

CAN REPORTING ON EDUCATIONAL INDICATORS SERVE AS A CATALYST FOR
THE IMPROVEMENT OF EDUCATIONAL ACHIEVEMENT?
-- A VISIONARY EXPLORATION --

Pascal D. Forgione, Jr. and Joan B. Baron, CT State Department of Education
Pascal D. Forgione, Jr., P.O. Box 2219, Hartford, CT 06145

A. BACKGROUND

Perhaps the landmark event in this arena of education indicators and systems for reporting on the condition of education was Secretary Bell's Wall Chart first released in 1982 (U.S. Department of Education, 1987A). This scoreboard ranking of states on a set of available statistics stimulated a great deal of discussion on the part of state education officials and citizens about the quality of education. Subsequently, over the past five years, two important national organizations of educational policymakers have endorsed the policy needed to establish high quality state-by-state educational indicators, i.e., the Council of Chief State School Officers (1984) and the National Governors' Association (1987). When combined with the host of national reports on education (like the Nation at Risk, 1983, and A Nation Prepared, 1986) and a number of recently conducted educational studies (such as the NAEP Reading and Writing Reports and the International Education Assessments of Mathematics and Science), it is not difficult to understand why educational issues continued to remain on the front burner in this country.

Let us use the Wall Chart as a point of entry into some of the major issues surrounding the collection and portrayal of educational indicators. The Wall Chart was divided into four sections: Student Performance, Resource Inputs, State Reforms, and Population Characteristics. There was, and continues to be, much appropriate disappointment with the Wall Chart's narrow and limited set of student outcome measures available for the U.S. Department of Education's state-by-state report card, i.e., SAT and ACT scores, and graduation rates. Following the pleas of state departments of education and other policymakers, the federal agency has indeed tried to present a context for looking at the educational achievement of this nation's students. In fact, we want to compliment OERI for its general interest in presenting responsible and appropriate data. Their willingness to canvass states for their ideas and their willingness to participate in symposia like this one is a very clear indication of their commitment to continuously improving the quality of data collected while remaining sensitive to the burdens of states and local agencies in collecting and aggregating those data.

What one portrays and reports sends clear messages. One of those messages is that the reported data are important. Another is that the context variables are in some way related to the student performance variables or they wouldn't be presented (see Figure 1). A result of these inferences is the notion that

if we somehow improve on those aspects over which we have control, that would in some way improve student achievement. That is, if we increased the resources and had more state reforms, education would improve. We will spend time in a later section exploring these assumptions. When it comes to the unalterable variables (see "inputs" under Figure 1), the assumptions change. People often wrongly assume that there is an unalterable causal relationship between certain socioeconomic and demographic factors and school achievement. Population characteristics such as the distribution of wealth and educational background of a district's (or state's) population have been shown to be related to education. The problem here is that because the relationship does exist between the wealth and educational background of parents and the achievement level of their students, this has been used as an excuse for poor performance in places where there is poverty and a less educated population.

But the impact of these background characteristics can be moderated by the effects of schooling (Brookover, *et.al.*, 1979). The challenge then is to present data on the alterable and unalterable variables in a more responsible way. It is our intent to suggest that it is possible to collect and portray information on educational indicators which will have the following characteristics:

1. The context data (i.e., school process variables) will be related to student achievement.

2. The alterable variables will be those that, because they have been demonstrated to be related to achievement, can serve as appropriate magnets for change. That is, if a district increases its attention to those variables, there is reason to believe that achievement will improve.

3. The unalterable variables will be presented in a framework of meaningful comparison groups which will permit states (or districts) to compare their performance with others like themselves. This is intended to serve a very real need for people to see how they compare with others who are like themselves. However, the framework will be structured so that within the context of the unalterable variables which structure the comparison groups, the achievement data will also be mapped against the alterable variables. Experience in a variety of settings has shown that when districts alter their resources, they can indeed make a difference in student achievement.

The messages that we hope will emerge through the data portrayal are these:

1. States (or districts) start out with different types of students who are easier or

harder to educate. Some enter school with considerable verbal fluency and high motivation levels; these students are ready to achieve. Others have language deficits and don't immediately see the relationship between school and future success; these students are not as ready to achieve.

2. Yet, despite initial differences on demographic variables, schools can make a difference. That is, not all districts with the same high wealth factor and high parent educational level will produce the same high student achievement level. Furthermore, not all districts with the same degree of poverty will have students who are succeeding at a low rate. The resource allocations of districts on the alterable variables that matter may indeed explain the differences in student achievement.

The challenge will be to create a reporting framework that portrays indicators in such a way that the reader can discern at a glance that the unalterable variables do not by any means tell the whole story. Such a reporting schema would not deny the importance of wealth and background variables, but it would reassert the importance of schooling. It would force one to reckon with the fact that the resources that are allocated to education and the instructional practices that are used have the potential to overwhelm the effects of disadvantageous backgrounds. This is the challenge in designing a new comprehensive indicator system for reporting on the condition of education in the Nation and across the states.

