DOCUMENT RESUME

ED 297 302 CS 009 256

AUTHOR Calfee, Robert
TITLE Indicators of Literacy.
INSTITUTION Center for Policy Research in Education.; Rand Corp., Santa Monica, Calif.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CPRE-JNE-04
PUB DATE Aug 88
GRANT OERI-G008690011
NOTE 51p.
AVAILABLE FROM The Rand Corporation, 1700 Main St., PO Box 2138, Santa Monica, CA 90406-2138.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Viewpoints (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; Communication (Thought Transfer); *Educational Policy; Elementary Secondary Education; *Evaluation Criteria; Futures (of Society); Grade 6; Language; *Literacy; Problem Solving; School Districts; Teacher Role
IDENTIFIERS Indicators

ABSTRACT

Predicated on the belief that current methods for assessing literacy give policymakers too narrow a view of performance and should be augmented by the informed professional judgment of classroom teachers, this report appraises how well schools in the United States are meeting the goal of a literate populace. The report seeks to improve understanding of what students need to know to employ language as a tool for problem-solving and communication, and how states and local school districts might measure whether or not their students are reaching such a level of critical literacy. First the report sketches a vision of literacy for future generations that is tied not to the printed page, but rather to a formal style of language that depends on structures, procedures, and strategies for effective use of language in thinking and communicating. Next, an approach is presented for achieving this level of critical literacy. Finally, the report (1) recommends that the professional judgments of classroom teachers be used to enhance more traditional, and often incomplete or misleading, assessments of student literacy; and (2) considers barriers to and benefits of the proposal to incorporate teacher judgments in the indicator system. Ten figures and one table of data are included, and 66 references are attached. (MS)

THE RAND CORPORATION

1700 Main Street, PO Box 2138
Santa Monica, CA 90406-2138

CPRE

CENTER FOR POLICY
RESEARCH IN EDUCATION

Indicators of Literacy

Robert Calfee

August 1988

40 Years
1948-1988

RAND

The work described in this report was funded by the Center for Policy Research in Education.

The Center for Policy Research in Education (CPRE), sponsored by the U.S. Department of Education, works to improve the quality of schooling through research on state and local education policy. Its studies seek to broaden the range of options from which policymakers can choose, and to strengthen the connections among policy, practice, and performance in American elementary and secondary education. From 1985 to 1987, RAND participated in CPRE, along with Rutgers University and the University of Wisconsin-Madison.

CPRE Joint Note Series

Joint Notes are issued by CPRE to faciliate the exchange of ideas among policymakers and researchers who share an interest in education policy. The views expressed in these reports are those of individual authors, and are not necessarily shared by the U.S. Department of Education or The RAND Corporation. This publication was funded in part by the U.S. Department of Education, Office of Educational Research and Improvement, Grant number OERI-G008690011.

CPRE

CENTER FOR POLICY
RESEARCH IN EDUCATION

JNE-04

Indicators of Literacy

Robert Calfee
Stanford University

August 1988

40 Years 1948-1988

RAND

PREFACE

One effect of the education reform movement of the past few years has been the heightening of policymakers' and educators' interest in obtaining better indicators of educational performance. They recognize that understanding trends in student achievement, why they occur, and how to fix problems or replicate successes depends on much more than student scores on standardized tests. Effective policy and practice require systematic data about the often complex outcomes of schooling, and about the teaching and learning environment experienced by different kinds of students.

This Note seeks to improve understanding of what students need to know to employ language as a tool for problem-solving and communication, and how states and local school districts might measure whether or not their students are reaching such a level of critical literacy. It recommends that the professional judgments of classroom teachers be used to enhance more traditional, and often incomplete or misleading, assessments of student literacy.

The analysis and recommendations presented here should be of interest to state policymakers and their staffs, local school board members, and professional educators.

Robert Calfee is a Professor of Education at Stanford University and an elected school board member in Palo Alto, California.

SUMMARY

A literate populace is essential to the well-being of any modern democratic state. Moreover, the graduates of the nation's schools must achieve a level of *critical literacy* sufficient to employ language as a tool for problem-solving and communicating. Minimum competency in *functional literacy* will not suffice.

How well are American schools meeting this goal? That is the key question for indicator systems that measure literacy outcomes. The answer depends on the clarity with which the goal is framed and the validity of the measures used. There is no lack of data. A plethora of reading measures floods the national press; technical reports flow from state departments of education and research institutions. Assessment of writing is less commonplace, but the situation is improving.

Despite the availability of information, however, *current methods for assessing literacy give policymakers too narrow a view of performance and should be augmented by the informed professional judgment of classroom teachers.* Multiple-choice tests of reading skill ask the student to do little more than recognize details and simple relationships, whereas genuine comprehension is a reconstructive activity. The student who can compose a personal narrative, a report, or an essay is not necessarily able to design and create a well-organized expository report. Current indicators are not only narrow; they are easily invalidated by "teaching to the test."

This Note sketches a vision of literacy for future generations that is tied not to the printed page, but rather to a formal style of language that depends on structures, procedures, and strategies for effective use of language in thinking and communicating. This capability is within the grasp of *all* youngsters, who deserve full access to this level of literacy regardless of their demographic characteristics.

Next, an approach is presented for achieving this level of critical literacy. Current findings are not very promising. Small gains in reading and writing in the early grades do not lead to improvements in later grades. Many young adults cannot handle even moderately complex literacy tasks, and instruction in the elementary years is not yielding transfer of knowledge and skill. The situation may, in fact, be much worse than we realize, given the narrowness of current indicators. Hence we recommend that teachers be asked for information. This proposal does not imply that the present system of indicators should be discarded, but rather that it should be augmented by the informed judgments of classroom teachers.

Finally, the Note considers the barriers to and benefits of this proposal. The main question is whether teachers are willing and able to make judgments about student achievement. We believe the answer is *yes* on both counts. Will these judgments be trustworthy and unbiased? Yes, if schools promote a level of professional competence by practitioners that ensures this outcome.

CONTENTS

FIGURES

TABLE

I. INTRODUCTION

Each year, school boards throughout the United States attempt to evaluate the performance of their districts' students on the basis of standardized achievement tests. Each year, these boards ponder the same questions: What do the numbers really mean? Are the tests measuring significant district goals for students? What about "nonbasic" areas such as writing, science, and citizenship? Why are the scores high or low? Will declining resources and increasing class sizes cause scores to drop? Are the students in the district realizing their potential? The questions remain largely unanswered.

School boards, like other policymaking bodies, rely on information to guide their deliberations. Sometimes the match between data and decision is close; often it is not. If the budget shows a deficit at year's end, it is necessary to make cuts. But if test scores drop for a particular cohort or school, the best course of action is less clear. Policymakers are generally overburdened and have little time for analysis and reflection. They want a "bottom line": If reading test scores are high, all is well; if they are low, any action that promises to raise them is a good thing. Contemporary reading tests have many virtues (e.g., consistency and cost-effectiveness), but they also have limitations (including narrow scope and a focus on low-level performance skills). Thus, we must ask whether high test scores really constitute trustworthy evidence that students are fully literate.

TWO THEMES AND A PROPOSAL

This Note focuses first on literacy, and then on the issue of *indicators*.[1] The first topic may seem to belabor the obvious. "Everyone" knows about literacy. A literate person can read and write and is probably quite good at both. But being good at a task does not necessarily mean that one *understands* the task. The value of an indicator system depends on how well it represents the construct and on the ability of the observer to make sense of the information. It helps if everyone is looking for the same thing, but there is no definite agreement on the practical meaning of literacy.

[1]Guthrie (1987) presents a brief account of *reading indicators*, focusing on standardized measures and "quantity" indices of schooling (e.g., time on task). Surveys by the Congressional Budget Office (1986, 1987) provide further background on recent trends in reading achievement. These reports illustrate the inconsistencies in standardized measures and the frustrations of trying to establish causal linkages from existing indices. Aside from the ubiquitous finding that poverty is correlated with achievement, few conclusions can be reached.

The issues of literacy and indicators can be framed as sets of distinctive questions:

- What do we expect of a student at the end of sixth grade, bound after a summer of adolescent freedom to confront the rigors of middle school? What is our vision of the student's state of literacy at this critical juncture? What should he or she know and do to succeed in high school and afterwards?
- What information about students' competence in reading and writing should be available to policymakers at various levels? How should the information be gathered? What policy questions should be informed by the data?

Present methods for assessing literacy give policymakers too narrow a view of performance and should be augmented by the informed professional judgment of classroom teachers. Most of what educational policymakers know about the literacy of students in their communities comes from group-administered multiple-choice tests mandated by central agencies. The information is objective, reliable, and cheap, but the range of assessment is limited, no direct link exists to guide practitioners in improving instruction, and those directly responsible for promoting literacy are left out of the "loop." Practitioners can give a broader view of student achievement than can a brief test. They can describe what students are learning and how they are being taught. This information is of value in its own right, and, equally important, its use contributes to the enhanced professionalization of the teacher.

Implementing this recommendation will not be simple. Elementary school teachers are seldom asked to make professional judgments, and some observers will question whether the typical practitioner is capable of this task. In fact, teachers make judgments and decisions that influence students daily, and we have little choice but to trust their assessments. The present proposal incorporates these judgments into the assessment of reading achievement and gives policymakers a clearer idea of the consistency of teacher evaluation.

Technical issues immediately come to mind: What information do we need? How can reliability and validity be assessed? What training do teachers need for the task? How can the data be combined to yield meaningful indicators? There are also political matters: How can the profession be engaged in the task? What can be done to reassure policymakers and the broader public about the trustworthiness of teacher judgments? These are not trivial questions, but the proposal offers sufficient benefits to warrant serious consideration.

TWO ASIDES

Before turning to a vision of literacy for future generations, two background issues must be briefly noted. First, *why does this study focus on the sixth grader?* The decision reflects the importance of this grade both academically and developmentally. Before a youngster leaves elementary school, he or she should have moved from *learning to read* to *reading to learn* (Chall, 1983). The goals of literacy at this grade level can be clearly defined: Sixth graders need skills and knowledge in the use of language that will guarantee success in secondary school.

Second, *is literacy the same as reading*? Literacy is sometimes viewed as a low-level skill, perhaps little more than the ability to read and write one's name or to handle the printed word in everyday life (Alexander, 1987: 20). In contrast, literacy defined as "using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential" (NAEP, 1986a: 3) goes well beyond the basic skills. In fact, the level of literacy required for the well-being of democratic society has increased substantially in recent decades (Venezky, Kaestle, and Sum, 1987: 5). Reading scores are important only as reflections of student achievement of the broader goals of literacy. This perspective meshes with the growing realization that *reading instruction* should be part of an integrated language program encompassing oral language development and writing (Anderson et al., 1985: 20ff). This broader view of language will be the point of departure in this Note, and it is assumed throughout that we are interested in a high level of competence in reading, writing, speaking, and listening—that is, a level of *critical literacy.*

II. A VISION OF LITERACY FOR THE YEAR 2000

"Although humans make sounds with their mouths and occasionally look at each other, there is no s id evidence that they actually communicate amo g themselves."
(© Sidney Harris, 1988, reprinted with permission)

What should be happening in today's kindergarten classrooms and today's schools to ensure that tomorrow's graduates will be able to handle the literacy demands they will face? We begin by describing a concept of the literate sixth grader.

