variations in individuals' reading grade levels within a given job. Elton Mayo was one of the researchers in the Hawthorne studies, which demonstrated that social relations in the workplace were more important than individual aptitude or skills in determining performance. He concluded, "The belief that the behavior of an individual within the factory can be predicted before employment upon the basis of a laborious and minute examination by tests of his technical and other capacities is mainly, if not wholly, mistaken" (quoted in Hale, 1982, p. 19).

Given the difficulties of measuring job-literacy demands, and given the limitations of the research to date, the reader should by now be aware that estimates of the workplace literacy gap projected into the future are hardly credible. Well-informed scholars have propounded quite different visions of whether job-literacy demands are increasing or decreasing. Those who wrote the recent educational reform reports, including the President's National Commission on Excellence, certainly argued that the nation was rapidly moving to a high-technology future that would require far different and more advanced skills than are currently possessed. Kirsch and Guthrie (1984), noted literacy analysts, have argued that jobs requiring little or no literacy are rapidly disappearing and that new, changing jobs require more skilled, literate people. By contrast, Bowles and Gintis (1976) and Bowles (1979) have argued that white-collar work has been proletarianized and that jobs are being "dumbed down." They estimate that most work requires only limited skills compared to those individuals possess, those schools could produce, or those that could be learned in restructured jobs. The federal commission that produced the 1970s report Work in America viewed the world differently from those who produced A Nation At Risk. They found a crisis in over-education, with credential effects being used unfairly in job screening, and recommended a vast expansion in worker control over jobs and working conditions (O'Toole, et al., 1973). Department of Labor projections suggest that most new jobs will be in retail and clerical work rather than in engineering or computers (U.S. Department of Labor, 1982). Nor will the expanded use of computers necessarily require a more skilled workforce (Levin and Rumberger, 1983). Much of the work associated with computers, such as data entry, is routine. Certain computer programs, such as grammar and spelling checkers, decrease the need for literacy. If the Work in America report is correct, any increase in the skills demanded by jobs would be welcome, for it would mean that the skills demanded were finally catching up with the skills possessed by an "over-educated" workforce. On the other hand, the vast sums of money spent by major corporations on

remedial training for workers suggest that many entry level workers lack basic skills (K&G, 1984).

Resolving this debate would require better research on the literacy gaps at various occupational levels. Probably both trends--dumbing down and rising literacy demands--are happening at different levels in the occupational pyramid and in different regions. Because we are not sanguine about the prospects of a valid assessment of the workplace literacy gap, we do not call for further research. Whether functional literacy is 10 percent or 35 percent, whether the population's reading grade level is seventh or ninth grade, and whether the job-literacy gap is growing smaller or larger is not as crucial as recognizing that substantial problems exist in the match of workers' literacy skills and jobs' reading demands. Workplaces must allow greater autonomy and skill development by workers; we need to face the unmet literacy needs of many people with expenditures of time, effort, and money at both the local, state, and national level. Legislation, referenda, volunteer programs, and employer-employee collaboration all must be adopted in an effort to change the ways people are trained and treated on the job.

III. CONCLUSION AND DISCUSSION

We have attempted to trace trends over time in literacy and reading ability in the United States during the past century. Our main conclusions, as the patient reader knows, are that the data is sketchy, the research is shaky, and the trends are murky. Problems of concept validity, representativeness of research samples, and noncomparability across time bedevil the attempt to discern trends.

As a consequence, present-day literacy policy should be argued on the basis of an assessment of our current condition, difficult enough in itself to make, and on the basis of shared goals for our schools and other educational institutions--not on the basis of alleged declines or rises in literacy skills. The trends are in doubt; the existence of literacy problems in our society is not.

As we have argued elsewhere (Stedman and Kaestle, 1985), we favor strenuous efforts to improve our society's record on basic literacy skills and on teaching all children critical reading and thinking skills.

As for historical trends, our counsel was skepticism and caution. What can we say to cut through this agnosticism?

First, during the twentieth century, self-reported outright illiteracy almost disappeared as a percentage of the whole population. The rates for various groups that previously had reported substantial outright illiteracy--women, blacks, immigrants, Southern whites--converged and shrank to below 5 percent. Still, when we look at numbers instead of percentages, even this self-reported illiteracy is troubling. In 1979, nearly a million Americans were rated as illiterate. On the other hand, it is no major accomplishment that outright self-reported illiteracy has been reduced so drastically during the past one hundred years.

Second, the big story in twentieth-century literacy is the rise in school attainment, not the relative effectiveness of our schools to teach children at a particular grade level. Although some children, clearly, attend school without learning much, it seems indisputable that a rise in schooling of the magnitude witnessed in the United States since 1880 has led to a much more literate population. In 1910, for example, almost one-fourth of the population had less than five years of schooling; by 1980 only 3.3 percent had so little. In 1910, almost half the population had less than an eighth-grade education; by 1980, that proportion was under 10 percent.

Third, we find our greatest difficulty in making a confident statement about the reading abilities of people at different points in time with the same amount of schooling. Most of the debate about reading achievement has centered on this measurement problem. Through the haze we squint, and we venture the conclusion that there probably has not been much of a decline in reading ability at a given

educational level during the twentieth century, even during the controversial 1970s.

Does this mean that things are rosy on the literacy front? Certainly not. The functional literacy tests showed clearly that a substantial portion of the society--from 15% to 35%--has difficulty coping with everyday reading tasks and materials. The job-literacy measures, for all their limitations, show that there are substantial mismatches between workers' literacy skills and job reading demands. Even if schools are doing more-or-less as well as they have in the past, they always have needed improvement in educating minorities and the poor and in teaching higher order skills. And if increased education is the only reason the population has kept up with the increasing literacy demands of our society, we have plenty to worry about. School attainment is no longer rising and school dropout rates are increasing. The solution to rising literacy demands in our society--once simple--is now more difficult. And even if the workplace per se is not truly demanding more reading ability, we shall nonetheless need much better reading skills across the whole spectrum of our population if we are to survive and improve as a democratic society in a highly technological age. Seen in this light, there is much to galvanize renewed efforts at literacy training, at all levels. We need no proof of a great decline to make us concerned.

Footnotes

[1]Copperman erred because if he was right that students started school at the same time, then the stricter retention policies of 1937 would have made the students of the same age more likely to have repeated a grade and thus to have received more schooling. Copperman noted that over 30 percent of the students had repeated one grade out of the first six during the late 1930s, while it was only 15 percent in the late 1950s (p. 33). Armbruster failed to present any evidence that students in 1957 started school earlier and did consider retention effects. He was merely speculating about the 1937-1957 contrast. Because his particular interpretation of the Gates study was the basis for his claim that there was a decline in reading achievement in the 1940s, we consider that assertion unfounded.

[2]The 1950s norms for two of the three tests that Elligett and Tocco used excluded private schools. Had they been included, the decline they found would have been greater. Age differences of the Indiana magnitude, however, still could offset this.

[3]It also should be noted that several of the tests that showed declines reported scores for total reading only, so we do not have separate scores for vocabulary and reading comprehension. Performance on the two can be quite different. The CTBS 1973-1980 renorming showed that 1973 students did equally well in these two areas, but that by 1980 students were doing much better on reading comprehension than on vocabulary. The poorer vocabulary performance dragged down their overall scores to below 1973 levels, but their reading comprehension scores appear to be greater than those of 1973 students (Analysis from CTB/McGraw-Hill, 1982a, p. 59t; CTB/McGraw-Hill, Equating Tables, 1982b).

[4]We compared the ratio of those who were "functionally illiterate" to those who were "marginally functional" since researchers did not break down their "functional illiterate" category any further. On Harris I, the ratio of those who scored below 70 percent to those who scored between 70 percent and 90 percent was one-eighth for sixteen- through twenty-four-year-olds and five-twelfths for those 50 years and older. On Harris II, comparing those below 80 percent to those between 80 and 90 percent, the ratios were one-eighth versus three-fourths. On the APL, the ratio of APL 1 to APL 2 (functionally incompetent to marginally competent) was three-sevenths for eighteen- through twenty-four-year-olds and two-thirds for sixty- through sixty-five-year-olds.

[5]Scoring was done with a correction for guessing, thus the number of right answers minus one-fourth the number of wrong ones. For someone who answered all questions, therefore, the cutoff was 13 correct answers (Donlon, McPeek, & Chathaw, 1973, p. 10).

REFERENCES

Acland, H. (1976, July). If reading scores are irrelevant, do we have anything better? Educational Technology, 25-29.

Admissions Testing Program of the College Board. (1983). National report on college-bound seniors, 1983. New York: College Entrance Examination Board.

Adult Performance Level Project. (1977). Final report: The Adult Performance Level study. Washington, D.C.: U.S. Office of Education.

Advisory Panel on the Scholastic Aptitude Test Score Decline. (1977). On further examination. New York: College Entrance Examination Board.

American College Test. (1985). Table. ACT score means & SD's for successive years of ACT-tested college-bound students. Iowa City: American College Test.

Armbruster, F. (1977). Our children's crippled future: How American education has failed. New York: Quadrangle Books.

Auwers, L. (1980). Reading the marks of the past: Exploring female literacy in colonial Windsor, Connecticut. Historical Methods, 13, 204-214.

Baer, M. F., & Roeber, E. C. (1977). Occupational information, 94-98. Chicago: Science Research Associates, Inc.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53(4), 571-585.

Berg, I., & Gorelick, S. (1971). Education and jobs: The great training robbery. Boston: Beacon Press.

Berk, R. A. (Ed.) (1982). Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Bloom, B. S. (1956). The 1955 normative study of the tests of general educational development. School Review, 64(3), 110-124.

Bode, R. K. (1981a). SRA achievement series (Technical Report, Number 2). Chicago: Science Research Associates.

Bode, R. K. (1981b). SRA achievement series (Technical Report, Number 3). Chicago: Science Research Associates.

Borkow, N. (198?). Analysis of test score trends: Implications for secondary school policy--A caution to secondary school administrators. Washington, D.C.: National Institute of Education.

Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 10(4), 291-299.

Bormuth, J. R. (1969). Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 4(3), 358-365.

Bormuth, J. R. (1973-4). Reading literacy: Its definition and assessment. Reading Research Quarterly, 9(1), 7-66.

Bormuth, J. R. (1978). The value and volume of literacy. Visible Language, 12, 118-161.

Bowles, S. (1979, April 1). Second thoughts on the capitalism-enlightenment connection: Are Americans over educated or are jobs dumb? Address to Conference on Libraries and Literacy National Commission on Li' raries and Information Services. Washington, D.C.

Bowles, S., & Gintis, H. (1976). Schooling in Capitalist America. New York: Basic Books.

Breland, H. M. (1976). The SAT score decline: A summary of related research. In appendix, Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Breland, H. M. (1977). Family configuration effects and the decline in college admissions test scores: A review of the Zajonc Hypothesis. In appendix, Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Bridge, R. G., Judd, C. W., & Moock, P. R. (1979). The determinants of educational outcomes. Ballinger Publishing Company.

Brimelow, P. (1983, September 19). What to do about America's schools. Fortune, 60-64.

Buros, O. (1978). Fifty years in testing. In Buros, (Ed.), The eighth mental measurements yearbook. Highland Park, NJ: The Gryphon Press.

Buros, O. (Ed.) (1978). The eighth mental measurements yearbook. Highland Park, NJ: The Gryphon Press.

Buswell, G. T. (1937). How adults read. Supplementary educational monographs published in conjunction with The school review and The Elementary School Journal, Number 45, August. Chicago: The University of Chicago.

Carroll, J. B., & Chall, J. S. (1975). Toward a literate society. New York: McGraw-Hill. Connecticut: Linnet Books.

Carver, R. P. (1972). Reading tests in 1970 versus 1980: Psychometric versus edumetric. The Reading Teacher, 26, 299-302.

Caughran, A. M., & Lindlof, J. A. (1972, March). Should the "Survival Literacy Study" survive? Journal of Reading, 429-435.

Cervero, R. M. (1980). Does the Texas adult performance level test measure functional competence? Adult Education, 30(3), 152-165.

Cleary, T. A., & McCandless, S. A. (1976). Summary of core changes (in other tests). In appendix, Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Coles, G. S. (1976). U.S. literacy statistics: How to succeed with hardly trying. Literacy Work, 5(2), 47-70.

Copperman, P. (1978). The literacy hoax: The decline of reading, writing, and learning in the public schools and what we can do about it. New York: William Morrow.

Copperman, P. (1979). The achievement decline of the 1970s. Phi Delta Kappan, June, 736-739.

CTB/McGraw-Hill. (n.d.) CAT A-C equating tables. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1974a). Comprehensive tests of basic skills, Form S. all levels (Technical Bulletin, No. 1). Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1974b). Technical report: All levels, California Achievement Tests, 1970 Edition. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1977). Comprehensive tests of basic skills, Forms S and T, all levels (Technical Bulletin, No. 2). Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1979). California achievement tests: Forms C and D level (Technical Bulletin 1), 10-19. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1982a). Comprehensive tests of basic skills: Preliminary technical report forms U and V. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1982b). Equating tables for CTBS U and V and CTBS S and T. Monterey, CA: CTB/McGraw-Hill.