B. ESTABLISHING A NEW FRAMEWORK FOR A COMPREHENSIVE INDICATORS SYSTEM

In this part of the paper we will discuss the issues which we feel are essential in selecting the indicators to be used in each of the three components of a comprehensive indicators system: unalterable variables (inputs or background), outcome measures (results of the educational process), and alterable variables (including both resources and instructional practices). Figure 1 provides a graphic representation of how the three components relate to one another (Shavelson, 1987).

In the following three sections we will briefly review some of the recent conceptual work that we believe is the most promising in the hope that statisticians reading this paper can help us determine how to select the best indicators from among the hundreds that have been suggested. It is also hoped that statisticians may be able to develop ways to combine some of these separate indicators into scales which incorporate the relative importance of their contributions to achievement. For the policymakers reading this, we hope to stimulate dialogue on issues like "corruptibility" of educational indicators. Oakes recently stated that "what you count may be what you get". In other words, what we collect data on may drive districts to try to inflate those numbers.

The challenge for policymakers is to select for counting only those variables which would tend to produce positive results if districts were to raise their numbers. This is a very complex area. Murnane (1987) cautions us that it is easy for indicators to be corrupted which will have the effect of reducing the relationship between that indicator and student achievement. He uses the example of course-taking patterns and numbers of students in mathematics. If you merely want to increase the number of students taking mathematics, you can do so by making the courses easier. This may ultimately have no effect on the general achievement of mathematics in that school. In fact, it is possible that a "watered-down curriculum" might have just the opposite effect.

We are not by any means suggesting that this is an easy task. But, it must be done. We are currently being inundated by indicators. Therefore, we feel that in the face of decisions that must be made about what to measure, count and report upon, it is important to establish a set of working criteria that may help us to achieve our greater end which is not the reporting on student achievement but the improvement of student achievement.

I. Unalterables Variables

There are countless studies which show that unalterable characteristics of student background are correlated with student achievement. These include demographic variables (type of community), socioeconomic variables (income levels represented in the community), and educational background variables (the amount of schooling of parents in the community). The net effect of these unalterable variables is that some students enter school ready and eager to learn and others need to be motivated after they arrive at school. Some students will learn with very little effort on the part of schools; others will need considerable nurturance and patience.

Figure 2 provides a synopsis of the class of variables over which schools have no control, but which have dramatic effects on the schooling process (U.S. Department of Education, 1987C). We are opposed to the development of different expectations for students of different backgrounds. We do not believe in adjusted achievement scores. We suggest that it might be useful to profile achievement data in such a way that acknowledges the differences in the backgrounds of students, but does not offer it as an excuse for low achievement. If we divide states into quartiles using these variables and display the achievement data next to the alterable variables for each state, we believe that some important insights might emerge. States (or other jurisdictions) could compare their achievement with others similar to them. Yet, it is likely that there will be many surprises. For example, there may well be some poor states with higher achievement than

other poor states and some wealthier states with lower achievement than other wealthy states. The alterable variables may help to explain these differences.

Figure 3 provides an example of a possible framework of meaningful comparison groups based on a fair and valid classification schema. States would be profiled within "like groups" which have been structured within the context of the unalterable variables and profiled in terms of their outcome standings (e.g., achievement results) and the alterable variables.

Our goal in designing a new comprehensive indicator system for reporting on the condition of education in the Nation and across the states should be to create a reporting framework that portrays indicators in such a way that the reader can discern at a glance that the unalterable variables do not by any means tell the whole story. Such a reporting schema should also focus the attention of policy makers and program officials on the importance of schooling variables and it should provide direction in the resource allocations of schools, districts and states on these alterable variables that matter.

II. Outcomes Variables

Let us now examine the three categories of outcomes that we feel deserve to be portrayed in a comprehensive state-by-state indicator system. We will treat these in the following order: student achievement, attitudes and attributes, and student participation.

Student Achievement. Tests send messages to test takers. They reflect the test-makers' decisions about what is important for students to know and be able to do. Students wisely look at a teacher's tests to guide their allocation of time and resources in preparing for future tests. For example, multiple-choice tests which require fine discriminations require one kind of studying; essay tests which require integration or evaluation require a different review strategy.

In a recent mini-debate between Popham (1987) and Bracey (1987) on the subject of measurement driven instruction, whereas the authors do not agree upon the desirability of instruction targeted at tests, they do agree that this is a prevalent situation in our nation's schools. Classroom instruction generally becomes measurement-driven when the stakes of testing are high--either for the individual students or for the public images of schools and school districts.

When a test becomes the basis of students' promotions to the next grade or an important factor in determining whether a student will graduate, teachers use the test as a magnet for instruction. The same can happen when standardized test scores are reported to Boards of Education; district or school goals are generally established to raise test scores. State testing programs which publicly report the scores of schools or districts often have the same impact.

Districts target resources to improve test scores (see Tirozzi et.al., 1985).