THE LITERATE SIXTH GRADER

The physical differences among a group of youngsters about to graduate from elementary school are startling. Some sixth graders are sophisticated young adults, and others are still children. They are all in the midst of significant maturational, social, and educational changes, and hence most of them are both confused and confusing. Their moods fluctuate, their interests shift, and they undergo metamorphosis daily.

School does little to simplify life at this time. By the end of sixth grade, students must be ready for the demands of secondary education. Middle and high schools are *departmentalized*, and teachers are subject-matter experts more than child specialists. The school day comes in fifty-minute blocks with fifty-minute teachers, and students who are not proficient and self-motivated can easily fall between the cracks. Success in reading and language arts is critical to success in general.

What are the ingredients of literacy at this level? The basic elements are well known (Anderson et al., 1985; Calfee and Drum, 1986): First, the student must be able to *decode* (i.e., recognize words in print) with fluency and *spell* with reasonable accuracy. These skills must have become automatic, so that little conscious attention is required. Given the complexity of English spelling, this may seem to an unrealistic expectation. Curriculum materials generally present spelling in a piecemeal and rote fashion, yielding little insight into the historical and structural features of the English language. Many students have been assigned to special education because of what are actually decoding problems.

Vocabulary development and *concept formation* constitute another critical element of literacy. Intelligence is measured by a person's store of word meanings and concepts. The successful sixth grader knows many words, but, more important, he or she has developed strategies for approaching novel words and ideas. A kindergartner knows 5,000 to 10,000 words. From kindergarten through twelfth grade, a child spends about 2,500 days in school. If the child learns 10 new words per day, he or she will still know only a fraction of the 500,000 words in the English language by the time of high school graduation. Learning to learn is essential. The student who asks the teacher at the end of a vocabulary lesson whether the test will contain only the words on the list and is answered in the affirmative has not only ask the wrong question, but has been given the wrong answer.

A critical achievement for academic survival beyond sixth grade is *comprehension.* The sixth grader needs patterns, procedures, and strategies for dealing with complex passages (Calfee and Chambliss, forthcoming). Secondary teachers expect their students to tell the difference between stories and technical writing. They expect them to know how to analyze stories (characters, setting, plot, theme) and how to dig beneath the surface to find the themes that distinguish good literature from "junk." In technical prose, writers use a handful of building blocks (compare-contrast, topic-expansion, the five W's of the journalist, and so on). The youngster working through a chapter in social studies or science will flounder without these building blocks.

The literate sixth grader not only possesses this array of intellectual tools, but can use them in strategic and conscious ways. Psychologists refer to *metacognitive awareness*, that is, knowing what you know, knowing how to use the knowledge, and knowing how to express all of this. Research shows that explicit strategies for monitoring and improving comprehension lead to greater understanding, and that students (especially those at risk in school) benefit from direct instruction in these skills (Brown, 1978; Barr et al., 1987; Snow and Lohman, 1984).

The sixth-grade graduate differs in many ways from the kindergartner. Some changes are developmental, but many are the result of seven years in the classroom (Fig. 1). Literacy transforms the student's thinking, learning, and communication, yielding an academic portrait that can summarized in four key words:

- *Recognize.* The student has the skills and knowledge to read a text, extract the core elements, and store in memory enough detail to recognize the material later on. Recognition is not passive, but neither does it require great mental effort. Most literate adults read the daily newspaper in this mode.
- *Reconstruct.* To attain a deeper level of understanding of demanding text, the student can virtually recreate the writer's work in composing a text. When a student takes an essay test on the War of 1812, more is required than words, sentences, and facts. Moreover, the ability to reconstruct written material is not limited to academic exercises. An employee who is told to prepare an analysis of a contract in time for the next day's meeting is being asked to do a reconstruction, not complete a multiple-choice test.
- *Produce.* An outstanding English teacher who was asked whether today's students were writing enough responded, "They aren't reading enough!" The interplay between comprehension and composition is crucial. The teacher had a point, but the proof of comprehension comes when students create their own works. The tasks may be as simple as the five-paragraph essay, or as complex as comparing and contrasting speeches by Washington and Lincoln.
- *Explain.* Explanation is the strategic domain. The literate sixth grader can tell when he or she is having trouble with a task and can shift into a higher-order gear to handle the problem. This capacity helps the individual to solve problems and also fosters communication with others.

LITERACY AS FORMAL LANGUAGE

What should be taught? What should be tested? These questions must be addressed in the context of goals and definitions. An elementary teacher discussing reading usually emphasizes *skills*. These are specific objectives, laid out lesson-by-lesson, practiced through worksheets, and assessed by end-of-unit tests based on textbook materials. Teachers emphasize skills because they are the focus of standardized tests. The underlying assumption is that specific skills gradually add up to literacy, but this assumption is coming

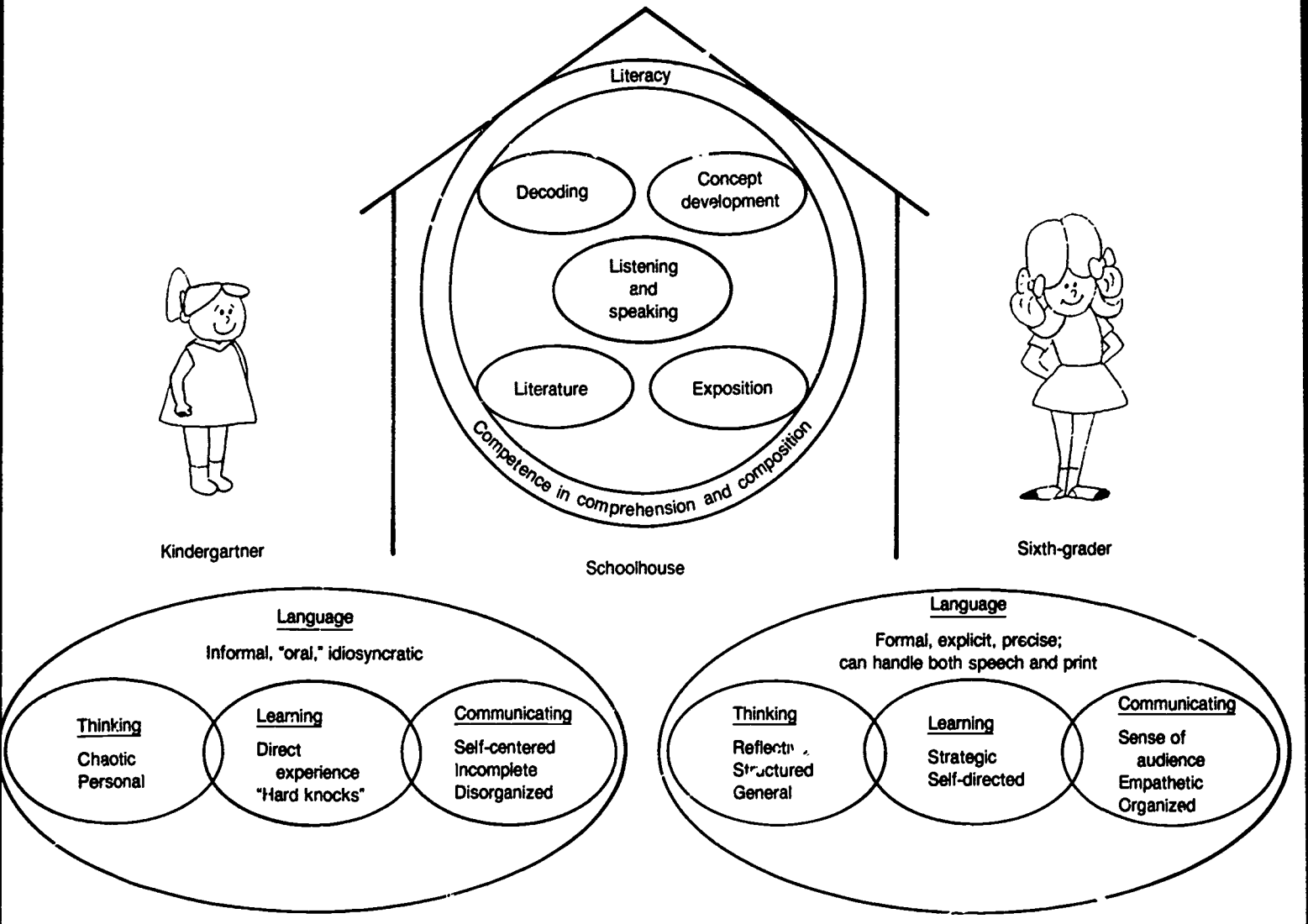


Fig. 1—The curriculum of *literacy*, showing the major facets that distinguish a kindergartner from a literate sixth grader

under increased scrutiny. While practice on clearly defined objectives is important, literacy is more than a collection of many "little things." Human beings need "big pictures"—coherence and structure are essential in human learning.

An alternative approach takes the view that differences between people who are more or less literate are only partly related to the *medium*, i.e., print versus *speech*; more critical is the *style* of language use, i.e., *natural* versus *formal*. Literate people can translate the printed word into the equivalent of speech, but they also interpret spoken messages differently than do uneducated people.

The difference between decoding and reading can be illustrated by considering the following paragraph:

> The procedure is quite simple. First you arrange things into different groups. One pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities, that is the next step. Otherwise you are pretty well set. After the procedure is completed, arrange the materials into different groups again. Then they will be used once more and the whole cycle will be repeated. At first the whole procedure will seem complicated. Soon it will become just another facet of life (after Bransford and Johnson, 1972).

Most people have trouble understanding this passage. The problem is not one of phonics; it is not difficult to decode the words. The vocabulary is reasonably familiar, and the sentences are not particularly complex. The problem is, first, that the topic is uncertain; the reader does not know what is being discussed. This is often true of technical material. Second, the passage is written in expository style, and the structure does not lead the reader easily toward understanding.

In fact, the topic of the paragraph is doing laundry. Knowing that, you can make some sense of the passage. This example illustrates the importance of knowing both *what is talked about* and *how it is talked about.*

The contrasts between *natural* and *formal* language, i.e., between "spoken" style (which may or may not be written) and "printed" style (which may or not be on paper), are summarized in Table 1. People naturally assume that others know what they are talking about, and that listeners will interrupt if they are uncertain.

An important distinction between natural and formal language is the degree of *explicitness*. In formal language, little is left to chance. The writer has a particular audience in mind and aims for clarity and coherence. The printed page cannot answer questions, so

Table 1

CHARACTERISTICS OF NATURAL AND FORMAL LANGUAGE

Natural Language	Formal Language
Implicit (conversation)	Highly explicit (discussion)
Context-bound	Context free
Unique, idiosyncratic, personal	Repeatable, memory-supported
Intuitive, emotive	Rational
Narrative/descriptive	Expository

SOURCE: Calfee and Drum, 1986.

the writer tries to tell the whole tale. Again, the *medium* is not the issue. Lecturers generally assume an unconversational audience and organize their presentations accordingly. The process is unnatural and places demands on both the listener and the speaker.

Table 1 points to several other distinctions. Committing a thought to writing entails greater *permanency*. A text reads the same wherever it is being read. The formal writer tends to be more *reflective* and *rational*. Talkers tell tales, stories, and jokes; writers prepare reports, instructions, and essays. These distinctions are not absolute, nor are they value judgments. The skilled orator, for example, blends the elements of a rhetorical argument with anecdotes.

Describing literacy as the ability to employ language as a tool for thinking and communicating has implications for both instruction and assessment. It supports an integrated approach in which reading and writing are complementary, and oral language development is as important as decoding skill. It means that the mastery of specific objectives is important only as those objectives support the larger goal of linguistic competence. The classroom environment is a microcosm of society in which to practice this competence.