Cuban, L. (1983, June). Effective schools: A friendly but cautionary note. Phi Delta Kappan, 695-696.

DeFleur, M. L., D'Antonio, W. V., & DeFleur, L. B. (1976). Sociology: Human society. Glenview, IL: Scott Foresman and Company.

Diehl, W. A., & Mikulecky, L. (1980). The nature of reading at work. Journal of Reading, 24(3), 221-228.

Dodge, A. F. (1935). Occupational ability patterns. New York: Bureau of Publications, Teachers College, Columbia University.

Donlon, T. F., McPeck, W. M., & Chathaw, L. R. (1968). Development of the brief test of literacy. Washington, D.C.: U.S. Government Printing Office.

Dvorak, B. J. (1935). Differential occupational ability patterns. University of Minnesota Employment Stabilization Research Institute, XII, 8. Minneapolis: The University of Minnesota Press.

Dvorak, B. (1956, July). Occupational testing. The labor market and employment security, 8-147.

Echternact, G. J. (1977). A comparative study of secondary schools with different score patterns. In appendix to Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Educational Testing Service. (n.d.) Step III: Manual and technical report. Reading, MA: Addison, Wesley Publishing Company, Inc.

Elligett, J., & Tocco, T. S. (1980, June). Reading achievement in 1979 vs. achievement in the fifties. Phi Delta Kappan, 698-699.

Eurich, A. C., & Kraetsch, G. A. (1982) A 50-year comparing of University of Minnesota freshmen reading performance. Journal of Educational Psychology, 74(5), 660-665.

Farr, R., Fay, L., & Negley, H. (1978). Then and now: Reading achievement in Indiana (1944-45 and 1976). Bloomington, IN: Indiana University. (ED 158 262).

Farr, R., Tuinman, J., & Rowls, M. (1974). Reading achievement in the United States: Then and now. Bloomington, IN: Indiana University. (ED 109 595).

Fine, S. A. (1968). The Use of the Dictionary of occupational titles as a source of estimates of educational and training requirements. The Journal of Human Resources, 3(3), 363-375.

Fischer, J. K., Haney, W., & David, L. (1980). APL revisited: Its uses and adaptation in states. Washington, D.C.: Government Printing Office.

Fisher, D. L. (1978). Functional literacy and the schools. Washington, D.C.: National Institute of Education.

Fisher, D. L. (1981). Functional literacy tests: A model of question-answering and an analysis of errors. Reading Research Quarterly, 16(3), 418-448.

Flanagan, J. C. (1976). Changes in school levels of achievement: Project talent ten and fifteen year retests. Educational Researcher, 5(8), 9-12.

Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32(3), 221-233.

Flynn, J. R. (1984). The Mean IQ of Americans: Massive gains, 1932 to 1978. Psychological Bulletin, 95, 29-51.

Folger, J. K., & Nam, C. B. (1967). Education of the American Population. Washington, D.C.: U,S. Government Printing Office.

Freeman, H. E., & Kassenbaum, G. G. (1956, May). The illiterate in American society: Some general hypothesis. Social Forces, 34 371-375.

Gadway, C. J., & Wilson, H. A. (1976). Functional literacy: Basic reading performance. Denver: Education Commission of the States. (ED 112 350).

Gates, A. I. (1961). Reading attainment in elementary schools: 1957 and 1937. New York: Teachers College.

Gebericn, J. R. (1952, March). The first of the three R's. Phi Delta Kappan, 33, 345-349.

Gilmore, W. J. (1982). Elementary literacy on the eve of the Industrial Revolution: Trends in rural New England, 1760-1830. Proceedings of the American Antiquarian Society, 92, 87-178.

Ginzberg, E., & Bray, D. W. (1953). The uneducated. New York: Columbia University Press.

Gould, S. J. (1981). The mismeasure of man. New York: W. W. Norton & Company.

Grant, W. V., & Eidon, L. J. (1982). Digest of Educational Statistics. Washington, D.C.: U.S. Government Printing Office. (Table 10, p. 16 level of school completed by persons age 25 and over and 25 to 29, by race: U.S. 1910 to 1981).

Grant, W. S. (1949). Comparative study of achievement in reading in 1916 and 1948. Grand Rapids school survey. Grand Rapids, Michigan Board of Education. Results briefly described in Gray, W. S. (1950). Summary of Reading Investigation July 1, 1948 to June 30, 1949, Journal of Educational Research, 43, 401-439.

Gray, W. S. (1956). How well do adults read? in N. B. Henry (Ed.), Adult Reading. The fifty-fifth yearbook of the National Society for the Study of Education, Part II. Chicago: The University of Chicago Press. Griffith, W. S., & Cervero, R. M. (1977). The Adult Performance Level Program: A serious and deliberate examination. Adult Education, 27(4), 209-224.

Green, D. R. (1982). Methods used by test publishers to "Debias" standardized tests: CTB/McGraw-Hill. Chapter 9, p. 229-240. In Berk, R. A. (Ed.) (1982) Handbook of Methods for Detecting Test Bias. Baltimore: The Johns Hopkins University Press.

Griffith, W. S., & Cervero, R. M. (1977). The Adult Performance Level program: A serious and deliberate examination. Adult Education, 27(4), 209-224.

Hale, M. (1982). History of employment testing. In Alexandra K. Wigdor & Wendell R. Garner, (Eds.), Ability testing: Uses, consequences, and controversies, Part II, Documentation Section. Washington: National Academy Press. 173-194.

Harman, D. (1970). Illiteracy: An overview. In Harvard Educational Review, 40(2), November, 226-243.

Harnischfeger, A., & Wiley, D. (1975). Achievement test score decline: Do we need to worry? St. Louis: CEMREL.

Harris, L., and Associates, Inc. (1970). Survival literacy study. Washington: The National Reading Council. (ED 068 813).

Harris, L., and Associates, Inc. (1971). The 1971 national reading difficulty index: A study of functional reading ability in the United States, for the National Reading Center. Washington, D.C.: The National Reading Center.

Heath, S. B. (1980). The functions and uses of literacy. Journal of Communication, Winter, 30(1), 123-133.

Heath, S. B. (1984). Ways with words: Language, life and work in communities and classrooms. Cambridge: Cambridge University Press.

Hieronymus, A. N. (1985, September 16). Preliminary information on the 1984 ITBS results. Personal Communication.

Hieronymus, A. N., Lindquist, E. F., and Hoover, H. D. (1982). Manual for school administrators: Iowa Tests of Basic Skills. Chicago: Riverside Publishing Company.

Hoffman, B. (1962). The tyranny of testing. New York: Collier Books.

Holt, J. (1976). Instead of education. New York: E. P. Dutton & Co.

Hunter, C. S. J., & Harman, D. (1979). Adult illiteracy in the United States: A report to the Ford Foundation. New York: McGraw-Hill Book Company.

Husen, T. (1967). International study of achievement in mathematics: A comparison between twelve countries. New York: Wiley.

Illiteracy and the scope of the problems in this country. Hearing before the House subcommittee on post secondary education and labor. September 21, 1982. (1984). Washington, D.C.: U.S. Government Printing Office.

Iowa Testing Program. (n.d.) Summary of national achievement trends, 1955-1981, grades 3 - 8, using median performance of 1955 fall as a base. Iowa Tests of Basic Skills. Mimeo provided by A. N. Hieronymus.

The Iowa Tests of Educational Development: A summary of changes in the ITED norms. (1971). Iowa City: The University of Iowa.

Jencks, C., Smith, M., Acland, H, Bans, M. J., Cohon, D., Gintis, H., Heyns, B., & Michelson, S. (1972). Inequality: A reassessment of the effect of family and schooling in America. New York: Harper & Row Publishers.

Jencks, C. (1980, December). Declining test scores: An assessment of six alternative explanations. Sociological Spectrum. Premier Issue, 1-15.

Kaestle, C. F. (1985). The history of literacy and the history of readers. In E. Gordon, (Ed.), Review of research in education. Washington D.C.: American Educational Research Association, 12.

Kirsch, I. (1985). NAEP profiles of literacy an assessment of young adults development plan April 1985. Princeton: National Assessment of Educational Progress.

Kirsch, I. S., & Guthrie, J. T. (1983). Adult reading practices for work and leisure. Draft.

Kirsch, I. S., & Guthrie, J. T. (1984). Adult reading practices for work and leisure. Adult Education Quarterly, 34(4), 213-232.

Kirsch, I., & Guthrie, J. T. (1977-8). The concept and measurement of functional literacy. Reading Research Quarterly, 13(14), 485-507.

Klare, G. R. (1963). The measurement of readability. Ames, Iowa: Iowa State University Press.

Kozol, J. (1985). Illiterate America. New York: Anchor Press, Doubleday.

Lerner, B. (1981). The minimum competence testing movement: Social, scientific, and legal implications. American Psychologist, 36 (10), 1057-1066.

Levin, H. M., & Rumberger, R. (1983, January 30). Hi-tech requires few brains. Washington Post, C5.

Lockridge, K. A. (1974). Literacy in colonial New England: An enquiry into the social context of literacy in the early modern west. New York: W. W. Norton.

Matras, J. (1975). Social inequality, stratification, and mobility. Englewood Cliffs, NJ: Prentice-Hall.

Meier, D. (1981, Fall). Why reading tests don't test reading. Dissent, 457-466.

Meier, D. (1984, Winter). "Getting tough" in the schools. Dissent, 61-70.

Mikulecky, L. (1980). Functional writing in the workplace. In L. Gentry Ed. Research and instruction in practical writing. Los Alamitos, CA: SWRL Educational Research and Development.

Mikulecky, L. (1981). The mismatch between school training and job literacy demands. The Vocational Guidance Quarterly, 30(2), 174-180.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. Reading Research Quarterly, 17(3), 400-419.

Mikulecky, L., & Winchestor, D. (1983). Job literacy and job performance among nurses at varying employment levels. Adult Education Quarterly, 34(1), 1-15.

Monteith, M. (1980, February). How well does the average American read? Some facts, figures, and opinions. Journal of Reading, 460-464.

Munday, L. A. (1979, March). Changing test scores: Basic skills development in 1977 compared with 1970. Phi Delta Kappan, 670-671.

Munday, L. A. (1979, May). Changing test scores, especially since 1970. Phi Delta Kappan, 496-499.

Murphy, R. T. (1973). Adult functional reading study. Princeton, NJ: Educational Testing Service. (ED 109 650).

Murphy, R. T. (1975a). Adult functional reading study. Supplement to final report. Princeton, NJ: Educational Testing Services. (ED 109 651).

Murphy, R. T. (1975b). Assessment of adult reading competence. In Duane M. Nielsen & H. F. Hjelm, (Eds.), Reading and Career Education. Newark, DE: International Reading Association.

Natziger, D. H., Thompson, R. B., Hiscox, M. D., & Owen, T. R. (1976). Tests of functional adult literacy: An evaluation of currently available instruments. Portland, OR: Northwest Regional Educational Laboratory.

National Assessment of Educational Progress. (1979). Changes in mathematical achievement, 1973-78. Report Number 09-MA-01. Denver, CO: Education Commission of the States.

National Assessment of Educational Progress. (1981). Three national assessments of reading: Changes in performance, 1970-1980. Report 11-R-01. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1982). The reading comprehension of American youth. Report 11-R-02. Denver, CO: Education Commission of the States.

National Assessment of Educational Progress. (1981). Three national assessments of science. Changes in achievement, 1969-1979. Report Number 08-S-35. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1982). Writing achievement, 1969-1979. Report Number 10-W-35. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1983). The third national mathematics assessment: Results, trends and issues. Report Number 13-MA-01. Denver: Education Commission of the States.

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. Report No.15-R-01. Princeton, NJ: Educational Testing Service.

National Center for Education Statistics. (1979). The Condition of Education. Washington, D.C.: Government Printing Office.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, D.C.: Government Printing Office.

New York Times. (1919). Editorial. February 19, 1919, p. 12, column 4.

Nielson, D. M., & Hjelan, H. F. (Eds.) (1975). Reading and career education. Newark, DE: International Reading Association.

Nitko, A. J. (1976). Exploring alternatives to current standardized tests. Proceedings of the 1976 National Testing Conference. Pittsburgh: University of Pittsburgh.

Northcutt, N. W. (1975). Functional literacy for adults. In Nielson, D. M. & Hjelm, H. F., (Eds.), Reading and Career Education. Newark, D.C.: International Reading Association.

Olneck, M. (1984). Terman redux. Contemporary Education Review, 3, 297-314.

Olson, D. R. (1977a). The languages of instruction: The literate bias of schooling. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.) Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum.

