DOCUMENT RESUME

ED 109 143 TM 004 525

AUTHOR Cohen, Monroe D., Ed.
TITLE Testing and Evaluation: New Views.
INSTITUTION Association for Childhood Education International, Washington, D.C.
PUB DATE 75
NOTE 68p.
AVAILABLE FROM Association for Childhood Education International, 3615 Wisconsin Ave., N.W., Washington, D.C. 20016 ($2.50)

EDRS PRICE MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS Academic Achievement; Classroom Environment; *Educational Assessment; Educational Environment; Educational Improvement; Elementary Education; *Evaluation Methods; Intellectual Development; Intelligence Tests; Learning; Participant Satisfaction; *Student Evaluation; Student Motivation; Student Teacher Relationship; Teacher Attitudes; Teacher Role; Test Construction; *Testing; Test Reliability; *Test Validity

ABSTRACT

The methods and procedures of education in our schools must be reexamined according to the writers who have contributed to this document. This is most strongly felt in the areas of testing and evaluation for it is these areas that may have the greatest impact on a child's educational career. This document is presented in three sections: (1) An Overview, (2) Testing Problems and Possibilities, and (3) Some Examples of Meaniful Evaluation. Each author has set forth what they see as inadequacies in the evaluation and testing procedures of our educational systems. The role of student and teacher are considered throughout the document. (DEP)

Testing and Evaluation:

ASSOCIATION FOR CHILDHOOD EDUCATION INTERNATIONAL

Testing and Evaluation: New Views

Association for Childhood Education International
3615 Wisconsin Avenue, N W
Washington, D C 20016

Vito Perrone, Coordinator
Monroe D Cohen, Editor
Lucy Prete Martin, Assistant Editor

3615 Wisconsin Avenue, N W , Washington, D C 20016

Library of Congress Cataloging in Publication Data
Main entry under title
Testing and evaluation
Bibliography p 62 .
1 Educational tests and measurements. 2 Reading — Ability testing I Perrone, Vito II Cohen, Monroe D , 1924- ed III Martin, Lucy Prete, ed IV Association for Childhood Education International
LB3051 T44 371 2'6 74-34211

ISBN 0-87173-000-6

1974-76 Biennial Bulletin Order

Assistance from the Rockefeller Brothers Fund in the planning and development of this publication is gratefully acknowledged

Distributed jointly by the Association for Childhood Education International, Washington, D C , and Citation Press, New York

Design by McGurk & Associates

Contents

PART I Overview

PART II Testing Problems and Possibilities

PART III Some Examples of Meaningful Evaluation

Part I
Overview

Introduction: A Time for Rethinking

4.

In November 1972 educators from several parts of the United States met at the University of North Dakota, to discuss some common concerns about the narrow accountability ethos that had begun to dominate schools and to share what many believed to be more sensible means of both documenting and assessing children's learning. Subsequent meetings, much sharing of evaluation information, and financial and moral support from the Rockefeller Brothers Fund have all contributed to keeping together what is now called the North Dakota Study Group on Evaluation. A major goal of the Study Group, beyond support for individual participants and programs, is to provide materials for teachers, parents, school administrators and governmental decision-makers (within State Education Agencies and the U S Office of Education) that might encourage reexamination of a range of evaluation issues and perspectives about schools and schooling. *Testing and Evaluation. New Views* represents one effort in this direction.[1]

What will be clear to readers of this bulletin is that the individual writers are deeply involved in schools. This accounts for their intensity. They view evaluation/testing issues as particularly important because they know these issues seriously affect children, parents and teachers, oftentimes adversely. All our writers can provide many examples of children who, as a result of standardized testing programs, were labeled negatively as learners. From firsthand experience our writers know of classrooms where normative testing and other narrow evaluation measures limited the quality of human relationships, where teachers felt forced to effect a segmented curriculum with children engaged much of the day completing paper-pencil "skill" worksheets in order to "improve" test scores, where the potential for children's growth as learners was minimal because so much time was spent on what was most measurable, not on what was most meaningful. It will also be clear that the writers believe strongly in evaluation, viewing it as a critical process for growth, not only for individual children and teachers but for educational programs as a whole.

The contributors share an open-classroom orientation — interactive rather than behaviorist — toward education. They believe that learning is a personal matter and has an intentional quality, that it varies for different children, that it proceeds best when children are actively engaged, that it takes place in a variety of environments in and out of the school and is enhanced in a supportive setting where children are taken seriously They favor an integrated curriculum and an active decision-making role for teachers[1]

The bulletin is organized into three sections (1) An Overview, (2) Testing Problems and Possibilities, and (3) Some Examples of Meaningful Evaluation In Part One, **Anne Bussis, Edward Chittenden** and **Marianne Amarel** set forth what they conceive as the inadequacies of experimental evaluation procedures for programs concerned with "considerations of process, content and context." Experimental procedures that stress standardization of treatment, behavioral outcomes and quantitative data analysis tend to dominate educational evaluation. By over-reliance on them we may, as James MacDonald has noted, "reduce all significant school related behavior to performative acts. In the process, we will say in effect that what takes place internally is either illusory or irrelevant to our concern "[2]

There is an alternative with a tradition predating the experimental mode and, from my perspective, more appropriate to many of the directions being established in open, child-centered educational programs Bussis, Chittenden and Amarel describe that alternative

In Part Two, the focus is on testing Although we would not want readers to believe that *evaluation means testing*, we realize that such a view is growing (How does one evaluate the progress of a first-grade classroom? Give the Metropolitan Achievement Test, of course!) **Susan Stodolsky** provides a balanced overview of "What Tests Do and Don't Do " While she does not argue that testing has no place in schools, she does suggest that "as we move from teacher-centered to child-centered classrooms, from group instruction to individualized instruction, from a fixed to a more fluid curriculum, the whole enterprise of testing must be reoriented and reassessed."

5.

Michael Patton extends Stodolsky's discussion by providing a guide to the statistical nomenclature accompanying testing Does everyone using a standardized test understand how "grade-level equivalency scores" are derived? Or the meaning of "reliability" and "validity"? (This past summer, in a graduate seminar relating to evaluation methodology, an experienced school principal, after reading a statement that "grade-level is simply the middle score — half must always be above and half below— asked, "Is that really true?") Patton notes that the testing industry is booming in America "Production of new tests is occurring so rapidly that even specialists appear to be overwhelmed." But how appropriate are the products? Are the tests free of serious error, bias and invalidity? Do they provide better information than teachers can gain through personal observation and interaction with children?

How would test-makers respond to the foregoing? I suspect they would agree that there *are* some problems, that there is room for improvement As **George Hein** indicates, "The questions could be better, the standardization could be based upon more representative samples of the population, the tests could be validated against criteria more appropriate than the ones used More imaginative use of the available technology could vastly improve even paper-and-pencil machine-graded examinations. A much broader range of activities could be standardized." But Hein also argues that "Reform Is Not Enough " He asks us to examine the politics of testing, the role testing plays in "sort[ing] and classify[ing] children for their assigned roles in society " Thomas Cottle, who is with the Children's Defense League, points out in *Social Policy* that "tracking," one outcome of testing in many

[1] The annotated bibliography prepared by Brenda Engel (pp 62-64) includes material written by members of the Study Group Early in 1975 a series of additional evaluation monographs will be available Selected titles are *Observation and Description An Alternative Methodology for the Investigation of Human Phenomena* by Patricia Carini, *Alternative Evaluation Research Paradigms* by Michael Patton, *An Open Education Perspective on Evaluation* by George Hein, and *A Handbook on Documentation* by Brenda Engel For further information about the series, write to Vito Perrone, Center for Teaching & Learning, University of North Dakota, Grand Forks, ND 58201

[2] es B MacDonald, "An Evaluation of Evaluation," *The Urban Review* 7 (1974) 9

school districts, is "masked by rational educational theory its political implications overlooked by some who think of it merely as an inevitable consequence of human differences "[3]

Deborah Meier, in a question that relates closely to some of those raised by Patton and Hein, asks, "What do we mean by reading competence?" To observe in classrooms the mass of skill sheets aimed at auditory discrimination, blending and syllabication, among others, one would have to conclude that reading is a skill (Meier suggests "a trick") that, for the most part, is learned in isolation. As Meier documents, little evidence is available to show that such activities will improve a child's comprehension — the capacity for turning the written page into something that makes sense Meier argues convincingly that standardized reading tests contribute heavily to the concentration on the skill sheets, tending in the process to distort the meaning of reading Because she feels that evaluation is important, Meier outlines some alternative means of assessing reading

Lois Barclay Murphy provides a synthesis to Part Two by bringing to readers' attention some of the research relating to a range of important school issues, for example, teacher judgment, children's intellectual growth and social development, and motivation The close of Murphy's sensitive statement provides an excellent transition to the final section of this bulletin, "Some Examples of Meaningful Evaluation," where *qualitative* data is a basic concern of the writers

Patricia Carini opens Part Three by discussing the documentation activity at the Prospect School of Bennington, Vermont "Taking Account of Process" is a concept and a practice more of us ought to learn It suggests a careful, systematic, documentation which assumes that learning cannot be "recorded and assessed as isolated elements independently of the meaning for the learner" In order to capture this broader view of learning, the processes underlying children's directions "must be observed over *time* to determine a pattern, a matrix of descriptions of the learner's involvement" Carini's examples, taken from the Prospect School's documentation, are explicit and illuminating In introducing some of Carini's methodology to teachers, I have often had the response, "You can't really expect us to do all of that!" I am convinced that significant evaluation that "takes account" of children's growth and the "standards that exist in a school" demands a level of intensity not yet characteristic of most schools

The way the Marcy Open School, Minneapolis, Minnesota, uses evaluation is outlined next by **Ruth Anne Aldrich**. Participants at Marcy raised the questions, "For what must a school be held accountable?" and "How can evaluation provide us with the information we need to develop an increasingly responsible program?" Attempting to respond to such questions has brought about a particular style of evaluation and a conclusion that "schools should be growing, evolving institutions aware of their successes and designing change for their failures" How many schools have attempted to organize such a process of internal evaluation? My experience suggests that it happens rarely Perhaps this lack accounts, in part, for the narrow evaluation patterns that exist

What could we learn if we listened to children? **Nancy Miller** describes her work with the Children's Interview which focuses on children's roles in the classroom and their contribution to their own learning, children's perceptions of the teacher's role and their relationship to the teacher, the contribution of classroom peer interaction to children's learning, and the children's view of the classroom as an overall learning environment and the ways in which they relate to that environment Miller, and others of us who have worked with the Children's Interview, have found it a powerful process for taking account of children and *their* learning

Another source for qualitative data on what is happening in a school or class-

[3] What Tracking Did to Ollie Taylor, *Social Policy* (July/August 1974) 21 Readers might also note that the NAACP and the Mexican American Defense League are engaged in several litigations in California that relate to the social effects of testing

room is the teacher Anne Bussis and **Edward Chittenden's** summary of their intensive Teacher Interview Study is, in large measure, an extension of their methodology article which is included in Part One "In what ways do teachers think about teaching? How do they conceive of the complex pattern of events that mark the school day? What assumptions do they hold about learning and development? What are the grounds for their planning, provisioning and evaluation?" Taking account of the teacher's "personal construct" is a process that is rarely a part of what schools view as evaluation But not to take account of such a process is to fail to understand the central role of the teacher in most classroom settings

We close with a highly selective annotated bibliography prepared by **Brenda Engel**. Those of us who have been struggling with the issues that dominate this bulletin are aware that much of the important literature is in unpublished form and hence not readily available Engel has included only accessible materials with the potential for extending our discussions

At the outset of this introduction, I commented that one of our major purposes is to encourage a rethinking of a range of evaluation issues We hope that this bulletin is helpful to such a process We invite your reaction

Vito Perrone
December 1974

7.

Alternative Ways in Educational Evaluation

Anne M. Bussis, Edward A. Chittenden, and Marianne Amarel

Elementary educators are caught up in debate over assumptions about ways that children learn best, about the teacher's role in curricular and instructional decision-making, and about the networks of human relationships that constitute a school Although the development of approaches for studying and evaluating educational programs is an old problem, these current debates have served to highlight the inadequacies of existing ways of categorizing and thinking about educational variation. Our purpose here is to describe problems in educational evaluation as we have come to identify them in the course of our studies with teachers, children and schools — and to suggest some alternative directions

The Problem of Educational Models

A look at the total field of elementary education suggests that the present degree of experimentation and variety among programs and approaches is greater than in the past Much of this experimentation has involved only surface change of a somewhat gimmicky nature, but change in some schools appears much more substantial, involving the development of new understandings regarding learning and teaching The literature about change in early education, however, has tended to blur rather than clarify the issues underlying these variations Part of the difficulty is that the literature has cast variation predominately into the language of "models" The model concept would seem an efficient way of describing the basic components of different programs, but in reality it has proved to be something of a trap for educator and researcher alike.

Educational "models" have two clear qualities First, they contain accounts of, or prescriptions for, methods—for ways of doing the model Second, they contain statements about intended outcomes—objectives to be accomplished Frequently both methods and outcomes are stated in concrete behavioral terms Models differ extensively in the kinds of materials and methods they prescribe, in the settings for which they are intended, in their degree of prescriptiveness, and in the scope of their ambition—i e , some models are for instruction in selected areas only, whereas others are for instruction more broadly conceived Some early education models are designed along behavior modification lines Some are based on child development theory (e g , a Piaget curriculum), some are models of components of the British "integrated day"

Inasmuch as model thinking is very much a part of most research and evaluation schemes, at the very outset the teachers participating in such evaluation have to contend with a set of rules that may not have been clearly explicated For example, they soon may be asked to specify exact teaching methods and desired student outcomes, or otherwise find themselves responding to a request to delineate the criteria of a "model classroom" To the extent that the teachers go along with, or cannot escape, the pressure to operate in this manner, they do indeed try to implement the model—whatever they perceive it to be They may try to dispense reinforcements in order to "modify behavior," run twenty-five tutorials in order to "individualize instruction," or set up interest corners in order to have an "open classroom" The defining criteria of methods and outcomes they have previously specified then become the yardstick of success—the yardstick by which they and the evaluator measure their efforts

Standards of Quality

In contrast to this specification of "method" and "outcome" that is associated with a "model" approach in evaluation, many educators today are emphasizing the development of standards of *quality* in learning considerations of process, content and context Such standards represent educational and psychological constructs more than behavioral criteria They are a frame of reference from which the teacher works and evaluates what has been accomplished, but they are not prescriptions of methods "to be followed" or outcomes "to be obtained" Standards, in the sense of constructs or frames of reference, are neither directly "followable" or "obtainable," since a construct (by definition) is of a more abstract nature It is a principle, an understanding derived through intellectual synthesis, that underlies the teacher's consideration of any particular procedure or learning Educators working within such a framework have as an objective the creation of conditions that promote quality learning, but they do not and cannot have strictly prescribed methods for achieving that objective—and they do not and will not have a limited, narrow set of behaviors in mind as the only guide for judging children's progress in learning Another difference worth noting is that standards cannot be stated exhaustively beforehand By their very nature, standards become better articulated over time and with experience, but they necessarily remain "open constructs," capable of absorbing new and unpredicted examples within their definition

Toward Clarifying the Practitioner's Frame of Reference

Research on educational models (e g , Head Start, Follow Through) suggests that teacher differences *within* educational programs is as great as or greater than variations *between* programs Although the problem has been recognized, the apparent solution that is attempted in conducting evaluation is to try to obtain greater clarity of program descriptions and to define criteria in the hope that teacher variability (within the given model) can be greatly reduced

The problem as we see it, however, is not so much one of trying to define precisely any particular educational *program*, but of defining those characteristics that reflect program variations at the level of teacher understandings and perceptions The following quotation from a study of preschool programs illustrates what we mean (DiLorenzo et al , 1969)

> *it was the naive assumption of the research staff responsible for the design of the study that prekindergarten programs for the disadvantaged existed in packages, to be picked off the shelves in the educational market place. Once the districts had made their choices, the program treatments would be inserted into the design Distinct programs did not exist Points of view . . did, and they determined the type of program which evolved.*

Analysis of the data from our teacher interview study thus far indicates that programs and classrooms that superficially seem to share common elements may be very different in intent and emphasis To distinguish between them, one needs to identify the basic or "prototypic" notions that predominate as the teacher's referents for instructional activity What are his or her learning priorities—beliefs about what children should learn to "care" about? What are the teacher's organisms—are they seen as constructors of reality or receptacles for knowledge, or some combination of both? What are his assumptions about the organization of knowledge and how it is best learned? What are perceptions about the use and potential value of material resources in the classroom? The answers to these and similar questions determine the nature of the construct systems that guide teaching behavior [1] Tentatively, at least, we have identified what seem to be quite different construct systems—all from a population of teachers working under the heading of "open education"

Although the classrooms of two different teachers might look rather alike at a given point in time, the notion of personal construct systems implies that they may be headed in quite different directions and that the teachers may expect different kinds of environmental support Such a conceptualization—one dealing with the practitioner's frame of reference—has important implications for evaluation and research, as well as practical value in helping to sort out the activities now going on under numerous labels

[1] See pp 58-61 of this publication

Toward Clarifying the Research/Evaluation Frame of Reference

Substituting the practitioner's frame of reference in place of precise specifications for educational programs is one step in a different direction, but it is not sufficient in and of itself We also need to look at the research/evaluation model

One basic problem with the conventional research/evaluation approach is that "student outcomes" are generally defined by behavioral criteria and are therefore statements about learning necessarily lifted out of the context of the total activity as it actually occurs in the classroom. Thus, an example, the outcome-statement "works independently on a project" does not take into account the purpose or quality of the particular project in question, whether the project is more appropriately a group or an individual endeavor, and so on This difficulty is the same one mentioned previously in our discussion of "model thinking," where we contended that some educators are now emphasizing standards of quality in learning Thus, they evaluate teaching in terms of educational and psychological constructs and not in terms of out-of-context behavioral criteria

To illustrate the idea of standards a bit more, standards of quality in the process of learning would include such factors as originality of work (the notion of "authorship"), purposeful effort, independence of effort A concern with quality in the *content* of learning would include consideration of what children produce (e g., writing, drawing), evidence that instruction deals with "powerful"[2] concepts (e g , graphing) as well as necessary skills (addition/subtraction) Standards relating to the *context* of learning center on desired qualities in the nature of human relationships (child/child as well as child/adult)—openness and honesty of encounters, respect for the efforts and feelings of others However standards are thought about and described, the important point is that different kinds be considered and applied in evaluating a particular learning situation

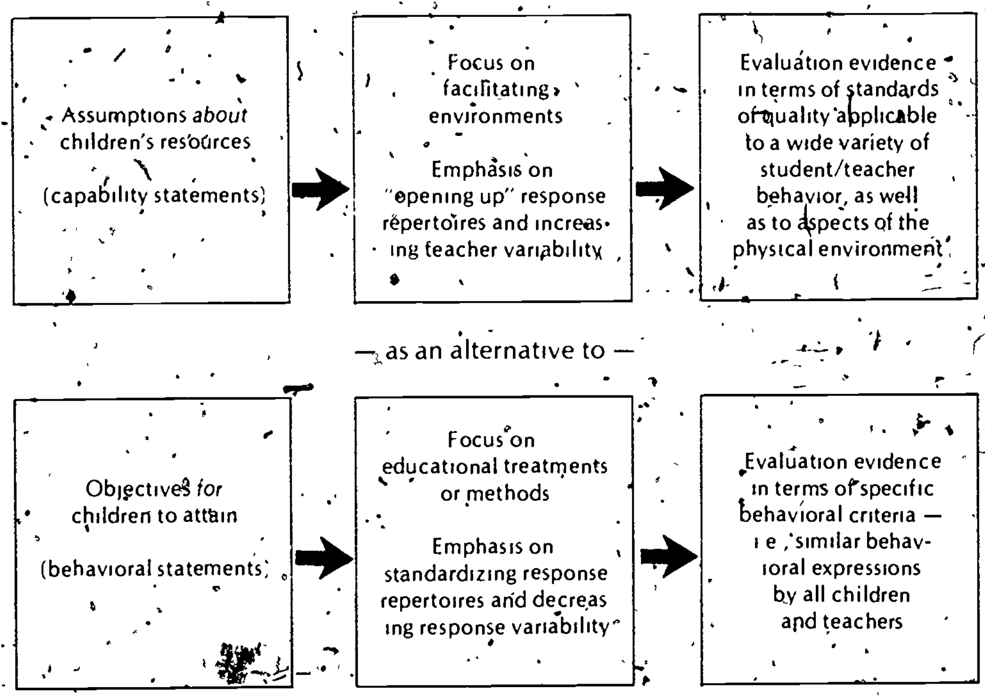
These standards that provide an evaluative framework for many teachers could, with clarification, serve to broaden the general approach to formal evaluation, because they suggest relevant evaluation data other than student behavior taken out of context

Another problem with the conventional research/evaluation approach is the implicit assumption that the educational treatment causes or produces the objective or outcome Rather than stating objectives for children that may be attained as a result of certain methods or treatments, however, many educators prefer to state assumptions about children's capabilities that may be realized in certain facilitating environments They assume, for example, that all children are capable of displaying intelligent effort, responsibility, concern for others, respect for self—given an instructional environment that elicits and supports such behavior Such capacities are not thought of as instilled or caused by a preconceived educational treatment but rather as drawn out and encouraged by a responsive and flexible educational program

The difference between stating objectives *for* children and assumptions *about* children's capabilities and resources may seem minor at first, but it has far-reaching implications. It is a difference that leads to (a) a concern with environments rather than treatments, (b) an emphasis on response variability among teachers rather than response uniformity, and (c) a focus on standards of quality in learning rather than behavioral criteria outside the context of purposeful action If research is to accommodate these priorities now being held by many educators, an

[2] See Flavell's (1970) discussion of "power" as a dimension of concepts

overhauling of our basic paradigm seems called for. As a step in this direction, we suggest the following:


Assumptions about children's resources
(capability statements)
Focus on facilitating environments
Emphasis on "opening up" response repertoires and increasing teacher variability
Evaluation evidence in terms of standards of quality applicable to a wide variety of student/teacher behavior, as well as to aspects of the physical environment
— as an alternative to —
Objectives for children to attain
(behavioral statements)
Focus on educational treatments or methods
Emphasis on standardizing response repertoires and decreasing response variability
Evaluation evidence in terms of specific behavioral criteria — i e , similar behavioral expressions by all children and teachers

Research and evaluation along the lines suggested in the diagram can draw upon a tradition within psychology that has emphasized the study of inner states such as belief systems, attitudes and understandings. In the United States this tradition is represented in the writings of Kelly (1955), Snygg and Combs (1949), as well as others With a few notable exceptions, this "phenomenological" tradition has not been recognized within educational research Instead the field has been dominated by the conventions of testing and measurement and by behavioral psychology. At the level of instrumentation, new approaches that fit with the phenomenological tradition could well include interviews, documentation of environments through observation, the systematic collection of work and language samples.