IS CRITICAL LITERACY AN IMPOSSIBLE DREAM?

A major goal of virtually all actors on the educational stage—teachers, administrators, parents, policymakers, and citizens—is to have the nation's students graduate from school as fully literate adults, not simply able to read in a mechanical sense, but possessing the critical ability to understand and evaluate the worth of what they are reading.

Anyone familiar with the demographics of today's kindergartners might be skeptical about the prospect of achieving this goal. Poverty, broken homes, and lack of parent education are established precursors of school failure, and the majority of the children entering elementary school during the next decade will come from poor homes with single parents who are high school dropouts. The match between the home and school language is another predictor of reading achievement. For many of today's kindergartners, action is more important than words. Moreover, many enter school speaking a language other than English.

The correlation of family and student background with reading achievement, a primary finding from present indicator systems, has led some policymakers to conclude that students and their families are the key to literacy. Anderson et al., in *Becoming a Nation of Readers*, begin their recommendations with the statement that "parents play roles of inestimable importance in laying the foundation for learning to read" (1985: 57). Policy based on links between home background and student achievement often assumes that (1) at-risk homes can change to support students' literacy skills, and (2) schools can do little to overcome the shortcomings of students from at-risk backgrounds.

Consider a different analysis of the problem, in which schools are not altogether clear about the goals of literacy, and reading instruction is shaped to match expectations based on the students' socioeconomic background. Both of these notions are supported by research findings. As to the first, reading experts continue to argue the merits of phonics versus comprehension and give little guidance to practitioners about the balance between them. There is much confusion and little clarity about the goals of literacy.

On the second aspect, a student's preparation for school entry is critical to academic placement. The child who enters school ill-prepared, who has not learned the alphabet and does not know classroom manners, is likely to be assigned to the lower track and likely to remain there. Readiness tests reveal whether a preschooler has been introduced to reading. Teacher observations are the index to socialization. Student placement is critical; once assigned to a particular learning group, a child's academic future is set (Calfee and Brown, 1979). The teacher's instructional decisionmaking only fine-tunes the materials (Barr and Dreeben, 1983; Fraatz, 1987).

What happens after the student is assigned to a group? Higher-ability groups tend to be taught in an interactive and conceptual fashion, with the emphasis on meaning and comprehension. In contrast, students in lower-ability groups have been shown to benefit from instruction that is structured, coherent, and explicit; they are more likely to be taught

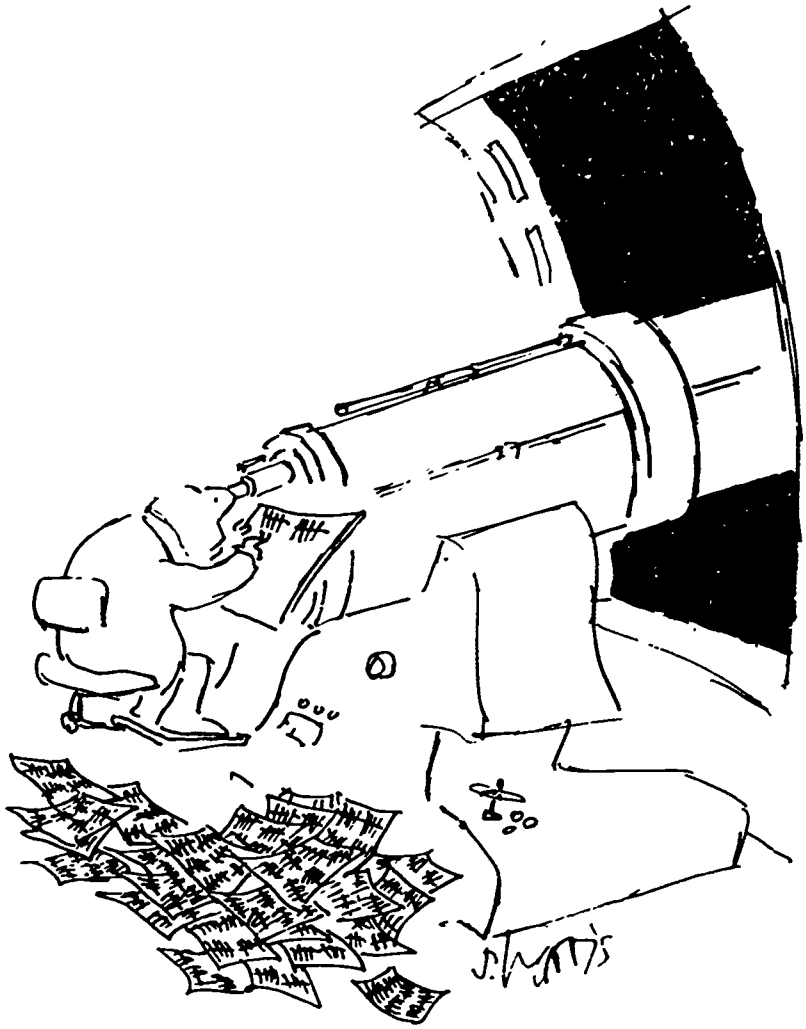
specific skills by rote, with an emphasis on phonics (Calfee, 1986; Applebee, Langer, and Mullis, 1988). Classroom policies amplify differences in students' home backgrounds, enhancing the "Matthew effect," i.e., the rich become richer and the poor become poorer (Stanovich, 1986).

Can this situation be altered? First, it is necessary to evaluate whether the picture is, in fact, accurate. It is based primarily on findings from small-scale studies, and present indicator systems focus on student performance and student background but neglect curriculum and instruction (McLean and Goldstein, 1988). Second, we need to know how local practices can be influenced to improve instruction (David, 1987). Small-scale experimental studies suggest that at-risk students can improve their literacy skills (Orasanu, 1986), but the information is piecemeal and scattered.

What might be accomplished by more effective use of existing resources? Some argue that the potential of many youngsters is limited, that we have reached a plateau in the level of student literacy and are unlikely to do much better. Carroll (1987), commenting on National Assessment of Educational Progress (NAEP) surveys of reading, concludes that "it now seems unrealistic to expect that at any time in the near future all or nearly all adults will attain the *adept* level [the capacity to reconstruct], even with the best instruction that anybody might devise." Stedman and Kaestle (1987), reviewing literacy levels over the last century, reach a similar conclusion.

But this is only one view. Data on current classroom practices are scattered and piecemeal, but they suggest the possibility of improvements for at-risk youngsters. Research on "effective schools" (Edmonds, 1985; Purkey and Smith, 1983) shows that students from disadvantaged backgrounds can do well on standardized tests. To be sure, the characteristics that distinguish an effective school are not yet well-defined (Stedman, 1987), and we should hold higher expectations for youngsters than multiple-choice mastery. Nonetheless, it is important to examine and assess the situations in which students succeed where failure is predicted.

III. IMPROVING INDICATORS OF LITERACY



Measurement can be an important part of assessment,
but in the service of significant matters . . .
(© Sidney Harris, 1988, reprinted with permission)

THE PRESENT PICTURE

Before proposing changes, we shall briefly examine the present state of student literacy and the strengths of the existing system. The NAEP is a major source of information about the status of literacy throughout the nation. Three recent NAEP reports are summarized in *Focus 20*, the newsletter of the Educational Testing Service (ETS). The picture is not very promising:

> [All age groups] were reading better in 1984 than in 1971. The gap between the performance of minority and disadvantaged urban youngsters and that of other youngsters narrowed [but remained substantial]. Only about 5 percent of 17-year-olds [in 1984] had acquired advanced reading skills and strategies, and 16 percent failed to reach even the intermediate level. (NAEP, 1985)

> The results . . . were disheartening. Few students could write analytically. . . . When asked to write to their principal suggesting a change in a school rule, only 22 per cent of the eleventh graders did an adequate job. (NAEP, 1986a)

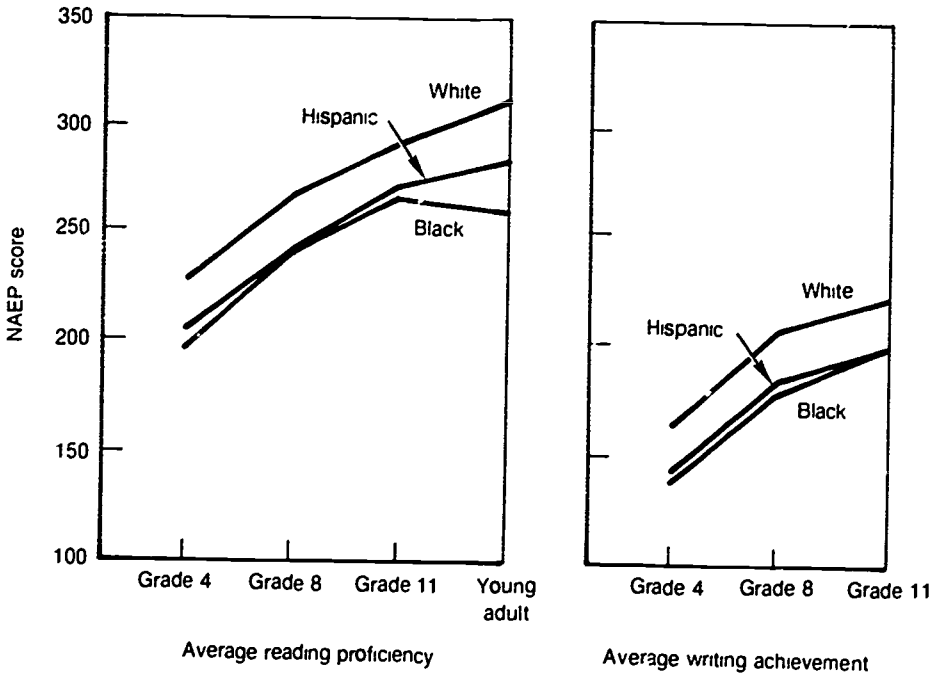
> Many were unable to do well on tasks of even moderate complexity. Fewer than 10 percent could master the most demanding tasks, such as interpreting a poem, using a bus schedule (sic), or estimating prices based on grocery unit-price labels. (NAEP, 1986b)

Figure 2 shows the achievement of students as a function of grade level and ethnic background. From these data, it is clear that school has an effect; students move from *basic* reading in fourth grade to *adept* reading in eleventh grade. Also, the status of reading is much better than that of writing. In view of the fact that writing may be a better index of reasoning skills than is a multiple-choice reading test, this pattern is disturbing. It suggests that high school graduates can deal with the surface demands of text, but are less able to think about its meaning. Blacks and Hispanics score 25 to 30 points lower than whites at all grades. The margin of error is only a few points, so these differences are significant. Since minority youngsters make up an increasing proportion of each year's kindergarten class, most policymakers are uneasy about the pattern, but they do not know what to do about it.

Figure 3 shows reading and writing achievement for the same group, with parent education as the independent variable. The overall trends are similar, of course, but parent education is shown to be an important factor. Its effects are about the same as those of ethnicity, tending to increase as students move from fourth grade to high school. However, the prospects in this area appear to be more hopeful, since the effects of success with one generation are likely to be passed on to the next. That is, a major effort to improve literacy should increase the number of homes in which parents are better educated. Unfortunately, failure to provide adequate education for one generation of students will likewise have detrimental effects that will extend into the future. But at least, courses of action can be formulated on the basis of this information.