O'Toole, J., Hansot, E., Herman, W., Herrick, N., Liebow, E., Lusignan, B., Richman, H., Sheppard, H., Stephansky, B., & Wright, J. (1973). Work in America: Report of a special task force to the secretary of Health, Education, and Welfare. Cambridge, MA: The MIT Press.

Rankin, E. F., & Culhane, J. W. (1969). Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 13(3), 193-198.

Ravitch, D. (1985). The schools we deserve. New York: Basic Books.

Reagan, R. (1985a, February 7). Remarks of the President to the National Association of Secondary School Principals 68th Annual Convention. Press Release. Washington, D.C.: The White House, 1985.

Resnick, D., & Resnick, L. (1977). The nature of literacy: An historical exploration. Harvard Educational Review, 43, 370-385.

Resnick, D., & Resnick, L. (1985). Standards, curriculum and performance: Historical and comparative perspective. Education Researcher, 14(4), 5-20.

Riverside Publishing Company. (1983). The development of the 1982 norms for the Iowa Tests of Basic Skills, Cognitive Abilities Test, and Tests of Achievement and Proficiency. Chicago: Riverside Publishing Company.

Robinson, T. (1985, March 4). "President Reagan sees education as important and improving; says students should feel 'reverence.'" Education Times, p. 1.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49(2), 252-279.

Schrader, W. B. (1968). Test data as social indicators. Princeton, NJ: Educational Testing Service.

Science Research Associates. (1978). Iowa tests of educational development (Technical Report). Table XIV, p. 13.

Science Research Associates. (1979). SRA Achievement Series Technical Report #1, Addendum for Fall 1978 Standardization. Chicago: Science Research ASsociates.

Science Research Associates. (1980). Test talk, 80-2: Comparing the results of testing with the 1971 and 1978 editions of the SRA achievement series. Chicago: Science Research Associates.

Selden, R. (1978). "Literacy: Current problems and current research," in National Council on Eductional Research, Fifth report (Washington, Department of Education, 1978-1979), pp. 31-40.

Silberman, C. (1970). Crisis in the classroom. New York: Vintage Books.

Silberman, C., ed. (1973). The open classroom reader. New York: Vintage Books.

Soltow, L., & Stevens, E. (1981). The rise of literacy and the common school in the United States. A socioeconomic analysis to 1870. Chicago: The University of Chicago Press.

Spenner, K. I. (1980). Occupational characteristics and classification systems. Sociological Methods and Research, 9 (2), 239-264.

Spufford, M. (1979). First steps in literacy: The reading and writing experiences of the humblest seventeenth-century spiritual autobiographers. Social History, 4, 407, 435.

Stanley, J. C., & Hopkins, K. D. (1972). Educational and Psychological Measurement and Evaluation. Englewood Cliffs, N.J.: Prentice-Hall.

Stead, W. H., & Masincup, W. E. (1942). The occupational research program of the United States employment service. Chicago: Public Administration Service.

Stead, W. H., & Shartle, C. L. (1940). Occupational counseling techniques: Their development and application. New York: American Book Company.

Stedman, L. S., & Kaestle, C. F. (1985). The test score decline is over: Now what? Phi Delta Kappan, 67(3), 204-210.

Stedman, L. C., & Smith, M. S. (1983). Recent reform proposals for American education. Contemporary Education Review, 2, 85-104.

Stewart, O. & Green, D. S. (1983, March). Test-taking skills for standardized tests of reading. The Reading Teacher, 634-638.

Sticht, T. G., (Ed.). (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., & McFann, H. H. (1975). Reading requirements for career entry. In Nielsen, D. M. & Hjelm, H. F. (Eds.), Reading and career education. Newark, DE: International Reading Association.

Tavris, C. (1976, April). The end of the IQ slump. Psychology Today, 69-74.

Taylor, W. L. (1953). "Cloze Procedure": A new tool for measuring readability. Journalism Quarterly, 30(4), 415-433.

Test Department, Harcourt Brace Jovanovich, Inc. (1971). Equivalent scores for Metropolitan Achievement Tests, 1970 edition, and Metropolitan Achievement Tests 1958 edition. Special report number 14. New York: Harcourt Brace Jovanovich, Inc.

Test Department, Harcourt Brace Jovanovich, Inc. (1983). Some comments on the relationship between scores on the 1973 and 1982 edition of Stanford Achievement Test. Special report number 4A. New York: Harcourt Brace Jovanovich, Inc.

The Psychological Corporation. (1978). Equivalent grade equivalent scores for metro '7% and metro '78. Special Report Number 20. Revised. New York: Harcourt Brace Jovanovich, Inc.

Thompson, R. (1983, June 24). Illiteracy in America. Editorial Research Reports, Washington, D.C.: Congressional Quarterly, Inc., 1(24), 475-492.

Thorndike, R. L., & Hagen, E. (1969). Measurement and evaluation in psychology and education. New York: John Wiley & Sons, Inc.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. Public Opinion Quarterly, 43(3), 359-377.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. American Psychologist, 3, 54-56.

Tully, A. (1972). Literacy levels and education development in rural Pennsylvania, 1729-1775. Pennsylvania History, 39, 301-312.

Tyler, R. W. (1930). High school pupils of today. Educational Research Bulletin, 9(15), 409-410.

Tyler, R. W., & White, S. H. (1979). Testing, teaching, and learning. Washington, D.C.: National Institute of Education.

U.S. Bureau of the Census. (1920). Fourteenth census of the United States, population: 1920. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1948). Illiteracy in the United States: October, 1947. Current population reports, series P-20, number 20. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1960). Literacy and educational attainment: March 1959. Current population reports, series P-20, number 99. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1971). Illiteracy in the United States: November 1969. Current Population Reports, series P-20, number 217. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1973). Census of the Population: 1970 Subject Reports. Final Report PC(2)-5B. Educational Attainment. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1980). Educational attainment in the United States: March 1979 and 1978. Current Population Reports, series P-20, number 356. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1981a). School enrollment, P-20, Number 374. Washington, D.C.: Government Printing Office.

U.S. Bureau of the Census. (1981b). U.S.A. statistics in brief, 1981. Washington, D.C.: Government Printing Office.

U.S. Bureau of the Census. (1982). 1980 Census of Population and Housing Supplementary Report: Provisional Estimates of Social, Economic, and Housing Characteristics. States and Selected Standard Metropolitan Statistical Aeras. PHC80-S1-1. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. (1982a). Ancestry and language in the United States: November 1979. Current Population Reports, series P-23, number 116. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Defense. (1984). Counselor's Manual for the Armed Services Vocational Aptitude Battery Form 14. Washington, D.C.: Office of the Assistant Secretary of Defense.

U.S. Department of Labor. (1965). Dictionary of Occupational Titles 1965 Volume II Occupational Classification and Industry Index. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Labor. (1979). Section II: Occupational aptitude pattern structure. Manual for the USES general aptitude test battery. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Labor. Bureau of Labor Statistics. (1974). Special Labor Force Report, Number 175: Educational attainment of workers, March 1974.

U.S. Department of Labor, Bureau of Labor Statistics. (1982). Economic projections to 1990 (Bulletin 2121). Washington, D.C.: Department of Labor.

Vinovskis, M. A., & Bernard, R. (1978). Beyond Catherine Beecher: Female education in the antebellum period. Signs, 3, 856-869.

Vogt, D. K. (1973). Literacy among youths 12-17 years United States. Rockville, MD: National Center for Health Statistics.

Wellborn, S. N. (1982, May 17). Ahead: A nation of illiterates? U.S. News & World Report, 53-57.

West, J., Diodato, L., & Sandberg, N. (1984). Trend study of high school offerings and enrollments; 1972-73 and 1981-82. Washington, D.C.: U.S. Government Printing Office. (ED 253 530).

Willig, A. C., Harnisch, D. L., Hill, K. T., & Maehr, M. L. (1983). Sociocultural and educational correlates of success-failure attributions and evaluation anxiety in the school setting for Black, Hispanic, and Anglo children. American Educational Research Journal, 20(3), 385-410.

Windle, C. (1959). The accuracy of census literacy statistics in Iran. Journal of the American Statistical Association 54, 578-81.

Winstor, S. (1930). Illiteracy in the United States from 1870 to 1920. Chapel Hill: University of North Carolina Press.

Witty, P., & Coomer, A. (1951). How successful is reading instruction today? Elementary English, 28(8), 451-459.

Wolfinger, R. E., & Rosenstone, S. T. (1980). Who Votes? New Haven, CT: Yale University Press, 115-118.

Yerkes, R. M. (1921). Psychological examining in the United States Army. Washington, D.C.: National Academy of Sciences.

Yoakum, C. S., & Yerkes, R. M. (1920). Army mental tests. New York: H. Holt and Company.

Zajonc, R. B., & Bargh, J. (1980). Birth order, family size, and decline of SAT scores. American Psychologist, 35, 662-668.

Chart 1

Local Then-and-Now Studies[1]

Subjects Other Than Reading Study	Then	Now	Grade	Location	No. of Students Then	No. of Students Now	Subjects	Results
Caldwell & Curtis	1845	1919	8	Boston, MA	530 Boston	12,000 Across Bottom 40%	Geography, History, Philosophy, Astronomy, etc.	+
Riley	1846	1906	9	Springfield, MA	245	709	Arithmetic, Spelling, Geography	+
Luther	1848	1947	8	Cleveland, OH	35	40 (10 Best from 4 Schools)	Mental, Written Arithmetic, American History, etc.	+
Fish	1850	1929	8	Boston, MA	20	200	Arithmetic, Grammar, Geography	+
Rogers	1923	1946	6	Chicago, IL	16,000	13,047	Arithmetic	-
Daugherty	1929	1947	4-7	Florida	Several Counties	Somewhat Different Area	Arithmetic, Spelling	+5th Spelling
Reading								
Grant	1916	1949	-	Grand Rapids, MI	5 Schools	5 Schools	Comprehension, Oral Reading, Speed of Silent Reading	+, =, =
Boss	1916	1938	1-8	St. Louis, MO	8,923	1,156 "Measured Sample"	Oral and Silent Reading	-
Woods	1924	1934	6	Los Angeles, CA	33 Elem. Schools	33 Elem. Schools	Reading	+6 Months
Worcester & Kline	1921	1947	3-8	Lincoln, NE	5,690	5,106	Silent Reading	+
Davis & Morgan	1927	1938	6	Santa Monica, CA	Grade 6	Grade 6	Reading	+2 Months
Krugman & Wrightstone	1935-41	1944-46	6-9,11	N.Y. City	>290,000	>242,000	Reading	+But N.S.
Tiegs	Before 1945	After 1945	4-11	6 Communities in 7 States	115,000	115,000	Vocabulary, Reading Comprehension, Total Reading	-1, +1.3, +1.7 Months
Finch & Gillenwater	1931	1948	6	Springfield, MO	144	198	Reading	+But, N.S. (p. 38)
Burke & Anderson	1939	1950	1-6	Ottawa, KS	162	216	Reading (+5 Other Subjects)	- (p. 40)
Miller & Lanton	1932	1952	3&5,8	Evanston, IL	1,828	1,828	3rd Reading Completion, Paragraph Meaning, Vocabulary 4th,5th,8th, Reading Comprehension, Vocabulary	Ranged from +2.5 to 8 months
Partlow	1933	1953	5-8	St. Catharines, Canada	All Pupils in City	All Pupils in City	Reading Completion, Vocabulary	5-8th R +, +, =, = V +, -, -, -
Fridan	1940	1956	1-7	Lafayette, IN	All Pupils in One Parochial School		Reading	1-5th, 7th +1.1 to 6 months, 6th-8.5 mo.
Bradfield	1928	1964	5	Rural CA	35	51	Reading	+But N.S.

[1]From Farr, Tuinman, & Rowls (1974), except Grant (1949).