It is one thing to analyze issues and point to directions that might alleviate some problems It is obviously quite another matter to translate ideas into reality Nonetheless, we feel fairly confident that advances can be made Many people have already made progress in devising more appropriate ways of assessing children, teachers and educational environments The problems and questions raised here are complex But if they are not addressed, we face the real possibility that a good deal of substantial progress in educational thinking and practice will go down the drain—because it is judged to be "not very effective" on the basis of inappropriate criteria

"Objectivity" and Decision-Making

A final problem in much of the current thinking about educational evaluation is that it is assumed to be an "objective technology." That behavioral science operates in an objective (in the sense of value-free) fashion or that evaluation leads objective (value-free) decision-making are myths that have been too long with us and far too widely perpetuated The latter myth is particularly destructive to the degree that people in education actually believe it, which many apparently do Decision-making is invariably a subjective, human activity involving value judgments (or weights) placed on whatever evidence is available to the decision-maker Depending on the extent to which parties to a decision agree that the available evidence has been impartially gathered and represents "important" information, people may or may not agree on the meaning of the evidence Even when there is virtual consensus on "the facts of the matter," such facts do not automatically lead to decisions regarding future action People make decisions, information does not

Biologist Rene Dubos, in describing the diverse reaction of fellow scientists to his book *Only One Earth The Care and Maintenance of a Small Planet*, provides an instructive example of the human reality of decision-making (1972, p 508)

> *Starting from the same set of scientific facts, the experts arrived at a multiplicity of conflicting conclusions with regard to the practical policies concerning the environment — policies, for example, about nuclear energy, pesticides, further industrialization of the world, et cetera. Their conflicts originated not from differences in knowledge or interpretation of facts, but from differences in the value judgments they put on these facts In this regard, experts display as much diversity as nations and individual persons, they differ not only in their approach to social and human goals, but even more in the selection of these goals.*

Although it is understandable that the term "evaluation" might gradually come to be applied to the activity of gathering information prior to decision-making, it is not at all clear why the human activity of actually "evaluating" the information has been so left out of the publicized picture. If the values that dictate educational decisions remain unexplicated — if, by default, they are the implicit values built into the information-gathering instruments — then we are indeed settling for more or less "impersonal decisions", but they are hardly "objective decisions." Perhaps we should use terminology such as *assessment* and *analysis* for information-gathering activity — and reserve the term *evaluation* specifically for decisions made about the information

Concluding Remarks

In summary, we have proposed that the assumptions inherent in the notions of educational treatment and behavioral outcomes are basic issues that need to be readdressed along with problems of instrumentation and data interpretation. While alternative models of educational evaluation do not as yet exist on any broadly accepted basis, the range of admissible techniques and strategies has broadened, and in some places, parents, school board members, administrators, and state and federal officials have supported alternative forms. Hence the need for the kinds of accounts described later in this publication — examples that can be looked to for guidance,

References

Di Lorenzo, L T , R T Salter, & J J Brady *Prekindergarten Programs for Educationally Disadvantaged Children* Washington, DC U S Office of Education, Department of Health, Education and Welfare, 1969

Dubos, R "The Despairing Optimist " *The American Scholar* 41, 4 (Autumn 1972) 508-12 Copyright © 1972 by the United Chapters of Phi Beta Kappa By permission of the publishers

Flavell, J H, "Concept Development " In *Carmichaels's Manual of Child Psychology* (3d ed), Vol I, Paul H Mussen, ed New York John Wiley, 1970

Kelly, G *The Psychology of Personal Constructs*, Vol 1 New York W W Norton, 1955

Snygg, D , & A W Combs, *Individual Behavior* (Rev Ed) New York Harper & Brothers, 1959

Part II Testing: Problems and Possibilities

What Tests Do and Don't Do

13.

Susan Silverman Stodolsky

Tests come in many shapes and sizes Test constructors have produced instruments for use in measuring a wide array of human characteristics (Buros, 1971, Johnson and Bommarito, 1971) Most tests children take while in school are teacher-made, that is, designed by their own teachers. Others are provided by textbook publishers In addition, a child in elementary school may be administered a group intelligence test, possibly some aptitude or interest measures, and a number of standardized achievement tests As George Weber notes in his pamphlet, *Uses and Abuses of Standardized Testing in the Schools* (1974).

> *Some standardized tests do not do a good job of what they claim to do, and for some testing purposes non-standardized tests are more appropriate or more efficient. But standardized tests are used by all our public schools. Important and even critical conclusions and decisions are made on the basis of their results. For example, on the basis of their results the public is told that reading achievement is going up or going down, an experimental program is deemed successful or unsuccessful, a child is placed in this or that class, and students gain admission to a particular college or fail to do so* (pp. 1-2).

Standardized test scores have been shown to play a crucial and often unwarranted role in determining further schooling and teacher attitudes and in affecting pupils' self-concepts For most of us, testing has become an expected if not accepted part of schooling

Teachers, administrators and parents ought to determine the appropriate roles for testing and judge the utility of testing in fostering the healthy growth and development of children In this paper, I will try to present an overview of the field of testing, including ways the field itself is trying to change, I will discuss testing in terms of (1) purposes for giving tests, (2) effectiveness of different types of tests in providing needed information for a given purpose, (3) determination of who is benefited by the test results or testing experience, (4) relationship between the instructional process and testing procedures, and (5) kinds of skills, learning and growth to be assessed. These issues seem to me to be central if a teacher wants to make an informed decision about a given testing procedure. They are also lively issues in the field of testing.

I hope, then, to answer questions like the following. What are the various purposes for which tests have been or can be constructed? What kinds of tests are or could be available? What kinds of information can tests provide? What are the major values and major limitations of testing?

Author is grateful to Mark Martin for assisting in searching the literature and for helpful suggestions

For purposes of this paper, when I refer to a test or testing, I mean *a systematic and deliberate way of sampling a student's behavior or thinking.* Ordinarily we think of a test as a paper-and-pencil device, but many other kinds of evidence-gathering procedures are available to the teacher and researcher

To simplify discussion of the purposes of testing and types of tests, I will restrict my attention to evaluation of academic achievement [1]

Recently a number of important distinctions have been made regarding possible functions of testing and ways tests are constructed, scored and interpreted. Historically tests have been administered mainly at the *beginning* of some learning experience (purpose prediction or selection) or at the *end* of a learning experience (purpose grading or classification)

There are situations in which using tests for prediction, selection and classification are justified It is important to recognize, however, that such usages assume that the success or failure of a child in school is a function of the child's initial characteristics, and that the educational environment or program is virtually fixed When standardized tests have been successful in predicting further achievement of students, instruction in the schools has usually persisted in sorting and ranking students in much the way they would be ranked on standardized tests. Even so, standardized tests are usually best at predicting future performance on similar instruments, *not* in predicting success in the life activities that might be associated with the area of achievement measured.

Ordinarily the purposes of prediction, selection and classification are best served by *norm-referenced* tests which *rank order* individuals or groups Michael Patton's article which follows describes the meaning (as well as some of the problems) of norm-referenced tests. Here, it is important to keep in mind that standardized achievement tests of this type are constructed by attempting to sample some domain of subject-matter content and learning processes that represent the objectives of a curriculum or group of curricula in use in the schools. Norm-referenced tests can reliably estimate the relative standing of a child with respect to the area measured By the very nature of the way they are constructed, however, these tests *cannot* provide the child, his teachers or parents information about what he has *specifically* learned or not learned in a given subject matter over a given period of time Since standardized tests are only samplings of course material, scores cannot be used to determine explicit instructional needs of children in any but the most global way.[2]

Formative and Summative Testing

Closely related to the purposes served by standardized achievement tests is the idea of *summative testing,* which is concerned with product measurement. A summative test is constructed in a manner similar to a standardized achievement test in that the questions are sampled from the course objectives and contents. Since the purpose of a summative test is usually grading, these instruments are not designed to provide detailed feedback to the student. Summative tests ask, "How

[1] For useful discussions of intelligence tests see Anastasi, 1961, and Kaye, 1973

[2] Standardized achievement tests are often used in comparing one educational program with another Norm referenced tests have been used as key measures in most large scale evaluations of intervention programs such as Head Start, Follow Through and Title I The use of standardized tests for comparative evaluation studies may be appropriate if the programs studied are all trying to achieve the objectives measured in the test, but this is not often the case (Stodolsky, 1972) Usually when norm referenced tests are used in comparative evaluations, they are better articulated with some programs than with others Also children in certain educational programs have more familiarity with test-like situations and are more able and willing to produce on demand (Shapiro 1973, Chittenden and Bussis, 1972) At best, results from the administration of standardized achievement tests give us a heavily confounded estimate of the actual academic achievement of children in different programs

well has the student mastered the material and processes involved in the learning units he has just studied?" In many educational contexts it is still believed desirable to inform a student about his progress relative to the expectations of a course Final exams constructed by teachers are the most common variety of summative tests.

The term *formative testing* has two uses in the literature, both concerned with providing feedback about the learning process In one usage (Scriven, 1967), it is testing to gather evidence while a curriculum is being developed, so that curriculum writers can improve materials and procedures In the other usage (Bloom, Hastings and Madaus, 1971), formative testing is used as students go through learning units in order to provide feedback to students and teachers about their progress and to suggest areas in which additional learning and practice are necessary Formative tests are part of a trend to use tests to provide meaningful feedback about student learning and growth

As defined by Bloom, formative tests are often used in conjunction with a mastery-learning strategy in which it is expected that virtually all students can master the unit being studied with sufficient time and learning aids (Block, 1971) In this context, students take formative tests when they have completed initial study of a learning unit. Since formative testing is an integral part of the instructional process, these tests are very different from norm-referenced tests. Formative tests cover a relatively narrow range of topics and need to provide sufficient information so that a student can determine future steps in learning based on the test results For example, a formative test dealing with a unit on learning long division would contain a number of items that would pinpoint the steps students had and had not mastered, whereas a summative test would not include such items but only long division problems

The use of formative testing for student feedback does not depend on adopting a mastery-learning strategy It does require a decision on the part of the teacher to use tests as *instructional tools* or *aids*. Similarly, when formative tests are used in the curriculum development process, their chief purpose is to provide detailed feedback about the curriculum and its effectiveness so that weaknesses can be improved

Criterion-Referenced Testing

Another recent distinction is that between *criterion-referenced* and *norm-referenced tests*. We have already discussed norm-referenced tests, whose major purpose is to allow one to interpret scores with respect to the relative standing of an individual or group. Criterion-referenced tests have been developed in order to provide information about student performance that has been difficult to obtain from standardized tests. "A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards" (Glaser and Nitko, 1971, p. 653).

Scores from criterion-referenced tests should be directly interpretable in terms of actual student behaviors and abilities. Either the items or tasks are precisely those one is interested in assessing, or the items on the test have been shown to directly relate to the behavior of interest. For example, on a criterion-referenced reading test one would be able to relate performance on a given set of items to the child's ability to read and comprehend passages of established difficulty or specific books. Such an interpretation would differ from that in a standardized test situation where the specific skills and abilities of a child scoring at a given grade level could not be well defined Criterion-referenced tests can be used for certifying performance (e.g., a lifesaving test), for direct instructional feedback or for summative purposes.

Diagnostic Testing

A last type of test that has relevance to teachers is *diagnostic testing* Typically, diagnostic testing attempts to provide an assessment of a student's present strengths and weaknesses with regard to a given area of achievement Diagnostic tests may be very similar, in construction and intent to formative and/or criterion-referenced tests However, an additional factor involved in diagnostic testing may be the desire to find the cause or etiology of difficulties in learning. For example, a child having difficulty learning to read might be assessed with regard to vision, hearing and other perceptual functions. Possible motivational or emotional factors associated with reading or learning might also be explored In diagnostic testing it may be necessary to go beyond course content in order to determine the best ways to begin remedying some difficulty in learning Sometimes direct instruction on prerequisite learning is quite sufficient, but at other times additional efforts are necessary

In my opinion the development of criterion-referenced, formative and diagnostic tests hold promise for teachers and students I would expect fewer dangers and limitations with such tests than one would expect with standardized achievement tests These newer forms of tests are meant primarily as aids in the instructional process and can be related much more closely to the actual classroom learning of the child than are standardized tests There are difficulties in constructing such tests, however, and we need more experience before we can feel confident that their promise as instructional aids will be fulfilled.

16.

Tests and the Instructional Setting

In considering the purposes for which tests might be used, a central issue is the relation between testing and the instructional program. Most currently available tests are geared toward instructional settings in which the curriculum is specified and in which each child is expected to master the same materials and objectives as his classmates. As we move from teacher-centered to child-centered classrooms, from group instruction to individualized instruction, from a fixed to a more fluid curriculum, the whole enterprise of testing must be reoriented and reassessed.

It is possible that the entire role of testing as we usually conceive it has little place in more informal educational environments. I believe that persons involved in changing educational environments must be able to specify the areas of human learning, growth and development they believe to be important All educators have an obligation to systematically document growth and change in children Informal educators should be pressuring those in the field of testing and in research positions to develop ways of assessing the important aspects of growth and development that informal education tries to foster I do not pretend that the task of assessing the goals of open education will be easily solved, the behaviors and outcomes are very difficult to measure and may not be open to classic psychometric approaches Nevertheless, I believe test publishers and evaluators can be responsive to educational change and might well have the resources to begin to solve some of the difficult measurement problems involved.

Undoubtedly, to specify and assess pupil growth in informal settings will be more difficult than in traditional schools. One major problem is how to deal with the fact that informal education fosters a diverse set of outcomes and activities (Karlson and Stodolsky, 1973). Nevertheless, I think the basic notions of formative and criterion-referenced testing could be applied to children in certain areas of an informal curriculum, such as in the development of math and science concepts, the development of reading skills, and with respect to certain general areas of problem-solving and critical thinking. Systematic use of work samples and observational schedules as well as interviews with children are also promising methods for the informal classroom. A recent article by Hawes (1974) offers a collection of many useful evaluation alternatives

More Broadly Conceived Tests Are Needed

The last area I would like to discuss in regard to testing is that of content or domains. Most of this paper has implicitly assumed that the major area of assessment in regard to children is academic, subject-matter achievement As teachers we have often been committed to more than cognitive outcomes for our children, but rarely do we systematically collect evidence about growth outside the cognitive area. As we begin to broaden our entire view of schooling and learning, we should simultaneously attempt to incorporate into our evaluations aspects of growth that are affective, social, emotional, fanciful, creative in nature I agree with others (e g , Chittenden) that one should attempt to evaluate these aspects of functioning in the context of subject matters or areas of activity Thus, rather than taking a trait-like approach, one would work within the concept of the child as a whole person and his activity as incorporating many facets of functioning. Despite the difficulties, we must attempt to assess children's development of interests, their capacity and modes of learning on their own, patterns of social interaction and the like

It seems to me that a continuing dialogue between teachers, researchers and persons involved in test construction could help broaden the options available to us all

From the viewpoint of informal education in particular, we need to alter the traditional situation in which standardized tests are often powerful shapers of curricular content To break out of such a pattern, teachers must be able to articulate their views about growth, to use their observations as illustrations of growth, and to choose consistently when their educational purpose is compatible with the purpose of a particular assessment procedure

References

Anastasi, A 'Psychological Tests Uses and Abuses" *Teachers College Record* 62 (1961) 389-93

Block, J H , Ed. *Mastery Learning Theory and Practice.* New York Holt, Rinehart & Winston, 1971

Bloom, B S , J T Hastings, & G F Madaus *Handbook on Formative and Summative Evaluation of Student Learning* New York McGraw-Hill, 1971

Buros, O K *The Seventh Mental Measurements Year Book* Highland Park, NJ Gryphon Press, 1971

Chittenden, E A , & A M Bussis "Open Education Research and Assessment Strategies " In *Open Education A Sourcebook for Parents and Teachers*, E B Nyquist & G R Hawes, eds New York Bantam Books, 1972 Pp 360-74

Glasser, R , & A J Nitko 'Measurement in Learning and Instruction " In *Educational Measurement*, 2d ED, R L Thorndike, ed Washington, DC American Council on Education, 1971 Pp 625-70

Hawes, G R 'Testing, Evaluation and Accountability Managing Open Education " *Nation's Schools* 93, 6 (1974) 33-47

Johnson, O G , & J W Bommarito *Tests and Measurements in Child Development A Handbook* San Francisco Jossey-Bass, 1971

Karlson, A L , & S S Stodolsky 'Predicting School Outcomes from Observations of Child Behavior in Classrooms " Paper presented at annual AERA meeting, New Orleans, Feb 1973

Kaye, K IQ A Conceptual Deterrent to Revolution In Education " *Elementary School Journal* 74 (1973) 9-23

Scriven, M The Methodology of Evaluation " *AERA Monograph Series on Curriculum Evaluation* 1 (1967) 523-49

Shapiro, E 'Educational Evaluation Rethinking the Criteria of Competence " *School Review* 81 (1973) 523-49

Stodolsky, S S Defining Treatment and Outcome in Early Childhood Education " In *Rethinking Urban Education*, H J Walberg & A T Kopan, eds San Francisco. Jossey-Bass, 1972 Pp 77-94

Weber, G *Uses and Abuses of Standardized Testing in the Schools* Washington, DC Council for Basic Education, *Occasional Paper #22*, 1974 Used by permission

Understanding the Gobble-dy-gook

Michael Quinn Patton[1]

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write. H G Wells

18. Standardized tests now pervade our lives in ways never dreamed possible by early pioneers in educational measurement only a century ago Production of new tests is occurring so rapidly that even specialists appear to be overwhelmed. The dimensions of this explosion are indicated by the fact that in 1972 the Educational Testing Service Collection held 680 different tests in just one category alone—reading *The Seventh Mental Measurements Yearbook* (Buros, 1972) is composed of two fat volumes describing the vast literature on tests and measurement. At the same time testing methodology and theory are becoming increasingly complex

At the 1972 Invitational Conference on Testing Problems, Henry S. Dyer noted these facts and put forth "*Dyer's First Law of Information Dilution,* which states *that, as knowledge expands, while the population of potential users of knowledge also expands, the probability approaches unity that everybody is ignorant of what anyone else knows In other words, the great majority of test users simply does not have the time to look up or catch up or keep up with the enormous number of tests and the mountainous literature that the testmakers continue to pile up.* (Dyer, 1973 91)"

Standardized tests have been developed for almost any cognitive, affective or social human trait you can think of, from intelligence to alienation, self-concept to maturity, moral development to creativity These tests are being used to select people into and out of a wide range of educational programs, private and public projects and a variety of jobs—often without knowledge or understanding of how the tests are being used on the part of those being tested Standardized testing has become a socio-political tool, deciding the fate of both individuals and entire educational programs (cf Manning, 1969, Kirst and Mosher, 1969, Cohen, 1970, McDill, McDill and Sprehe, 1972) Again to quote Dyer (1973:86)

> *The field of education has become strewn with politics, and educational testing has become an instrument, if not a weapon, in the political process. And this means that our worries today about the mishandling of tests and the misuse of test scores must embrace not only school personnel, but also politicans and the diverse and pluralistic constituencies they serve.*

[1] I am indebted to Dean Vito Perrone, Center for Teaching and Learning, University of North Dakota, and participants in the North Dakota Study Group on Open Education meeting in Chicago, November 9-10, 1973, for encouragement and helpful suggestions in writing this article Support from the National Institute of Mental Health in the form of a postdoctoral fellowship in Evaluation Methodology was also helpful Comments by Barbara French, University of Minnesota, on an earlier draft of this paper were particularly useful

This also means that, as H G Wells predicted, individual teachers, parents, students and administrators need to make basic statistical thinking a part of their personal survival kit As Darrell Huff (1954) noted in his primer, *How To Lie with Statistics*, the abusers already know the inadequacies and tricks of statistics, "honest men must learn them in self-defense" (p 9)

The Meaning of Numbers: Interpretation

The first thing to keep in mind when interpreting standardized test scores is that, even at their best *they are only rough indicators* of some human characteristic. Anne Anastasi (1973 xi) has noted the danger of focusing so much on tests and test scores that we lost sight of the actual behaviors that matter to us

> *The widespread misconceptions about the so-called IQ provide a particularly flagrant example of such a dissociation. One still hears the term "IQ" used as though it referred, not to a test score, but to a property of the organism*

In other words, the numbers that come out of standardized tests are not embedded in the genes or on the foreheads of students. They are only rough approximations of some characteristic at a specific point in the time under particular conditions. Test results are only one piece of information about a person or a group—a piece of information that must be interpreted in connection with other information we have about that person or group.