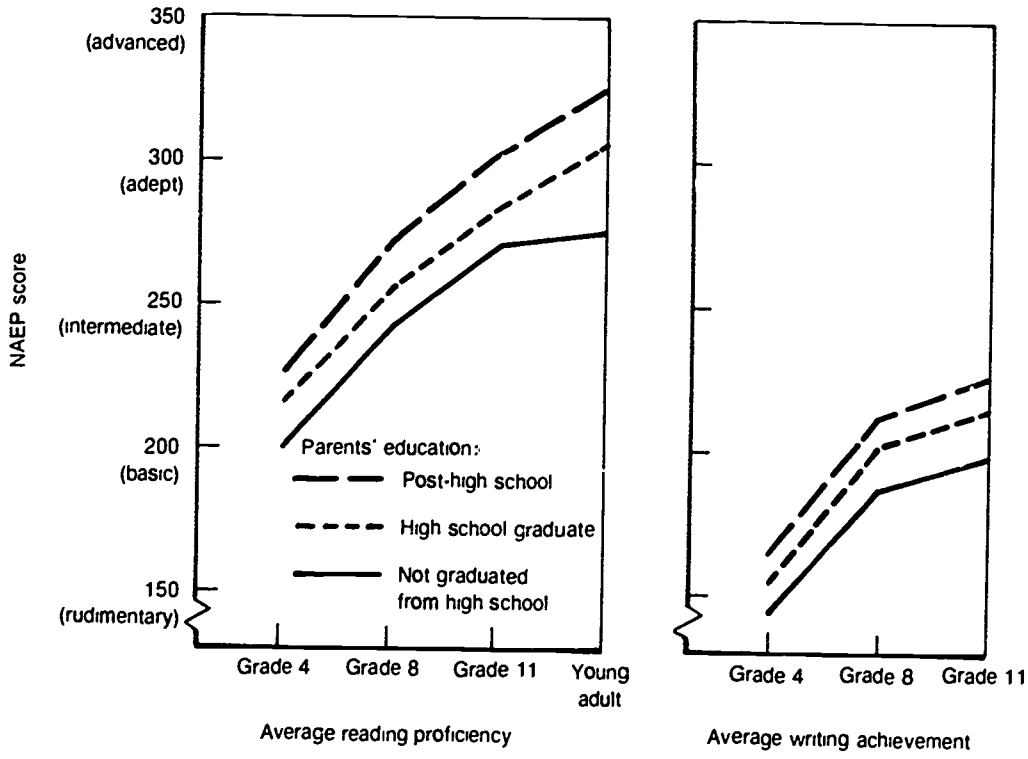
Figure 4 shows long-term trends in student achievement for students living in different communities.[1] The good news is that reading scores have risen slightly in recent

[1]Comparable data are not available for writing; the patterns reported in *Writing: Trends Across the Decade, 1974-1984* (NAEP, 1986b) are complex but generally consistent with the data presented above.



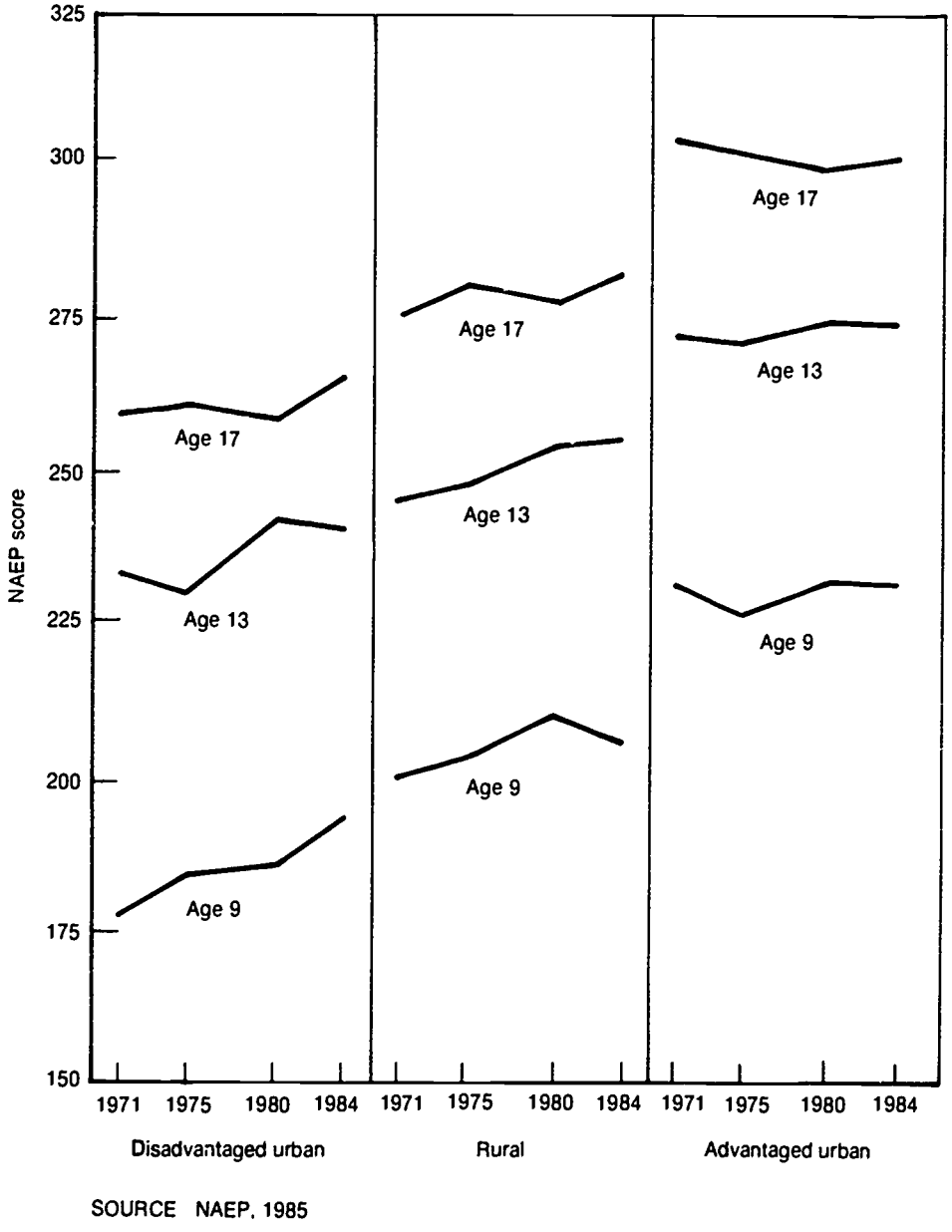
SOURCE: NAEP, 1987

Fig. 2—Reading and writing performance of students with different ethnic backgrounds



SOURCE: NAEP, 1987

Fig. 3—Reading and writing performance as a function of parents' level of education



SOURCE NAEP, 1985

Fig. 4—Reading performance assessments showing trends over time as a function of age

years. But this trend is limited to younger children in disadvantaged urban areas, where progress is a matter of improvement in the basics rather than in high-level "skills." A similar pattern emerges for the effects of parent education; noticeable improvements have been achieved over the past decade for young children from homes where parents have not completed high school, but otherwise there has been little progress. The bad news is the small proportion of high school students (17-year-olds) who are *adept* readers—40 percent of white students, 15 percent of Hispanics, and 10 percent of Blacks. At a time in their lives when school success demands fluency in comprehension, only about one in three students can handle the task.

The emphasis on basic skills seems to be most helpful for those younger students who are most at risk. There appears to be little payoff for older students or for performance in reasoning and communication. Schools must do something else, it appears, but the course of action is uncertain. One approach would be to do more of the same in the early grades. A second strategy would be to extend current programs into the later grades. Another alternative would be to strive for a more balanced program for at-risk students, with greater emphasis on higher-level reading skills. The data are of little help in considering these courses of action. This study focuses on the achievement of at-risk populations; a broader perspective would consider how to improve schooling across the board. It is difficult to feel satisfied when a substantial proportion of youngsters from relatively advantaged backgrounds leave high school barely adept at handling formal language.

Another NAEP report, *Who Reads Best* (Applebee, Langer, and Mullis, 1988), provides some insights into possible reasons for students' poor performance. Asked to give their perceptions of teaching practices, students reported a variety of strategies, but poor readers said that their instructors were "less likely to emphasize comprehension and critical thinking, and more likely to focus on decoding strategies." One might conclude that basic strategies were all that these students could handle, but the study also showed that poor readers "seem to be even more limited in their school reading experiences than in the reading they do on their own."

The researchers stressed the differences in the treatment of better and poorer readers. Even more disturbing is the low level of instructional support for the development of comprehension and critical thinking. The data for students at all grades indicated that fewer than 30 percent of teachers asked questions during reading, and reflection following a story was also rare (less than one-third of the students were asked to support their ideas; not one in seven reported discussion after a story). The NAEP assessment design asked students to

write about their reading—a novel and notable feature of the study. The writing was generally poor, especially in the area of exposition (less than 1 percent produced an *elaborated* written response to an article on marketing goods). Moreover, students at all grades and achievement levels showed little awareness of the technical terms for analyzing narratives (fewer than one in four eleventh graders mentioned character, setting, and plot in answer to the question, "What do you think about when you read?").

From the perspective of students' performance and perceptions, the current picture appears straightforward: Most youngsters lack the rhetorical tools needed to handle language; they are equipped neither to comprehend nor to compose. The most probable reason, based on their reports of the school experience, is that they are not instructed in these areas, nor do educational experiences engage them in applying the instruments of critical literacy.

CAN WE BELIEVE WHAT WE SEE?

Student literacy seems to be improving, although not very rapidly. Minorities and poor students remain far behind the middle-class majority. Absolute levels of competence are distressingly low.

These conclusions appear to be trustworthy, since by contemporary standards NAEP instruments are carefully designed and meet rigorous standards of reliability. The national samples are drawn to ensure representativeness. The margin of error for the various statistics is minuscule. The longitudinal trends from *The Reading Report Card* (NAEP, 1985) cover more than a decade and appear to be trustworthy.

The NAEP continues to improve its methods, both technical enhancements and reporting. For instance, "numbers" are now supplemented by a verbal scale; a score of around 300 means that one is "adept" at reading and can perform certain tasks.[2] A second advance is the portrayal of the findings through clear and engaging graphics.

The improvements appear to make a difference. Educators paid little attention to NAEP findings until recently, but the Study Group on National Assessment (Alexander, 1987) has proposed that the NAEP be expanded to provide state-by-state information. This proposal would mean testing more students more frequently. The Study Group also

[2]To be sure, the scale may not be what it seems. McLean and Goldstein (1988) say that any unidimensional portrayal of literacy is suspect and of limited value for educational decisionmaking. Unidimensional scales are generally reliable and stable, but the "anomaly" in the 1986 NAEP reading assessment, mentioned in the appendix to *Who Reads Best* (Applebee, Langer, and Mullis, 1988), illustrates the hazards of relying on a single indicator of any sort.

recommended adding "school variables" to the NAEP system, insofar as these have "significant effects on student achievement," but it expressed concern about the costs of including "exploratory variables." On the other hand, McLean and Goldstein (1988: 372) make the important point that "to have relevance for policy . . . assessment must use measures that are connected to teaching and learning." They point to the work of Great Britain's Applied Performance Unit as a demonstration of how to achieve this goal.