- = Decrease
+ = Increase
= = No Change
N.S. = Not Significant

Chart 2

State and National Then-and-Now Studies[2]

State Studies	Then	Now	Grade	Location	No. of Students Then	No. of Students Now	Subjects	Results
Tyler	1924	1930	HS	OH	Selected High Schools	Then and Now	Physics, Math, English	Mixed
In Witty & Coomer	1915	1947	HS	NY	Statewide	Then and Now	NY Regents	71% Pass Rate to 84%
Sligo in Armbruster	1934	1954	HS	IA	Selected High Schools	Then and Now	Algebra, General Science, English, History	-
In Farr, Tuinman, & Rowls	1940	1965	3-8	IA	38,000	Similar	Reading	+2/10 to Over 1 Grade
Farr, Fay, & Negley	1944	1976	6,10	IN	Volunteer Schools 15,206 11,424	Stratified Sample 8,000 7,000	Reading (Average of Various Subtests)	6th -2 Month 10th -2 Percentile Pts. After Age Adjustment +8 Months, +10 Pts.
Eurich & Kraetsch	1928	1978	1st yr. College	U.ofMinn.	1,313 Freshman 4,191 HS Seniors	865 In-Coming Students	Paragraph Comprehension Subtest. Vocabulary, Reading Comprehension, Rate	-
National Studies								
Gates	1937	1957	1-6	NATL	107,000	31,000 Equating Study, 12 School Districts	Reading (Average of Various Subtests)	-1/10 to -3/10 Grades; After Age Adj. 4-6th +4 to +6.7 Months
Tuddenham	1918	1943	Young Men	NATL	WWI Large Sample White Recruits	WWII Representative Draftee Sample	Army Alpha	+33 Percentile Pts.
Yerkes, Gray	WWI	WWII	Young Men	NATL	Millions of Draftees	Millions of Draftees	Rejection Rates, Illiteracy, Years of Schooling	Non-Comparable
Bloom	1943	1955	12	NATL	See Text	Equating Study	English Comp., Social Studies, Nat. Science, Literature, Math	+2 to 8 Percentile Pts.
Elligett & Tocco	1950's	1979	6	Pinellas Co., FL	-	18 6th Grade Classrooms Equating Study-1 School Dist.	Reading	-5 to -10 months

[2]Sources as listed. See Bibliography.

- = Decrease
+ = Increase

Chart 3

Gates' 1937 and 1957 Results

Number of Months by Which 1937 Exceeded 1957

Grade	Total Performance*	Comprehension
3.5	.7	.4
4.0	.9	.5
4.5	1.1	.5
5.0	1.4	1.0
5.5	1.8	1.5
6.0	2.1	2.0
6.5	2.4	2.0
6.9	2.5	2.5
7.5	2.8	3.0
8.7	3.4	4.5

Gates (1961), Table 3, p. 25

*Total performance included sections on vocabulary, understanding directions, speed, noting details, and others as well as the section on comprehension.

Number of Months by Which 1957 Exceeded 1937

Age in Years	Total Performance
6.6 to 7.6	0.0
7.6 to 8.1	1.0-1.5
8.1 to 8.6	2.0

Students of Same Age at Grade	Total Performance	Comprehension
4.0	5.0	7.5
4.5	5.1	6.0
5.0	5.7	6.0
5.5	6.7	6.5
6.0	5.4	5.0
6.5	4.0	5.0

Chart 4

Sample Form from Harris & Associates (1970)
Survival Literacy Study

Form III

Application for Driver's License

1. What is your name? _____

2. What is your weight? _____

3. What is your height? _____ feet, _____ inches

4. What is the color of your eyes? _____

5. List any visual, physical, or mental conditions that might impair your ability to drive safely? _____
_____

6. List any previous driver's license issued to you:
State _____ Year _____

7. How many times have you previously been examined for a driver's license? _____

8. What day of the week would be most convenient for you to take th driver's examination? _____

9. What hour of the day would be most convenient for you to take the driver's examination? _____

PLEASE MAKE SURE ALL QUESTIONS HAVE BEEN ANSWERED. IF YOU ARE NOT SURE OF AN ANSWER TO ANY ITEM, DRAW A LINE THROUGH THE SPACE PROVIDED FOR THE ANSWER.

(N.B. Drawing a line through the space was considered a correct answer)

Chart 5

Results of Harris & Associates, Survival Literacy Study, 1970

Population (16+):

	Functionally Illiterate (Less than 70% Correct)	Not Functionally Literate (Less than 90% Correct)
Public Assistance	*	3
Social Security	1	7
Driver's License	1	8
Bank Loan	2	11
Medicaid	9	34
Average of Forms	3%	13%
Millions	4.3	18.5

*Less than .5%

Chart 6

Items from the 1971 National Reading Difficulty Index

Example #1: From Telephone Dialing Section

People were given a card with the following information on it:

Area Codes for Some Cities

Place	Area Code
Evansville, Indiana	812
Oakland, California	415
Harrison, New York	914
Williamsport, Pennsylvania	717
Austin, Texas	512

They were asked two questions:

1. "Please look at this card and see if you can tell me the area code for Williamsport, Pennsylvania."

2. "Please look at the card again and tell me which city you would reach by dialing area code 812."

Example #2: From Housing Ads Section

People were given a card with the following:

"Attractive house in excellent condition. Three floors. Full basement. Large living room. Backyard with garden. Two-car garage."

They were then asked:

2a. "Would you tell me how the ad describes the living room of the house?"

2b. "How does the ad describe the backyard?"

2c. "How does the ad describe the basement?"

(The employment section was similar in nature to the housing ad sections; the application form was similar to the Survival Literacy example previously presented.)

Source: Harris and Associates, Inc. (1971).

Chart 7

The 1971 National Reading Difficulty Index Study

Task	# of Items	Percent of Population Getting All Items Correct	Only One Wrong	None or Only One Right
Telephone Rate, Directions	4	90%	8%	1%
Housing Ads	9	88%	7%	1%
Employment Ads	9	92%	6%	1%
Personal ID Information	10	93%	6%	*
Employment Information	4	85%	9%	3%
Income Information	3	77%	16%	7%
Housing Information	8	87%	11%	*
Automobile Information	3	97%	1%	2%
Medical Information	3	86%	10%	4%
Citizenship Information	6	87%	6%	*

*Less than 1%.

Overall Scores	# of Items	Percent of Population Getting 100%	96-99.9%	90-95.9%	80-89.9%	Less than 80%*
Weighted Text Items	59	43%	23%	19%	11%	4%

*No further breakdown was given.

Source: Harris and Associates, Inc. (1971)

Chart 8

Functional Illiteracy Rates

Study	Year	Sample Ages	Tasks	Criteria	Rate for Lower Criterion	Rate for Higher Criterion
Survival Literacy	1970	16+	Application Forms	70%, 90% correct	3%	13%
Reading Index	1971	16+	Application Forms Telephone; Ads	80%, 90% of weighted items	4%	15%
MAFL	1975	17	Everyday Reading	75% correct	12.6%	-
AFRS	1973	16+	Everyday Reading	None set	-	-
APL	1974	18-65	Functional Comp.	APL 1, APL 1 & 2	19.7%	53.6%

All were nationally representative samples.

111

Chart 9

The 1971 National Reading Difficulty Index Study

Overall Scores	# of Items	Percent of Population Getting 100%	96-99.9%	90-95.9%	Less than 90%
Weighted Test Items	59	43%	23%	19%	15%*
Unweighted Test Items	59	45%	35%	15%	5%**

*11% between 80 and 89.9%, 4% less than 80%
**No further breakdown was provided

Source: Harris and Associates, Inc. (1971)

Chart 10

Functional Literacy Test Items
(Results unadjusted for sampling and difficulty)

API

Source	Material	Task	% Can't
2	Business letter	Complete return address w/o error form, content	80%
2	Mail order	Calculate total cash price	75%
2	Odometer, fuel readings	Calculate mpg	73%
1	W-4, # of dependents	Enter # of tax exemptions	70%
3	Tax table	Find tax for given income	61%
1	Pamphlet on rights	Underline part applicable to situation described	58%
1	Given # of years of prison sentence	Given portion before parole; calculate the time to serve	53%
1	Tax rate for each item	Calculate total tax	49%
1	Map of 4 imaginary states with populat.	Which state has most senators?	49%
1	Calorie chart, meal	Calculate total # calories	44%
2	For sale ads	Calculate difference new, used	40%
1	Agency fee % of salary	Calculate # of months to pay	39%
1	Help Wanted Ad	Match personal characteristics and job requirements	38%
1	Home heating oil invoice	Given a certain % reduction calculate new allotment	37%
1	Question on rights	Arrest, detain without charges?	34%
1	Vocabulary question	Define "open shop"	33%
1	Hourly rate, hrs/week hours overtime	Calculate amount of pay for a given number of hours	33%
1	Graph from medicine ad	Interpret	33%
1,3	Airline schedule	Select flight to make appointment in another city	33%
1	Given 4 definitions	Pick right definition of "right of way"	30%
2	Menu	Meal for two under set amount	29%
1	Monthly gross, monthly deductions	Calculate yearly take-home pay	29%
1	Pharmacy receipt	Calculate change from $20	28%
1	Identification	Human body temperature	27%
1	Health insurance policy	Amount paid for given condition	27%
1	Earnings statement	Identify soc. security deduc	26%
1	Vocabulary question	Define "credit check"	26%
2	Earnings statement	Meanings "gross" and "net"	25%
1	3 cereal packages with price, weight	Determine "best buy" in terms of per unit price	25%
1	Paper and pencil	Address envelope; no return add	24%
1	Pencil & paper	Write school excuse note; forgot salutation	22%
4	Medicine label	Follow instructions "Take two pills twice a day"	21%
4	Housing pamphlet	How often termite inspection?	21%
1	YMCA swim lesson fee schedule	Calculate money for five children	21%

Chart 10 (Continued):

2	Repayment schedule	Monthly payment for given amount	20%
1	Equal Oppty Notice	Select correct definition	20%
1	Check cashing sign	Understand policy	20%
1	Letter to U.S. repres	Urge vote against a bill; forgot recommendation	20%
2	Help wanted ads	Identify public vs. private	17%
1	Earnings statement	Determine # of deductions	17%
1	Blank check	Fill it out properly	14%
1	Map of 4 imaginary states with popul	Determine which state has most congressmen	14%
1,3	Road map	Given a journey, name town where switch highways	14%
1	Letter to U.S. repres	Urge vote against a bill; did not properly identify bill	13%
1	Paper and pencil	Address envelope properly	13%
1	Question on rights	Should radicals, trouble-makers be allowed peaceful public mtgs.?	12%
1	Paper and pencil	Write school excuse note: incomprehensible message	7%
		did not identify child	7%
1	Map of 4 imaginary states with popul.	Determine population of a given state	6%
1	Paper and pencil	Write school excuse note; illegible	3%

SOURCES:

1- Adult Performance Level Project (1977)
2- Northcutt (1975)
3- Acland (1976)
4- Thompson (1983)

AFRS

5	Train schedule	Circle time that the 1:46 train from Trenton arrives in D.C.	33%
5	Picture of 5 garment tags	Circle two tags that indicate garment is 100% polyester	10%
5	Application for employment	Put X where name of someone to contact in emergency goes	7%
5	Picture of 6 mailing labels	Select one for mailing easily broken item	4%
5	Picture of 3 labelled jars	Select one safe to drink	.1%

5- Murphy (1975)

MAFL (17 year old students) (1975 results)

6	Auto insurance policy	Identify amount of coverage for bodily injury liability	82%
6	Book club membership form	Realize that no money had to be submitted with application	57%

Chart 10 (Continued):

6	Traffic ticket	Identify date by which fine due	51%
6	4 line passage on Colorado mountains	Identify sentence that doesn't apply	32%
6	Report card	Identify subject that is improving (special code)	32%
6	Billboard sign	Identify probable location	31%
6	Coupon	Identify applicable sizes	21%
6	Help wanted ad	Identify how to apply	9%
6	Coupon	Identify group at whom product is targetted	6%

6- Gadway & Wilson, (1976)

Harris I Survival Literacy

			% functional ill.
7	Application forms	Fill them out	3%

7- Harris and Associates, Inc. (1970)

Harris II National Reading Difficulty Index

			% less than 70% correct
8	Telephone Dialing	4 questions on rates and what to dial	
8	3 Housing ads	3 questions on each	
8	3 Employment ads	3 questions on each	
8	Application forms ads & telephone	Fill out forms, answer questions	4% functional illiterates (<80% correct)

8 - Harris and Associates, Inc. (1971)

Chart 11

ORAL DIRECTIONS

Item 1 Place a circle around the bottle of liquid that would be safe to drink.

Chart 12

Item 3 Look at the application for employment. Put an *X* in the space where you would write the name and address of someone to notify in case of emergency.

APPLICATION FOR EMPLOYMENT DATE

NAME

LOCAL ADDRESS TELEPHONE NO.

PERMANENT ADDRESS ZIP CODE

MARITAL STATUS SINGLE ☐ MARRIED ☐ DIVORCED ☐ WIDOWED ☐ SEPARATED ☐ HEIGHT WEIGHT

NAME & ADDRESS OF PERSON TO NOTIFY IN EMERGENC NO. OF CHILDREN AGES

FIRST NAME OF SPOUSE PLACE OF EMPLOYMENT

EDUCATION	NAME & LOCATION OF SCHOOL	FROM	TO	COURSE OR MAJOR	YEAR GRAD	DEGREE
HIGH SCHOOL						
COLLEGE						
BUSINESS OTHER SCHOOLS						
SPECIAL STUDY OR ART						

Source: Murphy, R. T. (1975b). Assessment of adult reading competence. In Duane M. Nielsen & H. F. Hjelm, (Eds.), Reading and Career Education. Newark, DE: International Reading Association. Reproduced by permission.