Now, because of recommendations made by the Council of Chief State School Officers (1984) there is the distinct possibility that there will be a national test with scores that are available to be reported for each state (National Academy of Education, 1987). Obviously, all states will want to score high on such a test. Yet, the different political and economic contexts in each state will determine just how important these test scores will be perceived to be by state policymakers. To the extent to which some states view these scores as more than "indicators" of achievement and make them magnets of instruction, the content of these tests will take on extraordinary importance (Forgione, 1987A).

How might a national test look? If we were back a few years, it might have looked like one of several dozen minimum competency tests that had been developed by states like Connecticut. Each state established an advisory committee to represent all of the important constituencies impacted by the tests. The central concern was curricular validity in those attempts to establish tests that reflected the nature of the 3 R's curricula in their schools. They wanted to be fair--to hold students and schools accountable for only those skills for which there was a strong consensus on their importance and their presence in school curricula. (This was particularly critical in preparing for any future law suits.) By the early eighties, states became quite sophisticated about conducting appropriate reviews for content, bias, and a host of psychometric properties that would enable the construction of multiple parallel test forms.

The minimum competency testing movement had resulted in great part from a concern for equity. States wanted to assert that all of its students had an equal opportunity to master the basic skills. During the past few years, in direct response to the Nation at Risk (1983) and the host of other national reports and international studies, the cry for equity was matched by a cry for excellence. Several states turned their attention to raising standards for their students. This manifested itself in several ways--ranging from more comprehensive lists of skills within the 3R's to a more comprehensive list of subjects tested. Some states required these new harder tests of all students; others provided them for those seeking a special diploma. This latter distinction is an important one, because it is considerably harder to design a test that will raise standards for all students. Furthermore, curricular validity becomes more and more difficult to achieve as the standards are raised. This is particularly true when a state does not have in place a statewide curriculum.

Since 1983, several new voices are being heard--each with a somewhat different view of what students ought to know. For the past

three years, educators in all content domains have been calling for the inclusion of higher order thinking skills on state tests. Many of them feel that it is important to measure the depth of students' knowledge of a subject as well as the breadth. These educators feel that it is extremely important for students to read critically, solve multi-step problems, and synthesize knowledge. Minimum competency tests did not generally emphasize such skills (see Kearney, 1986).

Since April, 1987, a second voice has been heard--best typified by Hirsch, (1987) in his recent book, "Cultural Literacy: What Every American Needs to Know". He and many others feel that there is a core of knowledge that all students should know and that these facts should be taught and tested.

A third voice, is represented by Sizer (1986) in Horace's Compromise. He calls for the development of rich sustained projects and activities that would permit students to provide detailed evidence about what they know and can do. Consistent with Sizer's vision, though not as rich, have been statewide performance tests with observable behavioral criteria. For example, more than a dozen states have successfully measured writing with a written essay. In Connecticut, we have used performance tests on samples of students in art and music, science, foreign language, business and office education, drafting, graphic arts, and small engines (see Baron, 1987). In addition, national performance tests have been conducted in Great Britain (Burstall, 1986) and most recently piloted by NAEP in Science.

The type of test that would result from each of these positions would be dramatically different. So would any instructional program targeted to improve test scores. Therefore, designing a national test is a staggering responsibility. Some important questions that would need early resolution are:

1. Will curricular validity be the critical concern or will a national test abandon the consensus approach and attempt to define the "shoulds" for education?

2. Which "shoulds" will they be? Who will be responsible for determining the philosophical underpinnings of the new national test?

In conclusion, our experience in Connecticut has taught us that the most pressing problems facing us in test development are conceptual rather than technical. The appropriate test methods are in place for building a valid and reliable instrument. As we approach the 1990's, we are limited more by our imagination in conceptualizing the test's content than by the technology available to us.

It is important to remember that we have been able to design and report results from minimum competency tests in a way that has allowed scores to improve over time. What is needed now is for a group of our most responsible educators and policymakers to

think deeply about the outcomes we desire from schooling. Ideally, we would seek to build tests that mirror those outcomes. Then, we would hope that the tests would indeed serve as magnets for instruction. Ravitch (1985) and Powell, Farrar and Cohen (1985) have asserted that Americans get the schools they want and/or deserve. It is our hope that our tests will be a responsible positive force in guiding that overwhelmingly important social phenomenon.

Attitudes and Attributes. There is general acceptance or acknowledgment of the important influence of students' attitudes and attributes on their learning. The time is ripe for the development of indicators of attitudes such as those called for in Connecticut's Common Core of Learning (1987) which proposes to address seven essential areas, including: (1) positive self-concept; (2) motivation and persistence; (3) responsibility and self reliance; (4) intellectual curiosity; (5) interpersonal relations; (6) sense of community; and (7) moral and ethical values.

Many of these are thoroughly consistent with those promulgated recently by cognitive psychologists (Sternberg, Bransford, Feuerstein); educational psychologists (Duckworth) and philosophers (Lipman, Ennis). A recent paper (Baron, 1986) provides a comprehensive synthesis of the dispositions related to higher order thinking skills. A significant refocusing on the importance of this area is provided in the extremely well written report from the Second International Mathematics Study: The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective (January 1987).