While the recent enhancements are commendable, they only begin to address the most fundamental issue, *validity*. As is clear in the *Standards for Educational and Psychological Testing*, validity boils down to "the degree to which the evidence supports the inferences that are made from the scores" (AERA, APA, and NCME, 1985; Messick, 1984).

For the policymaker, validity can be put into perspective by considering the following questions (which parallel Venezky's 1974 "canons" for reading assessment):

- Are you sure you know what you want to look at?
- Do you have a clear sense of why you are looking at it?
- Are you getting the right information from the right sources?
- . e you getting the information in a useful form?
- Do you understand the information?

If the answer to any of these questions is not clearly affirmative, validity is compromised. Validity is a commodity that is neither cheap nor simple. As stated by Messick (1988: 43): "The practical use of measurements for decision making is or ought to be *applied* science, recognizing that applied science always occurs in a political context. . . . The justification and defense of measurement and its validity is and may always be a rhetorical art." Angoff (1988) expands on the same point, arguing that simpleminded definitions of educational outcomes are inherently untrustworthy and urging for a broad range of evidence that allows conclusions to be based on *triangulation.*

Policymakers tend to take validity for granted, usually accepting face validity. If the indicator system seems to make sense for the purposes intended, the trustworthiness of the indicators is seldom questioned. Cronbach emphasizes the role and responsibility of the consumer of information in the validation process:

> Responsibility for valid use of a test rests on the person who interprets it. . . . He has to combine [various sources of objective information] with his other knowledge . . . to decide what interpretations are warranted. (Cronbach, 1969: 51.)

> Validators should indeed do what the detached scientist would do, as best they can within [short time] constraints. In validation, a vigorous, questing intellect has further importance for analyzing the values and rights embodied in—or sacrificed to—a testing program. (Cronbach, 1988.)

Two issues are especially important in the validity of educational indicators: *pressure* and *distance*. In this context, *pressure* is the tendency for information to be subjected to social and political influences. Pressure is most likely to have detrimental effects when concepts are vague and evidence comes from a narrow base. Participants under pressure will be driven to do whatever is necessary to improve performance *as measured by the indicator*. Frederiksen (1984: 199) describes the situation for standardized achievement testing:

> School constituencies are far more likely to pay attention to educational outcomes that are measured and reported than to possible outcomes that are not measured. . . . Accountability systems involving currently used tests are likely to improve the educational system only in the narrow sense that they perpetuate the teaching of what is measured and make it more effective.

At the other end of the line, policymakers under pressure to report that "things are getting better" are likely to accept positive findings and to disregard contradictory evidence from outside sources.

Distance refers to the gap between an indicator and the "real thing. In general, as distance increases, so does the risk of invalidity. For literacy, the most direct approach might be to have the student read several passages and explain the meaning and implications of each to the examiner. The multiple-choice examination is at a distance from this starting point in several ways: It requires the student only to recognize the correct answer; no explanation is called for. The passages are likely to be short. Misunderstandings of the task cannot be detected and remedied. Only one facet of the vision of the literate sixth grader is tapped by this approach.

Assessing writing may seem straightforward: simply ask the student to compose. But questions arise as to how the topic should be chosen; how long the writer should be allowed to write; how many revisions, if any, should be allowed; and whether it is the product, the process, or both that is important.

Problems that result from distance could be alleviated if policymakers allowed input from local data collection activities designed primarily for local functions but potentially informative for agencies "up the line." When districts, states, or the federal government create an indicator system, there is a tendency to determine centrally what needs to be known and how to collect information. The proposal put forth in this Note requires trust in local data collection and the creation of an overall design for integrating data from various levels. Giving up central control entails some risks, but when information is available from multiple sources, policymakers are in a better position to make valid judgments.

TEACHER JUDGMENTS: A BASIS FOR IMPROVED LITERACY INDICATORS

The primary recommendation of this monograph is that *judgments from the classroom teacher should be collected regularly to complement existing programs for obtaining objective information on student achievement and student demographics.* Oakes (1986: vii) defined an educational indicator as "a statistic that tells something about the performance or health of the education system." She delineated five issues: *level* ("bottom-up" or "top-down"), *fairness, scope, politics,* and *decisionmaking.* The present recommendation centers on the first of Oakes' five issues but has implications for all five. It is organized around the five questions posed earlier: *what, why, from whence, for whom,* and *to do what*?

Implementation of a proposal to have teachers play an important role in assessment of student achievement faces several barriers, including the question of teacher knowledge and skill, possibilities of bias, and availability of time to perform the task. The most basic issue may be that of whether teachers can be entrusted with the task. However, the teaching force must be capable in this area—their ability to make informed instructional decisions is based on their skill in assessing student performance (Calfee and Hiebert, 1988). In any case, this proposal is not novel; papers by Stiggins (1988), David (1987), and Romberg (1988) have each recommended greater reliance on locally generated indicators, and both Stiggins and Romberg point to informed teacher judgment as a crucial source of information.

What Should Be Measured?

Indicators should be chosen that assess *things that matter.* For literacy, the main point is that reading and writing are more than a collection of unrelated skills. The design of an indicator system for literacy should be driven by a coherent representation of curricular goals. The piecemeal listing of scope-and-sequence objectives is inadequate for this purpose, even if it simplifies life for the item writer.

Also, *things that matter* should be measured *in a variety of ways.* Reliance on single measures or on slight variations on a single theme flies in the face of expert advice (e.g., Cronbach, 1984: 339; AERA, APA, and NCME, 1985: 9). If diverse sources of information converge on a common question, the decisionmaker has a basis for triangulating the issues. Objective tests are one source of information, but the capable teacher is continuously collecting and merging information about student knowledge and performance from a broad array of tasks and situations. A teacher might make an assessment such as:

> Sally seems to have a limited vocabulary; she is at the 30th percentile on the standardized test. But she just said that there was a crab in the story; it was actually a lobster, which means that she remembered the lobster, knows that lobsters and crabs are similar, and confused the two. She needs to refine her word use but she may know more than I thought.

Finally, it is necessary to measure *things that influence the things that matter.* It makes sense to record student factors that are correlated with learning—socioeconomic status, minority and language indices, sex, and so on. But these data do not explain the forces that influence growth in literacy, nor do they provide a sound basis for action. It can be argued that these "predictors" are actually "effects"—teachers may treat the child identified in the primary grades as at-risk differently than they treat the middle-class student. Factors are needed that can serve as policy levers and that have fairly direct influence on significant outcomes. For instance, lengthening the instructional day or increasing the required courses may provide students with more opportunity for learning, but they have no influence over the quality of curriculum offerings.

What Is the Purpose of Collecting Information?

Once information is in hand, it can be used for a variety of purposes. It can be used, for example, to *monitor* the performance of a system. To be sure, the "facts" take on different meanings depending on one's knowledge of how the system works. A car has instruments that show how fast it is moving, how much gas is left in the tank, whether the oil pressure and electrical voltage are within limits, and so on. These are the "facts." The person who knows the system (both the automotive system and the indicators) sees more than the data. The combination of high oil pressure and high engine temperature makes the expert think about changing the oil filter. Higher reading scores and fewer students taking the test leads one to question whether schools have really improved reading instruction and to place more emphasis on the proportion of students taking the test.

Information can be a *guide for action*. Decisions can be based on the data; indeed, data may call for direct response. If the speedometer is reading 80 mph or the gas gauge registers empty, a driver should act promptly. In education, indicators both provoke action and guide decisionmaking. For instance, in *A Nation at Risk*, the National Commission on Excellence (1983) justified its call for reforms by presenting a troubling picture of low test performance, in particular, a decline in SAT scores. The Commission's advice about how to address the problems sprang not from the data, but from the creative and political impulses of its members. The low levels of writing revealed by the NAEP reports are distressing. Teachers might have told us about this problem some time ago, and possibly could have provided insights into how to improve the situation.

What Sources Should Be Used?

The question of what sources to use can be divided into two parts: *from whom* and *how*? *From whom* asks about the person or group responsible for generating the data. Reading indicators presently come directly from the student, with little input from those responsible for providing and evaluating the delivery of services.

How refers to the methods of data collection. Multiple-choice tests are favored by administrators, teachers, and policymakers for assessment of reading achievement, because they are cost-effective (Cole, 1988). Students also prefer them because they require less thought than essay tests or discussion. When policymakers ask about the curriculum, the information is usually in the form of course outlines or textbook titles. "Is phonics taught?" can be translated as "What decoding objectives are included in the scope-and-sequence chart for the reading series? Are these aligned with test objectives?" Deeper questions (How does the teacher handle phonics? What opportunities do students have to practice this skill? Is there an effort to integrate decoding and spelling?) are less often addressed. As Graham (1987) has noted, these latter questions are of central importance for improving education.

Who Needs to Know, and How Should Information Be Reported?

Presumably, the people who asked for the information are at the head of the *need-to-know* line. It is useful to distinguish between *internal* and *external* mandates for data (Haertel and Calfee, 1983; Cole, 1988). Internal indicators are collected close to the scene

of action for use by those with immediate responsibility for taking action. External indicators are requested by more remote individuals who may or may not have the authority and means for action.

The contrast can be illustrated by the case of reading achievement. The teacher has the most direct responsibility for assessing student knowledge and performance. In the ideal situation, classroom assessment is continuous, multifaceted, and interactive, with both formative and summative elements (Stiggins, Conklin, and Bridgeford, 1986). Student feedback is quick and directive. Students read, write, and talk, and the teacher draws from the flow of information to adjust instruction, to guide students, and to assign grades. This pattern exemplifies the *internal* model (Hiebert and Calfee, forthcoming).

The district and the state also mandate assessment of student achievement to meet public responsibilities. In earlier times, accountability was situated at the local level. The teacher was the primary actor, and grades were the primary indicators. Parents were the people who needed to know, and the teacher had the answer. As states have assumed greater accountability, they find themselves subject to constraints and pressures: indicators must be *efficient* (cheap), *standardized* (to eliminate bias), *simple* (easy to read), and *generalizable* (to cover different districts). Standardized achievement tests evolved to meet these requirements.

The *how* question deals with the reporting of the data. Whatever the mandate, many actors want access to the findings. Reporting presents several challenges (Boruch and Wohlstetter, 1983), including, for example, amount of detail. An audience can be overwhelmed with mounds of information; a principal may decide that spelling four-syllable nonaffixed words is a school priority, and literature is lost in the shuffle.

Standards and context are also part of *how*. Knowing that students answered only 53.6 percent of the questions on a vocabulary test means little by itself. California uses "bands" to guide expectations. Demographic factors are used to project the range of student performance typical of schools serving similar populations. If most of the students in a school are poor, if their parents are not well educated, and if their first language is other than English, the band will be low. Failure can then be success; the students may perform quite poorly, yet the district may appear to be fulfilling its mission. Unfortunately, these youngsters must later face a competitive marketplace that values absolute, not relative, competence.

How to Make Sense of the Information

Indicators serve pragmatic ends, so the need to evaluate and interpret them may seem superfluous. But while understanding most often comes into play when things are going wrong, it can also help to anticipate problems. Present approaches to teaching literacy may have served us well in past generations, but they do not appear to be working now, and if we cannot understand the implications of the data provided by literacy indicators, the situation may get worse.

A major barrier to valid interpretation of literacy indicators is the lack of adequate conceptualization:

> Clarifying the underlying concepts and investigating whether particular indicators provide reliable information about the concepts will make the indicators more useful. . . . [The] difficulty of the validation task [is related to] the extreme difficulty of defining the underlying concepts [and] developing operational measures of the concepts. (Murnane, 1987: 24.)