Test scores then are neither good nor bad. They are pieces of information that are subject to considerable error—and that are more or less useful depending on how they are gathered, interpreted, applied, abused and used In this context let us look at some of the more frequent types of scores reported.

Norms

> *Let a parent read, as many have done in such places as Sunday rotogravure sections, that "a child" learns to sit erect at the age of so many months and he thinks at once of his own child. Let his child fail to sit by the specified age and the parent must conclude that his offspring is "retarded" or "subnormal" or something equally invidious. Since half the children are bound to fail to sit by the time mentioned, a good many parents are made unhappy. Of course, speaking mathematically, this unhappiness is balanced by the joy of the other fifty percent of parents in discovering that their children are "advanced" But harm can come of the efforts of the unhappy parents to force their children to conform to the norms and thus be backward no longer . . . . Hardly anyone is normal in any way . . . . Confusing "normal" with "desirable" makes it all the worse.* Darrell Huff (1954 44-5)

"Norms" are scores that provide a comparison for interpreting how one pupil, school or school system compares to some other group Norms provide information about how children from some comparison group actually performed on a particular test, *not how they ought to have performed*. Norms can be reported in several ways — percentile ranks, grade equivalents, stanines, scaled scores, difficulty coefficients and quartile points, among others The important point to keep in mind is that all of these ways of reporting results are based on the same basic information Some are simpler to understand than others, some are more useful for special purposes than others—but *all of them are ways of looking at how an individual or group performed compared to how the norm group performed*. Such "norm-referenced" tests constitute the most common type of standardized t and are the only type we shall discuss here

Norms are used for interpreting tests because raw scores (i e , the total number of correct answers on a test) have very little meaning in themselves Different tests have different numbers of items and are usually designed so that high raw scores are unlikely For example, in the ninth-grade mathematics test, Minnesota High School Achievement Examinations, 36 correct answers out of 70 items places a student in the 99th percentile Or consider the Stanford Achievement Test in reading (1964 edition) where a beginning third-grade child need answer correctly only 33 of 60 items to be exactly at grade level in paragraph meaning For the test results to have meaning, raw scores must be translated and placed in a context, usually involving comparison to some norm group. All major standardized tests include norms for a representative national sample of pupils For many tests, norms are also available for the state, the county, the city, and/or even a specific local school

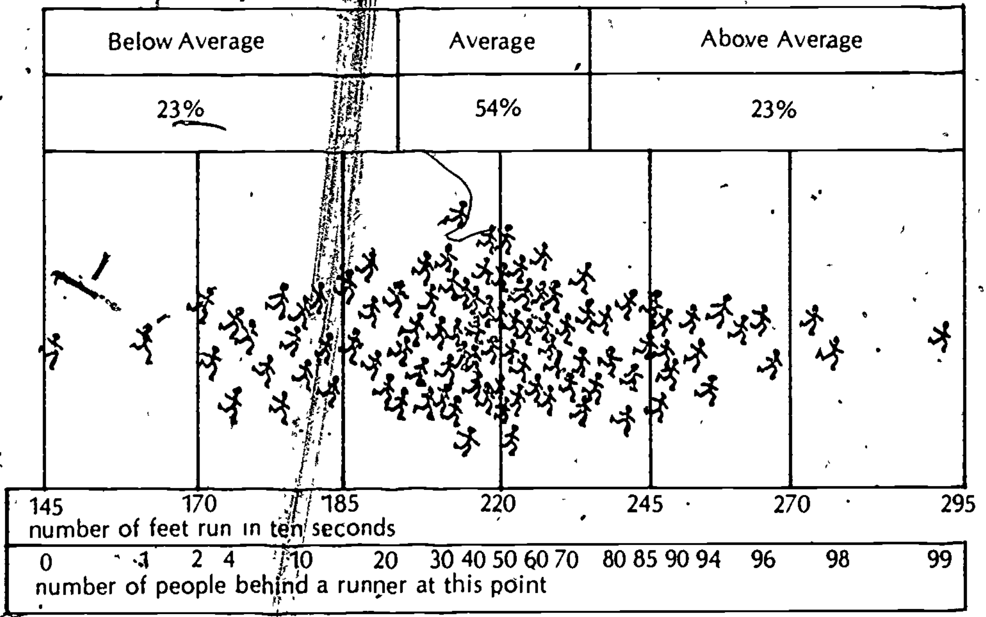
Standardized test scores are somewhat easier to interpret when you know a bit about how they are designed. Norm-referenced tests are designed so that a national, representative sample of students taking the test will fall along what is called a "normal" curve On such a curve, most students will bunch near the middle scores on the scale (or near the average score), with those performing at higher or lower levels spaced more toward the extremes of the scale above or below the average Diagram I shows a normal curve for 100 runners in a 10-second foot race.[2]

20. Two points are particularly important with regard to this diagram as it applies to standardized tests First, and most important, *the tests are designed so that about half of the children score "below average."* In the diagram, 220 feet run in 10 seconds is the average About half of the children run slower than this, and about

DIAGRAM I

ILLUSTRATION OF A NORMAL CURVE

10-SECOND FOOT RACE
100 Runners



half as fast or faster A runner going 220 feet in 10 seconds is at the 50th percentile, or right at grade level for this group of runners Grade level is simply the middle score—half of the children in the norm group must be at grade level or below and half at grade level or above This fact should be kept in mind when interpreting results for different children

Grade-level equivalency scores again reflect how a national sample of students at different grade levels *actually* performed on the test, not how children at a particular grade level *ought* to perform The diagram shows how a group of children do run, not how they ought to run Who indeed can say how fast all children in grade three should run? Specifying how children at a particular grade level *ought* to perform would require a statement of specific educational values and objectives in terms of learning theory and developmental psychology A consensus among educators on issues of values, objectives and principles of educational psychology for children at different grade levels in school is still lacking

Thus, standardized tests reflect how children *typically* perform on particular items, with tests designed so that half of the norm group scores below the average score. These tests are strictly comparative measures They do not indicate what a child should be able to achieve in any absolute sense. We shall return later to this point, which is crucial to an understanding of both the uses and abuses of standardized tests

Another way of reporting scores (also illustrated by Diagram I) is simply to divide the diagram (or normal curve) into parts and report what part the student falls in. 21.
Diagram I indicates a division into thirds, with the lowest 23 percent designated "below average," the upper 23 percent "above average," and the middle 54 percent "average." *This division is completely arbitrary.* The diagram and scores could be divided into fourths (quartile scores), into fifths, or into however many parts we felt would be useful

Percentile ranks are a way of dividing tests into 100 parts based on 100 percentile points Percentile ranks do *not* give the percentage of correct answers, rather, they show what percentage of pupils in the norm group (national, state or local, whichever norm group is used) scored *at or below* a certain level For example, on the Stanford Achievement Test in reading (paragraph meaning) mentioned earlier, a beginning third-grade student with a score of 33 out of 60 would have a percentile rank of 42 percent This means that 42 percent of the children who took the test at the beginning of the third grade in the Stanford national sample scored 33 or lower on the test, 58 percent got more than 33 out of 60 correct answers.

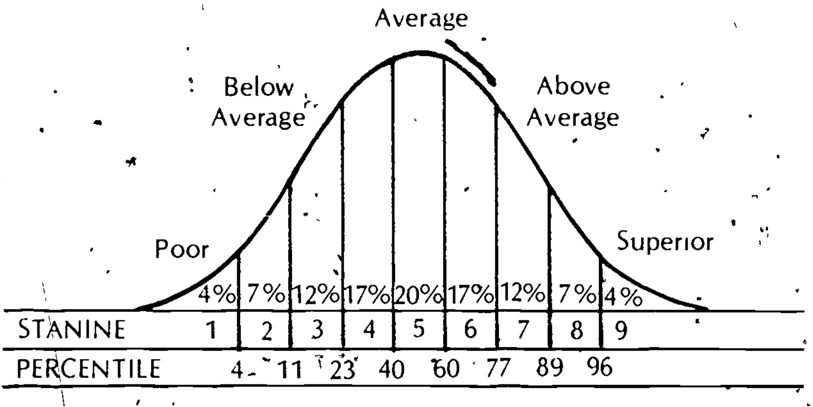
Percentile scores are shown along the bottom of Diagram I as "the number of people behind a runner at this point." This means that, for example, after 10 seconds of running, 50 percent of the children have run 220 feet, thus, a raw score of 220 feet places a student at the 50th percentile with half the students being faster than that and half being slower.

Another common scoring system used by teachers is the "stanine" system A stanine scoring system divides normalized test scores into nine parts (hence the name—*sta* for standard, *nine* for nine-point scale or division). As shown in Diagram II, a score of five is average on a stanine scale with the scale divided so that 40 percent are below average and 40 percent above. (There are statistical features of stanine scores related to how they are computed that give them some additional specialized uses.)

[2] I am indebted to Dr Ralph H Johnson, Director of Guidance and Assessment Services, Minneapolis Public Schools, for [illegible]ding this diagram

DIAGRAM II

THE STANINE REPORTING SYSTEM



The important point about these various ways of reporting test scores is that *the number of divisions used is totally arbitrary* depending upon particular uses, personal preferences and, as often as not, tradition. Halves, thirds, quartiles, fifths, 100 percentile points—each of these reporting systems has advantages and disadvantages All are based on different ways of comparing test results to a norm group

Range Scores

One of the greatest abuses of standardized test scores is the tendency to focus on a single result (e.g , the student is *at* the 40th percentile) instead of on a range of scores Thinking about a range of scores (e g , 35th to 45th percentile) is important because it calls attention to the fact that all test scores are subject to error. For many reasons, *all tests involve some measurement error* Henry Dyer (1973) tells of trying to explain to a governmental official that test scores, even on the most reliable tests, have enough measurement error that they must be used extremely cautiously The government official, who happened to be an enthusiastic proponent of performance contracting, responded that test makers should "get on the ball" and start producing tests that "are 100 percent reliable under all conditions "

Dyer's comments on this conversation are particularly relevant to an understanding of error in tests. He asks

How does one get across the shocking truth that 100 percent reliability in a test is a fiction that, in the nature of the case, is unrealizable? How does one convey the notion that the test reliability problem is not one of reducing measurement error to absolute zero, but of minimizing it as far as practicable and doing one's best to estimate whatever amount of error remains, so that one may act cautiously and wisely in a world where all knowledge is approximate and not even death and taxes are any longer certain? (p. 87)

All reputable test materials include a "standard error or measurement," which is an index of the precision of measurement for individual students This number should be both added to and subtracted from the actual raw score to set the raw score range In our Stanford Achievement Test example the standard error is 2.5 for grade three. Thus, for our student who scored 33 out of 60 the range for the raw score would be 30 5 to 35 5 On paragraph meaning, the student grade equivalent is between 2 8 and 3 1, and the percentile ranking is between the 32d and 50th percentile

The range avoids putting too much emphasis on a single score—an important caution since so many pupils cluster around the middle of the test that the answer to one question can raise or lower a percentile ranking considerably For example, continuing to use the same Stanford Achievement Test example, if our student who got 33 out of 60 correct (42d percentile) had answered just one more item correctly, he or she would have jumped 8 percentile points to the 50th percentile, on the other hand, missing one more item (a score of 32 out of 60) would have dropped the student to the 36th percentile Thus *a difference of one question right or wrong covers a range of 14 percentile points.*

Grade-equivalent scores sometimes allow for similar large jumps at the extremes on the scale In the Stanford Achievement Test cited earlier, 55 correct answers out of 60 is equivalent to the beginning of sixth grade (6 0), one additional correct answer moves the grade equivalency to 6 4, nearly halfway through sixth grade, yet another additional correct answer and the grade equivalent becomes 6 9, while 58 answers out of 60 is more than halfway through seventh grade. By answering correctly three additional questions at the upper end of the scale, a student can jump a grade and a half in achievement.

A good deal of misunderstanding about and abuse of test scores, norms or averages can be avoided if the focus of attention is on a range of results, not a single number Moreover, adding and subtracting the standard error of measurement to a score assures us only that, statistically speaking, we can expect the range to include the true score two-thirds of the time Test statistics represent *probabilities*, not certain results: One-third of the time there is a chance of error even with a range based on the standard error of measurement

Error

Round numbers are always false.
Samuel Johnson

Sources of error are many. The health of the child on the day the test is given can affect the score Whether or not the pupil had breakfast can make a difference Noise in the classroom, a sudden fire drill, whether or not the teacher or a stranger gives the test, a broken pencil, and any number of similar disturbances can change a test score The mental state of the child—depression, boredom, elation, a conflict at home, a fight with another student, anxiety about the test, a low self-concept—all of these factors affect how well the student performs. Simple mechanical errors such as marking the wrong box on the test sheet by accident, accidentally skipping a question, or missing a word while reading are common problems for all of us Students who have trouble reading will perform poorly on reading tests, but they are also likely to perform poorly on social studies, science and arithmetic tests, because all of these tests require reading Thus the test may considerably underestimate the real knowledge of the child.

Some children perform better on tests because they have been taught how to take written tests Some children are simply better test-takers than other children because of their background or personality or how seriously they treat the idea of the test. Some schools make children sit all day long taking test after test, sometimes for an entire week. Other schools give the tests for only a half day or two hours a day to minimize fatigue and boredom Some children like to take tests, some do not Some teachers help children with difficult words, or even read the tests along with the children, others do not Some schools devote their curriculum, or at least some school time, to teaching students what is in the tests Other schools, notably alternative schools—open classrooms, free schools, street academies — place little emphasis on test-taking and paper-and-pencil skills, thus giving students less experience in the rigor and tricks of taking tests

All these sources of error—and we have scarcely scratched the surface of such possibilities—can seriously affect an individual child's score Moreover, they have virtually nothing to do with how "good" the test is, how carefully it was prepared, and how valid its content is for a given child or group *Intrinsic to the nature of standardized testing, these errors are always present to some extent and are largely uncontrollable. They are the reason that statisticians can never develop a test that is 100 percent reliable.*

The errors are more or less serious depending on how a test is used. When looking at test scores for large groups, we can expect that because of such errors some students will perform above their true level and others students will perform below their true score For most groups, statisticans believe that these errors cancel each other. The overly high scores of some students compensate for the overly low scores of others so that the group result is relatively accurate The larger the group tested, the more likely this is to be true

24. However, for a specific individual, no other scores are available to make up for the error in his/her score The only hope is that the questions the student answered wrong because of error will be compensated for by the questions he/she got right either accidentally or by guessing This type of error compensation is much less reliable in correcting for error than the situation described for large groups The least reliable result is one individual's answer on a single question Nothing can compensate for error in this case Thus, *one must be extremely cautious about making too much of results for individuals particularly on single, specific test questions and short tests*

Bias and Invalidity

It ain't so much the things we don't know that get us in trouble. It's the things we know that ain't so. Artemus Ward

All the above sources of error have virtually nothing to do with the actual content of tests. Even the best of tests, carefully prepared by the best-trained professionals, are subject to the kinds of errors we described above. Unfortunately, the tests themselves can also be biased or invalid Whether or not a test is valid depends on whether or not it measures what it is supposed to. Many tests are biased in favor of white, middle-class people so that the tests are as much a measure of ethnic, racial or social class origins as of reading, arithmetic or intelligence. Rural children will have difficulty with a test aimed at and based on city life, and city children will have trouble with tests containing a rural bias Many people feel that all standardized tests are culturally biased to some extent. This severe problem of cultural bias is the center of much controversy among educators (cf J McV. Hunt, 1973, Lavin, 1973, Willie, 1973, Meier, 1973) and is a major factor to be considered in attempting to interpret the meaningfulness of test scores

Another source of content error, bias and invalidity is simply poor preparation of test items Questions that are ambiguous and unclear, have more than one possible right answer, or are based more on ideology than on logic or facts are far more common than one might suspect Such questions can seriously affect test results and are extremely unfair to students and others taking or using tests

Good test questions are extremely difficult to write The writer may be unaware of his/her own bias Unfortunately the only way to determine if the test questions are really fair for your child or your classroom situation is to discuss very carefully every item on a test with every child, to understand what the items *mean* to the children Even this method is subject to error depending upon how aware you are of your own biases Some examples may be helpful to illustrate common sources of test invalidity

Is there more than one right answer? Consider the following question from the Metropolitan Achievement Test (MAT) used to test reading for children seven years old

To *keep* means to □ carry □ hold

After this test Derrick said, "When I want to *keep* something, I *carry* it." "No," said Yvette, "when I want to *keep* something, I *hold* it." In reporting these children's comments Deborah Meier, Herb Mack and Ann Cook (1973) suggest that "these two remarks tell us that children differ in the way they reason" Different wavs of reasoning make test items ambiguous and unclear. Such items do not appear to be good measures of reading ability.

Is the question more ideological than factual? Consider this item from a ninth-grade social studies test (Minnesota High School Achievement Examinations, 1974):

Prejudices are most frequently the result of
- A being born of foreign parents
- B living in communities with mixed racial groups
- C inadequate information
- D personal experiences
- E parental influences on children

Most sociologists, myself included, would argue that all these answers are true under certain circumstances Few sociologists, myself included, would agree with the testmakers that "C" is the best *factual* answer.

Is the question meaningful and relevant or trivial and esoteric? Consider this question, again from the same ninth-grade social studies test

The person with the lowest level of mental ability would properly be classified as
- A an imbecile
- B an idiot
- C. a moron
- D a slow learner
- E normal

It would seem to me that one could seriously question the relevance of this question in a social studies test. The question tests rather trivial factual knowledge about vocabulary and technical distinctions.

These examples point to merely a few of the problems in writing good test questions that actually measure what they are supposed to measure. Tedious though it may be to examine all the items on a test, if you are to properly interpret it, you must find out whether or not you feel the test questions are valid and meaningful One good way to do so is to ask the students who took the test what they meant when they answered the questions.

A Final Note

But the very hairs of your head are all numbered.
Matthew 10·30

We began by noting that standardized testing now pervades our lives Standardized tests have been developed for almost every human characteristic one can name. But testing is still a developing art subject to considerable error. Test results are only one piece of information that can help us understand and clarify our personal observations But test results are easily abused, misunderstood and misused, especially when applied in making vital decisions about the lives of individuals

Given the state of the art, *standardized tests are no substitute for your own carefully considered observations about children you personally know.*

References

Anastasi, Anne "Preface" *Assessment in a Pluralistic Society*. Proceedings of the 1972 Invitational Conference on Testing Problems Princeton, NJ Educational Testing Service, 1973

Buros, Oscar K (ed) *The Seventh Mental Measurements Yearbook* Highland Park, NJ. Gryphon Press, 1972

Cohen, D.K "Politics and Research Evaluation of Social Action Programs in Education." *Review of Educational Research* 40 (1970) 213-38.