Chart 13

NEW YORK – WASHINGTON

	New York, N.Y. (Penn. Sta.) Leave	Newark, N.J. Leave	Trenton, N.J. Leave	North Philadelphia, Pa. Leave	Philadelphia (Penn Central Sta.—30th St.) Leave	Wilmington, Del. Leave	Baltimore, Md. Arrive	Capital Beltway, Md. Arrive	Washington, D.C. Arrive
177 Mondays thru Saturdays	3:23 AM	3:35	4 27	5 00	5 09	5 48	6.59	—	7.50 AM
131 Mondays thru Saturdays	6:30 AM	6.45	7 32	8 00	8.10	8 38	9:39	—	10 20 AM
101 Metroliner Mondays thru Fridays	7.30 AM	ε 7.42	—	—	8.43	—	9 49	ε10:13	10:25 AM
133 Daily	8:00 AM	8 16	8 59	9 26	9:35	10:18	11:18	—	12 00 Noon
103 Metroliner Daily	8 30 AM	ε 8.42	9 18	—	9:46	10:10	10:57	—	11:30 AM
135 Daily	9.30 AM	9:46	10:31	11 00	11:10	11 45	12.48	—	1:30 PM
137 Daily	10:45 AM	11 01	11 45	12:12	12:21	12:51	1:51	—	2:40 PM
105 Metroliner Daily	11:30 AM	ε11.42	12:18	—	12:46	1:18	1:57	—	2:30 PM
171 Daily	12:45 PM	1:01	1.46	2:13	2:22	2:53	2:58	ε 4.27	4.45 PM
107 Metroliner Daily	1:00 PM	ε 1:12	—	—	2:13	2:36	3:2[illegible]	ε 3.47	4:00 PM
163 Runs Feb. 12 and 15 only	2:00 PM	2:16	3:00	3:29	3:40	4:09	5:18	—	5:50 PM
173 Daily	3:00 PM	3:16	4:00	4:29	4:39	5:08	6:09	—	6.50 PM
109 Metroliner Daily	4:15 PM	ε 4:27	5:03	—	5:31	5:55	6:42	—	7:15 PM
165 Runs Feb. 12 and 15 only	4:30 PM	ε 4 45	ε 5 20	5:55	6:15	6:43	7:44	ε 8:10	8:30 PM
111 Metroliner Sundays thru Fridays	5:00 PM	—	—	—	6:10	6:33	7:17	ε 7:40	7:55 PM
175 Daily	5:45 PM	6:01	ε 6.43	7:15	7:24	7:52	8:53	—	9:35 PM
159 Sundays only	6:30 PM	6.46	7.29	7:57	8:07	8:36	9:40	ε10:09	10:25 PM
139 Mondays thru Saturdays	6:30 PM	6.45	7.33	8:07	8:26	8:54	10:00	ε10:40	10:55 PM
155 Daily	7:30 PM	7:46	8.29	8:57	9:06	9:39	10:40	—	11:20 PM
113 Metroliner Sundays thru Fridays	8:30 PM	ε 8:42	—	—	9:43	10:06	ε10:53	ε11:16	11:30 PM
147 Daily	9:00 PM	9:15	10:04	10:41	11:01	11:29	12:37	—	1.35 AM
161 Sundays and Feb. 15 but not run Feb. 16	10:00 PM	10:16	11:05	11:33	11:46	12:22	1 29	—	2.15 AM

WASHINGTON – NEW YORK

	Washington, D.C. Leave	Capital Beltway, Md. Leave	Baltimore, Md. Leave	Wilmington, Del. Leave	Philadelphia, Pa. (Penn Central Sta.—30th St.) Leave	North Philadelphia, Pa. Leave	Trenton, N.J. Leave	Newark, N.J. Arrive	New York, N.Y. Arrive
140 Daily	2:25 AM	—	ε 3:03	ε 4.16	ε 4 50	ε 5.23	ε 5 50	ε 6.39	7:00 AM
170 Daily	6.55 AM	—	7.36	8:37	9:10	9:2[illegible]	9 50	10:38	10:55 AM
100 Metroliner Mondays thru Fridays	7:30 AM	ε 7.40	8:06	8:51	9:15	—	—	ε10:16	10:30 AM
102 Metroliner Daily	8:30 AM	—	9:02	9:07	10:13	—	10:39	ε11:16	11:30 AM
126 Daily	8:40 AM	ε 8 54	9:29	10:29	10:58	11 07	11:35	12:22	12:38 PM
172 Daily	10:00 AM	—	10:40	11:41	12:14	12:23	12:51	[illegible]	1:30 PM
130 Daily	11:40 AM	—	12:21	1:35	2:03	2:15	2:43	3 30	3:45 PM
104 Metroliner Daily	12:00 Noon	ε12:11	12:37	—	1:42	—	2:09	ε 2:46	3:00 PM
106 Metroliner Daily	1:00 PM	—	1:32	2:17	2:43	—	3:09	ε 3:46	4:00 PM
174 Daily	1:40 PM	—	2:21	3.22	4:00	4:10	4.39	5:24	5.40 PM
132 Daily	3:00 PM	ε 3:16	3.45	4.45	5:13	[illegible]	5:51	6:34	6.50 PM
152 Daily	4:00 PM	ε 4:14	4:44	5:52	6:10	6:28	ε 6:55	7.40	7:55 PM
108 Metroliner Daily	4:30 PM	—	5:02	5.47	6:13	—	6:39	ε 7:17	7:30 PM
154 Sundays thru Fridays	5:00 PM	—	5.47	6:50	7:19	7:29	7:56	8.40	8:55 PM
110 Metroliner Sundays thru Fridays	6:00 PM	—	6:22	7:17	7:43	—	8:09	ε 8.46	9:00 PM
166 Saturdays, Sundays and Feb. 15 but not run Feb. 16	6:05 PM	—	6:45	7:50	8:19	8.29	9:03	9 55	10:10 PM
158 Daily	7:25 PM	ε 7.38	8:11	9:12	9:42	9:58	10:26	11:20	11:35 PM
112 Metroliner Sundays thru Fridays	8:30 PM	—	9:03	9:47	10:13	—	10:39	ε11:16	11:30 PM
176 Daily	10:15 PM	—	10:50	12:04	12 51	1 01	1:31	2.24	2.49 AM

[illegible]

Source: Murphy, R. T. (1975b). Assessment of adult reading competence. In Duane M. Nielsen & H. F. Hjelm, (Eds.), Reading and Career Education. Newark, DE: International Reading Association. Reproduced by permission.

Chart 14

Look at the description of a group plan for blood donations. Circle whom you should call if you want to become a donor under the group plan.

The Red Cross Blood Program - Our Group Plan

If you are a member of the Red Cross Blood Program, you and members of your family are entitled to receive blood free at any hospital. In order to obtain the blood, you must have the Group Program Chairman, Joan Knapp, sign an authorization form. The form may be signed either before or after the administration of blood.

Who makes this program possible? The Plan requires that at least 20 percent of our employees donate blood during the year.

Can you become a donor? Any staff member between the ages of 18 and 66 is eligible. However, those between 18 and 21 must have the written permission of their parent or guardian.

If you have any questions about the program, or if you are willing to become a blood donor, please call Rex Jackson at 231-0027.

Source: Fisher (1981)

Chart 15

Functional Illiteracy by Age Cohorts

APL 1974	% Functional Incompetent	Years in K-12 School
18-29	16	1951-1974
30-39	11	1941-1962
40-49	19	1933-1952
50-59	28	
60-65	35	

Harris I 1970	< 90%	< 70% Correct (F. ill)	Years in K-12 School
16-24	9	1	1952-1972
25-29	11	2	1947-1963
30-49	11	2	
50+	17	5	

Harris II 1971	Unwtd. < 90%	Wtd. < 80%
16-24	2	1
25-30	6	2
31-49	7	3
50+	10	9

AFRS 1973	% Incorrect	Years in K-12 School
16-19	28	1960-1975
20-29	23	1950-1971
30-59	26	
60+	39	

MAFL	% F. Ill	Years in K-12 School
1971	83.7%	1960-1972
1974	85.6%	1963-1975
1975	85.9%	1964-1976

Unwtd. = Unweighted
Wtd. = Weighted

Chart 16

Sample Items from the Brief Test of Literacy

Example:

It was a beautiful gift, wrapped with bright red paper and tied with silver string. It was small, but very heavy. No one knew who had brought it, but it had Mr. Jones' name on top. Mr. Jones just smiled and said, "I'll open it when I get home."

01. Whose name was on the top of the gift?

a) Mr. Jones
b) Mr. Pike
c) Willy
d) The postman
e) No one knew

02. In what color paper was the gift wrapped?

a) red
b) silver
c) green
d) orange
e) yellow

03. Where was the gift going to be opened?

a) Where it was found
b) At the police station
c) In the car
d) At the office
e) At home

Source: Vogt (1973).

Chart 17

Reading Grade Levels for Texts

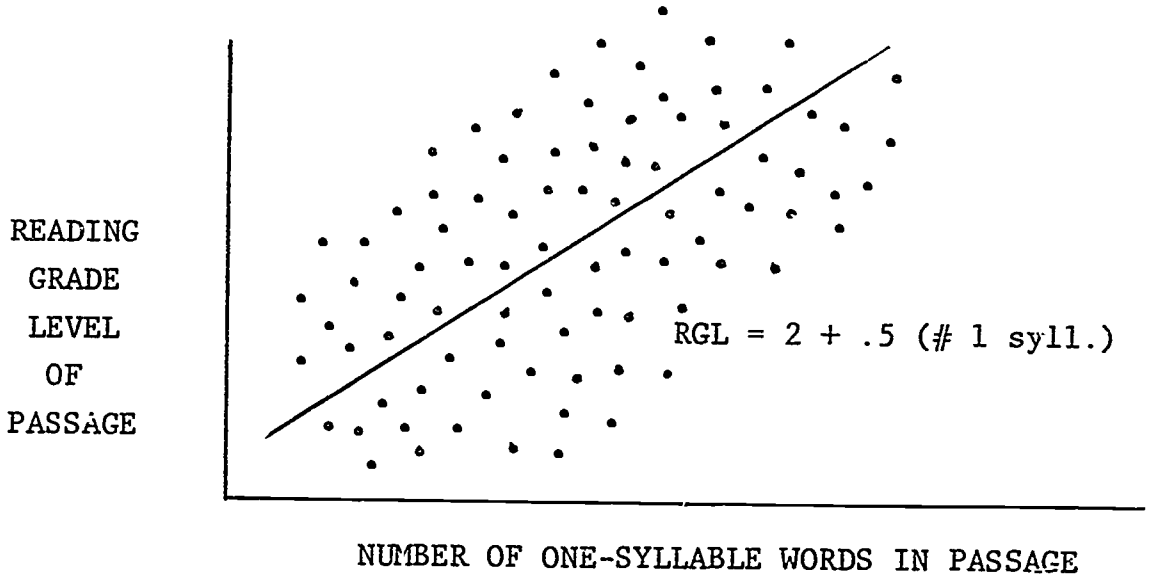

Chart 18

Cloze Procedure--Rosalynn Carter's Girlhood

Her early childhood was ---- by her father, a ----, handsome, curly-haired school bus ---- named William Edgar Smith. ---- adored him even though ---- would sometimes spank her, ---- tell her not to ----. "And I wouldn't," she ----. "But later I would ---- to the outdoor privvy ---- cry and cry there ---- alone."

She writes that ---- childhood ended at the ---- of 13 when her ----, suffering from leukemia, told ---- and his three younger ---- that he was dying. ---- family's economic existence after ---- death was difficult. She ---- she lost her enthusiasm ---- self-confidence, but gradually righted ---- and eventually attended college, ---- because it was her ---- dying wish.

As a ---- sophomore, she says she ---- in love with Jimmy's ---- on visits to the ---- home. As children in ----, she writes, "I don't ---- ever having said a ---- to him except . . .

Source: Fanlund (1984, p. 18).

Chart 19

Percent of American Workers in Major Occupational Groups, 1900 and 1970

	1900	1970
Total Labor Force (In 1000's)	29,030	78,627
White-Collar Workers	17.6	48.3
Professional & Technical Workers	4.2	14.2
Managers, Officials, & Proprietors	5.8	10.5
Clerical Workers	3.0	17.4
Sales Workers	4.5	6.2
Blue-Collar Workers	35.8	35.3
Craftsmen & Foremen	10.5	12.9
Operatives	12.8	17.7
Non-farm Laborers	12.5	4.7
Service Workers	9.0	12.1
Private Household Workers	5.4	
Other Service Workers	3.6	
Farmworkers	37.5	4.0

Source: DeFleur, M. L.; D'Antonio, W. V.; & DeFleur, L. B. (1976). Sociology: Human Society. Glenview, IL: Scott, Foresman and Company, p. 231.