The single most important contribution of researchers and scholars to the development of indicators of student learning may be a clearer conception of the outcomes of schooling and the factors that lead to variation in those outcomes. The argument of this Note has now come full circle: Informed policymaking requires clarity about what needs to be known, to guide the design and implementation of indicator systems, to help the policymaker interpret the information, and to guide action.

IMPLEMENTING THE PROPOSED INDICATOR SYSTEM

Implementing an indicator system that includes information from practitioners at the local school site requires turning the usual process "inside out." The proposal put forth in this study starts from the perspective that schools are the abode of professionals who are a source of informed judgment (Calfee, 1987). This view assumes that teachers are the country's "brain builders," with principals (head teachers) leading the enterprise. Midrange professionals (teachers and principals) should play a central role in communicating their judgments to the broader community (policymakers and other interested parties). An approach for gathering survey data for this purpose is outlined below.

A Teacher Survey of Sixth-Grade Literacy: The Plan

Figure 5 presents a sample survey instrument for sixth-grade teachers. It is assumed in the following discussion that the survey is administered at the end of the school year. The teacher is asked to provide information in the following categories:

- Assessment of students' achievement in literacy: reading, writing, and oral language.
- Information about the curriculum and instruction being provided for students.
- Personal information: teacher's background, instructional philosophy, and present assignment.
- Class description, including the range of variability and how the class is organized for reading and writing.
- Perceptions of the students' background characteristics, including categorization into three groups based on general performance level (not necessarily instructional groupings).
- Description of student performance and instructional program against the rubrics of an integrated literacy program, with the primary categories (oral language, reading, and writing) divided into subcategories.
- An estimate for each group of achievement at the beginning and the end of the year, along with judgments about the curriculum, instructional methods, and techniques of assessment.

Figure 5 is intended only as an overall design; the actual form might cover a few pages. The choice of categories will vary, depending on policy needs. This example focuses on how teachers organize groups for instruction and on their plans for each group. The decision to arrange the curriculum as an integrated program fits the vision of literacy presented at the beginning of this Note, but other alternatives exist.

The Teacher's Task

The sixth-grade teacher would complete the survey once a year, at about the time standardized tests are administered. The first entry on the form, *Background*, covers how long the teacher has been teaching; the teacher's certification; special training, if any; present assignment and how long the teacher has been in it; priorities and aspirations as a professional; categorical programs active in the teacher's class; and availability of aides or parent volunteers.

Teacher profile	General background (experience and training) Instructional philosophy Present assignment																					
Class profile	Background																					
	Achievement in literacy	Top Group:																		Mid Group*		Low Group*
		Performance						Curriculum														
		Fall			Spring			Emphasis			Materials							Instruction		Assessment			
		Exceeds	Adequate	Limited	Exceeds	Adequate	Limited	High	Medium	Low	Textbooks	Workbooks	Computers	Teacher-made								
	Oral language																					
	Reading																					
	Story comprehension																					
	Vocabulary development																					
	Expository comprehension																					
	Phonics and word attack																					
	Writing																					
	Journal/informal																					
	Stories																					
	Reports																					
	Spelling and grammar																					

*The same performance, curriculum, instruction, and assessment categories shown for the Top Group would be repeated for the Mid Group and Low Group.

Fig. 5—Sample literacy survey instrument for sixth-grade teachers

The second entry sorts students into three ability levels:[3] What are the characteristics of the students at the upper, middle, and lower levels in this year's class? (An interesting policy decision concerns whether the classification should be based on entry level, exit level, or both.) Information is requested about the ratio of boys and girls in each group, ethnic and socioeconomic features, family makeup, attendance, and the like.

The third item calls for assessments of literacy. The choice of goals is important because they communicate to practitioners a set of policy decisions in a context that demands engagement. Curriculum frameworks are designed to communicate curriculum decisions, but they can easily remain on the shelf.

The fourth item calls for judgments about *achievement, curriculum, instructional approach,* and *methods of classroom assessment* for each student group. What was this group like at the beginning of the school year (above, at, or below the teacher's expectation for adequate performance at this level)? This question is repeated for each entry in the left-hand column. What was the decoding curriculum? Did it receive primary emphasis? What materials were employed? What were the primary instructional approaches (discussion, individualized study, cooperative learning)? How was student performance monitored (curriculum tests, observation)? Other categories might be included in the survey—attitude, motivation, social and behavioral performance, and so on. The form might even ask whether the students seemed to enjoy reading.

The finished surveys would be sent together with objective tests to the mandating agency for statistical analysis and reporting. The package might include a school profile completed by the principal and other school-level personnel (e.g., resource teachers). The school-level survey would complement the teachers' assessments, with an emphasis on curriculum, instruction, assessment, and student achievement, as seen from the administrative perspective.

The Reporting Agency's Task

How can the data obtained by the surveys be transformed into "pictures" that speak directly to the needs and concerns of the various policy audiences? The following examples illustrate how the central issue might be addressed.

---

[3]This practice poses problems and is not generally recommended (Calfee and Brown, 1979); the purpose of categorization here is to find out how the teacher deals with variations in ability.

One issue of immediate interest to policymakers is *whether teacher judgments about student achievement match the results from objective tests.* Can teachers tell how students are doing? Assuming that the teacher uses the NAEP rubrics (*rudimentary, basic, intermediate, adept,* and *advanced*), how well do teachers and tests agree? The answer can take a variety of forms. For example, teachers may be more pessimistic or optimistic than objective test findings (Fig. 6). The policymaker would have to decide which source to use as a guide.

Teacher assessments may or may not parallel test scores (Fig. 7). If they do not, it is necessary to know what their instructional decisions are based on. Assuming that the teachers' judgments are not random, what *are* they noting that is not reflected in the objective tests? How does this information relate to student assignments and achievements?

Objective tests tend to be highly correlated with each other; teacher judgments may give a more differentiated picture of student strengths and weaknesses. For example, it is difficult to assess oral language with objective tests, and teachers may be able to provide insights into the link between reading problems and language limitations (Fig. 8).

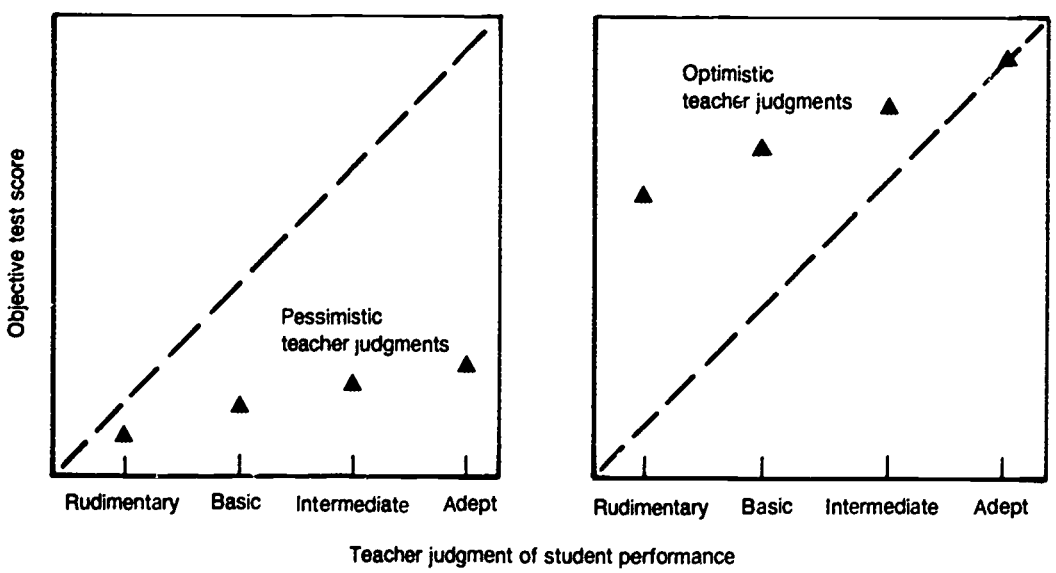


Fig. 6—Optimistic and pessimistic teacher assessments of student performance

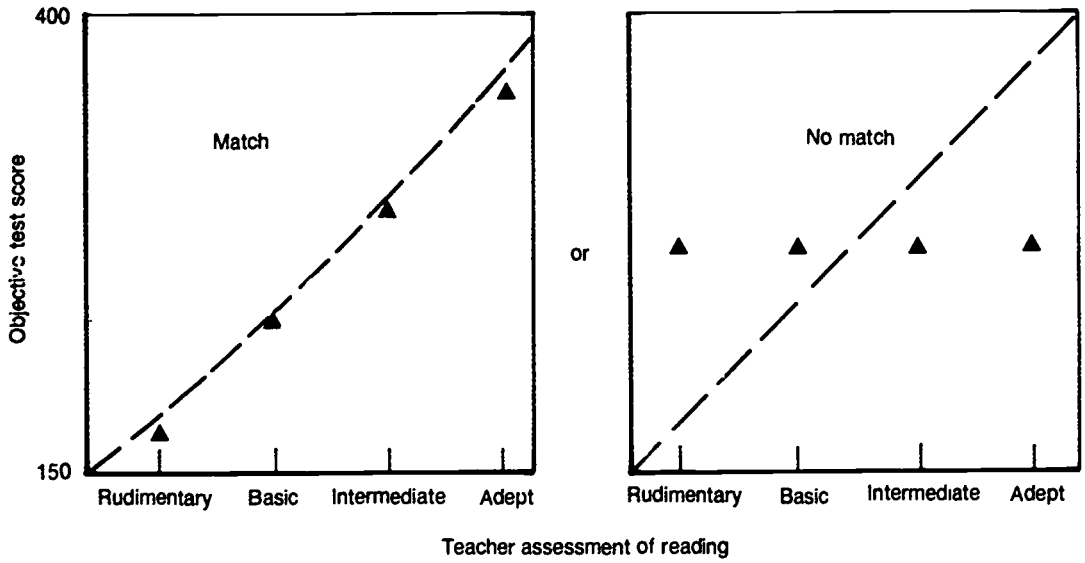


Fig. 7—Hypothetical relationship betw ı objective test results and teacher judgments of student progress

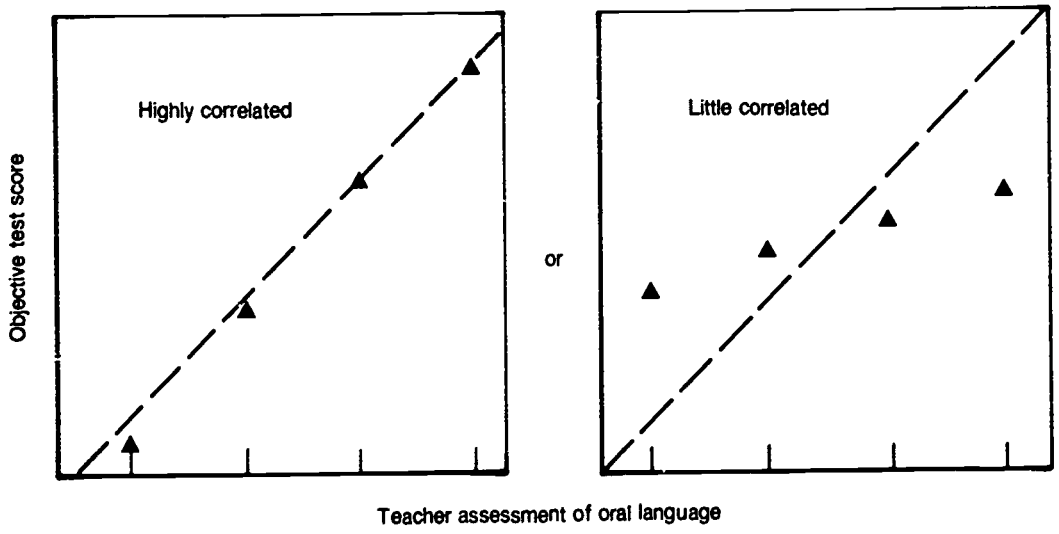


Fig. 8—Illustrative teacher judgments of oral language in relation to reading performance