Dyer, Henry S "Recycling the Problems in Testing" *Assessment in a Pluralistic Society* Proceedings of the 1972 Invitational Conference on Testing Problems Princeton, NJ Educational Testing Service, 1973 Reprinted by permission

Huff, Darrell *How To Lie with Statistics* New York W W Norton, 1954 Copyright 1954 by Darrell Huff & Irving Geis Reprinted by permission

Hunt, J McV "Heredity, Environment, and Class or Ethnic Differences." *Assessment in a Pluralistic Society* Proceedings of the 1972 Invitational Conference on Testing Problems Princeton, NJ Educational Testing Service, 1973

Krist, M W, & E K Mosher "Politics of Education" *Review of Educational Research* 39 (1969) 623-40

Lavin, David E "Sociological Determinants of Academic Performance" In *The School in Society Studies in the Sociology of Education*, Sam D Sieber & David E Wilder, eds New York Free Press, 1973 Pp 78-98

Manning, W H. "The Functions and Uses of Educational Measurement" *Toward a Theory of Achievement* Proceedings of the 1969 Invitational Conference on Testing Problems Princeton, NJ Educational Testing Service, 1969

McDill, Edward L, Mary S McDill, & J Timothy Sprehe "Evaluation in Practice Compensatory Education" In *Evaluating Social Programs Theory, Practice, and Politics*, Peter H Rossi & Walter Williams, eds New York Seminar Press, 1972 Pp 141-85

Meier, Deborah *Reading Failure and the Tests*. New York Workshop Center for Open Education, City College, 1973

Meier, Deborah, Herb Mack, & Ann Cook *Reading Tests. Do They Hurt Your Child?* New York Workshop Center for Open Education, City Colleges and Community Resources Institute, 1973

Minnesota High School Achievement Examinations, Form E, H Rev. Circle Pines, MN American Guidance Service, 1974

Nunally, James C *Psychometric Theory* New York McGraw-Hill, 1967

Profiles of Performance in the Minneapolis Public Schools Minneapolis, MN Division of Planning and Support Services and Department of Guidance and Assessment Services, Minneapolis Public Schools, 1973

Stanford Achievement Test. Primary Battery and Directions for Administering. New York Harcourt, Brace & World, 1964

Stanford Achievement Test Advanced Battery Teacher's Guide for Interpreting New York Harcourt Brace Jovanovich, 1974

Willie, Charles V. "A Theoretical Approach to Cultural and Biological Differences" *Assessment in a Pluralistic Society* Proceedings of the 1972 Invitational Conference on Testing Problems Princeton, NJ Educational Testing Service, 1973.

Standardized Testing: Reform Is Not Enough!

George E. Hein

Evaluation is an integral part of the political processes of our society.
Ernest R. House (1973)

A number of recent publications have sharply criticized the standardized achievement tests that form the basis for most evaluation of student progress in American education today. James McDonald (1974) has called evaluation "the major disaster area in education", the Council for Basic Education (Weber, 1974) and the National Council of Teachers of English (Venetzky, 1974) have published pamphlets highly critical of present standardized testing procedures. 27.

An obvious response to such criticism is to undertake a major program to reform or revise the tests. There is certainly much room for improvement. The questions could be better, the standardization could be based on more representative samples of the population, and the tests could be validated against criteria more appropriate than the ones used. More imaginative use of the available technology could vastly improve even paper-and-pencil machine-graded examinations. The whole notion that the scoring and administration of all six of the most widely used achievement tests is done on a basis of total questions right in each area without any further modification is really quite absurd. Why not a choice of questions, or questions that relate to a wider range of skill, or the possibility of more than one correct answer in some cases? There is also no reason why achievement must be tested only by paper-and-pencil measures. A much broader range of activities could be standardized.

Unfortunately, any effort to reform the tests avoids a thorough analysis of the reasons why the tests are so bad now. To assume that achieving better standardized tests is simply a matter of making changes in the tests themselves is, I believe, to hold a naive view of the education world and the society. It is highly unlikely that all the people who put tests together, suggest the questions, write the language, try them out on children, standardize them and finally publish and sell them are all totally unperceptive and uneducated, and that testing practice will be greatly improved if the technical competence of its practitioners is increased. We must recognize that the tests and their use are deeply imbedded in the fabric of American society and must be challenged on political grounds, not modified at the technical level.

Lessons from Curriculum Reform Efforts

Any proposal for a major effort to produce new testing mechanisms is niscent of the programs launched almost twenty years ago to produce new science and math curricula Scientists and mathematicians who turned their attention to schools in the late 1950s were horrified at the state of the situation the curriculum was simply bad, they said—full of error, wrong concepts, incorrect statements, too much stress on rote learning, simple drill, etc They set out to reform education and produced high quality, innovative and up-to-date curricula

One of the major learning experiences for those involved in curriculum reform was that providing new curricula, although a necessary condition for better school experiences for children, was hardly a sufficient change. In fact, much of the new curricula was fitted neatly into existing school structures (indeed, it was designed to do so) and, instead of the curricula changing the schools, the schools absorbed the new curricula without much modification in the educational program offered most children In many cases the more innovative characteristics of the new curricula and materials were simply ignored While the New Math has had wide acceptance in the schools, it would be hard to recognize its influence on day-to-day classroom practices (Sarason, 1974) and difficult to discern it on the items appearing on the standardized tests.

There is obviously some merit in developing more reasonable and wide-ranging approaches to standardized testing, as long as one neither expects the task to be simple nor hopes to change education by this means alone. The area of developing alternative tests is wide open, remarkably little work has been done on it. The standardized achievement tests and their companions, the widely used intelligence tests, so dominate the field that little else has been explored and certainly few other approaches have been carried very far

The Case of the Automobile Industry

An analogy can be made to the automobile industry. At one point in the early development of automobiles in the United States, many designs and approaches to the problem of mechanical energy-driven vehicles were explored, engines powered by electricity, steam and other fossil fuels (such as diesel fuel) competed with those developed to use gasoline. The gasoline-powered internal combustion engine was so successful, it spread so widely over the market, that many other technologies were simply abandoned. Today we know a great deal about the gasoline engine that uses rather a lot of gasoline and very little about the alternatives. The recent sharp rise in fuel costs and increasingly serious concerns over the automobile's role in pollution make us painfully aware of the social costs of this unbalanced technological progress

But another component of this analogy is not quite so innocent. The automobile industry evolved policies that channeled and directed research, production and expenditures in the direction of private automobile travel and away from mass transit systems which generally used different forms of locomotion. At the same time, these decisions benefited a particular sector of private industry They had profound effects on our society. A recent Senate Subcommittee report states flatly (*Boston Globe*, March 10, 1974):

> *GM, Ford, and Chrysler reshaped American ground transportation to serve corporate wants instead of social needs. This study suggests that a monopoly in ground vehicle production has led inevitably to a breakdown of the nation's ground transportation.*
>
> *Beginning in the 1920s General Motors began to buy up rail and electric urban transportation systems and then replaced them with buses or diesel locomotives which it manufactured.*

The same report also documents (*Boston Globe*, March 3, 1974) that changes in styling in the automobile industry through the years were not necessarily related to improvements in technology.

We must recognize parallels in the continuing use of large-scale standardized testing programs in our school systems. The companies that produce standardized tests are analogous to the big three automobile manufacturers they dominate their market and dictate what is and is not acceptable Their outlook is limited by what they have found successful Commercial self-interest makes it very unlikely they will launch speculative new projects that might undercut their own positions And, like the big three automobile manufacturers, the publishers who produce testing programs are not isolated from the rest of society They have connections in schools of education, foundations and government that reinforce each other and thus tend to maintain the *status quo*—just as the automobile industry has connections in research institutes, regulatory agencies and government

Analysis of Costs

One strong argument that is continually made for maintaining the present evaluation system is based on relative costs. It is simply a great deal cheaper in dollars and cents to give the MAT (or one form of the Iowa Test) to every child in the school system than it would be, for example, to introduce some sort of individual observation system to determine the status of each child. But the total expenses are so different that no direct comparisons can be made.

The cost of feeding the present testing machine is quite small in comparison with setting up another one, but that does not mean the total investment in it is small Besides the cost of the millions of test booklets, which are not reusable, there are a number of school personnel, especially in large city systems (but smaller ones as well), whose sole job is to organize, administer and interpret the test programs Teachers and children spend a good deal of time giving and taking tests In some Follow Through sites as many as six weeks of the spring term are essentially lost for instructional purposes while the classes go through the agony of taking the various required tests dictated by the school district, the state and the federal program The whole experience disrupts instructional activities for a month and a half (that is about 18 percent of the *total* school year). In addition, a thorough analysis of costs must consider the human and social factors. The tests affect the content and approach of educational programs, they tyrannize teachers and demoralize students Also, part of the cost is the incredible inefficiency of testing schedules. Typically, children are tested sometime in the fall and spring, and the comparative results are released very late in that year or, often, in the next year Teachers cannot even use the tests for their own teaching purposes—they can only be used by outsiders, for purposes other than assisting instruction.

29.

Social Implications of Testing

None of the problems mentioned above would be seriously modified by the availability of better standardized tests.

The major function of the present testing programs is not to determine how much children know, to diagnose their learning stages, and to assist them in their growth and development but, rather, to sort and classify children for their assigned roles in society For this task the present system, with its shoddy and discriminatory tests, works quite well and almost independently of the quality of the tests themselves! As Henry Dyer stated recently (1973):

The widespread use of tests for purposes of selection for deciding from Kindergarten on up who will pass and who will fail, who will be winners and who will be losers, is not likely to go away in a hurry. For, whether we like it or not, it has become indigenous to the kind of competitive culture that characterizes our social institutions, including our educational institutions.

In a historical analysis of the standardized testing movement, Karier (1972) has described how the use of tests to classify children developed in the 1920s and 1930s reflected the prejudices of our society prevalent at that time Unfortunately, the same views still influence decisions today In a standard psychology text published in 1970 (Bernard), individuals are classified according to IQ into categories designated as follows.

Intellectual Category	Educational Potential	Live Work Potential	IQ Range	% of Population
Feeble-minded	Uneducable	Typically dependent	70	1
Borderline	Special Schooling	Routine jobs	70-80	2
Slow learners	Special classes	Day laborers, routine jobs	80-90	16
Normal	High school, but perhaps with difficulty	Laborers, semi-skilled jobs, clerical work, some semi-professional	90-110	50
Superior	High School, some college	Skilled work, professional work for some	110-120	16
Very superior	College	Skilled work, professional career	120-130	2
Gifted	Graduate School	Professional, creative	130	1

The social implications of the view that only 2 percent of the population, selected on the basis of a series of standardized paper and pencil tests, definitely has the capacity for college work and professional careers and that at best 16 percent of the population has the potential for "some college" represents a political judgment that would not be affected significantly even if the tests used were less subject to technical criticism.

The possibility of high test scores is held out to parents as a way of providing a great future for their children when, in fact, it would take a very high score indeed to change significantly the life chances of poor children The complex relation of school to economics, to college admission and to the job market is closely related to larger social and political issues, including the prejudices of our society (Berg, 1970).

What the tests encourage is a lottery concept of education. It is true that an unusually high-achieving child from deprived circumstances—a child who does very well on the standardized tests—can break out of the bounds of the class in which he or she lives and actually change status. But the odds against such an occurrence are enormous. This kind of case—and there are some all the time—has the same effect on redistribution of classes in society that the lottery has on redistribution of income. The lottery in Massachusetts, for example, provides about a 13 million to one chance of winning $1,000,000. That means that, after 13 million tickets are sold, one person may drastically change his or her economic status

There are just enough winners of smaller amounts so that many people can support the illusion that they too may be a winner, that they too can change their status But, of course, the actual number is so small that the few who break out of poverty by winning the sweepstakes is insignificant for any change in class alignment Exactly the same reasoning supports the concept that good reading scores will help populations break out of poverty or oppression The actual number of children who can change their status as a result of school success is trivial compared to the total population of those condemned to poor jobs and continuing poverty

We can recognize that the tests do not necessarily reflect accurately children's abilities and knowledge of individuals, by noting the number of exceptions to expected results Every person active in education has his or her own store of anecdotes about Jane who did poorly on an MAT but could do the work, of Frankie who could read only on the second-grade level and, after two months of help, could read on the sixth-grade level, of Janice whose IQ rose 25 points in a year In some cases where people have looked at children carefully and worked with them sensitively, whole classes and groups have increased their IQs (Rayder et al , 1973) or their grade level achievement phenomenally over relatively short periods of time In *Reading, How To* (1973) Kohl reports the case of Lillian, a child whose performance improved so much that it required the threat of a law suit to force the school to accept the results of three reading tests

Summary

Before we advocate major programs to improve standardized tests, we have to recognize their role in the American educational scene. The tests are one component in the sorting system of American schools. They contribute one element, but not the only one, one necessary condition, but not a sufficient one, to see to it that the schools continue the society as it is We have to be aware of the use of testing in the society, of the social and political role of the education system, and of the investment in the present evaluation structure Only if we are willing to reexamine the entire structure and nature of the education of our children can we hope to achieve more equitable schooling that assists each child to develop his or her full potential

References

Berg, I *Education and Jobs The Great Training Robbery* New York Praeger, 1970

Bernard, H W *Human Development in Western Culture*, 3d ed Boston Allyn & Bacon, 1970 Reprinted by permission of the publisher

Boston Globe, Mar 3, 1974

__________, Mar 10, 1974 Courtesy of *The Boston Globe*

Dyer, H S "Testing Little Children Some Old Problems in New Settings " *Childhood Education* 49, 7 (Apr 1973) 362-67

House, Ernst R *School Evaluation, The Politics and Process* Berkeley, CA McCutchan, 1973

Karier, C J "Testing for Order and Control in the Corporate Liberal State " *Educational Theory* 22 (1972) 159-80

Kohl, H *Reading, How To* New York Dutton, 1973 P 20.

McDonald, James B "An Evaluation of Evaluation " *The Urban Review* 7, 3 (1974)

Rayder, N , B Body, & C Nimnicht "Assessing Follow Through " San Francisco Far West Laboratory for Educational Research and Development, 1973

Rothchild, Emma *Paradise Lost The Decline of the Auto-Industrial Age* New York Random House, 1973

Sarason, S B *The Culture of the School and the Problem of Change* Boston Allyn & Bacon, 1974 Ch 4

Venetzky, R L *Testing and Reading* Urbana, IL National Council of Teachers of English, 1974

Weber, George "Uses and Abuses of Standardized Testing in the Schools " *Occasional Paper No 22* Washington, DC Council for Basic Education, 1974

32.

Another Look at What's Wrong with Reading Tests

Deborah Meier

A teacher listening to two children read the passage, "He lived in a big house," notes that one read, "He lived in a big apartment" and the other, "He lived in a big horse "[1] Are both equally wrong? In fact, many teachers, like many tests, are prone to consider the latter mistake less serious, since it's "off" by only a single consonant rather than a whole word.

In our frantic effort to teach children how to beat the testing game we have lost sight of the purpose of reading. to turn the written page into something that makes sense. Tests (formal or informal) are, at best, only symptomatic, a roundabout way of getting some hints as to what students are doing and whether schools are helping them. But the nature of the tests we have devised and our single-minded focus on them have led to a decline in concern for the real act of reading with its power to explain, to influence and to move. (For example, amidst the presumably enormous concern to improve reading in New York City, neighborhood libraries have been cut back to a few hours a day and weekend use has been eliminated.) In order to understand how this is so, we need clarification regarding both the act of reading and the nature of testing. Particularly, we need (1) a definition of what we mean by reading competence, (2) a closer look at the implicit underlying definition of reading that is embodied in current reading tests, and (3) some alternative means of assessing reading that would document better what *is* and direct attention toward what could be

Toward a Definition of Reading

We pretend—to parents, teachers and children—that it is enough that a child be drilled into changing a set of visual symbols into oral ones. We act as though this action were reading. (It is such a commonly accepted definition of reading that the parents of fluent readers sometimes complain that schools are not teaching their children phonics) We call this behavior "decoding" although, as linguist Frank Smith points out, actually it is merely translating from one code (visual) into another code (oral).[2] Such a skill, useful as it may be, is a very trivial one. For example, I can do it for Spanish without being able to understand a thing I am saying.

We are in fact faced with a vast number of students who have made this first translation into oral reading. They can decode, yet they are still at sea We are all in this fix sometimes when reading something we find difficult The problem we face is not "breaking the visual to oral code." Our difficulty lies in the subject matter itself or the language used to describe it. At such times we cannot translate the visual or the oral symbols into significant meanings. Turning those marks on the page into meaning is what constitutes reading. To do this means bringing a lot *into* the act of reading quite aside from what we know about the visual marks on the written page.

When a college professor complains that his students these days do not even know how to read, he naturally means (although he may not realize it) that such students *cannot make sense of* the reading material he minimally expects of a college freshman The distinction between a 12.9 and a 6.9 reader (using the grade-level equivalents of the standardized reading measures), after all, is not that one *can* read and the other cannot. The difference lies in *what* they can get meaning from or the different sorts of meaning they take from the same material. What changes—or what should change—over time is what we bring *to* the act of reading

Many experts have suggested as a definition of reading skill qua reading skill (literacy), the closing of the gap between what one can make sense of orally and what one can comprehend visually.[3] Given such a closure, the school's task should be to help children make sense out of more of the world.

What Is Happening Back in the Real World?

In the absence of an acceptance of the kind of definition of reading outlined briefly above, the tests themselves have become a kind of implicit definition. When we ask a teacher or parent about a child's reading, they all increasingly fall back on the jargon of test scores. All common-sense judgments are abandoned. Even children begin to judge their reading as though it were merely an extension of testing. And no wonder, when even the best intentioned of us begin to urge children to read *in order* to raise their test scores.

Those children who come to reading easily are the least injured by this, although they too are encouraged to focus narrowly on keeping ahead on the tests. Upper-grade children who are already fluent readers, especially if they are in inner-city schools, are often kept busy filling in blanks and drilling on subskills that might appear on tests while the content of the world is skipped over as a luxury. Good books are used merely to teach test skills. "getting the meaning" or "inferential thinking."

But the children who badly need the teacher's assistance are the most seriously handicapped. While we hammer away at skill tasks that appear on tests, we often deprive these youngsters of the kind of knowledge and language experience that they badly need *to bring* to their reading. Even the skill-tasks themselves are often justified by the teacher only on the grounds that they are necessary for the tests. They too may leave children as much in the dark as ever about how to use their own natural intelligence to work out the relationships between the visual symbols and the world of meaning Worse still, these tasks convince some children not to trust their own intelligent hypotheses, thus making it virtually impossible for them to develop fluency.

The task of making sense of the written word is a procedure very similar to one all children accomplished just a few years earlier—when they learned to talk. It is well to remember that the children who enter school, including the most disadvantaged, have only recently constructed and verified a set of bewilderingly complex rules that "summarize the relationships and regularities underlying language "[4] They succeed "even though adults are far from any understanding of what these relationships and regularities are, let alone how to impart them through formal instruction."[5] They used a process of trial and error—*plenty* of error! To encourage such experimentation we merely supported the never-stop noisemaking and monologue-like conversation of small children. We responded to the sense of what they were saying whenever appropriate. We did not categorize them at each successive stage, isolate the sounds or rules for their practice, count their errors or rict them on the basis of some prior logical notion of "sequence." We did not need a test to know if they were progressing or whether indeed they knew "how to talk" when they came to school We have no comparable grade-level oral language standards—important as oral language is Nor do we confuse good talking with conscious knowledge of the way language is put together, intriguing as the latter may be

Reading, like talking, appears to be logically impossible only when we persist in acting as though one needed to know all about it in order to do it We appear to believe that learning to read requires both a superhuman memory and a quite impossible memory retrieval system We appear to do so in a period of educational history in which we simultaneously admire the research of Jean Piaget, which leads us to conclude that young children's thinking is still very concrete and utilizes a form of logic that seems indeed illogical to adults Yet we try to teach children to read as though they were indeed highly sophisticated and self-conscious computer programers (and as though our system of written language was "computerable") That it sometimes appears to work is a credit to the flexibility and tenacity of human intelligence—to pick out what it needs and discard the rest That it so often does not lead to anything resembling the real act of reading is hardly surprising

But we persist in this view of our task since the one thing such an approach may indeed succeed in doing, at least in the short turn, is raise reading scores .