After describing a set of deceptive explanations for the relatively poor performance of students from the United States, the authors provide some insights gained from interviews with teachers in Japan, the United States and elsewhere. The authors contend that much of the success of the Japanese education program could be attributed to the optimistic attitudes of the Japanese teachers about the capabilities of their students to learn.

Student Participation. A major focus of school and policy officials has been the elusive "dropout rate" statistic. Significant resources are being proposed at the national, state and local levels to address this important but ill-defined area. As proxy variables for this issue a variety of data elements have been produced and reported (see U.S. Department of Education, 1987B). Much of these data are questionable in value and highly unreliable (i.e., persistence rates (Grades 9 to Grade 12; Grade 9 to Graduates), otherwise termed the graduation rate). The converse of this statistic is the loss rate, or dropout rate, as used by many jurisdictions (including the Wall Chart). This is an indicator which needs a considerable amount of thought, especially if states are to report data that can be compared with one another.

III. Alterable Variables

One assumption underlying this section is that schools can and do make a difference in the achievement of children. Schools have control over how they spend their money, who they hire, who they tenure, what materials and instructional strategies are used in classrooms, what kinds of professional development is provided, what kinds of aids are provided to teachers, etc. These are the alterable variables. It doesn't matter whether the children come from poor or affluent homes and live in cities or suburbs. In every school district, there is an underlying belief system about how children learn best and which children can learn what things. The alterable variables are manifestations of those beliefs.

Another assumption is that school can change. They can reorder their educational priorities, add or delete courses, buy new books, update equipment, provide different staff development opportunities, emphasize the inclusion of learning strategies. With strong instructional leadership and a sensible and adequate program, teachers can both alter their beliefs about how all children can learn and subsequently teach them the skills required to function successfully in school and prepare for the world of work. We are not suggesting that these changes are easy to produce. Nor, are we suggesting that they are easy to measure.

It appears that at the same time we have both too many and too few indicators. In this information age, local and state education officials feel that we are on the verge of being inundated with requests for data. Educational indicator systems are proliferating at an almost incredible rate.

The increased sophistication and capabilities of computers and information management systems have made it possible to collect large amounts of data on virtually everything that can be counted. The problem is simple: it takes valuable time and energy to collect and report data for the computers. It is our contention that at this time, we need to pull in the reins on the burgeoning number of separate indicators. We need to ask ourselves two questions:

1. How do we select from among all of the currently available indicators the ones that are useful to understanding and enhancing our educational systems?; and

2. How can we prioritize our needs in order to target resources for the development of measures that currently do not exist but appear to be important for understanding educational outcomes?

It may seem strange that at the same time we are advocating both the elimination and the creation of educational indicators. One of the most important influences in our thinking has been the recently published, The Underachieving Curriculum (McKnight, et. al., 1987). After presenting the data from the latest international study of mathematics, the authors included a chapter entitled, Deceptive Explanations. In this chapter, the

authors demonstrate how many of our favorite explanations for poor achievement (i.e., time for instruction, class size, the comprehensiveness of our educational system, the preparation and status of teachers, and the quality of mathematics teaching) do not explain why students in the United States achieved so poorly in mathematics. Through the painstaking collection of data related to the amount of time spent in classrooms on each mathematics topic assessed on the test and through interviews with American and Japanese teachers about their beliefs about the capabilities of their students to learn and the ways in which the schools structure their professional development, the authors have provided us with some critically important insights about the optimistic beliefs that Japanese teachers have in their students' abilities to learn. This could only have come from thousands of hours of coding data on what is actually happening in classrooms and the beliefs, philosophies and staff development opportunities of the teachers responsible for students' learning. In order to attain these insights, there were large numbers of variables that didn't emerge as important and thousands of data elements that were collected, coded and analyzed which proved to be fruitless. The present task for statisticians, researchers, and educators today is to go back to the drawing board and use information that we have in order to develop strategies for collecting data on those variables that we believe make a difference in the quality of education.

A good starting point in this quest is the work of Oakes at the Rand Corporation who has produced a thoughtful and comprehensive review of the conceptual and measurement issues related to the construction of school quality indicators. Oakes suggests three major categories of process indicators which have been found to be related to student outcomes: Access to Knowledge, Press for Achievement, and Professional Conditions for Teaching. Within each category, she further subdivides the variables into three types: Resources, Structures, and Culture. For example, in the grouping called Access to Knowledge, Oakes (pp. 31-32) is concerned with the issue of whether students of all abilities have sufficient opportunities to learn. The Resource variables include the availability of sufficient instructional materials, laboratories, computers, and equipment; teachers' qualifications and experience for the courses they teach, and the availability of discretionary funds for supplies, materials, trips, speakers, etc. Under Structure, Oakes includes instructional time in days per year and hours per day, the emphasis the school places on different curriculum areas as measured by course offerings and staffing patterns, the procedures schools use to assign students of different abilities to classes and the types of assignments they receive within classes, as well as the academic enrichment and supports available to students. In the

Culture section, Oakes is concerned with opportunities for staff development, parent involvement and staff perceptions about the importance of learning for all students.