Chart 20

Different Literacy Measures for Roughly Comparable Age Groupings and Years Among Youth

Measure	Age Grouping	Years	Percent Illiterate
Census	14-17 16-24	1969	.3
Educ. 0-4 Years[1]	14-15 16-24	1969	.4 - .7
Survival Literacy	16-24	1970	1.4 (f. ill.) Forms
Educ. Less Than 8 Years[2]	16-17 18-19	1970 1970	3.9 3.3
4th Grade Reading Level[3]	16-17	1966-70	4.8 7 Word Passages
MAFL	17	1974	13.2 (f. ill.) F.L. Battery
APL	18-29	1975	16.0
8th Grade Reading Level[4]	10th Graders 12th Graders	1970 1970	20-26 Test Norms (CTBS, STEP) 13.0 STEP II Norm
Educ. Less Than 12 Years[2]	20-21 22-24	1970 1970	21.0 21.6

[1] U.S. Bureau of the Census (1971)

[2] U.S. Bureau of the Census (1973)

[3] Vogt (1973)

[4] Fisher (1978), p. 37-38.

N.B. Adult Functional Reading Study results cannot be included--they did not calculate the percentage of illiterates... their figure of 28% is the percent of wrong answers.

Survival Literacy (Harris and Associates, 1970)
M[illegible]adway & Wilson, 1976)
A[illegible]ult Performance Level Project, 1977)

APPENDIX A: LITERACY GAPS ACROSS GROUPS

Male-Female Differences

On most measures of literacy, at most points in this century, women have outperformed men. Typically, the differences are not great. On crude literacy, for example, among 14- to 24-year-olds in 1910, 5 percent of women and 6.3 percent of men were illiterate. By 1969, the one-percentage-point difference had almost vanished, but women still had a slightly lower rate, .2 percent to .3 percent (This was likely within the sampling error range). The 1930 data showed an unusual and unexplained reversal, with the male illiteracy rate at 1.7 percent and the women's at 2.7 percent. The pattern for men from 1930 to 1950 also was unusual, showing one of the few increases in illiteracy for any subgroup or regions described. Again, the data were for 14-to 24-year-olds.

On educational attainment in 1910, 21.6 percent of the women had not completed fifth grade, whereas 25.9 percent of the men had not. By 1979, the gap had closed, but it still favored women, 3.7 percent to 3.2 percent. Among the young, 25 to 29, the negligible difference favored men, .9 percent to 1.0 percent. For the high school completion standard, again the differences were small and until recently favored women.

Most of the functional literacy tests display the same gender gap. On the Survival Literacy test, for example, 3 percent of the men were functionally illiterate whereas only 2 percent of the women were. For those who failed to reach functional literacy, the rates were 14 percent and 11 percent. On the MAFL, the average seventeen-year-old girl scored 86.5 percent in 1975, whereas the boy scored 85.2 percent. The 1.3 point gap was down from a 2.9 gap in 1971, so the gender gap had closed by more than half. In 1975, 89.1 percent of the girls were functionally literate, whereas 85.4 percent of the boys were. The only exception among the functional literacy tests was the APL, which found a male advantage, 17 percent vs. 23 percent, falling into APL 1, the functional incompetent category. (Could this be due to the method of construction? If the items were designed to favor higher occupational status, in which men are overrepresented, fewer would be in APL 1 category!)

The data for reading performance from standardized tests is almost as clear-cut in demonstrating a female advantage. On the verbal portion of the SAT from 1967 through 1971, women outscored men, although this reverses in 1972 and stays that way through 1975, the last year for which we have data. On the PSAT, from 1959 to 1972, girls outscored boys in all years but 1965-66. In 1973-74 and 1974-75, the boys outscored the girls. Again, 1974-75 is latest data we have gathered. On the ACT from 1965 to 1974, female students outscored males by large and stable margins.

As measured by reading grade levels, female students again outperformed males. The Brief Test of Literacy, for example, determined that 6.7 percent of boys aged 12 to 17 read below the beginning fourth-grade level, whereas only 2.8 percent of the girls did.

Black-White Differences

On all measures of literacy, at all points during this past century, whites have outperformed nonwhites. In 1880, the differences were vast. On the lower-level measures such as crude literacy and few years of schooling, the differences have virtually disappeared, while on the higher-level measures such as functional literacy tests and high school completion, they remain large. In 1880, for example, only 9.4 percent of whites were illiterate compared to 70 percent of nonwhites and blacks.[1] By 1979, the white rate was .4 percent, while that of blacks was 1.6 percent (for those 14 years and older). Crude literacy had become nearly universal, and the black-white literacy gap had nearly vanished. The current difference is almost entirely due to elderly illiterate black people. Among youth, aged 14 to 24, the rates are .18 percent for whites and .23 percent for blacks, well within sampling error.

On educational attainment, among those aged 25 to 29, 12.9 percent of whites in 1920 had fewer than five years of schooling compared to 44.6 percent of blacks. By 1981, the rates were identical at .7 percent. For the high school completion standard, the gap was 15 percentage points in 1920; 78 percent of the whites had not graduated compared to 93.7 percent of the blacks. In 1981, the gap remained large: 12.4 percent of the whites reported they had not completed high school compared to 21.3% of the blacks.

The functional literacy tests also display consistent, large racial gaps. On the Survival Literacy test, for example, 8 percent of blacks were functionally illiterate, whereas only 2 percent of the whites were. For those who failed to reach functional literacy, the percentages were 22 percent for nonwhites and 12 percent for whites. On the MAFL, the average 17-year-old black scored 74.1 percent correct, while the average white scored 87.8 percent. This 13.7 point gap was down slightly from 1971, when it was 15.6. Functional illiteracy was defined on this test as failure to achieve a score of 75 percent correct. By this standard, 41.6 percent of black 17-year-olds were functionally illiterate, while 9.2 percent of white 17-year-olds were. Of course, many black respondents scored near the 75 percent cutoff. However, the test originally was designed so that all 17-year-olds could answer all questions correctly. Thus, a 75-percent-correct standard does not seem unreasonable. Nevertheless, as on the Survival Literacy test, it might have helped to have had three or more categories: functionally literate, in-between, and functionally illiterate. In this case, on the MAFL, data on how many failed to reach the 60 percent cutoff also were presented. The percentage of black illiterates drops dramatically, from 41.6 percent

to 14.7 percent, but still remains far greater than the white illiteracy rate of 1.2 percent. On the APL, less than 20 percent of the whites were found to be functionally incompetent, while over 40 percent of the blacks were, and over 50 percent of those with Spanish surnames were.

In terms of general reading performance, on the NAEP 1970-1980 reading testing, the average black 9-, 13-, and 17-year-old scored below the average white at those age levels, and while there were improvements for black 9- and 13-year-olds over the decade, blacks remained behind the whites (17-point gap for 9-year-olds closed to 10 points, a 17-point gap for 13-year-olds closed to 13, the gap for 17-year-olds remained about 19 points). On the Brief Test of Literacy, which determined the percentage of 12- to 17-year-olds unable to read at beginning fourth-grade level, 15 percent of the black youth were considered literate compared to 3.2 percent of the whites.

Certainly a degree of sociological sensitivity is necessary in interpreting these gaps. Blacks are disadvantaged in this society in several ways that are known to affect educational performance: a higher incidence of poverty, lower levels of parental education and income, and concentrations in the South. Still, large racial differences persist even after controlling for, or accounting for these other factors. The Survival Literacy study, for example, showed that a white-nonwhite gap existed even when controlling for income. Among those earning less than $5,000, whites had a functional literacy rate of 4 percent, while nonwhites rate was 8 percent double--that of whites. The Literacy Among Youth study (Brief Test of Literacy, Vogt, 1973) found an overall literacy rate of 15 percent for blacks and 3.2 percent for whites. A breakdown comparing blacks and whites at the same educational levels found that large differences persisted. Among those with parental educational of one to eight years, the black illiteracy rate was 18.2 percent compared to 6.5 percent for whites. For those whose parents had 9 to 12 years of school, the black rate was 12 percent, while that of white youth was 2.3 percent. Only for those with parents who had more than twelve years of school did the percentage point difference shrink greatly--the black rate was 1.8 percent, the white .6 percent. The regional breakdown showed similar patterns in the South, which had the most unfavorable illiteracy rate overall; whites had a 5.9 percent rate while blacks had a rate of 20.7 percent. A smaller, but nonetheless substantial, gap persisted in the Midwest, which had the most favorable overall illiteracy rates. Illiteracy rates for whites were 1.7 percent compared to 9 percent for blacks.

These comparisons dealt with only two variables at a time, so that the racial correlation may mask other underlying variables. Reporting on a multivariate analysis of racial differences in cognitive tests, however, Jencks et al. (1972) concluded that less than one-third of the differences could be accounted for in terms of economic background (p. 82). These analyses of contributions by other factors to the racial gaps suggest, therefore, that while a

substantial portion of the black-white differential in illiteracy can be explained in terms of regional, educational, and economic differences, a major portion cannot be.

APPENDIX B: THE MAGNITUDE OF THE DECLINE IN READING ACHIEVEMENT

As the text notes, the decline has been described in dramatic terms. But what were these declines in concrete terms? When we focus on the actual tests, a less catastrophic view of the decline emerges.

We must keep in mind what "reading" tests are. Reading comprehension tests primarily measure students' ability to process rapidly a series of short prose passages, answering 40 to 50 multiple-choice questions on them in 35+ minutes. The test, at least at the high school level, has nothing to do with the placement or academic prospects of most of the students--there is little reason to believe that students should feel extrinsically motivated to do well. The tests concentrate on literal comprehension, and certain important reading skills, often called "critical reading skills," such as understanding motivations for actions, distinguishing fact from opinion, and determining point of view, are only minimally represented. Other important reading skills are not measured at all. There is, for example, no testing of recall for information read, yet reading for retention is an important skill. One essential element of reading that is deliberately avoided is the process of integrating new information. Test designers strive to make their questions independent of the respondent's prior information. But everyday reading is a process of relating prior knowledge to the text at hand. The more the test-maker succeeds in making the test knowledge independent, the less like everyday reading the test becomes.

Given the limited nature of such tests and their motivationally bland quality, changes in performance seem less worrisome. How great a change occurred? A year's decline sounds large until one realizes that standardized tests are designed in such a way that shifts of only a few percent produce dramatic differences in grade equivalents and percentile rankings.

The tests are designed so that students in each progressively higher grade score a few percentage points higher. Thus, ninth graders on the CAT in 1970 answered 52 percent of the reading comprehension exercises correctly, tenth graders 57 percent, eleventh graders 64 percent, and twelfth graders 66 percent (CTB/McGraw-Hill, 1974). Although the percentages vary from test to test, as do the differences between grades, a number of tests show eleventh and twelfth graders differing by 1.5 to 3 percentage points, lower grades such as seventh and eighth differing by 4.5 to 7 percentage points.[2] The crucial point is this: The differences between grades are only a few percentage points. On the 1978 SRA, for example, twelfth graders scored 72 percent in reading comprehension, while eleventh graders scored 69 percent (Bode, 1981a, p. 33). A year drop amounted to only 3 percentage points. The percentages correct for eighth graders and seventh graders on the CAT in 1970, level 4, were 58 percent correct and 52 percent--a 6 percentage point difference. The reading decline

for eighth graders, however, was 3 months, or one-third of a year. This translates into a drop of only 2 percentage points (Derived from CTB/McGraw-Hill, n.d.; 1974; 1979). On the SRA, CTBS, and CAT, at least, the declines typically amounted to only a few percentage points.

Drops of this size do not seem very troubling. Knowing they have occurred is also not very useful because we do not know which aspects of reading performance have changed. The SRA 1978, for example, consists of grasping details, summarizing, perceiving relationships, drawing conclusions, and understanding the author. Which skills deteriorated? Were students performing much worse on one or two of the skills, but maintaining on the others? Standardized tests consist of different types of stimulus materials. Word passages, for example, on the CAT-70, came in four types: general, social studies, science, and mathematics. Did students do substantially worse on a particular passage type? Did comprehension decrease, for example, on reading mathematics? If so, ability to handle a particular type of material, rather than general reading ability, changed. Further questions arise. Did students do worse on difficult items or easy ones? On items that strongly differentiated students across grades or on those that adjacent grades performed similarly? If, as various analysts have suggested, items gain their discriminatory power not by tapping reading skill, but by measuring vocabulary, reasoning, or general knowledge, then reading-test-score changes would reflect changes in these skills rather than reading performance (see, e.g., Carver 1972).

The NAEP, of course, is designed to answer questions about changes in performance on particular skills. We know, for example, that during the 1970s, 17-year-olds' inferential comprehension skills declined. Their performance, however, dropped from 64 percent correct to 62 percent (National Assessment of Educational Progress, 1981). It seems hard to attach much significance to such a decline. This can be illustrated by an alternative method of expressing the decline in which the current performance level is expressed as a proportion of the prior one. Expressed in this way, the average 17-year-old in 1980 scored at 97 percent of the 1971 level on inferential comprehension (62/64 = .97). Overall, there was no change in reading. The test-score declines can be similarly expressed and, on many tests, depending upon the grade, fell in the 90 to 97 percent range.