Even when children get past the first roadblock and begin to read with some fluency, new obstacles appear in the form of new test-related demands For what is especially vicious about this test-dominated approach is that it never ends. There are always more tasks that will make you read (i e., test) even better At no point along the way is one allowed to say, "Goodness, he reads! On to other things "

For example, in a study of the Stanford Reading Diagnostic using inner-city Philadelphia seventh-graders, it appeared that many children scored well on the reading comprehension subsection (the only part that comes close to measuring reading per se) but were pulled down by low scores on subsections measuring auditory discrimation, blending and syllabication [6] Since the youngsters' final scores reflected the sum of all the parts (and since teachers, children and programs are judged by such final scores), many odd classroom procedures naturally follow

Conscientious teachers give such children remediation tasks on the subskills they tested poorly on They do so regardless of any evidence that this will lead to improvement in comprehension Children are drilled on recognizing similarities between certain isolated sounds ("Which word," the teacher might ask, "has the same sound in the middle as 'rat'—*table, run, camp* or *seat?*") They spend hours learning rules to help them decide whether to break "riddle" into "ri-ddle," "rid-dle" or "ridd-le " Considerable energy is also spent helping students decipher test instructions, since all these skills will be in vain if the student cannot demonstrate them successfully on the test. Little time remains for reading or for other content areas The child has been trapped Any other course seems to court disaster, being labeled "below grade."

The Tests

It is critical to recognize that a test is based on assumptions regarding what is being measured and why The difficulty with our reading tests is that we have accepted the machinery of the tests without having questioned whether we agree with their implicit or explicit definition of reading or of reading progress Even less so have we agreed on how such reading is acquired (In fact, the testmakers deny having either a definition or a theory. They are merely measurement men.)

The tests are constructed not from an explicit theory of reading but out of an eclectic potpourri of items whose justification lies in the fact that they have a high degree of correlation with later school success, are consistent with other similar tests, and produce a normal curve The midpoint along this curve then constitutes what the layman mistakenly assumes is the "should" of reading Does this sound too slipshod, unfair, unlikely as a description?[7]

In the summer of 1973 a group of respected reading and testing experts met together in Georgetown, Washington, D.C , under the auspices of the International Reading Association They had a hard time agreeing about much, particularly about what to do or say to the public at large But there was, as one participant noted in summarizing the conference,[8] virtually unanimous agreement that all the existing normative-based reading tests were without a theoretical rationale, had "little relevance for instruction and were not designed to measure or record educational improvement." The experts agreed that the tests "both mask and distort the real issues involved in the acquisition of reading skill" and that there is today "no definitive knowledge regarding either the sequential learnings or component skills that children must acquire in order to read successfully." They further endorsed the notion that "especially in the acquisition period" reading tests should be "program specific," testing "only what has been directly taught or indirectly fostered."

Alternatives

There are many alternative forms of assessment No perfect ones exist for all purposes and all programs[9] For example, the use of individually administered reading inventories such as the Spache[10] or the Silvaroli[11] are a step in the right direction if we want rough comparative data on individual skill They also can provide some diagnostic information, although so can any teacher who gives attention to a child's reading. Kenneth Goodman's Miscue Inventory[12] is better as a tool for gaining insight into a child's individual approach, although too detailed and complex for everyday use

Short, program-specific tests designed to fit the activities of a particular classroom or program are manifold They often come with commercial reading systems, or can be quickly whipped up by a teacher to see if what has been specifically taught has been learned

Good anecdotal documentation of observed student language and reading is done by many good teachers and researchers and could yield rich information that is both diagnostic and suggestive if we chose to spend our time and money that way.[13]

For obtaining general data on larger populations, programs or trends, particularly beyond the stage of minimal reading acquisition, random sampling techniques applied even to the existing normative-type tests would be preferable Random sampling would also make alternative, more individualized methods financially feasible and could thereby provide far richer data. Incidentally, it would also avoid the enormous and unbeatable problem of cheating that is encouraged under present circumstances, since it removes both the opportunity and the incentive to coach for the test or cheat during it. The English have, for example, given a short individually administered reading test to a sample population every ten years[14] While the English system of testing has also received criticism for archaic language as well as methodology, it has at least attempted to develop comparative longitudinal data without distorting the educational process by the evaluation of it.

The problem of assessment, in fact, is so equally well met by other and even cheaper methods than the current mass testings that one is led to conclude that e may be method to this madness[15]

Conclusion

Most standardized reading tests play a negative role They discourage us from using schools to help children become readers Those with other resources learn anyway, those least advantaged are as usual stumped

The tests encourage us to fall for the notion that reading is mostly just a "trick," a useful one for "getting ahead" in the "real world" rather than a means of expanding our understanding of it

It is well to remember that schools cannot get everyone above the median It stands to reason, given other facts of life, that the least advantaged will *more likely* fall below the median than above This natural fact of life is reinforced by the nature of any standardized instrument weighted, as it must be, toward the culture and associative patterns of the mainstream child [16]

Still schools could succeed in making almost all children good readers This appears otherwise only so long as we define good reading as a point on the normal curve If we fall for that frame of reference, it is indeed a logical contradiction to seek improvement

To say there is no intrinsic necessity for the poor to turn off written language is not to pretend that we could alter the class structure, achieve economic and social equality, produce vast changes in patterns of mobility, or "even" reverse the locations of socioeconomic groupings on a normative scale by making all children competent readers But merely because we cannot achieve all this just by helping children be readers does not mean it is not worth doing

36.

References

[1] This marvelous example was raised by Frank Smith in a seminar on reading held by the City College Workshop in Open Education in May 1974

[2] Frank Smith *Psycholinguistics and Reading*, 1974, and *Understanding Reading*, 1973 (New York Holt, Rinehart & Winston) Two superb books describing what is—"musts" for those who take this issue seriously

[3] This definition was proposed by one of the working groups at the Georgetown IRA Conference, and primarily was pressed by L Gleitman (University of Pennsylvania) and T Sticht (Human Resources Research Organization) It also appears in Herbert Kóhl's recent book, *Reading—How To* (New York E P Dutton, 1973)

[4-5] Frank Smith, *op cit*

[6] Virginia Allen, Triple "T" Project Monograph Series #1, "What Does a Reading Test Test?" (Philadelphia Temple University College of Education, 1974), pp 1-21

[7] Which is not to say that technically the tests are not often carefully and conscientiously, even painstakingly, well put together! The criticism is largely of the rationale, not the skill with which this rationale of testing is pursued

[8] Report by Anne Bussis of Educational Testing Service, "Memo For the Record," Dec 1973

[9] See excellent summary of this issue, "Consumer Awareness in Test Reviews," by Roger Farr & Wm Ellery (Washington, DC Linguistics Institute, Georgetown University, Aug 1973), mimeo

[10] Diagnostic Reading Scales, devised by George D Spache (published by Calif Test Bureau of McGraw-Hill, 1963) Also see Vivienne Garfinkle's guide to this material (published by Project Follow Through, Bank St College of Education, New York, Feb 1, 1972)

[11] Nicholas Silvaroli, *Classroom Reading Inventory*, 2d Ed (Dubuque, IA W C Brown, 1973)

[12] Yetta Goodman & Carolyn Burke, *Reading-Miscue Inventory Procedure for Diagnosis and Evaluation* (New York Macmillan, 1972)

[13] The Workshop Center for Open Education, City College of New York, (6 Shepard Hall, Convent Ave & 140th St , New York City 10031) has published various brochures on reading profiles and other informal evaluation tools See also studies by Patricia Carini of The Prospect School, North Bennington, VT—e.g., "A Methodology for Evaluating Innovative Programs," June 1969

[14] Note this remarkable rationale for the English system, published in a Department of Education pamphlet (No 50) in 1966 Perhaps the strongest arguments in favour of the present test are first that it implies a definition of reading ability that is in accordance with common sense, and secondly that it takes no more than ten minutes of the pupil's time This economy is valuable, since the business of the school is not to test but to teach Moreover it is not ten minutes of every pupil's time that is needed "

[15] A number of provocative essays on the history of normative-type tests raise important questions regarding the larger social function of looking at children in this particular way See, for example, "Testing for Order and Control," by Clarence J Karier, *Educational Theory*, Spring 1972, pp 159-80

[16] See "Reading Failure and the Tests," by Deborah Meier, City College Workshop Center, An Occasional Paper, Feb 1973 See also "What's Wrong with the Tests?" by Deborah Meier, Notes from City College Advisory Service to Open Corridors, Mar 1972, pp. 3-17

The Stranglehold of Norms on the Individual Child

Lois Barclay Murphy

Our children are choking in a stranglehold of norms. What do I mean by "stranglehold"? I am talking about the stifling, asphyxiating, smothering effect that comes from pigeonholing children in terms of test scores, of normative categories of pathology and nonconformity to social demands. The breath of learning requires oxygen for mental growth and respect for the integrity of the child's individual psychic metabolism as well as his physical idiosyncrasies. Confinement to a narrow, tight, constrained mental and emotional environment limited by statistically based norms and unrealistically restricted expectations can starve as well as frustrate the child, just as the Berkeley rat cages interfered with optimal development of brain tissue and problem-solving in little rats (Rosenzweig, 1972).

The strangling and starvation go on in schools, hospitals, clinics, families — wherever we freeze our expectations of a child in terms of the belief that tests are the truth, the whole truth and the final permanent truth about a child's potentialities Judgments are too often weighted on the negative side. Instead of asking about how sick, or bad, or problem-ridden a child is, we *could* ask. "How is he dealing with the complex life situation he is in, with its remoteness from his style, his longings, his realities, his needs, his particular balance or imbalance of strengths and vulnerabilities?" We *could* ask "How can we support his integrity, build on his resources, help him in ways he wants and is ready to be helped, wherever we find him?"

37

How and When To Look at Tests

Some wise teachers say "I don't look at the tests until I've had time to get acquainted with the child and discover what he responds to, what skills he has, what he is interested in, what his tempo is, what he is hiding and what he lets us see (I usually need a month or six weeks for that). Then I look at the test results in relation to the way he is functioning Sometimes he does better than the tests would lead one to expect, sometimes not so well. If the latter, I can explore different avenues of reaching him, to bring out his better level, and I try to find out what is hurting, or why he cannot bring to the group what he was able to bring in the test situation If he does better in the classroom than he did on the tests, I try to find out what bugged him in the tests, or what they did not give him a chance to show"

Tests can enrich the teacher's insight or the therapist's understanding only when the details are looked at in relation to the child's experience in the test situation— what he was coping with and how.

Speed norms can be especially misleading, as we know from studies of American Indian children (Klineberg, 1928) and from many children to whom strangeness in a situation slows them down as they try to grasp what is going on.

Age norms have been useful in gross distinctions between severely retarded and

Adapted from a presentation to the American Orthopsychiatric Association at its 1972 Annual Conference in Detroit, igan First published in *Childhood Education*, April 1973, pp 343-49.

adequately endowed children and, more broadly still, between capacities of children and adults But they have led to rigid assumptions regarding age-appropriate behavior and unrealistic pressures on children to behave like a twelve-year-old or a fourteen-year-old, as I shall document later, or even at younger age levels to "be a big boy " Pressures to meet arbitrary sex-role standards especially ignore the wide variety of growing-up patterns exhibited by children in longitudinal studies Mental health norms and concepts of problem behavior are often inconsistent with what we know of the vicissitudes accompanying the process of growing up, the long struggle with remarkably unique patterns of vulnerability and strength, and the very important "Toynbee effect," the response to challenges evoked by the confrontation of one's specific checkerboard of weaknesses and resources with the environmental checkerboard of stresses and supports. To a large extent, each child's development is a mystery story whose outcome we cannot really predict The complexity of the developmental process with the emerging capacities, drives, investments, conflicts is still far beyond our complete comprehension, at our present primitive stage of understanding

Some Key Sources on Human Development

The most important volume on human development to appear to date became available in time for me to use some of the findings to document this thesis I refer to *The Course of Human Development* (1971) by Jones, Bayley, Macfarlane, and Honzik, of the Institute of Human Development at the University of California. This book contains edited versions of over sixty papers on physical, mental, emotional and social development of the children, now adults, in three major longitudinal studies at Berkeley begun in the 1920s and continued into the present. It is not only a gold mine but *the* major gold mine of solid findings on development, with implications everyone working responsibly with children must take into account

Here I can share with you only a few highlights, which I believe should be taken most seriously and which, if we do, should challenge and turn around some of our most rigid assumptions about behavior and development.

Physical Development and Behavior

Let us start with the reports about relations between *physical* development and behavior, some of which should be familiar but generally are not, from the volume by Stolz and Stolz (1951), the paper by M. C. Jones and N Bayley (1950) and by H E. Jones (1971). These related articles document ways the 20 percent of girls who mature most early get into difficulties as a result their social and sexual drives are precocious in relation to their intellectual development, they may attract older boys, be full of adolescent fantasies too early and be out of harness with their slower more-average peers Since boys generally mature more slowly than girls, the *slowest* 20 percent of boys are also out on a limb, left to feel inadequate with both boys and girls, isolated, rejected, and of course they develop some form of coping and defense-techniques to deal with their situation. Similar dilemmas are experienced by children who may not be in the *extreme* 20 percent, but whose growth pattern is variable.

In other words, it behooves us to watch closely to see exactly what the developmental situation and problem are for the child before we complain that he or she is not "acting appropriately for his age." How can they fit into a norm that is inappropriate to their individual patterns of development? Physical measures such as height were much more predictive over a long age-span than mental test measures,

Much of the support for the thesis Dr Murphy presents herein comes from longitudinal series of studies of normal children, initiated at the Menninger Foundation in Topeka Kansas by S K Escalona and M Leitchan 1948, and carried further under Dr Murphy's direction until 1969 as well as from her prior research at Sarah Lawrence College Support was contributed for the Topeka studies from the Pfeiffer and Neumeyer Foundations and from the National Institute of Mental Health, and for the latter studies by grants from the Josiah Macy Jr Foundation

while these were better than personality measures Correlations for height between ages three and eighteen years were in the 70s, whereas mental-test measures were around 40 But few personality measures reached .40 It is worthwhile to look closely at IQs Macfarlane (1971) observes that for the eight tests given between ages six and eighteen, only 15 percent of the children showed a range of less than 10 IQ points, 58 percent showed a change of 15 IQ points or more. One-third of the children showed a range of 20 IQ points or more Ten percent showed a range of 30 points or more These results are in line with our Topeka findings (Moriarty, 1966) and, in addition, data from the Fels Research Institute (Sontag, Baker and Nelson, 1958), as well as other studies such as Nancy Bayley's analyses of sequences in IQ (1949).

Misconceptions About Mental Development

Macfarlane concludes that "little reliance can be placed on one test." Beyond this, she notes the striking finding that a number of men with poor records both on mental tests and school grades, right through high school, came as adults into positions requiring creativity and high intelligence. For example, one man had an average IQ through his developmental years around 100, he was held over three times in elementary school, and finally graduated from high school at age twenty-one without college recommendations He left the community, made up his high school deficiencies and now is a highly talented architect. Currently he is living out a normal life through his children, being active in his community, and finding life exciting and satisfying after thinking of himself as "a listless oddball" during childhood and adolescence

Social Development

Macfarlane also tells us that, of the children studied by a large research staff with different theoretical biases, close to 50 percent turned out to be more stable and effective adults than any staff member had predicted, 20 percent were less substantial than predicted, and scarcely one-third turned out as predicted

Among the 10 percent who turned out *far* better than predicted were two who presistently spent their energies in defiance of regulations, getting marginal or failing grades throughout their schooling, and finally getting expelled at ages fifteen and sixteen. Macfarlane finds both of them to be wise, understanding parents now, who appreciate the complexities of life, moreover, they are humorous and compassionate.

In reflecting on the factors contributing to erroneous predictions, Macfarlane remarks that no one becomes mature without the pains and confusions of maturing experiences. Even experiences that looked traumatic at the time are now regarded by subjects as forcing them to come to terms with what they wanted and did not want out of their lives, and to shift their behavior in the direction of goals they clarified. Also, many times, behavior considered unpromising by clinical examiners, such as overdependence, was converted into nurturance by adulthood and *not* overprotection — since these people wanted their children to avoid the overdependence they themselves had experienced. There were also "late bloomers" who blossomed only after they got away from their families and were released to be themselves. Macfarlane emphasizes the capacity of these young people to drop early habits and behavior that got in their way as adults, and to develop new patterns on a trial-and-error basis. She also emphasizes the tendency of a clinical staff to overweight pathogenic aspects of behavior seen in childhood to give too little weight to the maturity-inducing aspects.

Some Basic Findings on Individuality

I have begun with the most familiar areas for clinical and educational workers. Let me go on to some very fundamental findings from the medical, physiological and biochemical areas Another important volume for clinical workers must be *The Biology of Human Variation*. Here Dr Sontag (1966) of the Fels Research Institute reports evidence that the heart rate of the fetus, influenced as it may be by the fetal environment, tends to persist to adulthood. And we are beginning to have data on the relation of behavior to heart functioning and other autonomic reactivity patterns Granted that the autonomic nervous system interacts with the central nervous system and is influenced by cognitive functioning — as biofeedback studies are showing (Hefferline and Bruno, 1971) — the influence from autonomic functioning to behavior urgently demands attention. In fact, it may well be that we will understand more about variability in IQ when we study the relation of these variations to autonomic reactivity.

Still another basic contribution to our thinking is that of Roger Williams in his book on *Biochemical Individuality* (1956). Here he documents the extraordinary variations in individual structure and needs — from sizes and shapes of stomach and intestines to the most extreme differences in needs for each of the different vitamins Such findings simply add another dimension to the extensive and solid data on individuality of growth patterns from infancy to adulthood (Shirley, 1931, Olson, 1943; and others)

40.

Misuse of Statistics

We really don't need any more support for the necessity of recognizing individuality at every level but, just to cap the climax, I will mention the comments of one of the world's foremost statisticians, C. Radhakrishma Rao (1965), to the effect that much of our statistical thinking is unsound when we draw conclusions from average scores in large groups, while ignoring the scores of counterbalancing subgroups that contribute to the averages. I emphasize this point because the averages are used as a basis for norms — norms which we have already seen are misleading, as in the data on variations in ages of maturation at puberty.

Rethinking Frozen Concepts

We have been talking about individuality, plasticity and the capacity for self-directed change. Macfarlane also emphasized the trap of pathology-oriented thinking. Surely the data amassed over the last fifty years *demand* that we rethink our frozen concepts, loosen up, confront the realities of child development and come up with some better, more realistic if less quick and easy concepts. Reliance on the IQ has stultified our thinking about potentialities of children. Reliance on pathology-dominated concepts of drives has distorted our thinking, even polluted it

What are the alternatives, possible ways of thinking about children that might be more fair to the child and his potential development?

Using Data Qualitatively

First, we need to recognize that IQ tests, personality tests, and the rest are limited We can use what they tell us about what the child does at this moment under these confining, distracting, uncomfortable, frightening, boring, uninspiring conditions They don't tell us what the child does under other conditions, or what he *might* do if comfortable, stimulated or inspired, healthier, or less bogged down in family anxieties As one little boy in the Topeka studies implied, they do not even necessarily tell what the child can do right now. He asked, "Why don't you ask me to do what I can do?"

Along with this, we need to recognize that the *specifics* of the test may be much more illuminating than the IQ, which is the average of all the functions. One little girl who barely passed seventh-year-level tests on routine items passed twelfth-year-level tests involving insight and comprehension. Although she was retarded in reading at that time, she has become an expressive and original writer and a remarkably intuitive and creative mother She was considered a "slow learner" in the second grade, while all the time she was storing away observations and reflections in her independent sensitive way. I could give other examples, from our Topeka studies (Murphy and Moriarty): a girl with an early IQ of 100, who was considered in college the most outstanding candidate for a music scholarship, and is now having her graduate year of practice teaching in preparation for a career as a music teacher In her case, the incidental observations of her social awareness could have provided a better basis for prediction of later development.

Understanding Motivation — Some Examples

Along with the *qualitative* use of data from tests of all sorts and the observations that can be made during, before and after tests, we need a drastically new approach to motivations and drives. Perhaps we will be on more solid ground if we ask. "What is this child's situation, the positives and negatives (roadblocks, frustrations, etc.) for him; what can we learn about the positives and negatives of the equipment he brings to dealing with the situation — the areas of strength and of vulnerability, and in terms of his plus and minus resources what is he *trying to do with his situation?"* In Juvenile Court one Monday morning I saw a ghetto boy brought in for picking up some discarded metal tubing in a construction area; I don't know just what he was going to do with it, but no one asked him. Here is a boy for whom the city provided no play space, nothing to explore or to create with. He finds something he might use: Bravo! A boy with some initiative, some active drive to pick up a crumb of possible value in the arid ghetto where he lives. Let us get him into a shop for metal- and wood-working where he can try out his ideas instead of sending him to Detention where he'll learn fascinating criminal techniques.

Here is another six-year-old boy with an extremely disturbed mother. I happened to be nearby when I heard him yell to a friend, "Bill, let's get the hell out of here, Mom's starting to go on a rampage." His drive to survival was being expressed in utterly healthy and sensible escape. At school his teacher reported that he didn't seem to trust any adults and was not "learning." Of course, he *had* learned some very basic things, and was using his learning well in terms of what he had experienced of life so far.