In Press for Achievement, Oakes (pp. 33-34) is concerned with how schools organize their staff, time, curriculum and materials to support the belief that all students can learn. The Structures include student participation in long-term projects, papers and research activities, opportunities for school-wide recognition of accomplishments, graduation requirements, and student participation in challenging study as measured by enrollment in challenging courses, and average course completion rates. The Culture variables are more diverse and include graduation and attendance rates as well as student attitudes toward achievement and staff perceptions about the importance they and their school place on student achievement. Because of their breadth and their importance, these are listed in Figure 4.

In Professional Conditions for Teaching, Oakes (pp. 35-36) is concerned with how schools provide teachers with the supports regarded as important in order to be successful on the two categories described above. The Resource variables include teacher salaries, pupil load, class size, funding for school-based staff-development activities, and clerical support available for teachers' noninstructional tasks. The Structures include the amount of teacher-time scheduled for teaching, non-teaching work, school-wide staff-development activities, and special teacher-developed projects (e.g., curriculum development, instructional improvement, collaborative research, etc.). The Culture variables include a set of staff perceptions related to the school's goals and the nature and level of staff involvement in curriculum and instruction. These are listed in Figure 4.

Oakes' lists of indicators have been presented to be illustrative of the types of process variables that could make successful indicators of school success. They should be viewed as illustrative rather than final as there are undoubtedly other process variables to be considered. Whatever the operational set will be, we are certain that it will include many indicators that will be difficult to measure. They almost certainly will require some in-depth interviews with teachers as well as classroom observations. But, if policymakers and statisticians are tempted to shy away from investing the time and creative energies required to reach an understanding of Access to Knowledge and Press for Achievement, they should turn again to The Underachieving Curriculum in Mathematics Report to remind themselves of the importance of collecting time consuming yet invaluable quantitative and qualitative data. Otherwise, our favorite deceptive explanations are likely to be around for many years to come.

Oakes and others are sensitive to the fact that many of these variables are currently difficult to measure and some of those that are easy to measure can be corrupted. Murnane, in a recent paper (1987) points out how easy it is to corrupt an indicator if high stakes are placed on them. For example, the traditional strong, positive relationship between number/amount of math courses taken and math achievement is presently at risk of being watered down by the proliferation of less challenging math courses. This has arisen because in order to comply with new state and local course taking standards, schools have put new courses into place or restructured or renamed their mathematics course offerings without significantly improving the quality of the mathematics curriculum offered to children. Therefore, it is important to consider the impact of putting these indicators into place.

One area which needs a great deal of work in the development of a comprehensive education indicator system is in the measurement of teaching quality (i.e., a teacher's capability to teach one's subject matter). Up until now, we have only been able to address the teacher quality component through proxy measures and variables, such as educational background, areas of certification, years of experience, etc. (see Figure 1A). Shavelson (1987) provides a conceptual schema for distinguishing between the input (teacher quality) and the process (teaching quality) components in a comprehensive model of the education system (see Figure 1). Recently, the Connecticut State Department of Education has begun to develop behavioral protocols for use in a clinical performance assessment of Connecticut beginning teachers as part of the Connecticut Teacher Licensure Assessment Program. Prototype protocols have been developed in the areas of secondary mathematics and special education (see Forgiione, 1987B). These assessments will be used to evaluate both the subject matter and pedagogical understandings of new beginning teachers entering the profession in Connecticut (Connecticut State Board of Education, 1986). Figure 5 describes the four components and associated competencies incorporated in the proposed format for the structured interviews that Connecticut is planning to use as part of an assessment center evaluation of novice teachers beginning in school year 1990-91. For example, nine specific categories of subject matter-pedagogical understandings have been identified that will be probed through use of the secondary mathematics protocol: (1) sequence topics within a unit; (2) identify major concepts and skills within a unit; (3) sequence learning activities within a lesson; (4) identify prerequisite skills and concepts; (5) analyze textual material; (6) identify alternative approaches to teaching a topic; (7) recognize common errors and suggest methods of clarification; (8) identify importance/significance of

topics; and (9) evaluate student performance.. It is Connecticut's expectation that such performance testing can ultimately be used as a more direct measure (i.e., indicator) of teaching quality in Connecticut.

C. DEVELOPING A COMMON UNDERSTANDING

An immediate use of a reconceptualization such as the one we are advocating is in the upcoming NELS, NAEP, and CCSSO assessments. Each of these studies is committed to developing more sophisticated indicators of both the context and outcomes of schooling. These national developmental efforts should assist states and locals in developing new and better indicators that are directly related to student achievement and amendable to the actions of education policy makers.