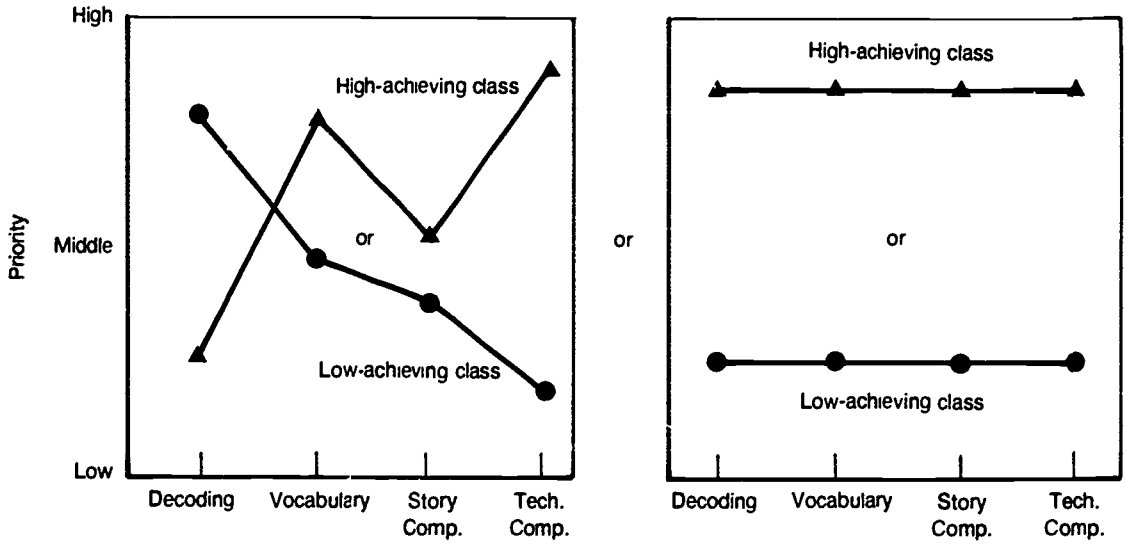


Fig. 9—Patterns of teacher priorities as a function of class achievement level

A second question concerns the *relation of curriculum and instruction to student achievement.* For simplicity, let us assume that teachers and tests provide similar assessments of student performance. One survey question asks teachers to indicate the priority they give to various aspects of reading. The first thing policymakers need to know is whether teacher priorities are informative. For example, the left-hand panel in Fig. 9 contrasts two profiles. One seems appropriate for a sixth-grade class: strong emphasis on vocabulary development and expository comprehension, and less emphasis on decoding and stories. The other picture seems more suitable for earlier grades. But the profiles do show a pattern. In contrast, the two profiles on the right-hand side are not informative; one pattern says "I do everything," and the other says "I don't do much of anything."

If teachers were found to fit one of the patterns in the left-hand panel, this would be consistent with research showing that students with reading difficulties are placed in groups where decoding is emphasized, while more successful students are given more challenging tasks. Figure 10 displays different patterns in achievement patterns for each of the programs. Each panel shows two outcomes. The first is based on the principle that students learn what they are taught. If a teacher emphasizes the basic skills needed for fluent decoding, students will make progress in this area. If the teacher stresses comprehension,

students will show improvement in that domain. The second pattern is based on the idea that some children will become readers and others will not; regardless of the instructional program, the first group excels across the range of skill and knowledge, while the second group falls behind. It should be emphasized that these are only examples; they are not intended as positions or predictions. The basic point is that teacher judgments may help to explain the mechanisms by which students succeed or fail.

The above examples demonstrate a systematic approach to tracing the relationship between teacher judgments and student test performance. Another part of the puzzle is the role of student background. Similar profiles can be developed for specific populations ranging from those most at risk to those predicted to do relatively well. The critical questions revolve around patterns that explain success and failure and provide a basis for policy actions. For example, some observers think that children from poor families are directed on entry to school into instructional programs that are self-fulfilling prophecies. If this is the case, this pattern should appear in the indicators sketched above. On the other hand, equal treatment for children independent of socioeconomic level should also be apparent.

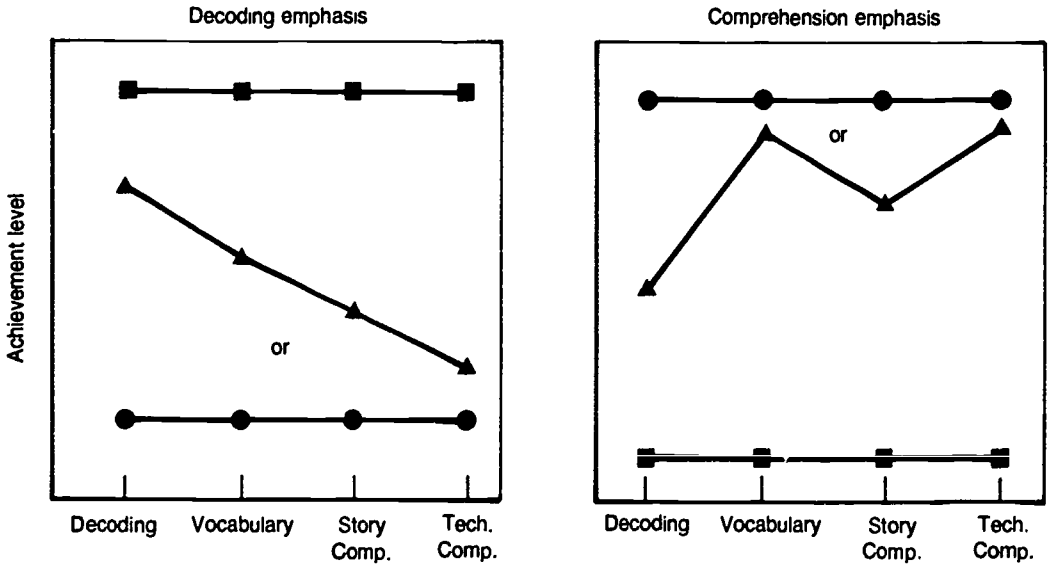


Fig. 10—Hypothetical influence of curriculum emphasis on student performance

The Policymaker's Task

This proposal places different demands on the policymaker. At present, reading scores are released once a year. A state superintendent, district superintendent, or principal looks at the scores to see whether they have improved or at least stayed the same, and whether they meet the "band" predictions. If scores have declined or are below the predicted range, the policymaker wants to know what is causing the trouble and must decide what action to take. These problems then go to curriculum assistants or to a committee of teachers.

The changes proposed here would make life more complicated, because they bring into the open issues that are seldom considered by educational administrators, i.e., what is being taught and how is it being taught. The additional information would provide a sounder base for decision-making, but only if policymakers know how to use it. It is no longer enough to sort schools and districts into successes, failures, and those in-between. The shift in responsibility is consonant with calls for administrators to play a greater role as instructional leaders. The basic concepts of education are clearly within the intellectual reach of administrators. A sounder approach to educational indicators would place instructional issues at the top of their agendas.

WORK TO BE DONE

The proposal outlined above carries implications for the redesign of objective tests and curriculum frameworks, and it raises both technical and political questions that are not addressed in this Note. Solutions to these questions do not seem beyond reach, but all would have to be considered eventually.

Consider the implications for existing multiple-choice tests. First, the curriculum design for teacher judgments should be paralleled in the design of objective tests. Second, student assessments should include formats other than "the one right answer" if teachers are assessing other dimensions of student understanding. For example, if teachers evaluate strategic approaches to reading, objective tests should tap this domain. If the teacher profile gives separate assessments of competence in decoding and oral language competence, test developers may need to create group tests of oral language. Teachers have a wide spectrum of evidence for judging achievement in literacy—reading, writing, speaking, and listening; development of a parallel assessment based on objective measures is a major challenge.

Illinois and Michigan are both experimenting with innovative test content and format, including strategic skills (Wixon and Peters, 1987; Cross and Paris, 1987). The surveys conducted by the Assessment of Performance Unit (Gorman et al., 1982) assessed student skill and knowledge across literature, reference works, and subject disciplines such as social studies and science. The range of tasks was broad, and students had to explain the reasoning behind their answers. The materials were "real," i.e., students were provided booklets with well-formed stories or complete expositions about a given theme (e.g., space). The influence of practitioners was apparent in the design of the assessment and the language of the report: "In devising questions, the guiding principle is that they should be questions that an experienced teacher would be likely to ask pupils, taking into account the subject matter, form and function of the text" (15). The report's description incorporated excerpts from students' written responses, attitudes toward reading and writing, and comments about implications for instruction.

IV. CONCLUDING REMARKS

Finally, we consider barriers to implementing the proposal and discuss its potential benefits. We reiterate that existing literacy indicators (objective tests of reading and writing) serve an important purpose and are *not* excluded in this recommendation. The proposal to incorporate teacher judgments in the indicator system is intended to enhance the present approach, both by encompassing a broader array of outcomes and by gathering information that allows a sounder basis for action.

BARRIERS

The major barrier to implementation of the proposal centers around the ability and willingness of practicing professionals to handle the task. Can teachers make the required judgments in an informed and responsible fashion? Will bias and prejudice influence their ratings? Will they complete the surveys without compulsion? Such questions imply distrust of teachers; moreover, research on the fallibility of human judgment suggests that people seldom act as rational decisionmakers, but are influenced by the way issues are framed, the local context, and personal experience (Kahneman, Slovic, and Tversky, 1982; Nisbett and Ross, 1980). However, evidence suggests that teachers approach such tasks responsibly. The National Education Association (NEA) conducts regular surveys of a sample of its membership (e.g., NEA, 1987); participation is voluntary, but 75 percent of the sample complete the survey. A lengthy ETS questionnaire of compensatory reading practices yielded a similar return rate (Calfee and Drum, 1979). Each year, teachers in California complete the California Basic Educational Data System forms; school funding is related to the information, and the return rate is virtually 100 percent. The federal Office of Education conducts regular surveys (no money is involved); its 1985 project covered 3,000 schools and 10,000 teachers and had a return rate of 85 percent (Hammer and Batcher, 1986). Willingness to participate seems to depend more on the quality of the survey and the intelligent use of the findings than on any inherent interest or lack of interest on the part of practitioners.

The next issue is that of *trustworthiness*. Can teachers furnish judgments that are reliable and valid? The evidence here is mixed. Teachers are able to predict student achievement fairly accurately (Shavelson, 1984), but they have trouble agreeing on the diagnosis of specific reading problems unless they are given training beyond what is

routinely available (Vinsonhaler et al., 1983). These conclusions are based on spotty evidence, however. Relatively little research is available on the consistency and validity of teacher assessments, so we must rely on informed judgments. Preparing a profile of the strengths and weaknesses of a class of students is a demanding technical task for which teachers receive little training. Research on complex decisionmaking suggests that judgments are often intuitive and heuristic, balancing the consistency of rational methods and the wisdom of experience. We might expect that novices would behave this way, but professionals and experts should be more articulate about the process. Staff development will be a necessity (Stiggins, 1988; Hiebert and Calfee, forthcoming).