Another six-year-old boy who rejected a very rigid teacher was placed in a different school and reported, "This new teacher understands children much better than that other teacher." He was correct — just one instance of how important it is to listen to the observations, judgments and points of view of children.

The examples I have given illustrate the child's *integrity* as an autonomous growing person, appraising his environment, finding ways to survive in it, developing whatever coping methods and defenses he can devise to get along with the business of growing up and get along in the situation in which he finds himself. The extensive documentation of transitoriness of fears (Jersild, 1935) behavior problems (Macfarlene, Allen and Honzik, 1971) and even changes in body-build (Stolz and Stolz, 1951) attest to the extent of the child's plasticity and capacity for change, and for progress in mastery and outgrowing earlier patterns he does not need any longer Topeka (Murphy and Moriarty) and Berkeley (Macfarlane, "Perspectives ," 1971) studies document the positive outcomes that can emerge from the child's mastery of his vulnerability and the stresses he successively faces

Discovering Coping Strategies

The obvious conclusion is that we need to focus on and better understand the nature of ongoing current coping struggles, how to support them, how to help the child to extract the strength and insight that successive experiences may make available to him. We need to understand the positive strategic values of withdrawal in certain situations, and be very cautious about talking about a "withdrawn child " Similarly we need to respect and value children's protests, resistances, attempts to change or control situations, and all the other active coping efforts that can give us cues to what the child finds intolerable, unsuitable, boring, distasteful or threatening to his integrity I am not offering a new scale or a new test to freeze your thinking once again. I am pleading instead that each clinician, each teacher, use all of the available resources along with his own fresh look at the child in his situation in order to discover the meaning of the child's behavior from the child's own point of view.

References

Bayley, N "Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years " *Journal of Genetic Psychology* 75 (1949) 165-96

Hefferline, R F , & Louis J J Bruno 'Bio-Feedback Controlling the Uncontrollable " *In The Psychology of Private Events*, Ralph Jacobs & Lewis B Sachs, eds New York Academic Press, 1971

Honzik, M "Perspectives on the Longitudinal Studies " in *The Course of Human Development*, M C Jones, N Bayley, J Macfarlane, & M Honzik Waltham, MA Xerox College Publishing, 1971

Jersild, A T , & F B Holmes *Children's Fears* Child Development Monograph No 20 358, 1935

Jones, H E "Physical Maturing Among Girls as Related to Behavior " In *The Course of Human Development*, M C Jones, N Bayley, J Macfarlane, & M Honzik Waltham, MA Xerox College Publishing, 1971

Jones, M C , & N Bayley "Physical Maturing Among Boys as Related to Behavior " *Journal of Educational Psychology* 41 (1950) 129-48

Klineberg, O "Racial Differences in Speed and Accuracy" *Journal of Association and Social Psychology* 22 (1928) 273-77

Macfarlane, J "From Infancy to Adulthood " In *The Course of Human Development*, pp 406-9

______ 'Perspectives on Personality Consistency and Change from the Guidance Study " In *The Course of Human Development*, pp 410-15

______ "The Impact of Early and Late Maturation in Boys and Girls Illustrations from Life Records of Individuals " In *The Course of Human Development*, pp. 426-32

Moriarty, A *Constancy and IQ Change*. Springfield, IL Thomas, 1966

Murphy, L , & A Moriarty *Vulnerability, Coping and Development* Yale University Press, in press for 1975 publication

Olson, W C , & B Q Hughes 'Growth of the Child as a Whole " In *Child Behavior and Development*, R A Parker *et al* , eds , pp 199-208 New York McGraw-Hill, 1943

Rao, C R 'Perspectives on the Conference " In *Classification in Psychiatry and Psychopathology* Proceedings of a Conference in Washington, DC, November 1965 U S Dept of HEW, P H S , p 560 Chevy Chase, MD National Institute of Mental Health

Rosenzweig, M R , Edward I Bennett, & M C Diamond "Brain Changes in Response to Experience " *Scientific American* 226, 2 (1972): 22-29

Shirley, M M *The First Two Years*, 3 vols Minneapolis University of Minnesota Press, 1931

Sontag, L W 'Implications of Fetal Behavior and Environment for Adult Personalities " In *Biology of Human Variation*, E M Weyer, Editor-in-Chief Annals of the New York Academy of Sciences 134 (1966), Art 2

Sontag, L W , C T Baker, & V L Nelson *Mental Growth and Personality Development A Longitudinal Study* Monographs of The Society for Research in Child Development 23 (2, Whole No 68), 1958

Stolz, H R , & L M Stolz *Somatic Development of Adolescent Boys* New York Macmillan, 1951

Williams, Roger *Biochemical Individuality* Austin University of Texas Press, Reprint 1969 (original edition John Wiley & Sons, New York, 1956).

Part III
Some Examples of Meaningful Evaluation

The Prospect School: Taking Account of Process

43.

Patricia F. Carini

In the past, many schools have recorded only achieved outcomes. Thus, teachers and parents have typically accepted isolated end by-products of the learning process as representing learners' knowledge. If a child could state an answer to an arithmetic problem, the process by which he reached the solution — whether by guessing, counting on the fingers, or engaging in logical derivation — was neither discussed nor recorded.

To the extent that process was given consideration, as in the instance of asking a child to correct a wrong answer by correcting the process he used to reach that answer, process itself was construed to be a correct procedure. Therefore, as in the assessment and recording of end products, process too has usually been measured in terms of correctness rather than in terms of productivity, flexibility, meaningfulness or spontaneity.

For children in classrooms where their activity is minimal except for verbal responses, such as assessment is virtually all that is available to a teacher. And, in any event, end products are most readily available for measurement.

When classrooms are structured as environments that invite direct, active involvement in exploring concrete material, however, the *processes* underlying the child's organization of the world around him become more apparent to the

Reprinted from *Childhood Education*, April 1973, pp 350-56 insightful teacher With this awareness, the teacher may develop reluctance to maintain the old forms of achievement-oriented record-keeping and testing.

A crucial difference in recording and documenting *process* rather than *achievement* is that the former must be observed *over time* to determine a pattern, a matrix of descriptions of the learner's involvement. Using measures of end-achievements for assessment purposes, on the other hand, assumes that learning can be recorded and assessed as isolated elements independently of the meaning for the learner.

Speaking of the ongoingness of the learning situation, the importance of continuity, John Dewey (1938, p. 38) said

> *if experience arouses curiosity, strengthens initiative, and sets up desires and purposes that are sufficiently intense to carry a person over dead places in the future, continuity works in a very different way. Every experience is a moving force Its value can be judged only on the ground of what it moves toward and into . . . It is then the business of the educator to see in what direction an experience is heading. . . . Failure to take the moving force of an experience into account so as to judge and direct it on the ground of what it is moving into means disloyalty to the principle of experience itself.*

And Alfred North Whitehead (1919, p. 25), who also grasped thoroughly the organic nature of the complex we call "school," counseled that to assess the school we should not test the achievement of its students but sample the program according to its stated goals and philosophy

> *Primarily it is the schools and not the scholars which should be inspected. Each school should grant its own leaving certificates, based on its own curriculum. The standards of these schools should be sampled and corrected. But the first requisite for instituting educational reform is the school as a unit, with its approved curriculum based on its own needs, and evolved by its own staff.*

In any human process we only see, truly *encounter*, what is available to us from our own point of view. We must first of all acknowledge the relativity of the event — be it the learning process, knowledge itself, or the school — and the subjectivity of our own assessment.

The "objective" measure, such as a score or computerized datum, always is rooted in someone's point of view about what is worthy of validation. If all we can articulate from our point of view about a child's experience with arithmetic is the correctness of his information, then we can respond to and record only that aspect of his experience.

If we perceive the most important skill in reading to be word-recognition, we will find an "independent measure" of word-recognition and by applying it we will validate not the reading process but the single dimension that (unacknowledgedly) we deem to have greatest significance But by being willing to forego certainty and by accepting our point of view as part of the datum, we can accrue descriptive patterns that will permit the stable characteristics, the stable themes of an event to emerge.

* * *

The Prospect School, as a demonstration school for the state of Vermont, has had to confront the necessity for record-keeping, documentation and evaluation since it was founded in 1965 Beginning with an original group of twenty-five five- and six-year-olds from widely varied economic backgrounds, the school has evolved to include approximately ninety-five children, grades K-9. The population continues to reflect diverse economic backgrounds. Staff presently includes five teachers, and it is their records that provide the basic data for an ongoing research and documentation program. The design for the record-keeping, the research and the documentation, and the collation of the longitudinal data have been largely the responsibility of a small research staff (originally one person and, since 1971, three persons) serving in an adjunct relationship to the school.

What those of us working at The Prospect School have done within our program is to construe our record-keeping as a consciously temporal and subjective process In practice, we consciously examine and record processes — e g , social development, expressiveness, reading — descriptively so that any given process is available for interpretation over *time* according to the way it contributes to the child's total development or to the evolution of the learning environment. That is, the availability of descriptive records provides the basis for an ongoing examination and interpretation from a variety of points of view of such diverse processes as the physical and intellectual development of the individual child, the patterns in learning to read among a group of children, the relationship of early arithmetic skill to social development, the contribution of the individual's interests to the evolution of the total curriculum, etc. While the primary objective of these records is to contribute to the continuity of the individual child's learning experience, their secondary objective is to provide a documentary account of the evolving school program Finally, from the data and insight accrued through the records and documentation we have also designed instruments (Carini, Blake, Carini, 1969) to make independent and longitudinal assessments of underlying processes in children's problem-solving and thinking Through this instrumentation we hope to learn more of children's spontaneous formulations of their experience to better enable us to provide a learning environment that will support their continuing growth.

The basic records (Carini and Carter, 1971, Carter and Carini, 1972) we have kept to provide a continuing description of each child and to provide a documentation of program include the following:

- —Children's work, e g , drawings, photos, etc
- —Children's journals[1]
- —Children's notebooks or written work
- —Teachers' weekly records
- —Teachers' reports to parents
- —Teachers' assessment of children's work in math, reading, activities
- —Curriculum trees
- —Sociograms

How these records contribute to the understanding of each child's involvement in the learning process and to the documentation of the total program can be gauged from the following excerpts from the records kept of the involvement of a group of older children in an ongoing project (Carter & Carini, 1972):

Teacher's Records of Group Involvement in the Merck Forest Project [Excerpts]

September 24 — Today was beautiful We took the group to the Merck Forest — Chris drove David Sobel's van, beautiful weather and the kids seemed to enjoy it — only Heidi mentioned that she didn't like the trips too much but after lunch she played her recorder with some of the other girls and seemed happier

Hugh (Putnam) introduced the forest and we broke up

—Chris took Morris, Alec, Ned and Per up to Mount Antone

—Hugh led a group for the grouse expedition and wildlife—Elizabeth, Louise, Penny, Dru, Anna, Karl, Jacob and me.

—Charlie had a group learning about trees—Emily and Priscilla

—Only Heidi didn't choose an activity

After lunch — which everyone brought for themselves (except Morris and Jacob forgot) — David read some of *The Living Forest.*

his form of record-keeping applies generally only to children aged eleven or older

October 28 — I went to the Merck with seven students — Ned, Alec, Karl, Louise, Morris, Elizabeth, Per Chris also drove — it was a warm sunny day
— Per and Ned hiked with Chris to the hunting lodge
— Louise, Morris and I helped blaze a trail
— Karl and Alec did surveying
After lunch Louise and I made tea from yellow birch and mint We got back late but all went well.

November 12 — Went to Merck Forest — lots (four inches) of snow Looked for and discussed animal tracks in the snow Hugh came (Chris went with us), Karl, Emily, Louise, Ned and Alec

December 2 — Chris and I took a group of nine to the Merck Forest — three feet of snow with bright sun we hiked to the lodge on snow shoes and cross-country skiis — all seemed to enjoy it despite cold and exhaustion
etc

Teacher's Records on an Individual Child's Study of Mushrooms [Excerpts]

September 20 — At the Merck Forest Elizabeth got really interested in mushrooms — wants to study them next time out

September 27 — (Elizabeth) — The mushroom project has really captured her interest She has just been up to her armpits in work since we came back She and Anna and Dru made some really beautiful posters explaining what they had found.

Elizabeth's Identification of Mushroom [Excerpts from Her Notebook]

1 Growing in rich soil 2½ inches high, sort of shady spot Thick stem about 1 inch in diameter it is, while gills are white getting down toward the edge Also curves inward making cup shade, is about 2½ to 3 inches across the cap The color gets darker as it cups in, feels hard, the stem is white. When broken open, it is (or we think it is) a tricholoma They grow next to pine trees mostly This one was growing with birches and maples Animals may like eating this kind of mushroom

2 2¾ inches tall, the cap is 3 inches across It is a brown color, is very roundy The top flat part is darker than the rest The gills are white, the stem is pinkish white UNIDENTIFIED

Children's Observations on the Merck Forest [Excerpts from Journals]

Karl — no date — There is a stream on a mountain It's far There is a place where I went that has a pool deep and clear There are woods all around

Dru — October 7 — Went to Merck Forest, went to identify mushrooms

Hike & Hike & Hike
finally at
the lean-to

Death caps &
Gas teromycetes
are in store for
us today

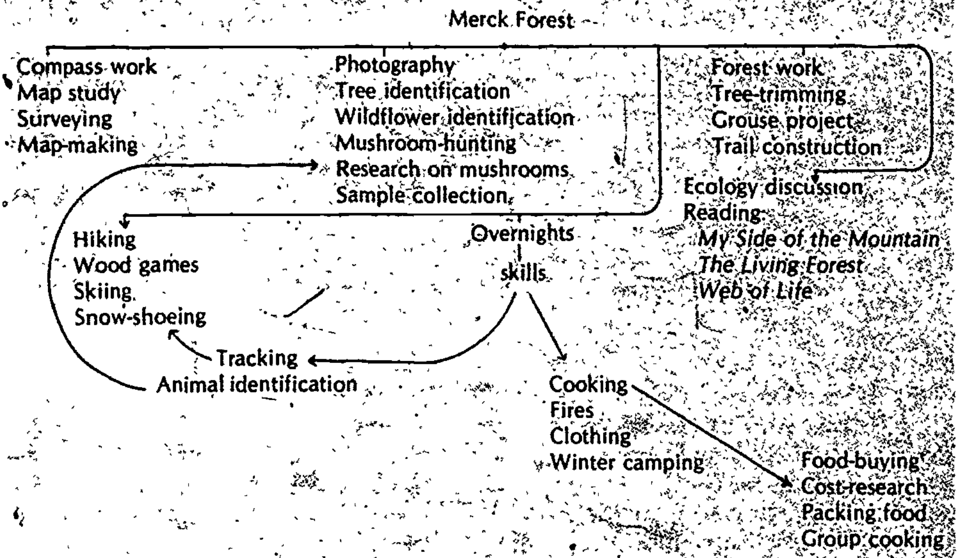
Penny — December 2 — Most of us went to the Merck today We had to go a different way today because the other was not plowed When we got there it was a very different place That's what I had thought, it was a big hill of white, and it had pine and spruce trees all over, and they had snow weighting down their branches The whole view was so beautiful There were some things in Mary's car that were Dave's that I wore They were like boots in one way and slippers in the other and they were very warm It took us a long time to get up to the hunting lodge. When we got there we only stayed long enough to eat lunch and then we had to go back to the parking lot But we had a lot of fun On the way back I wore snowshoes

Karl — March 21 — Last Wednesday, Thursday, and Friday we went to the Merck Forest on a three-night overnight I don't know what to say about it I went out with Per and Jacob They were throwing snow-balls at me Today I wrote a letter to the National Outdoor Leadership School about going on one of their expeditions If I got to go I would be able to hike and fish and all kinds of good things for five weeks And this afternoon I don't know what I will do but I hope I can keep busy reading or writing or something I also hope I can write in my journal tomorrow.

Implications

The final record reported can serve to identify a comprehensive curriculum available to the total group at the Prospect School in math (surveying, maps, shopping), natural science (mushroom research, grouse project, animal identification, etc), and physical education (trail construction, snow-shoeing, skiing, etc.). Even from these brief excerpts, the pattern of interests and

Schematization of the Curriculum Evolved Through the Merck Forest Project [Taken from the Teacher's Records]



47.

involvements for individual children is provocative. Does Elizabeth's precision in classification of the mushroom reveal an absorption in close observation, an interest in mushrooms themselves, or perhaps both? Can this interest be continued, given an inner connectedness, through close work with microscopes or more work with identification? In fact, the interest did extend to microscopic study, and if one were to look *back* upon many years of records of Elizabeth's involvements one would find a balance struck between the precise and mathematical on the one hand and the fantastic on the other — the latter expressed through reading patterns, writing and productions in clay and papier-mache

Thus, each project reported in the comprehensive record reflects the specific interests and involvements of individuals like Elizabeth. That Karl, a child new to the group, both writes "There is a stream on a mountain. It's far . There is a place where I went that has a pool deep and clear . . ", blazes trails, writes to the National Leadership School, and acknowledges his uncertainty as a member of the group in his tangential relationship to Jacob and Per gives depth and meaning to Dewey's (1938, p 38) statement that

Every experience is a moving force. It can be judged only on the ground of what it moves toward and into . . . It is then the business of the educator to see in what direction an experience is heading . . . Failure to direct it on the ground of what it is moving into means disloyalty to the principle of experience itself.

In what ways should Karl be supported? What will extend and deepen his interests without distorting them? One function of records is to provide insight to teachers that enables them to promote continuity of the learner's experience Documentation in comprehensive form of that continuing interest and involvement, in turn, constitutes a description of the program so that, in Whitehead's words, "the standards of the school can be sampled and corrected"

Records like those above, together with brief daily recording of skills (e.g., reading and numbers), have permitted us to document our curriculum over the past seven years, to report precisely to parents and others on growth of individual children, to study patterns of relationship among children, and to discover patterns of occurrences that reveal styles of learning; e.g., the bright, late-conserving boy who is slow to learn the technical skills in reading.

From these close descriptive observations and recordings within the classroom, we have also developed hypotheses and instrumentation for making longitudinal assessments of processes in language, thinking and problem-solving (1967, 1968). Our process of data-analysis over a four-year period, in combination with observations of children carrying out spontaneous activities, has resulted in analysis of the problem-solving tasks and resolutions of the tasks to form a scale (Carini, 1972). This scale, at present only partially complete, is the single most significant result of The Prospect School evaluation to date, as it has potential to assess the following dimensions in the child's relationship to the world: a) what any given task demands of the child; b) what the complexity and the availability of the perceptual or conceptual material are to the child; and c) what level of differentiation is reflected in the child's resolution of the tasks. This kind of assessment to provide a definition of the limits and plasticity of a developmental stage is needed if, as Wohlwill (1968) points out, we are to specify a developmental timetable.

Equally, this kind of assessment has potentiality for replacing tests of correctness with tasks that provide descriptive and diagnostic information about a given child's approach to a problem. Thus, a child might be asked to give as many ways as he can think of to make ten. By his formulation of the problem as limited, for example, to one operation (addition) and that operation carried out randomly (8+2, 5+5, 3+7), we are informed of his capacity to resolve a task requiring the formulation of a conceptual framework independently of perceptual materials. Were we to reformulate the same task to provide the child with the operations and the logical relationships (9+____, 8+____, 7+____, 10—____ = 9, 10—____ = 2), we might anticipate a higher level of resolution.

Thus both our records and testing can be addressed to process and description. And if at times we also find reason to record or test specific knowledge end-products, then this isolated information can be embedded into and given its appropriate weight within the total mosaic of the person's learning experience.

We are encouraged by our progress so far, but look forward to learning more with and from the children of The Prospect School.

References

Carini, Patricia F *Outline of Research and Evaluation Design* North Bennington, Vt. The Prospect School, 1972 (mimeo)

Carini, Patricia F , Joan B Blake, & Louis P Carini Progress Report III. *A Methodology for Evaluating Innovative Programs.* North Bennington, Vt.. The Prospect School, 1969 (mimeo).

Carini, Patricia F , & Jane Carter, Eds. *Record Keeping.* North Bennington, VT. The Prospect School, 1971 (mimeo)

Carter, Jane, & Patricia F. Carini, eds *Documentation of the Middle School.* North Bennington, VT, The Prospect School, 1972 (Xerox).

Dewey, John *Experience and Education.* Kappa Delta Pi Lecture Series West Lafayette, IN. Kappa Delta Pi, 1938 Used by permission of Kappa Delta Pi, an Honor Society in Education.

Wohlwill, Joachim F "Piaget's System as a Source of Empirical Research." In *Logical Thinking in Children*, I Siegel & F Hooper, eds New York· Holt, Rinehart & Winston, 1968.

Whitehead, Alfred N. *The Aims of Education.* New York: Mentor, 1929.