There is a need for leadership -- both policy and technical -- to strengthen the current emphasis on educational indicators so that the decade of the 1990's will find the education profession in control of its destiny through a strong and reciprocal relationship among research, policy, and practice. We view the current interest in indicators as an opportunity to strengthen these relationships and to reinforce the public's perception that our profession and discipline is worthy of strong financial and moral support. The opportunity is before us -- let us reach for it and pursue the challenge to develop a set of a few indicators that are really sensitive to what is happening in schools and classrooms and that can have a catalytic effect on higher student achievement. Our ultimate objective in Connecticut is the improvement of student achievement. It is hoped that the development of better indicators for reporting on student achievement in the Nation and across the states will move us toward that goal.

REFERENCES

- Baron, J. (1986). Being disposed to thinking: A typology of attitudes and dispositions related to acquiring and using thinking skills. Proceedings of University of Massachusetts Staff Seminar, Boston Harbor Campus. Summer 1986.
- Baron, J. (1987). Performance testing in the Connecticut Assessment of Educational Progress (CAEP) Program, 1980-87. Presented at the Education Commission of the States, Assessment and Evaluation Conference. Boulder, CO. June 1987.
- Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice and The muddles of measurement-driven instruction. Phi Delta Kappan. May 1987. 68:9 683-686 and 689-690.
- Burstall, C. (1986). Innovative focus of assessment: A United Kingdom perspective. Educational Measurement: Issues and Practices, 5:1 17-22.
- Carnegie Forum on Education and the Economy (1986). A nation prepared: Teachers for the 21st century. The Report of the Task Force on Teaching as a Profession. Washington, D.C. May 1986.
- Council of Chief State School Officers (1984). Education evaluation and assessment in the United States: Position paper and recommendation for action. Washington, DC. November 1984.
- Connecticut State Board of Education (1987). Connecticut's common core of learning. Hartford, CT. January 1987.
- Connecticut State Board of Education (1986). Proposed format for structured interview protocol. Connecticut Teacher Licensure Assessment Program. Hartford, CT. November 1986.
- Forgione, Pascal D. (1987A). A state's perspective on the Alexander-James report. Presented at the Education Commission of the States, Assessment and Evaluation Conference. Boulder, CO. June 1987.
- Forgione, Pascal D. (1987B). Second wave of teacher reform: New prototype assessment strategies for state teacher licensure assessment programs. Presented at the Education Commission of the States, Assessment and Evaluation Conference. Boulder, CO. June 1987.
- Hirsch, E. D. Jr. (1987). Cultural literacy: What every American needs to know. Boston, MA: Houghton Mifflin Co.
- Kearney, C. P. (1986). Assessing higher order skills. TME Report 90. ERIC Clearinghouse on Tests, Measurement and Evaluation. Princeton, NJ: Educational Testing Service.
- McKnight, C. G., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swatford, J. O., Travers, K. J. (1987). The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes Publishing Co.
- Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? Presented at the annual meeting of the American Educational Research Association. Washington, D.C. April 1987.
- National Academy of Education (1987). The nation's report card: Improving the assessment of student achievement. Report of the Alexander-James Study Group. Washington, D.C.
- National Governors' Association (1987). Results in education: 1987, 1988, 1989, 1990, 1991. The Governors' 1991 Report on Education. Washington, D.C. 1987.
- National Commission on Excellence in Education (1983). A nation at risk. Washington, D. C.: U.S. Government Printing Office.
- Oakes, J. (1986). Educational indicators: A guide for policymakers. Center for Policy Research in Education, OPE-01.
- Oakes, J. (1987). Conceptual and measurement issues in the construction of school quality indicators. Presented at the award meeting of the American Educational Research Association. Washington, D.C. April 1987.

Popham, W. J. (1987). The Merits of measurement-driven instruction and Muddle-minded emotionalism. *Phi Delta Kappan*. May 1987. 68:9 679-682; 687-688.

Powell, A.G. Farrar C. and Cohen D. C. (1985). *The shopping mall high school*. Boston, MA: Houghton Mifflin.

Ravitch D. (1985). *The school we serve: Reflection on the educational crisis of our times*. New York, NY: NY Basic Books.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N., with Picus, L. (1987). *Indicator systems for monitoring mathematics and science education*. Santa Monica, CA: RAND Corporation.

Sizer, T. (1984). *Horace's compromise: The dilemma of the American high school*. Boston, MA: Houghton Mifflin.

Tirozzi, G. N., Baron, J. B., Forgione, P. D. and Rindone, D.A. (1985). How testing is changing education in Connecticut.

Education Measurement: Issues and Practice. Washington, D.C. Summer 1985. 12-16.

U.S. Department of Education (1987A). *State education statistics: Student performance, resource inputs, state reforms and population characteristics, 1982 and 1986*. Office of Planning and Budget and Evaluation. Washington, D.C. February, 1987.

U.S. Department of Education (1987B). *Counting dropouts: Discussion points* (Working Paper). Presented at Center for Education Statistics, OERI meeting. Boulder, CO. June 1987.

U.S. Department of Education (1987C). *Issues about state-by-state comparisons* (Working Paper). Presented at Center for Education Statistics, OERI meeting. Boulder, CO. June 1987.