Further complicating the interpretation of the test-score decline is the fact that performance on reading comprehension tests reflects skills other than reading comprehension. These include: test-taking skills, such as test-taking experience and knowledge of test-taking strategy; attitudinal factors, such as motivation for the test and interest in the passages; intellectual preparation, such as vocabulary, reasoning skills, and general knowledge; efficiency factors, such as carelessness, fatigue, and the ability to work quickly; and the nature of the test, such as the racial and cultural bias of items, the types of stimulus materials, and an improper--e.g. excessive--use of the mix of skills, such as vocabulary on reading tests and reading on math tests.[3]

Consequently, changes during the 1970s in student test-taking skills, attitudes, intellectual preparation, and efficiency would have affected test-score performance. The important point here is that there is a difference between everyday reading ability and test performance. To the extent that standardized tests of reading measure skills other than reading comprehension, assertions about a decline in reading are dubious.

Imagine for a moment that we are comparing two nations' reading performances. Would we really want to argue that Nation X whose twelfth graders score 66 percent on a reading test is so superior to Nation Y, whose students score 62 percent that Nation Y should revamp its curriculum and reorganize its educational institutions? Or that somehow Nation Y won't produce as many scholars, or make as many contributions to the letters, or maintain its democracy because of what amounts to an item difference in performance? We don't think so. What if Nation Y had undergone all sorts of social unrest, including bombings, student demonstrations broken up by government attacks, and the forced resignation of its highest leaders because of corruption and abuse of powers? Followed by strong currents of fundamentalism and materialism? Under such turmoil, a relatively minor difference is a testament to academic resilience, not a symptom of impending catastrophe.

We are not trying to deny that there may have been some decline in reading skills. We are emphasizing that reading is far more than what is measured by the tests, and that what is measured by the tests, is far more than reading. Furthermore, because the declines in performance typically were only a few percentage points, we believe that the magnitude of the decline has been overstated and its importance exaggerated.

APPENDIX C: THE EQUATING AND NORMING OF STANDARDIZED TESTS

The simplest approach to comparing the performance of American students over time would be to draw a representative sample of the population every few years and administer the same test to the students. This is essentially what the National Assessment of Educational Progress does.[4] The results provide a gauge of changes in actual performance of the nation's students. Only one major standardized test, the ITED in its 1971-1978 comparison, has produced results on the basis of repeated administrations of the same test.

Using norming results from standardized tests to make generalizations about changes in performance is far more complicated, and some test publishers warn against it (The Psychological Corporation, 1978, cover; Test Department, 1983, 1, 2). As should become evident, it is not an exaggeration to describe the NAEP approach which is direct, clear, and parsimonious as Copernican, and the standardized test approach which is indirect, obscure, and convoluted as Ptolemaic.

Test publishers change virtually their entire tests every five to ten years, and thus a simple comparison of test results is impossible. (The new tests consists of almost all new items, often new skills, and sometimes new time limits.) The tests themselves may have gotten harder or easier, accounting for differences in results. In order to make comparisons across editions of their tests, publishers perform equating studies. A description of the equating process is warranted, for it illustrates the difficulties and dangers involved in drawing hard conclusions from norming studies.

Publishers often administer the earlier test and the current test to a group of students. Thus, a group of 1977 students is given the 1970 test and the 1977 test.[5] The results of such testing produce an equating scale in which scores on the 1977 test are equated with those on the 1970 test. By looking up the 1977 norm (the mean or median score) on the scale, one can find the 1970 equivalent and determine whether the 1977 average was higher or lower than that of 1970. Obviously the NAEP approach is more straightforward, particularly when one dissects the equating process.

The major limitation is that the group on which the equating is done (the group that takes both the 1970 and 1977 tests) is <u>not</u> representative of the nation's student population. Rather than giving both tests to the entire norming group, supposedly representative of the nation, test publishers give both tests to a much smaller group to save time and resources. The number of school districts involved are too few to claim representativeness, and the number of students involved are too few to make reliable judgments of the population's performance. The 1970-1977 CAT equating study, for example, was based on only three school districts, which included only 111 twelfth

graders and 396 eleventh graders.[6] The 1968-1973 CTBS study involved only 125 to 256 students per grade at the eighth- through tenth-grade levels (CTB/McGraw-Hill, 1973, p. 7, p. 24). Other equating studies of this type were similarly limited.[7]

Such limited numbers cannot adequately indicate how the nation's average student would have performed on the two tests. Consequently, we must view the results of equating studies with suspicion.

The second major limitation is the equating method itself. The equating process maps scores from one test to the other on a point-to-point basis, giving the appearance, for example, that a 1977 score is uniquely equivalent to a 1970 score. In fact, students in the equating group who received a particular 1977 score vary in their scores on the 1970 test, and with such small samples, the variation can be great. Equivalent scores should thus be expressed as a range of scores. Making matters worse, publishers do not actually compare how well students at a particular level on the 1977 test performed on the 1970 test. Instead they consider the performance on the two tests separately. A ranking is made of the scores for each test, and scores are then matched on the basis of their percentile or normal curve positions. Whatever the equating group's average score on the 1977 test, for example, it is matched to its average score on the 1970 test--even though students in the equating group who scored around the group's 1977 average may not have performed at the group's average on the 1970 test. The process, therefore, does not take advantage of the fact that the same students took both the old and new versions of the tests.[8] Variations creep in further. The 1977 norm score comes from a sample of the student population and thus, much like political polling in which a candidate's standing is given as a percentage ± a few points, it also has a sampling range around it. Reliability variations (the fact that students do not make the same score if they take the same test a second time) add to this range. All this increases the range variation around the equated score, and means that one does not actually look up a given score and find a corresponding score, but rather a range is equated to a range. The equating process is thus fraught with error.

Other equating approaches used by the major standardized test publishers also are questionable. The MAT and SAT (Stanford Achievement Tests) generally have been equated by administering the old test and the new test to two different groups of pupils matched on the Otis-Lennon Mental Ability Test (See, e.g., The Psychological Corporation, 1978). This method seems particularly circuitous and Ptolemaic. Because the same students do not take both tests, any possibility of direct equating is lost, and another major source of error due to the matching is introduced. (Mismatching, SES not matched, etc.) We also have the same problem with the unrepresentativeness of the equating group. The 1970-77 equating of the ITBS was done by comparing the test results of schools that used the 1970 edition in 1977 and the 1978 edition in 1978 (Hieronymus, Lindquist, & Hoover, 1982, p. 111). In other words, the equating was across two different groups of pupils that took two different tests.

ITBS conducted two such studies, one of which involved only 31 schools! Not only does such a method introduce extra error due to demographic and ability differences between the groups (at least on the MAT and SAT methods, there is a bridge provided by Otis-Lennon), but we do not know how students who scored at particular levels on the old test would have done on the new test.

A final, somewhat tortured, method appeared in the 1973-82 SAT comparison (Test Department, 1983). Some of the items from 1973 tests were included with new items being tested in the try-out program for the 1982 edition. How students in the try-out group did on the new items can be compared to how they did on the old ones. The try-out group thus serves as an equating group, but with the limitation that only a subset of items from the old test are being equated. How this subset relates to the entire old test introduces another source of error.

Equating studies are thus a thin reed upon which to generalize about changing reading performance.

The next problem is that the norming samples that are supposedly being equated are not truly representative of the student population, and thus the results cannot be said to directly reflect national performance, even for a given year. Student performance for the norming year is not tested by drawing a random, representative sample of students who are representative of their population in terms of race, ethnicity, and family background. For practical reasons, the unit of sampling is the school district rather than the student. School districts are stratified according to geographical region and size, and sometimes by community type and/or social-economic demographic index as well. In this way, the country's school districts are divided into a variety of cells, and the sampling is done from each cell (See chart next page).

This can be a reasonable approach when the categorizations are numerous and produce homogeneous cells. The problem, however, is that districts were usually divided into a few broad categories. On the 1970 CAT, for example, the community type "urban" category referred to those cities of 25,000 people or more. As one test reviewer noted, this is "interesting," for it puts Jonesboro, Arkansas and New York City in the same category (In Buros, 1978, p. 720). Demographic social indexes are an improvement but are dependent upon the breadth of the other categories. Poor, small town districts, for example, could be in the same cell as poor, large, urban districts. Sampling from such cells, therefore, does not guarantee proper representation of the nation's school districts and hence cannot do so for its students, either.

Thus, even if a standardization is based upon thousands of students (well over the 1500+ Gallup and Harris use to sample the population), it can still be unrepresentative because the unit of sampling is the school district rather than the student. As Stanley and Hopkins note: "The size of the standardization sample is much

less critical than its representativeness" (Stanley and Hopkins, 1972, p. 85). Even though they tested thousands, the 1970 CAT was criticized for underrepresenting urban minorities, the 1970 MAT for overrepresenting them (In Buros, 1978, p. 39, 70). This has been improved upon in some standardizations by deliberately sampling from the country's largest city school districts.

There are other problems with the sampling. Typically, only Catholic private schools are included--independent and other sectarian private schools are ignored. Furthermore, since the testing takes place through schools, the sample does not capture the performance of those out of school. This is a major problem as one reaches the high school level, since dropouts are not accounted for. These factors further weaken the representativeness of the norming sample. (Over time, any high school norm comparisons must account for changes in the composition and performance of those who remain in school.)

After sampling, there is no systematic process by which the test producer ensures representativeness. In 1978, SRA did weight their final sample reflect the proper proportions of various ethnic, racial, and sexual groups in the nation (Science Research Associates, 1979, p. 5). This effort at correction may not produce accurate norms, however, because weighting a portion of the original sample can exaggerate the impact of certain groups if the sample was not properly representative. Consider, for example, adjusting the black percentage. If urban poor blacks had been underrepresented in the sample, weighting blacks exaggerates the contribution of well-to-do blacks in the final norming. Weighting, therefore, cannot substitute for proper sampling in the first place.

The only check the test producers make on representativeness is to administer a questionnaire completed by school administrators on the characteristics of the student population attending their schools. The answers are checked against national statistics. This is a perfunctory check, however, serving to rationalize the process, because there is never any adjustment made in the norming results. Although such a check must be highly unreliable (the answers reflect the administrators' opinions of their school's demograhics rather than hard knowledge), the check is used as justification for representativeness (CTB/McGraw-Hill, 1979, p. 62; 1982, p. 84). Furthermore, there are serious discrepancies between questionnaire data and nationally gathered census and survey data (whether these are real or reflect the unreliability of administrators' responses cannot be determined). Students from poor families and less well-educated families are often seriously underrepresented. National statistics in the late 1970s showed 10 percent of the population's families had incomes below $5,000, whereas on the 1978 SRA, for example, only 4.9 percent of the standardization sample came from such families (Bode, 1981b, p. 25-26). National statistics showed 33.4 percent of the population without high school degrees compared to only 10.6 percent of the family heads in the 1978 SRA standardization, 22 percent of the 1980 CTBS standardization, and 23.3 percent on the 1977 CAT (Bode,

1981b, p. 26; CTB/McGraw-Hill, 1982, p. 93; CTB/McGraw-Hill, 1979, p. 63).

Student self-descriptions would presumably be more accurate than administrators' reports, yet these, too, show discrepancies. National statistics showed 7 percent Hispanics in 1978 compared to only 4.4 percent on the SRA (Bode, 1981b, p. 25, 26). Overrepresentation occurs as well. The 1973 CTBS, for example, showed 7.9 percent Hispanics, and 16.7 percent blacks, whereas the 1970 national figures were 5.1 percent and 14.9 percent (CTB/McGraw, 1974a, p. 66).

The result is that the supposedly national samples are skewed in various ways and thus are not truly representative.

Furthermore, there is the problem of response rate. A substantial portion of the schools that are invited to participate in standardizations refuse. On the 1970 CAT, for example, 40 percent of the public schools districts chosen declined (CTB/McGraw-Hill, 1974, p. 39). In the Southeast, the proportion was 50 percent. There is the danger that turn-down is related to school achievement. Replacements are made from the same cell, which presents its own problems since the districts in each cell, as noted, can be quite heterogeneous and thus nonequivalent substitution can occur. This danger of a biased sample due to participation turn-down has been noted by numerous educational psychologists (See, e.g., Stanley and Hopkins, 1972, p. 84-5).