Marcy Open School: Feeding Back to Decision-Makers

Ruth Anne Aldrich

The basic issue behind any form of evaluation is accountability. Participants in Marcy Open School of Minneapolis, Minnesota, have actively considered the questions, "For what must a school be held accountable?" and "How can evaluation provide us with the information we need to develop an increasingly responsible program?"

49.

Such questions, which have obvious importance for assessing the effectiveness of *any* educational venture, have particular significance for those of us working to develop informal, open education approaches. The Open School at Marcy is a part of Southeast Alternatives, a federally funded five-year project (currently administered under the National Institute of Education). This Experimental School Project of the Minneapolis Public School District seeks to provide comprehensive change in education by offering a number of alternative school programs to children, parents and teachers. Marcy's open education program is one of the four elementary alternatives from which parents and children can choose. It features flexible curriculum, scheduling and age grouping for up to 330 children, ages five to eleven, with emphasis on helping the children learn to think and to make independent judgments.

Evaluation

Evaluation of the Southeast Alternative Project is both external and internal. A program of summative evaluation is being developed by a group known as the Minneapolis Evaluation Team (MET) which reports directly to the National Institute of Education. In this article we wish to focus on the program of *formative* evaluation provided by an internal evaluation group.

For the past three years I have been working as the internal evaluator of Marcy Open School. My role is to provide information to decision-makers that will help them improve program. Decision-makers may be individual classroom teachers who seek to identify and solve problems within their own classrooms, or they may be the staff as a whole or the Marcy Advisory Council—composed of parents, teachers and administrators—who are responsible for decisions concerning program and structure. I work with the individual or the group involved, to identify the program. As I gather relevant information, it is given to the people involved to be used as a tool toward making informed decisions.

Accountability

A major task the staff and Marcy Advisory Council have requested is the evaluation of general program achievement of its goals and its accountability for those goals.

To be accountable means to be held responsible for something over which one has control Schools have control over the goals they identify and over the environment they construct to achieve those goals. Whether or not they do achieve their goals with particular children is influenced by many factors that are not within the control of the school By the end of sixth grade, children have spent only approximately 7 per cent of their lives in school During that time families, peer groups and other societal agents have had a large influence upon a child's motivation for and ability to learn. Ultimately, only the children themselves can be responsible for what they learn.

Historically, the responsibility of schools has been misplaced so that they have been held accountable for what children learn. Several unfortunate effects result, most important of which is that schools become defensive about those things they can't control and are actually relieved of the burden for those things they *can* control—namely, the environments they create for children.* The Marcy Advisory Council and the Marcy staff as a whole have accepted a position of responsibility for the quality of the environment they create for children in the school.

Given this definition of responsibility for environment, and my role as internal evaluator, the general design for that evaluation is as follows: 1) selection by school participants of priority goals, 2) assessment of the environment of the school as it relates to those goals, 3) assessment of children's responses to that environment, and 4) feedback of information to relevant decision-makers.

I will describe further the implementation of this design at Marcy:

Selection of Priority Goals

Throughout the first year of Marcy Open School's existence (1971-72), the staff and parents identified and then later revised a list of seventeen goals for children. The Marcy Advisory Council, staff and Evaluation Committee (a standing committee of the Advisory Council) have chosen three of those goals as being of highest priority for evaluation

Goal 1: We want girls and boys to speak, listen, write, read and to deal with mathematical concepts effectively and confidently

Goal 2: We expect that children will take more responsibility for their own learning in all areas—social, academic, physical

Goal 3: We hope that children will increase their understanding of their individual rights and the rights of others

These three goals are accepted as being generally of greatest importance for the school. The other fourteen goals have not been abandoned, but for the school year 1973-74 were not considered as a focus for general evaluation.

Assessment of the School Environment

A school's creation of an environment includes the arrangements provided for use of time and space, the materials and activities made available for children, and the nature of the interactions that take place between adults and children. These are all dimensions over which the school has direct control and that should be consciously designed to facilitate children's growth in goal-areas. At Marcy Open School my work as internal evaluator has been to collect information about each

* Ruth Anne Aldrich, "Innovative Evaluation of Education," *Theory into Practice* 9 (Feb 1974) 1-4

of these dimensions through use of classroom observations, mapping, photography, teacher questionnaires and children's interviews. I have sought information about the environment as it relates to each of the three goal-areas stated earlier.

Time: How much time is given to formal instruction in goal-related activities instruction in reading skills, writing techniques, math skills, group discussion skills and techniques of responsibility? How much time is available for the children to informally use the skills they are learning reading for enjoyment, being read to by adults or other children, writing about experiences, applying concepts of individual rights in informal interactions with others, and discovering effects of being responsible for projects?

Space: How much space is available for goal-related activities? How readily available is that space? How undisturbed is that space?

Materials: What materials are available to the child expressive materials, books, magazines, writing equipment, listening equipment and recordkeeping materials? What range of ability and subject do those materials reflect?

Activities: What activities are provided to encourage growth in goal-related areas. language development, mathematical concepts of balance, design, calculations, understanding of the effects of not following-through on commitments, and increasing sensitivity toward self and others? How are those activities chosen?

Interactions: What is the nature of interactions between adults and children? Is there an expectation that children express themselves verbally, in writing and through artistic expression? Is an expectation communicated to children that they take responsibility for their own actions? Are they allowed to fail and to learn from that failure? Do adults express a respect for the rights of children and communicate an expectation that children will respect the rights of others?

This list is not exhaustive, but all of its dimensions are clearly within the control of the school—and school people should make conscious decisions about them.

Assessment of Children's Responses

Though the school must not be held directly accountable for what a child learns, because of other influences described earlier, a part of its accountability is in knowing how children are responding to the environment it has created and in modifying the environment if the children's responses are unsatisfactory. There is a distinct difference between knowledge of what a child learns and knowledge of how a child is responding. The question is not one of what a child *can do*, but instead of what the child *does*. This difference is reflected in Marcy's goals which state the desire that children will read effectively and confidently, rather than that they will know *how* to read, and the expectation that children will take responsibility for learning, rather than that they will know *how* to take responsibility. The evaluation, therefore, must look at the question of what children actually do within the school environment.

To facilitate gathering this information, I have selected a sample of 20 percent of the September 1973 enrollment at Marcy School from among children of each age group, children of racial minorities and of the racial majority, and children categorized as special education. Through classroom observation, children's interviews, photography and collection of classroom and school records, the following information has been made available for each of those children:

Goal-relatedness of activities during one day in October and one day in April
Participation in and products from various school interest-centers
Samples of weekly or monthly classroom-activity records
Participation in special education and counseling programs
Growth in language and math skills over a two-year period
Growth in affective characteristics throughout each one-year period
Standardized test scores in reading and math
Excerpts from end-of-year-reports to parents
Collected samples of art work and writing

This information has provided a profile of the involvements and growth in goal-areas of the sample of children

Feedback of Information to Decision-Makers

For evaluation to serve an ongoing formative function, we need to consider the information as it is collected, rather than at some pre-specified endpoint in time. Thus, feedback must be a constant process All information that I collect at Marcy is given to the teachers involved, as soon as feasible to do so The form of communication may be either written or verbal Specific details are included, identifying children, activities and times so that the information can be used in appropriate and meaningful ways for planning

In addition I make larger summary reports available to the total Marcy staff and Marcy Advisory Council A preliminary report is presented in midyear, and a report summarizing all the information collected for the year is presented in May In each case I generalize information so that individual classrooms and children are not identifiable to the reader, but I include sufficient detail so that decisions can be made on the basis of the data

Having received the information, the decision-makers themselves (be they individual teachers, schoolwide committees, Advisory Council members or parents) have responsibility to judge the success of the school in providing adequately for children's growth They are also responsible for making decisions about possible modifications of program and how to best achieve them Such decisions might involve rearrangements of space, changed grouping of children, or sharpening of teachers' skills through staff development. 52.

Conclusion

This model of school evaluation has been implemented in an open school. The implications of such a process are not limited, however, to open or informal education Schools should be accountable for what they provide for children, no matter what the structure of the program might be. Regardless of the setting, evaluation can serve the function of reflecting information about that environment

Schools should be growing, evolving institutions aware of their successes and designing change for their failures Through a realistic definition of accountability and an active program of evaluation, that process can become a reality for all schools

Children's Interviews

Nancy Ann Miller

Over the past several years the staff at the University of North Dakota's Center for Teaching and Learning (formerly the New School for Behavioral Studies in Education) has been working to develop new forms of evaluation As sponsors of a Follow Through approach, we have directed much of our evaluation effort toward developing interview instruments, with particular emphasis on community participation

This article will attempt to provide a general description of a children's interview, "And What Do You Think?," and to describe ways it can be used as a feedback tool for persons attempting to understand the classroom interactions of teachers and children

Development of the Interview

Work on the interview[1] first began in Chicago, 1970-71,[2] by a research staff having input into the process of "opening" four inner-city classrooms in two large Chicago elementary schools The staff soon realized that much of our discussion of where children were or what would be good for them was based on our interpretation of their classroom actions with little testing of our speculations against what the children themselves perceived about who they were and what they needed

We began evolving the questions by informally talking with children right in the classroom about what they were doing, wanted to do but couldn't, and so on. A useful reference in the process was the approach used by Piaget[3] when talking with children about their conceptions of natural phenomena—particularly his emphasis on so formulating a question that it allows children to respond from their own experiential and conceptual frameworks

Our first interviews were conducted by three staff members who had spent many hours in the classrooms as participant observers over a period of six to seven months prior to the interviewing Through their varied exchanges with the children, the questions and sequence were revised many times and, as a matter of fact, continue to be revised up to this time.

The present interview consists of approximately twenty open-ended questions about the child's activities and involvement in school, teacher-child relationships, peer interaction, and the child's view of the classroom as an overall learning environment

Children are usually interviewed outside the classroom, one at a time, to avoid disruption Interviews are taped initially and then later transcribed so that they can be analyzed more easily Length varies from thirty to sixty minutes We have found twelve children (from a class of twenty-five to thirty children) to be a manageable classroom-sample size—small enough to allow completion of the interviews and large enough to provide a good picture of the interaction patterns and activities in the classroom as seen from the children's perspective Confidentiality of the interviews needs to be stongly emphasized Throughout the interviewing process, care is taken to avoid identification of individual children and parents.

The effects of such variables as the setting of the interview and age, sex, race and familiarity of the interviewer must be considered. However, our findings tend to show that most important to the quality of the interview is the ability of the interviewer to structure relevant questions and to listen intently and nonjudgmentally to the child These skills, which require training and practice, are also those necessary for teachers in their interactions with children—and are too often overlooked in teacher-preparation programs

Uses of the Interview

The present children's interview has been used extensively at the Center for Teaching and Learning in both Follow Through classrooms and a wide range of other classrooms (grades 2-7) across North Dakota. It has served primarily to provide useful information to teaching staff—most productively when its results are given in a summarized feedback, along with results of teacher- and parent-interviews, in a team setting where discussion and clarification are possible [4]

An added benefit beyond the feedback to teachers is the opportunity for those in teacher preparation to take part in the interviewing. An often reported result of this experience is an increased sensitizing to children's experiences in the classroom

Interviewers and teachers are often surprised at the depth and range of the children's perceptions To increase this kind of two-way understanding and exchange between teacher and child is the major objective of the interviewing process

It is interesting that some children in the interviews have expressed a desire to know more about what the teacher thinks and feels For example, one third-grade child, whose teacher was under pressure to reverse the label of having the noisiest classroom in the school, responded to the question, "Tell me something you would like to know more about," with "How he [teacher] feels "

One area of the interview deals with how the child prerceives the teacher in terms of what he she does, likes to do or the kind of exchanges the child has with the teacher The following responses are from two children in two different classrooms to the question, "Tell me what your teacher does in your classroom "

> *She goes around and helps people, like if they have a problem and can t figure out their math, she comes over and she'll get the soma cube box and show them how to divide a fraction And when she plans with us writes on the board when we plan And, like when we have discussions she usually leads them and tells us some good stuff When we have projects, she comes up to us and brings us a book and shows us the right pages to look at and stuff like that*

> *He works [unintelligible] and he checks papers and people who are messing around, he tells them to quit and that and if they re throwing stuff, he hollers at them so they quit throwing it Sometimes when they do stuff they re not supposed to they have to sit down in their desk and that He checks quite a bit of papers and he checks the math books and we have like—every week we have like--I m in a B book and somebody's in an A book—they have sheets that you have to work with, and like one day he s got a B thing that goes up to the chalkboard, next day he s got a C, and so on like that*

When these same children were asked what their teachers like to do best and also if and when they talked to their teachers, their responses were

> **First child:** *Well, not to go over and like tell people to get to work all of the time She likes to help the different people that need help and that way she can go around and help everybody and give them ideas for planning and show them how to do a fraction a lot easier and—but a lot of times a lot of kids get noisy and she has to go over and tell them and she doesn't like that very well because then she can't help other people and show them what to do and easier ways and stuff*
>
> **Second child:** *I don't know You don't talk to him much, just if you have to get some work or something Sometimes you go up and talk to him about how you're working in your books and he tells you to come up and talk to him about what page you re going to work and all that*

Asked if the teacher talked to them about what they were doing, the first child responded

> *Yes, she talks to you If you're working on your SCS, or like building she comes over to you, and says, "Well, do you need a book on it, or do you need materials on it?" Some guys are making an ear harp—and she asked them, "Do you need some screws?" and the person says, Yes, and she asked some person to go uptown and get some screws for him, some other kids*

To the same question, the second child responded:

> *He just calls us up anytime and he asks us what you re going to do today and that stuff and I don t know—once a week or so he does it*

These radically different pictures of a teacher's role do not necessarily indicate that all the children in these two rooms perceive their teachers in the same way Patterns often emerge, however, and teachers may discover that, their own perception of their role in the classroom differs from that the children have.

Some of the exchanges are very warming The following dialogue leaves out the thoughtful pauses of the third-grade girl when responding to the question, "When is your teacher the happiest?"

> *[She] likes us to do our work—when we're good.*
> "How can you tell when your teacher's happy?"
> *I look at her eyes and I can tell that she's mad or happy.*

"How can you tell by the eyes?"

When they're happy their eyes look happy like they have a smile on their eyes When mad they look like a ball of fire—sometimes they get red

"They really get red?"

Uhhh-uhh

"Could I learn this—how to tell how people feel by looking in their eyes?"

Well my mother taught me [laughs] she could tell sometimes if I was lying by looking in my eyes But you could learn it too

"How?"

By looking straight in their eyes like you're going to hypnotize them

One can also gain information about how the children perceive sex roles When asked, "Are there some things only boys can do in your classroom?," children gave the following responses

They [boys] know more than you do but I know more math than they do

Only boys wrestle Why? *Because girls get hurt easier than boys can*

Saw wood Why? *Afraid we [girls] might cut ourselves or something*

Fewer responses are given to the question of what things only girls can do, but one such response was

The girls can knit they only sew by hand They [boys] usually have to be told to do their work and the girls don't

Although many factors influence children's perceptions of sex roles, teachers should be aware of the role their own actions play in perpetuating stereotypes that limit a child's growth

Another important aspect of classroom living is the nature of the interaction among the children and whether they view each other as resources and helpmates Most children prefer to work with others and, when asked how doing so helps them typical responses are

Like if you're having difficulties maybe they would know, help me understand better

Well, like if the other kid doesn't know something you can tell them and explain it and if you don't know something they can sort of explain it to you You can learn more things that way. You can help—you can help him learn stuff he didn't know and he can help you learn stuff you didn't know

Some children will say that whether they like to work with others depends on the specific activity involved One child, however, preferred working alone because, *Then I can talk to myself out loud*

Given the obvious relationship between the information one gets and the form and manner of asking a question, the interview can be useful too in suggesting questions teachers *should* be asking about their classrooms The problem of structuring the question so it does not suggest a response or limit children in responding from their own experience is indeed hard to appreciate unless one has consciously worked at it Here the process of the interview could be useful to both teachers and children—simply reflecting about information needed and finding questions to help elicit it Children can and should be more involved in this process of asking questions about their environment In collecting information, it is easy to get into the pattern of knowing the answer before one asks the question

We have also found in the interviewing that some children take much convincing that we really want to know what they think Generally, the children though sometimes shy or nervous will respond quite sincerely to the questions if the interviewer is genuinely interested Otherwise, some children will play the game of giving us what we want to hear [5]

Core purposes of the children's interview are to encourage children to say what they think and to stimulate those who listen to be affected by what is said. Unfortunately, many classrooms do not have time for such dialogue, as there is too much "work" to be done. Much of what happens in schools is based on "what should be done" as determined by someone other than the person doing it. Asking children what they think and responding to their answers will not only help them clarify their thoughts, feelings and needs but extend and strengthen them. In addition to practicing expression of their thoughts, the children are also themselves learning to ask questions and hopefully to influence change

In many classrooms children are *not* encouraged to question or to express their confusions Eliciting responses can be difficult; interest, acceptance and guidance are crucial ingredients of a supportive environment The following responses are from children asked what they would like to know more about or projects they are involved in

"Tell me something you would like to know more about?"

I would like to know everything in the whole world Do movie stars have to know everything there is to know?

"Do you think you could know everything?"

Well you would have to have a big encyclopedia with everything there is to know—but it would be so big that you would need a ladder to climb to turn its pages and it would take over 100 people to lift it

"I saw a lot of projects going on in your classroom Can you tell me what a project is?"

[Another child] *Well, like if you're working on the stock market you write reports on it and you read about it and want to know more about it and then like you have like a thousand and you give that to everybody and you have them buy stocks and you show them when—like you can be a stockbroker and show the other guy how to do it and plan and make a graph if the stocks go up or down and write reports on it—how it's working out And, you can show the other kids how to do it and then make stocks and it really goes nice for a while*

56. We need to learn more about space and how it affects us A typical response to the question, "Do you have a favorite place in the classroom?," is.

Yeah—one of the carpeted corners in the room Then it's really fun if you go back and there's not much noise there and it's a real nice cozy place that you can work in and there's music back there and you can work pretty fast It's really fun

While many children are bothered by noise and interruption, the answer is not simply to restrict talking and movement but to consider spatial arrangements that provide some isolation and facilitate quiet movement

Another point our interview experience has reinforced is that some children need a surprisingly long time to respond to a question and do not respond well to pushing One must work at "listening" and responding nonjudgmentally. Children often have a good understanding of their needs but have difficulty knowing how to find help in the classroom. For example, one third-grade girl said early in the interview that she wanted to learn more about math. Later, she replied to the question, "What would you like for the teacher to stop doing?," with *Giving us so much math* When asked to clarify this discrepancy, she responded that the teacher gave them a lot of math but never really showed them how to do it This perception may or may not have been shared by the other children in the class, but it is important to know about especially since we can continue along a direction for a long time before knowing how ineffective we are.

Teachers can use the questions from the interview informally in the classroom as a way of providing for continual feedback in conferencing and small-group discussions For some children the one-to-one exchange is the most comfortable setting, but they also need to develop the ability to express themselves in a group setting Questions can also be built into inquiry sequences. For example, the teacher asks for ideas for changing the room and then, in a nonjudgmental fashion, receives suggestions, guides toward a consensus upon one suggestion (or the number feasible), and has a group actually plan and carry out the change. This sequence involves getting feedback about the classroom plus helping the children become more articulate and able to plan and act. It is only one simple example of stimulating exchange between teacher and children and of involving children in classroom planning and change.

The interviewing process itself might be a profitable learning experience for teachers who go to another classroom and do some interviewing. Again, the problem of confidentiality has to be very carefully considered Certainly, visiting other classrooms and informally talking with children about what they are doing, their likes, dislikes, etc , can be a fascinating learning experience. If interviewing of a representative sample of children in a classroom were to be done, persons not directly involved in the classroom might more productively do the interviewing This kind of interchange too is a sensitive matter and should be approached so that it is a positive feedback process for the teacher. The interviewing should be done in the framework of a supportive staff development process

Presently members of our New School staff are attempting to relate the children's interviews to teacher's interviews, scaled on classroom dimensions described by Patton [6] Hopefully this would help us better understand the relationship between the classroom structure and happenings as perceived by the teacher as well as by the child A study is also in progress of the variance of children's responses in different classroom settings and the relationship between the responses and the setting.

Evaluation and accountability, though integral to learning and teaching, have become multiple-headed monsters in education. Often more money and energy are spent on judging success and failure and on producing packaged success than on supporting teachers and children in the difficult task of learning In this atmosphere, neither children nor teachers can admit their weaknesses for someone is always very willing and ready to hand them judgmental criticism

If the process of evaluation is to be a positive stimulus in learning, it cannot be a continually one-directional process with one side always setting the goals and the process for attaining them and having the power to determine success or failure

The kind of exchange the children's interview hopes to stimulate calls for increasing the active involvement of the children in their own learning and helping the teacher to better understand their experiences The children's interview would be sadly misused if it were ever to evolve into a "standardized" instrument used for accountability purposes As an evaluation tool, however, it can be used positively to help teachers and children "take a reading" of where they are in the process of learning and teaching.