COMPREHENSIVE MODEL OF THE EDUCATIONAL SYSTEM

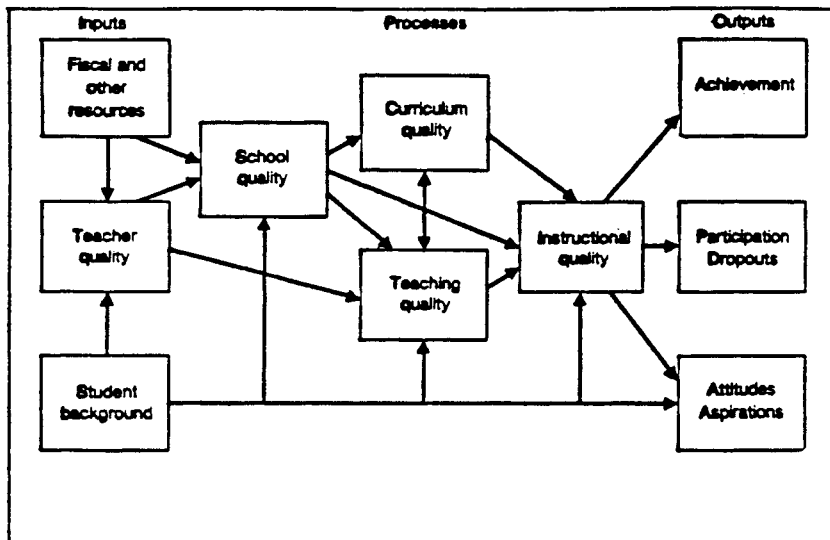
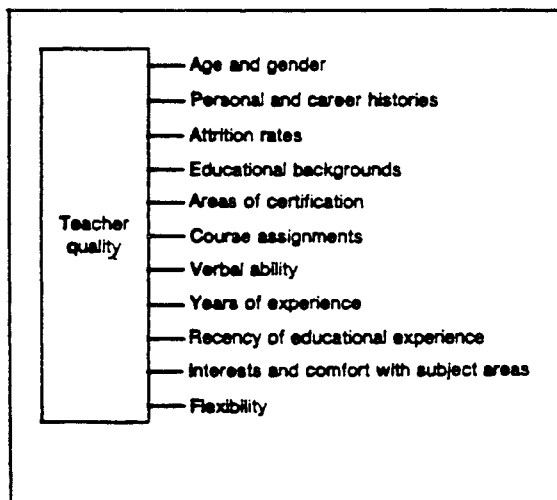


FIGURE 1A

TEACHER QUALITY COMPONENT OF THE COMPREHENSIVE MODEL



SOURCE: Shavelson, R., McDonnell, L., Oakes, J., Carey, N., with Picus, L. (1987). *Indicator Systems for Monitoring Mathematics and Science Education*. RAND Corporation. Santa Monica, CA. August 1987, pp. 10 and 12, and Appendix A (pp. 59-63).

Jeannie Oakes, *Educational Indicators: A Guide for Policymakers*, Center for Policy Research in Education, OPE-01, October 1986, p. 9.

FIGURE 2

DEVELOPING A CLASSIFICATION SCHEMA FOR CLUSTERING JURISDICTIONS FAIRLY: POSSIBLE COMPONENTS AND VARIABLES TO BE CONSIDERED

TYPES OF UNALTERABLE VARIABLES

Component	Level of Complexity	Variables	Data Source
A. SES/Demographic at community (state) level	Composite	1. Educational attainment of adults 25 years and older	Decennial Census
		2. Avg. occupational level of adults 16 years and older	
		3. Housing mobility past 10 years	
		4. Avg. number of people per housing unit	
		5. Percent urban population	
		6. Avg. family income	
		7. Percent unemployed experienced adults, 16 years and older	
		8. Percent families in poverty	
B. Fiscal -Capacity	Single	Per capita income	BEA
	Composite	Representative Tax Base	ACIR
-Cost of Providing Services	Composite	CPI	BLS
C. Student Needs	Single	LEP	Decennial Census
	Composite	1. LEP lunch eligible	Decennial Census/ U.S. Dept. of Ed.
		2. Poverty/free lunch eligible	
		3. Educational attainment of mother	
		4. Single parent households	
5. Handicapped			
D. School and District Organization and Structure	Composite	1. Average school size	School and Staffing Survey/ CCD/Census
		2. Urbanicity/Density	
		3. Number of LEAs	
		4. Proportion private school enrollment	
		5. Percent school-age population	

FIGURE 3

PROFILING OF STATE-BY-STATE INDICATORS: A POSSIBLE FRAMEWORK OF MEANINGFUL COMPARISON GROUPS BASED ON A FAIR AND VALID CLASSIFICATION SCHEMA