Changes in sampling between normings can also undermine equatings. The sampling procedure for a given test varies from norming to norming. The number of geographic regions may vary, and various factors, such as demographic indexes, large city sampling, and community type get dropped or introduced. Some of these changes definitely account for a substantial portion of changes in the achievement levels between normings. The late 1950s to mid-1960s comparisons, cited by test-score-decline reviewers, were based on norming samples that in the 1950s included only public schools, but in the 1960s added private schools, thus raising scores. Much of the apparent increase during that period, therefore, can be attributed to the inclusion of the private schools (See Schrader, 1968). Similarly, apparent declines on some of the major standardized tests correspond to changes in the norming sample. The 1968 to 1973 CTBS and the 1970 to 1977 CAT declines were likely created in part by the inclusion of large city sampling in the later years. Such large city districts have a higher proportion of minority and poor students and thus have lower achievement, and contributing to the decline. The 1977 CAT norming sample, for example, included Washington, D.C., Baltimore, Brooklyn, and Dallas, whereas the 1970 sample included only Chicago among large city districts (cf. CTB/McGraw-Hill, 1979, p. 151-2 and CTB/McGraw-Hill, 1974b, p. 103-4). The portion of the difference in norming achievement levels that can be accounted for by changes in sampling cannot be calculated without information on student family income and race, and the corresponding achievement test scores of such groups.

When one considers that a given norm is best with sampling error, reliability error, and nonrepresentativeness due to sampling procedure and participation patterns--and that these problems are even more acute in the equating samples--it is a wonder that anyone ever dared to use norming and equating studies to catalog achievement trends. As the publishers of the MAT 1970-78 comparison stated about their equating study:

> These data are not appropriate for making generalizations concerning changes over time in the relative achievement of American students in the basic skills areas (The Psychological Corporation, 1978, p. 1).

APPENDIX D: NAEP VS. STANDARDIZED TESTS

Although we could not make a direct comparison of NAEP test questions with reading-achievement test items, we did compare the skills the tests were supposed to measure and how many items were devoted to each skill. We looked at two high school reading comprehension tests: the 1970 California Achievement Test and the 1978 SRA Test. The NAEP reading test was focused on the same types of reading skills as the standardized tests: literal and inferential comprehension. Both had similar proportions of the higher order or inferential comprehension items. On the NAEP test for 17-year-olds, for example, 25 of 71 exercises, or 35 percent, tested inferential comprehension (National Assessment of Educational Progress, 1981, p. 4). On the 1978 SRA test, 13 of 50 reading comprehension exercises, or 26 percent, measured drawing conclusions (Bode, 1981a, p. 33). On the CAT, 20 of 45, or 44 percent, measured generalizations and inferences.[9]

The actual testing results show the NAEP is just as hard as the standardized tests. On the NAEP, the average 17-year-old answered correctly about 69 percent of the exercises. On the 1970 CAT reading comprehension test, eleventh graders answered correctly 64 percent of the items, twelfth graders, 66 percent (CTB/McGraw-Hill, 1974, p. 49). On the 1978 spring SRA, eleventh graders averaged 69 percent correct, while twelfth graders averaged 72 percent (Bode, 1981a, p. 37). Inferential comprehension items actually were harder on the NAEP! The average 17-year-old answered 64 percent of such items correctly in 1971, 62 percent in 1980 (National Assessment of Educational Progress, 1981, p. 23). On the CAT, eleventh and twelfth graders answered 65 percent and 66 percent of such items, while on the SRA, they answered 72 percent and 74 percent (Calculation from CTB/McGraw-Hill, 1974, p. 25; Bode, 1981a; p. 33). As to the claim that the NAEP reading test was at the elementary school level, it is worth noting that, on the items their test shared with 17-year-olds, nine-year-olds answered only 30 percent correctly. Seventeen-year-olds answered 68 percent (National Assessment of Educational Progress, 1981, p. 63).

Although an analysis of actual questions might show that the tests measured different skills, the argument that the NAEP involves easier skills cannot be maintained. Furthermore, anyone who categorically asserts there was a major reading decline in the 1970s must explain the improvement on the SAT-MAT 1973-78 comparison and the stable or rising scores on the other test results mentioned, including NAEP. They must also acknowledge that on the skills NAEP describes as reference, literal comprehension, and functional literacy, there was no decline among 17-year-olds during the decade.

APPENDIX E: READABILITY OF THE ROSALYNN CARTER PASSAGE

On the cloze test of the Rosalynn Carter passage, the answers were:

dominated, tall, driver, she, he, then, cry, writes, go, and, all, her, age, father, Rosalynn, children, Her, his, says, and herself, primarily, father's, college, fell, picture, Carter, Plains, remember, word

Using the FORCAST formula, this passage would be assigned a 10.3 reading grade level. So those readers who completed 35 to 40 percent would be reading at a tenth-grade reading level. Higher completion rates mean a higher grade level. For comparison, and to point out another limitation of readability formulas, Hamlet's soliloquy would be assigned 8.1, based upon the first 150 words, even though it is obviously more difficult, sophisticated, and loaded with meaning than the Rosalynn Carter passage.

FOOTNOTES TO APPENDICES

[1]This was for those ten years and older. Although not available for 1880, the black illiteracy rate was likely similar. The 1870 and 1890 nonwhite and black rates were virtually identical.

[2]A year's difference is an artificial construction of the test designers. On the 1970 CAT the difference between eleventh- and twelfth-grader performance is 1.4 percentage points; on the 1978 SRA, it is 3 points. On the 1970 CAT the tenth-eleventh difference is 7 percentage points, whereas it is only 4.5 on the 1977 CAT (cf. CTB/McGraw-Hill, 1974; Bode, 1981b, p. 33; CTB/McGraw-Hill, 1979).

[3]The effects of these factors on test scores can be substantial. A great deal of research, for example, has demonstrated that test anxiety can lower scores and test preparation can raise them (Willig et al., 1983; Bangert-Drowns et al., 1983; Stewart and Green, 1983; Sarnacki, 1979). Bangert-Drowns et al., in a systematic review of studies of test coaching on achievement tests, found coaching could raise students' scores by one-fourth of a standard deviation. This can amount from a half-grade level to over a full-grade level, depending upon the subject matter, grade, and test (See, e.g., Science Research Associates, 1979, p. 19-23; CTB/McGraw-Hill, 1974b, p. 48). Even something as apparently trivial as whether the answer sheets are in the test booklets or on separate sheets has been shown to change scores by as much as three months (Bridge, Judd, and Moock, 1979, p. 49).

Test designers have attempted to control for the effects of bias and speededness, but their commendable efforts have not been as successful as one might have hoped. During the past decade, for example, they have used minority review panels and elaborate statistical procedures to eliminate biased test items. Nevertheless, criticisms of the tests for their cultural and racial bias have persisted (Nitko, 1976; Tyler and White, 1979), and even those detecting bias for the publishers have questioned their methods. Green (1982), for example, who is the research director for the CTBS and CAT tests, writes: "In fact, the evidence about the effectiveness of these procedures is so thin that one might well wonder how we got into this" (p. 234). Several studies have shown that panel members are consistently unable to identify which items are harder for minority students than for whites (Berk, 1982, p. 21). The statistical methods cannot detect bias that pervades the test because they use the test itself as a reference point (Articles in Berk, 1982, p. 23, 135, 173).

Speededness controls have been weak. The 1978 SRA was checked by completion rates but the results were troubling. For fourth to sixth graders, for example, 6 percent to 11 percent of the students failed to complete the reading comprehension section, and 8 percent to 18 percent failed to complete math computation (Bode, 1981b, p. 3, appendix E), even with a liberal definition of "complete" (90% of the

items). Furthermore, many of those who did complete the test may have hurried through, not been able to double-check their answers, or been anxious because of the time limit, hence hurting their scores. The check on speededness, therefore, should not have been completion rates, but whether students' scores would have improved had more time been given. Designers of the 1970 CAT checked speededness this way, but they tested only 30 students for a given level of the test, and the most additional time they allowed on any section was only nine minutes. For most sections, only a couple of extra minutes were given. The time limits that were eventually adopted for several sections were actually shorter than the time at which students were still improving (CTB/McGraw-Hill, 1974). The MAT also has been questioned on speededness grounds. The intermediate level tests (sixth-eighth) were found to be "too highly speeded or too hard, or both" (Wolf in Buros, 1978, p. 69). Since answering correctly only one or two more questions can raise grade level scores from several months to half a year, the speededness problem can have a large impact on test scores.

[4]About half the items are released publicly, and new ones replace these. New skills and items are also added from testing-to-testing, but any comparisons made about performance over time involve identical items.

[5]Half the group takes the 1970 first, the other half the 1977 first, in order to balance practice and fatigue effects.

[6]For the equating analysis, they combined eleventh and twelfth grades. Tenth-grade equating involved students, but ninth involved only 281; eighth; 549; and seventh, 567 (CTB/McGraw-Hill, 1974b, p. 91; CTB/McGraw-Hill, 1979).

[7]1963-1970 CAT, only four school districts; 1973-1982 CTBS equating, one of the better ones, still involved only a fraction of the districts and students that were tested in the 1980 norming study: 44 vs. 119 districts; 3,752 students for ninth-twelfth, vs. 12,000+ fall norming and 23,000+ spring norming (CTB/McGraw-Hill, 1982, pp. 59, 77, 78). The equating group was not a norming sample and apparently was culled from willing, routine CTBS customers (p. 59).

[8]In fact, what is often done is to give half the equating sample one test, and the other half the other test, so there is no direct linkage at all.

[9]Three other CAT generalization items were part of reference skills and were not included in the above count (CTB/McGraw-Hill, 1974b, p. 4). Two differences among the tests should be mentioned, although we feel their impact was minor. Unlike most standardized tests, the NAEP included reference skills items in its reading test. But so did the CAT, and in a similar proportion. If reference items are considered "advanced comprehension skills," as the CAT publishers described them (CTB/McGraw-Hill, 1974b, p. 4), then this would increase the proportion of NAEP higher-order-skills items to 51 percent and the CAT

to 57 percent. The SRA included seven "understanding the author" items which, if considered higher-order skills, would up its proportion of such items to 40 percent (Bode, 1981a, p. 33). In any case, accounting for these other items still leaves the tests' emphasis on higher-order items in the same ballpark.

Research Staff

LEARNING AND DEVELOPMENT AREA

B. Bradford Brown
Assistant Professor
Educational Psychology

Anne M. Donnellan
Associate Professor
Studies in Behavioral Disabilities

William Epstein
Professor
Psychology

Arthur M. Glenberg
Associate Professor
Psychology

William M. Reynolds*
Professor
Educational Psychology

Laurence Steinberg*
Professor
Child and Family Studies

CLASSROOM PROCESSESS AREA

Thomas P. Carpenter
Professor
Curriculum and Instruction

Elizabeth H. Fennema
Professor
Curriculum and Instruction

Penelope L. Peterson
Professor
Educational Psychology

SCHOOL PROCESSES AREA

William H. Clune*#
Professor
Law

Gary D. Gaddy*
Assistant Professor
Journalism and Mass Communication

Adam Gamoran*
Assistant Professor
Sociology

Carl A. Grant
Professor
Curriculum and Instruction

Herbert J. Klausmeier
Founding WCER Director and V.A.C. Henmon Professor
Educational Psychology

Mary H. Metz*
Associate Professor
Educational Policy Studies

Fred M. Newmann*
Secondary Center Director and Professor
Curriculum and Instruction

P. Martin Nystrand*
Associate Professor
English

Janice H. Patterson*#
Assistant Scientist

Allan J. Pitman
Lecturer
School of Education
Deakin University

Stewart C. Purkey*#
Assistant Professor
Lawrence University

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller*
Professor
Educational Administration

Robert A. Rutter*
Assistant Scientist

Gary G. Wehlage*
Professor
Curriculum and Instruction

Kenneth M. Zeichner+
Associate Professor
Curriculum and Instruction

SOCIAL POLICY AREA

William H. Clune*#
Professor
Law

W. Lee Hansen°
Professor
Economics

Carl F. Kaestle
Professor
Educational Policy Studies and History

Joseph F. Kauffman°
Professor
Educational Administration

Cora B. Marrett*
Professor
Sociology and Afro-American Studies

Michael R. Olneck
Associate Professor
Educational Policy Studies and Sociology

Thomas A. Romberg
Professor
Curriculum and Instruction

Francis K. Schrag*
Professor
Educational Policy Studies

Marshall S. Smith*#
WCER Director and Professor
Educational Policy Studies and Educational Psychology

Jacob O. Stampen°
Assistant Professor
Educational Administration

Research Support Staff

Jacob Evanson
Statistical Data Analyst

Janice Gratch
Project Specialist

Susan D. Pittelman
Project Coordinator

Deborah M. Stewart
Administrative Program Manager

Dan G. Woolpert
Program Coordinator

*affiliated with the National Center on Effective Secondary Schools, University of Wisconsin-Madison
#affiliated with the Center for Policy Research in Education, Rutgers University
°affiliated with the Center on Postsecondary Management and Governance, University of Maryland
+affiliated with the Center on Teacher Education, Michigan State University