Footnotes

[1] Most recent copy of interview is available from author

[2] Work began on the interview in Chicago under a Ford Foundation grant directed by Daniel Scheinfeld

[3] J Piaget, *A Child's Conception of the World* (Totowa, NJ Littlefield Adams), 1967

[4] 3d Quarter Report, *Teacher, Child, Parent Interviews*, submitted to National Institute of Education, Center for Teaching and Learning, University of North Dakota, July 1974

[5] The role of questions in the classroom is further emphasized by Francis Hunkins in his book, *Questioning Strategies and Techniques* (Boston Allyn & Bacon, 1972), in which he discusses both the importance and strategies of questioning in the classroom Too often classroom questions are based on a hidden agenda of right and wrong answers and are one-directional or teacher-to-child

[6] Patton, "Structure and Diffusion of Open Education A Theoretical Perspective and an Empirical Assessment," olished doctoral dissertation, University of Wisconsin, 1973.

Reflection in Teaching

Anne M. Bussis and Edward A. Chittenden

In what ways do teachers think about teaching? How do they conceive of the complex pattern of events that mark the school day? What assumptions do they hold about learning and development? What are the grounds for their planning, provisioning and evaluation? These and similar questions about teachers' beliefs and understandings become increasingly important with change in the direction of more complex classrooms and greater teacher responsibility for curricular decisions

During the past few years we have been interviewing teachers in order to study some of these questions Although in-depth interview methodology is not common in educational research, it fits well with a phenomenological view of man The phenomenological tradition in psychology historically has emphasized attitudes, beliefs, understandings, values and perceptions as major determinants of human behavior Applied to education, this view places greater emphasis on the importance of a teacher's internal perspective (thinking/valuing) than on the importance of a particular method or startegy in determining what happens in the classroom. Depending on the particular theorist one reads, this internal perspective has been referred to as "life space," "assumptive world," "belief system," "reference system,' and the like George Kelly's (1955) notion of "personal construct system" seems particularly appropriate, however, because it so clearly suggests an image of man as activist — an image that is central to all phenomenological theory (see the methodology article, pp 7-12).

Teaching-Learning Constructs

A personal construct means what the phrase implies — a personal construction or representation of some aspect of reality that is the result of an individual's interpretation of his world. A construct may be likened in some respects to a concept, it refers to objects or events that a person categorizes in his mind as somehow similar in meaning It is unlike a concept in that its boundaries — the range of experience to which it applies — are personally defined on the basis of each individual's past history. But constructs are not merely ways of interpreting and labeling what has happened They are also ways of predicting and anticipating events, as forerunners of action For example, the teacher who construes block-building primarily as large muscle exercise will make different predictions

The authors of this article have been involved in several projects related to questions of the role of teachers in educational change These projects have involved interviewing, visiting classrooms and participating in workshops The comments included herein are based, to a considerable extent, on the experience and findings of their principal interview study The full report of this study, funded by the Ford Foundation, is now being written

Reprinted with minor modifications from *Notes from Workshop Center for Open Education*, Spring 1974, pp 2 7 Copyright © 1974 by the Workshop Center for Open Education, Shepard Hall, City College, Convent Ave & 140th St, New York, NY 10031 Reprinted by permission

about this activity and undoubtedly act in different ways from one who construes it as the child's concrete representation of thought

To the extent that a person is open to feedback about the consequences of his action, predictions via constructs will sometimes prove correct and sometimes be found wanting Thus, the revision of constructs is seen as a function of a person's willingness to act on his own best judgment and his openness to feedback from the environment Simply "having a new idea or feeling," while important in its own right, is relatively inconsequential for affecting behavioral change Translating an idea into action and experiencing its consequences count for much more and constitute the basis of personal (as opposed to "academic") knowledge and learning This last assumption points up the obvious importance of experience in shaping personal constructs and suggests that, if significant progress in teaching is to occur, teachers need a quality of experience supportive of personal exploration, experimentation and reflection

What we have been interested in studying, then, are the personal constructs of teachers regarding the teaching/learning process, as well as their perceptions of major supportive and inhibiting influences on their professional development Our interviews were semi-structured and as informal as possible, encouraging the teacher to stress and repeat whatever priorities and concerns were uppermost on his or her mind The questions were open-ended and designed to elicit judgments, opinions and reflection During the first part of the interview, questions relating to the teaching process were discussed in some depth — such topics as room arrangement and the value of different materials, the organization of the day, the nature of instructional planning, the role of children's interests and emotions in learning, how to evaluate children's learning, and so on The second portion of the interview centered on the teacher's perception of supportive and nonsupportive influences on his or her professional development, including the role of advisers, other teachers, administrators, paraprofessionals, parents, workshops, course work and school policies.

Constructing Surface Content

One of the most interesting problems of the study has been to interpret the notion of "curriculm" in a psychological rather than logical way, in order to reflect the broad range of teacher understandings and meanings The important questions from a phenomenological view are How does the teacher conceive the curriculum? What is the teacher's personal "curriculum construct"? In attempting to deal with these questions, we have distinguished two levels of curriculum At one level, curriculum refers to the variety of activities the teacher plans for and encourages as well as those he/she may merely permit or tolerate Because this is what an observer would see going on in the classroom, we have thought of this as the *surface content* of curriculum.

Organizing Content

At a deeper level, curriculum has an *organizing content* which consists of the learning priorities and concerns a teacher holds for children. To oversimplify matters, what does the teacher want children in his or her classroom to know, do, feel, think or care about? What qualities of learning are valued and are trying to be promoted? As it turned out, these priorities and concerns were not too difficult to identify from the recurrent themes that permeated the interview - and they ranged from quite comprehensive ones to relatively narrow and conventional ones. For example, a concern with children knowing "what they are about — and why" (i e., a concern with the qualities of intention and reflection) was considered a comprehensive priority, whereas emphasis on children demonstrating basic skills and facts expected at a particular grade level was considered relatively narrow We should point out that "comprehensiveness", in this respect refers not only to the extent to which a priority engages the totality of children's cognitive/emotional resources, but also to the subsuming power of the priority Thus, a concern for intention and reflection generally subsumes a concern for children's acquiring essential facts and skills although these are not viewed as tied to a specific grade level

Making Connections

Having distinguished between activities in the classroom (surface content) on the one hand, and learning priorities (organizing content) on the other, another set of questions deals with the connections and interconnections between them First, does the teacher perceive *many*, *few* or *any* connections between his or her priorities and what is going on? For what purposes—in the teacher's mind—are children building a block castle, or looking at leaves through a magnifying glass, or making books filled with their own stories? Second, does the teacher conceive of a particular set of activities as serving only one priority, with a separate set serving another, and so on? Or, are activities viewed from many perspectives and seen as potentially valuable for a number of learning priorities? The nature of these connections and interconnections theoretically becomes a critical factor in the degree of psychological organization or structure that pervades a classroom

Inferring Priorities

60. It should be pointed out that teachers were never asked directly about curricular priorities Rather, these were inferred from the substance of comments made in response to the many questions thoughout the interview. In all, seventeen priorities were identified, eleven of these having a cognitive emphasis and six having more of a personal/social emphasis. Not only did teachers vary considerably in the number and nature of priorities for which they were coded, but also in the degree to which they seemed consciously aware of having priorities at all One particularly interesting finding is the way in which the curriculum construct appears to relate to a teacher's feeling of confidence Greatest uncertainty was expressed by those who planned for a wide variety of activities but whose priorities tended to be dominated by basic skills and good behavior concerns These teachers were experimenting with surface curricular changes, but had difficulty seeing connections between many of these activities and their major concerns While they believed in an abstract way that worthwhile learning should be going on during these activities, they were struggling to understand it and were frequently worried about it In contrast, teachers who planned a wide variety of activities and who held comprehensive curricular priorities could more often see connections between what was going on and what they were trying to promote — as could teachers holding relatively narrow priorities and not engaging in much surface curriculum experimentation

Examining Intuition

Obviously, we cannot do justice to the observation about teachers' curricular thinking and feelings of confidence or uncertainty in the short space of this article This is not the intent. Rather, our purpose is to point up some questions one can ask about teachers' thinking and to raise a basic issue How important is it for teachers to be able to analyze, reflect upon and articulate their basic assumptions about teaching? Although there are differences of opinion on this matter, it seems to us that analysis and articulation of the teaching/learning environment are important in at least two respects. First, analysis and articulation are critical components of the teacher's ability to communicate to others to administrators, to parents, to other teachers (and, in a much more subtle and complex way, to children) This certainly is the most commonly mentioned and widely debated sense in which analytic articulation can be seen as important But second, and less commonly discussed, analysis would seem important for the teacher's evaluation of his/her own efforts—especially when things start to go poorly or to stagnate What conscious frame of reference can the teacher bring to bear in an attempt to analyze what is happening? Can he/she look at the relationship of curricular concerns to surface content and begin to sort out priorities? When called for, can teachers examine the words and not the uses? We are not advocating that a teacher should be able to formulate a rationale or purpose for everything he/she does in the classroom, and we certainly are not denying that the immediacy and complexity of teaching demand heavy reliance on common sense and intuition The point is, can the intuition later be examined in a reflective way?

Toward Ongoing Professional Development

This issue has a direct bearing on one's view of professional development Perhaps the most prevalent notion of teacher development is one that implies that the engagement of a teacher's critical and conceptual faculties will be most intense during preservice training and the initial two or three years of experience, and after a certain level of mastery and efficiency has been acquired, inservice education is more a matter of maintenance and retreading In a recent study (Zahorik, 1973) in which teachers were asked to throw off the constraints of their *actual* teaching situation and to imagine an *ideal* teaching situation, findings suggested that the options many teachers actually perceive in teaching are options between two or three accepted methods to achieve a given goal As advocated by open education, however, when the options broaden to include not only non-traditional activities and methods, but the very goals themselves, then curricular decision-making becomes considerably more complex Commensurate with this view of teaching is a conception of professional development as ongoing — with the goal being to sustain the critical reflection and conceptual growth of teachers

REFERENCES

Kelly George *The Psychology of Personal Constructs*, Vol 1 New York W W Norton 1955

Zahorik J What Good Teaching Is *Journal of Educational Research* 66 (1973) 435-40

Selected Bibliography

Brenda S. Engel

On Evaluation in General

Combs; Arthur W. *Educational Accountability Beyond Behavioral Objectives*. Washington, DC Association for Supervision and Curriculum Development, 1972 A critique of behavioral objectives and learning theory as inadequate bases for evaluation Importance of holistic view of education, "personal meaning" of curriculum as key to student behavior, implications for teacher accountability.

Duckworth, Eleanor "The Bat-Poet Knows, Evaluation in Informal Education." *Music Educators Journal*, Apr 1974 70-72 A brief article emphasizing the need for adults to try to understand children's work, suggestions for questions a teacher might ask himself about a child and music Randall Jarrell's poem, "The Bat Man Knows," is offered as a parable

62. *Evaluation Reconsidered* New York Workshop Center for Open Education, 1973 (an occasional paper) Various articles including "Toward the Finer Specificity," by Lillian Weber, "The Horizontal Dimensions of Learning," by Anne Bussis & Edward Chittenden, "Toward a Shared Appraisal," by Charity James, "Documentation, an Alternative Approach to Accountability," by Patricia Carini, "Evaluating African Science Case in Point," by Eleanor Duckworth, "Report from North Dakota," by Vito Perrone

Hickey, M E "Evaluation in Alternative Education " *NASSP Bulletin*, Sept. 1973. 103-109 Some reasons for the difficulties common to evaluating alternative programs, suggestions for implementing broader, more comprehensive kinds of assessment

Notes from Workshop Center for Open Education. New York Workshop Center for Open Education, Dec 1972 Seven articles including "On Accountability," by Lillian Weber & Celia Houghton, "Record Keeping," by Bonnie Brownstein, "A Teacher's Log," by Janet Arndorfer, "Parent-Teacher Conferences," by Ann Hazlewood, "Parent Interviews," by Michael Patton

Scriven, Michael *The Methodology of Evaluation*. Bloomington Indiana University — Social Science Consortium, 1966 A distinction made between "goals" (a final judgement of worth) and "roles" (uses along the way) of evaluation, in reference here to curricular materials, detailed analysis of pros and cons of formative and summative models.

Shapiro, Edna "Educational Evaluation, Rethinking the Criteria of Competence " *School Review*, August 1973 523-48. A critique of generalizable program evaluations and an argument for research designed for particular situations — also for evaluation in the service of program improvement rather than as justification.

Stake, Robert E. "The Countenance of Educational Evaluation " *Teachers College Record*, Apr 1967 523-40 A clarification of some of the issues involved in educational evaluation the nature of the evaluation (descriptive and/or judgmental), what is to be scrutinized (antecedents, transaction, outcomes), bases for judgment comparison of programs or absolute standards) and purposes or uses.

This bibliography is by no means a general listing of references in the field of evaluation and testing The references cited have been narrowly selected for their particular relevance to the contents of this publication

Zimiles, Herbert "An Analysis of Current Issues in the Evaluation of Educational Programs" *Disadvantaged Child*, Vol II Headstart and Early Intervention, 1968 547-54 Operational evaluation (as opposed to, or preliminary to, outcome evaluation) suggested for innovative programs as more goals-related and process-oriented

____________ "A Radical and Regressive Solution to the Problem of Evaluation" New York Bank St College, 1973 Diagnosis of methodological weaknesses in recent early childhood evaluation designs, recommendation to shift emphasis from impact studies to assessment of educational environments

On Open Education

Bussis, Anne M, & Edward A Chittenden *Analysis of an Approach to Open Education* Princeton, NJ Educational Testing Service, 1970 A preliminary, basic statement by a research group in response to need for methods of assessing programs in open education, conceptual framework clarified and implications drawn for evaluation and research

Walberg, Herbert J, & Susan Christie Thomas *Characteristics of Open Education* Newton, MA T D R Associates, 1971 An examination of open education concepts taking off from the conceptual framework presented by Chittenden and Bussis in *Analysis of an Approach to Open Education* (see above listing) Characteristics identified through analysis of literature on the subject and questionnaires Includes several teacher/classroom assessment instruments

On Testing

deRivera, Margaret "Academic Achievement Tests and the Survival of Open Education" Newton, MA Education Development Center, 1972 An analysis of exactly how standardized tests serve to undermine and threaten the future of progressive school programs based on experience with the Philadelphia Follow Through

Karier, Clarence J "Testing for Order and Control in the Corporate Liberal State" *Educational Theory* 22, Spring 1972 108-136 An historical account of cooperation between large private foundations and the standardized testing establishment in twentieth-century United States in maintaining the "meritocracy" and controlling the socioeconomic strucstructure

McClelland, David C "Testing for Competence Rather Than for Intelligence" *American Psychologist*, Jan 1973 1-14 A critique of current test practices and inferences drawn from them that perpetuate a myth of meritocracy, suggestions for alternative kinds of tests

McGarvey, Jack "Standardized Tests, 5 Steps to Change" *Learning*, Apr 1974 24-26 An optimistic view of general disenchantment with standardized tests in pockets across the country, along with an outline of five steps to be taken to assist change

Meier, Deborah *Reading Failure and the Tests*. New York. Workshop Center for Open Education, 1973 (an occasional paper) An indictment of standardized, normative group reading tests, based on seven cultural biases which are identified and examined, an alternative suggested of longitudinal data-collecting

Meier, Deborah, Ann Cook, & Herb Mack *Reading Tests, Do They Help or Hurt Your Child?* New York Community Resources Institute and Workshop Center for Open Education Examples from actual reading tests selected to demonstrate some of the confused thinking, cultural bias and ambiguous illustrations common to such tests Comments by the authors

Piaget, Jean "The Right to Education in the Modern World." *Freedom and Culture* (UNESCO) New York Columbia University Press, 1951 Pp 69-116 A strong indictment (pp 84-86) of academic examinations within the context of a general statement on educational rights

Silberman, Charles *Crisis in the Classroom*. New York Random House, 1970. Mention of testing in this standard work includes a brief discussion of attitudes toward testing in gland and testing as a rating method in the U S A

Other Methods of Evaluation

Aldrich, Ruth Anne "Innovative Evaluation of Education" *Theory Into Practice*, Vol XIII, Feb 1974 1-4 School accountability defined as responsibility for instructional environment rather than for learning outcomes

Carini, Patrici F "Documentation, an Alternative Approach to Program Accountability" North Bennington, VT Prospect School An argument for self-reflective, process-oriented documentation based primarily on biographical/historical method Samples of record-keeping from the Prospect School included

Cohen, Dorothy H , & Virginia Stein *Observing and Recording the Behavior of Young Children* New York Teachers College Press, 1972 What to look at and how to record anecdotally, directed toward nursery and kindergarten observations

Duckworth, Eleanor R *A Comparison Study for Evaluating Primary School Science in Africa* Newton, MA Education Development Center for African Primary Science Program A description of the problems, procedures and results of the actual evaluation "A Field and Planning Study for the Appraisal of Reading in Open Classrooms" Princeton, NJ Education Testing Service, 1973 An analysis of the inadequacy of conventional reading tests and plans for the development of more useful and appropriate means of assessment By the Early Education group at ETS.

Hawes, Gene R "Managing Open Education Testing, Evaluation and Accountability *Nation s Schools*, June 1974 33-47 A description, with examples, of some contemporary alternatives to traditional evaluation

Perrone, Vito, & Warren Strandberg. "A Perspective on Accountability" *Teachers College Record*, Feb 1972 347-55 An argument for a broader, more inclusive basis for accountability (instead of the usual "hard data") in order to respond to the educational value of a variety of experiences

Documents

Marcy Open School, 1973-1974 Goal Evaluation Minneapolis Southeast Alternatives, Aug 1974 $1 A formative evaluation report by the school documentor, Ruth Anne Aldrich, contains description of school, rationale for evaluation design and various kinds of data on goals-related program

Final Report The St Paul Open School St Paul, MN , 1973 An outside professional report of one school year (72-73) based on a variety of instruments (cognitive and affective), observations and interviews Conclusions and recommendations set apart from findings [illegible] and sympathetic look at children [illegible] ego-building needs, how adults can [illegible]

COMM[illegible] through improving relationsh[illegible] parent-community [illegible]

[illegible] theory and practice. Describes models of [illegible] evaluation, helpful literature and open [illegible]

NEW VIEWS [illegible] educators offer practical and original ideas for relating school [illegible] projects, extensive classified bibliography. Joint publi[illegible]

[illegible] of how nine teachers engaged in quest to open up curriculum [illegible] for learning. Coordinated by Minnie P. Berson. [illegible] pp. [illegible]

[illegible] parenting concept beyond parenthood: involvement of other persons [illegible] community. Changing roles, single parents, child care abroad [illegible]

TEACHING FOR SOCIAL VALUES IN SOCIAL STUDIES. Looks at value clarification and formation as [illegible] of social studies instruction. Maxine Dunfee and Claudia Crump write a [illegible] 72 pp. $2.75

THAT ALL CHILDREN MAY LEARN WE MUST LEARN. LOOKING FORWARD TO TEACHING. Collection of [illegible] articles from Childhood Education, focus on the developing child as a person [illegible] producing healthy climates. 96 pp. $3

TRANSITIONAL YEARS: MIDDLE SCHOOL PORTFOLIO. Specific ideas for ways of working with the [illegible] classroom organization, curriculum, evaluation covered in fourteen useful [illegible]

WHEN CHILDREN MOVE FROM SCHOOL TO SCHOOL. Impact on children of changing homes, schools [illegible] and helps teachers and parents can provide. Has a child's [illegible]

YOUNG DEPRIVED CHILDREN AND THEIR EDUCATIONAL NEEDS. By Barbara Biber. What are [illegible] of deprivation on young children? School role in making up for [illegible]

[illegible] ASSOCIATION FOR CHILDHOOD EDUCATION INTERNATIONAL, 3615 Wisconsin Avenue, N.W., Washington, D.C. [illegible] Please include check or money order payable to ACEI for cost of [illegible] Only orders over $10 may be billed. A complete publications catalog including [illegible] will be sent [illegible] request.

Why are traditional evaluation procedures inadequate for educational programs concerned with process, content and context"? This timely and significant publication outlines a new frame of reference for meaningful evaluation

Contents

PART I Overview

PART II Testing Problems and Possibilities

PART III Some Examples of Meaningful Evaluation

[illegible] jointly by the
[illegible] for Childhood Education International 3615 Wisconsin Ave N.W.
[illegible] D.C. [illegible] and Citation Press 50 W 44th St New York N.Y. 10036
[illegible]