	OUTCOME VARIABLES	ALTERABLE VARIABLES
!UNALTERABLE!	:Attitudes :Student :	:Conditions:
!VARIABLE!	!Achieve-:and :Partici-:	!Access to: Achieve-:for :
!GROUPS*	!ment :Attributes: pation : etc.!	!Knowledge: ment :Teaching : etc.!
!Top	:	:
!Quartile	:	:
!Group	:	:
!o State _b	:	:
!o State _e	:	:
!o State _l	:	:
!Third	:	:
!Quartile	:	:
!Group	:	:
!o State _a	:	:
!o State _h	:	:
!o State _j	:	:
!Second	:	:
!Quartile	:	:
!Group	:	:
!o State _c	:	:
!o State _d	:	:
!o State _j	:	:
!Bottom	:	:
!Quartile	:	:
!Group	:	:
!o State _f	:	:
!o State _g	:	:
!o State _k	:	:

*Example illustrates the distribution of twelve (12) state (symbols "a" thru "l") across four classes (quartiles #1, #2, #3, and #4) generated based on a set of hypothetical "unalterable" variables, such as the components and variables listed previously in Figure 2.

SOURCE: U.S. Department of Education (1987C). Working Paper on Issues about State-by-State Comparisons. Presented at USED/OERI Meeting. Boulder, CO. June 1987.

FIGURE 4

OPERATIONALIZING A SET OF CONTEXT INDICATORS
THAT PROVIDE USEFUL INSIGHTS INTO SCHOOL QUALITY¹

PRESS FOR ACHIEVEMENT: CULTURE

1. Student attendance rate.
2. Student graduation rate.
3. Student attitudes toward achievement (important to achieve, o.k. to just get by, not worth the effort, etc.).
4. Students' attitudes toward academic subjects (importance, relevance, usefulness, level of difficulty, level of interest).
5. Student attitudes toward high-achieving peers.
6. Student attitudes toward homework (important, waste of time, etc.)
7. Student aspirations, e.g., percent of students with college plans.
8. Student perceptions the school's valuing and commitment to their achievement (e.g., level of agreement with statements like the following: "This school really cares about my learning" or "Teachers at this school work hard to get kids to learn").
9. Staff perceptions of school-wide expectations for student achievement.
10. Staff perceptions of school emphasis on achievement.
11. Staff perceptions of school behavioral climate (discipline).
12. Staff perceptions of instructional leadership -- the extent to which a significant person or group at the school advocates and supports instruction, academics, learning, etc.
13. Staff perceptions of staff development activities -- the degree of focus on teaching and learning.
14. Staff perceptions of the extent to which student learning is important to teacher evaluation.
15. Staff perceptions of the degree to which noninstructional matters interfere with teaching and learning.

PROFESSIONAL CONDITIONS FOR TEACHING: CULTURE

1. Staff perception of the extent of professional consensus on school goals.
2. Staff perception of the extent to which the staff participates together in activities toward the attainment of school goals.
3. Staff perception of the level of staff involvement with school improvement efforts.
4. Staff perceptions of the extent to which teachers plan, teach, and/or evaluate collaboratively.
5. Staff perceptions of the extent to which they share in school-wide decisionmaking.
6. Staff perceptions of the principal's commitment to matters to curriculum and instruction.
7. Staff perceptions of the autonomy and flexibility they are provided in implementing curriculum and instruction.
8. Staff perceptions of the level of support for professional risk-taking and experimentation.
9. Staff perception of the degree of commitment among the teaching and administrative staff.
10. Staff perception of the degree to which the staff believes they are capable of achieving the schools' goals.

¹In the context of this paper, these are the alterable variables.

SOURCE: Jeannie Oakes, "Conceptual and Measurement Issues in the Construction of School Quality Indicators", Paper prepared for AERA Symposium, Washington, D.C., April 1987, Pages 33-34 and 35-36.

FIGURE 5

PROPOSED FORMAT FOR STRUCTURED INTERVIEW PROTOCOL

CONNECTICUT TEACHER ASSESSMENT PROGRAM

I. UNIT STRUCTURE: DISCUSSING UNIT TOPICS PRINTED ON CARDS

- o Demonstrate knowledge of an appropriate sequential order (including prerequisite knowledge)
- o Recognize variations of teaching order
- o Recognize possible omissions or additions needed for unit's coherency and/or completeness
- o Demonstrate ability to judge level of difficulty (appropriateness for learner)

II. LESSON: DISCUSSING RESOURCE MATERIALS

- o Demonstrate ability to recognize teaching objectives and appropriateness of materials for those objectives
- o Demonstrate ability to correctly sequence steps of a lesson
- o Demonstrate ability to evaluate and adapt lesson (students' comprehension, remedial strategies, and extension strategies)

III. ALTERNATIVE APPROACHES: DISCUSSING VARIOUS METHODS OF PRESENTATION

- o Demonstrate ability to assess strengths and weaknesses of teaching strategies

IV. EVALUATING STUDENT PERFORMANCE: DISCUSSING STUDENT RESPONSES

- o Demonstrate ability to assess students' work
- o Demonstrate knowledge of methods of remediation
- o Demonstrate ability to adapt a lesson based on students' needs

SOURCE: Connecticut State Board of Education (1986). Proposed Format for Structured Interview Protocol. Connecticut Teacher Licensure Assessment Program, Hartford, CT. November 1986.