4 *Emerging Models for Educational Evaluation*[1]

In this chapter I will develop three ideas that seem to me to hold promise for improving the process of educational evaluation. These ideas deal with some of the issues surrounding the character and functions of educational objectives, the variety of outcomes that one might reasonably expect schooling to yield, and a set of methods that appear useful for evaluating the character and effects of school programs. That the relationship between objectives and evaluation is, at least theoretically, an intimate one is clear. Almost all writers on education generally and curriculum theory particularly emphasize the point that evaluation procedures should be related to the objectives one has formulated. Thus, any modification in either the content or the form of objectives can have important implications for the method and goals of evaluation. This paper describes some ways in which objectives can be conceived and provides the conditions necessary for expanding modes of educational evaluation.

I would like to say at the outset that the ideas I will develop here are in their infancy. I have not written about them in detail in professional journals; they are glimmers that the task of writing this paper has brightened. They are collectively, as Joseph Schwab might say, an invitation to inquiry rather than a rhetoric of conclusions.[2]

In many ways, the development of new and better ways to evaluate is counter to at least one of the major developments on the educational scene. I am of course referring to the rash of books and articles that have recently appeared chastizing the schools for being test ridden, impersonal, oppressive, indifferent to students as people, bureaucratic, and mindless. *How Children Fail, Compulsory Miseducation, The Open Classroom, The Lives of Children,* and *The Way It's Spozed to Be* are only a few of these books. And their diagnoses of the ills of schooling are antithetical to those found in the books published

one and two decades ago, which in their own way also chastized the schools. *Retreat from Learning*, *Quackery in the Public School*, *Why Swiss Schools Are Better Than Ours*, *Educational Wastelands*: the titles tell their stories. The schools, those who run them, those who prepare teachers for them, and those who work in them have borne the brunt of the attacks.

Yet it is because of these attacks and the kind of passionate reform they urge that the need for more adequate methods of educational evaluation becomes even more important. The educational conservatives of the late forties and early fifties introduced the theme that was to herald the curriculum reform movement. That movement, developed with the help of the National Science Foundation and the U.S. Office of Education, provided what some of the conservative critics wanted.[3] It provided a no-nonsense curriculum developed by scholars, implemented by teachers, and geared, at least initially, to the production of young scientists and mathematicians. Educational reform in the mid-fifties was seen primarily as curriculum reform.

That educational reform could be achieved merely through curriculum reform is a seductive aspiration. I wish it were that easy. If we have learned anything from the curriculum reform movement, it is that the problems that pervade our schools go well beyond problems of curriculum. This is not to say that the curriculum of a school, by which I mean the program it provides to students, is unimportant. School programs are important. But one must also realize, as the Holts, Friedenbergs, Dennisons, Goodmans, and Kohls have brought to our attention, that other aspects of the school are also important and, some claim, much more important than the formal curriculum.[4]

The radical critics have made salient the idea that the type of relationship existing between teacher and student is critical. They have pointed out that the organizational structure of the school teaches as surely as the lesson in a workbook or test. The reward system of the school – the covert, muted one – speaks loudly. In short, they have injected a dimension that was generally neglected in the cool, cognitive approach taken by curriculum reformers during the fifties and sixties.

The language and the perceptiveness of some of the reformers are persuasive and in many parts of the country are being heeded. In California, the free schools have become a movement. While their lives are short – their average tenure is about a year and a half – people continue to establish schools that are intended to provide radical alternatives to the public schools. New journals concerned with free

schools are being published. And in Berkeley, California, free-school people have elected a candidate for the school board.

Yet it is precisely because the language of the radical reformers is persuasive and because the movement for alternative schools is growing that the need for sound and careful evaluation is important. There is no virtue climbing aboard alluring bandwagons only to find after some pain and dismay that we have indeed been taken for a ride. Rhetoric is not enough when the policies that such rhetoric yields can affect millions of students and teachers.

In calling attention to the growing dissatisfaction with American public schools and the alternatives that are being created, I do not wish to imply that my concern with developing more adequate models for evaluation rests solely upon the desire to assess the consequences of these so-called free schools. That task is important, but it is only a part of the problem. The school programs to which 52 million children are exposed annually are the prime subject that warrants attention. The need for more adequate methods of evaluation is surely as great for this population as it is for that segment of the school population attending free schools.