How much variability and inconsistency in teachers' judgments about student performance and about the instructional context that supports student achievement can we expect? Obtaining data that give clearer insights into these questions is a major research task. The plain fact is that teachers do make decisions about students, and these decisions influence student assignment to educational programs. But we do not know how well this part of the system is working.

The third issue is *bias*. Part of the original rationale behind the use of objective tests was the concern that teachers were subject to prejudice and that students needed a forum to demonstrate their competence free of considerations of ethnicity, sex, or other personal characteristics. The social context in the United States has changed substantially in the past fifty years, but concern about bias cannot be dismissed. On the other hand, objective tests provide little guarantee of equal opportunity unless they are supported by the efforts of the front-line practitioners. If prejudice is intruding into classrooms in any form, it needs to be identified, understood, and eliminated.

Inertia is perhaps the most significant barrier to change. We are accustomed to reports that portray achievement as a function of student background, and we all know what to expect. Yet the positive features of the proposal seem to outweigh the costs.

BENEFITS

There are, of course, limits to the ability of teachers to carry out the tasks called for in the recommendation, but the basic issue is simple: Quality education presupposes teachers who can assess student progress and evaluate the adequacy of instruction to meet student needs. The inclusion of teacher judgments in a system of literacy indicators serves several purposes: It gives policymakers information about the consistency of teacher assessments with other sources of information; it yields data on patterns of instructional practice as perceived by teachers; it provides a more informed basis for understanding the linkages

between classroom practice and achievement; and it suggests directions for improving practice.

The barriers discussed above place additional demands on the training and certification of teachers. Most professions have a substantial basis for rendering judgments: a rich technical language, explicit standards, and collegial review. Lacking these resources, elementary teachers are subject to the idiosyncracies and inconsistencies noted in the research on novice decisionmaking. On the other hand, research shows that reasoned judgments can be taught, and that professionals can be distinguished by this capacity (Nisbett et al., 1987).

Educational policymakers in general have the same "wish list":

- They want to know how their students are doing. Objective tests give one outlook on this issue, but they are too narrow and are subject to influence. Confirmation from other sources is needed.
- They want to know what is going on in the classroom—what is being taught, and how it is being taught. Administrators are one source of information, but they are busy and have little time for classroom observation.
- They want to change things for the better. They want to know whether they should alter the curriculum, whether different instructional practices are needed, whether teachers need additional training, and if so, in what areas. They want to know how community resources can be brought into play.
- They want assurance that the practitioners are competent. One would like to think that the nation's schools are *"centers of inquiry,"* learning communities where intellectual growth is the norm (Schaefer, 1967).

In a sense, the policymaker faces the same dilemma as the teacher. The most direct approach to a task is to do it yourself. But the teacher cannot use this approach; success depends upon creating an environment that encourages and guides students to the goal. An essential part of the process is careful monitoring of how well students are doing, of whether they can explain what they are doing, of whether they can identify and deal with problems.

We need to do more than improve reading scores on standardized tests. Literacy is the foundation for thinking and communication in modern society, and indicators of achievement and progress in this domain are of fundamental importance for all other areas of schooling. The advice of the people who guide children to literate understanding is likely to be the keystone of informed policy, and it seems reasonable to embrace those people as partners in this enterprise.

REFERENCES

AERA, APA, & NCME. (1985). *Standards for educational and psychological testing.* Washington: APA.

Alexander, L. (1987). *The Nation's Report Card: Improving the Assessment of Student Achievement.* Cambridge, MA: National Academy of Education.

Anderson, R. C., E. H. Hiebert, J. A. Scott, and I. A. G. Wilkinson (1985). *Becoming a Nation of Readers.* Washington, DC: National Institute of Education.

Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer and H. I. Braun (Eds.), *Test validity.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Applebee, A. N., J. A. Langer, and I. V. S. Mullis (1988). *Who Reads Best: Factors Related to Reading Achievement in Grades 3, 7, and 11.* Princeton, NJ: Educational Testing Service.

Barr, R., and R. Dreeben, with N. Wiratchai (1983). *How Schools Work.* Chicago: University of Chicago Press.

Barr, R., C. L. Z. Blachowicz, B. E. Johnson, D. Morris, J. H. Mosenthal, and D. M. Ogle (Eds.) (1987). Metacognition and reading (Special issue). *Journal of Reading Behavior, 19*(3).

Boruch, R. F., and P. Wohlstetter (1983). On educational indicators: A review of the "Condition of Education." *Contemporary Educational Review, 2,* 13-22.

Bransford, J. D., and M. K. Johnson (1972). Contextual prerequisites for learning. Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717-726.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Calfee, R. C. (1986). Curriculum and instruction: Reading. In B. I. Williams, P. A. Richmond, and B. J. Mason (Eds.), *Designs for Compensatory Education.* Washington, DC: Research and Evaluation Associates.

----- (1987). The school as a context for assessment of literacy. *The Reading Teacher, 40,* 738-743.

Calfee, R. C., and R. Brown (1979). Grouping students for instruction. In D. L. Duke (Ed.), *Classroom Management.* (NSSE 78th Yearbook, Part II). Chicago: University of Chicago Press.

Calfee, R. C., & Chambliss, M. (1987). The structural design features of large texts. *Educational Psychologist*, 22(3 & 4), 357-378.

Calfee, R. C., & Drum, P. A. (1986). Research on teaching reading. In Wittrock, M. C. (Ed.), *Handbook of research on teaching* (3rd Ed). (pp. 804-849). New York: Macmillan.

----- (Eds.) (1979). *Teaching reading in compensatory classes.* Newark, DE: International Reading Association.

Calfee, R. C., & Hiebert, E. (1988). The teacher's role in using assessment to improve learning. In Bunderson, C. V. (Ed.), *Assessment in the service of learning.* Princeton, NJ: Educational Testing Service.

Carroll, J. B. (1987). The rational assessments in reading: Are we misreading the findings? *Phi Delta Kappan, 68*(6), 424-30.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Cole, N. (1988). A realist's appraisal of the prospects for unifying instruction and assessment. In Bunderson, C. V. (Ed.), *Assessment in the service of learning.* Princeton, NJ: Educational Testing Service.

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

----- (1987). *Educational achievement: explanations and implications of recent trends.* Washington, DC: Government Printing Office.

Cronbach, L. J. (1969). Validation of educational measures. *Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a theory of achievement measurement.* Princeton, NJ: Educational Testing Service.

----- (1984). *Essentials of psychological testing* (4th Ed.). New York: Harper & Row, Publishers.

----- (1988). Five perspectives on the validity argument. In Wainer, H., & Braun, H. (Eds.), *Test validity.* Hillsdale NJ: Lawrence Erlbaum Associates.

Cross, D. R., & Paris, S. G. (1987). Assessment of reading comprehension: Matching test purposes and test properties. *Educational Psychologist, 22,* 313-332.

David, J. L. (1987). *Improving Education with locally developed indicators.* New Brunswick, NJ: Center for Policy Research in Education/Rutgers University.

Edmonds, R. (1985). Characteristics of effective schools. In Osborn, J., Wilson, P. T., & Anderson, R. C. (Eds.), *Reading education: Foundations for a literate America.* Lexington, MA: Lexington Books.

Fraatz, J. M. B. (1987). *The politics of reading*. New York: Teachers College Press.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*, 193-202.

Gorman, T. P., White, J., Orchard, L., & Tate, A. (1982). *Language performance in schools*. London: Her Majesty's Stationary Office.

Graham, P. A. (1987). "Achievement for at-risk students." Unpublished manuscript. Cambridge, MA: Faculty of Education, Harvard University.

Guthrie, J. T. (1987). *Indicators of reading education*. New Brunswick, NJ: Center for Policy Research in Education/Rutgers University.

Haertel, E., & Calfee, R. C. (1983). School achievement: Thinking about what to test. *Journal of Educational Measurement, 20*, 119-130.

Hammer, C. H., & Batcher, M. K. (1986). *E.D. Tabs: The 1985 Public School Survey early tabulations*. Washington, DC: Office Of Educational Research and Improvement.

Hiebert, E. H., & Calfee, R. C. (In press). What research has to say to the classroom teacher: Assessment. In Farstrup, A., & Samuels, S. J. (Eds.), *What research has to say about reading instruction*. Newark DE: International Reading Association.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

McLean, L. D., & Goldstein, H. (1988). The U.S. national assessments in reading: Reading too much into the findings. *Phi Delta Kappan, 69*(1), 369-372.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*, 215-237.

----- (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In Wainer, H., & Braun, H. I. (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis, 9*, 101-116.

NAEP (1985). *The reading report card: Progress toward excellence in our schools*. Princeton, NJ: Educational Testing Service.

----- (1986a). *The writing report card*. Princeton, NJ: Educational Testing Services.

----- (1986b). *Writing: Trends across the decade, 1974-1984*. Princeton, NJ: Educational Testing Service.

----- (1987). *Learning to be literate in America*. Princeton, NJ: Educational Testing Service.

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

National Education Association (1987). *Status of the American public school teacher, 1985-86.* Washington, DC: National Education Association.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238,* 625-631.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Oakes, J. (1986). *Educational indicators: A guide for policy makers.* Santa Monica, CA: Center for Policy Research in Education/The RAND Corporation.

Orasanu, J. (1986). Reading comprehension: From research to practice. Hillsdale NJ: Lawrence Erlbaum Associates.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal, 83,* 427-452.

Romberg, T. A. (1988). *Changes in school mathematics: Curricular changes, instructional changes, and indicators of change.* New Brunswick, NJ: Center for Policy Research in Education/Rutgers University.

Schaefer, R. J. (1967). *The school as a center of inquiry.* New York: Harper and Row.

Shavelson, R. J. (1984). Teachers' judgments. In Husen, T., & Postlethwaite, T. N. (Eds.), *International encyclopedia of education.* Oxford, England: Pergamon Press.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347-376.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of reading. *Reading Research Quarterly, 21,* 360-406.

Stedman, L. C. (1987). It's time we changed the effective schools formula. *Phi Delta Kappan, 69*(1), 215-224.

Stedman, L. C., and C. F. Kaestle (1987). Literacy and reading performance in the United States, from 1880 to the present. *Reading Research Quarterly, 22,* 8-46.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69*(1), 363-368.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement, 5,* 5-17.

Trevelyan, W. M. (1942). *English social history.* London: Longmans, Green, and Company.

Venezky, R. L. (1974). *Testing in reading*. Champaign, IL: NCTE, ERIC Clearinghouse on Reading and Communication Skills.

Venezky, R. L., Kaestle, C. F., & Sum, A. M. (1987). *The subtle danger: Reflections on the literacy abilities of America's young adults*. Educational Testing Service: Center for the Assessment of Educational Progress, Report No. 16-NAEP-01.

Visonhaler, J. F., Weinshank, A. B., Polin, R. M., & Wagner, C. C. (1983). *Improving diagnostic reliability in reading through training* (Research Series 176). East Lansing, MI: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Service No. ED 237 934).

Wixon, K. K., & Peters, C. W. (1987). Comprehension assessment: Implementing an interactive view of reading. *Educational Psychologist, 22*, 333-356.

This note may be obtained directly from the Publications Department, The RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138. Telephone (213) 393-0411, Extension 6686.