There is no doubt in my mind that the evaluation movement in education, especially as fostered by the efforts of those who worked on school surveys during the period from 1910 to 1920 and in psychometrics and test development during the First World War, made important contributions to the scientific study of education.[5] At a time when there was a need for more precise and sophisticated conceptions of schooling and teacher training, statistics and other quantitative methods were appropriate and useful tools. One should not forget that the first department of education in an American university was established in 1873; in 1920, education as a formal field of study and practice was still in its childhood. Through the efforts of such men as E.L. Thorndike, John Watson, Harold Rugg, and Charles Hubbard Judd,[6] the tools of research and, more importantly, the conceptions underlying research became a part of the armamentarium of the evaluator. Educational evaluation has grown up within the general field of educational research, and it is only recently that efforts have been made to distinguish between the two.

There was a period in the development of education when the family resemblance between educational research and educational evaluation was a virtue, but educational evaluation employing the premises and practices of educational research has some important limita-

tions. I would like to identify a few of these limitations in order to provide a context for the ideas I will develop later.

First, in the efforts that are made to evaluate the effects of a new program or method of teaching, inadequate attention is often given to distinguishing between findings that are statistically significant and those that are educationally significant. Differences between experimental and control groups can have no educational significance in spite of the fact that the differences are not random ones.[7]

Second, there has been a tendency to reduce educational problems into forms that fit research paradigms instead of finding research and evaluation procedures that fit the problems. The power of the ideas of correlation, analysis of variance, and random selection and assignment is enormous. But there is a tendency to conceive of research questions within the parameters such ideas provide instead of raising interesting questions and inventing fresh ways to answer them.

Third, there has been an overwhelming tendency to attempt to evaluate the effects of programs on student behavior, with very little attention paid to the assessment and description of the environment which creates such effects. This observation has been made most cogently by Lee Shulman in a recent issue of the *Review of Educational Research*:

> The language of education and the behavioral sciences is in need of a set of terms for describing environments that is as articulated, specific and functional as those already possessed for characterizing individuals.
>
> An example that is familiar to all educators is the continued use of such gross terms as 'deprived' or 'disadvantaged' to characterize the environments of many minority-group children. Labeling the setting as 'disadvantaged,' of course, communicates little that is meaningful about the characteristics of that environment. Educators seem unable to progress beyond such a simple dichotomy as 'advantaged-disadvantaged,' Reviewers and critics of research have long realized that even those few categories which attempt to describe environments, such as social class, have been remarkably ineffectual in pinpointing the educationally relevant differences in the backgrounds of individuals.[8]

Not only do I agree with Shulman's observations, but a review I did of the last two years of the *American Educational Research Association [AERA] Journal* indicated that experimental studies reported in those

volumes provide about three and a half to four times as much space to reporting the findings of the studies as they do to describing the treatment. Somewhere between one and a half and two and a half inches of copy is devoted to describing what it was that the experimenter manipulated. How can one be expected to replicate experimental studies when such a paucity of information concerning the characteristics of the treatment is provided?

A fourth characteristic of many of the efforts to evaluate the effects of schooling is the failure to recognize the difference between what students will do and what they can do. For example, achievement tests are given in contexts in which students know they must perform well. The set that is induced by the test administrator, the form the test takes, and the setting in which it is administered tend to elicit not what students will do in typical situations in their lives but how they can perform under artificial; circumstances. One revealing example of such a situation occurred annually during my years in elementary school. When I was a student at Lawson Elementary School during the 1940s I was expected, as were all of my classmates, to take a handwriting test on a yearly basis. Each year from grade 3 on the teacher would write on the blackboard in her finest cursive form the following: 'This is a sample of my handwriting on January 24, 1943. If it is not as good as it should be for a student of my age and grade, I will try with my teacher's help to improve by this time next year.'

I remember vividly sitting on the hard oak seat attached to a desk screwed into the floor, eight deep in rows of six. The white, slick, lined paper was before me with the blue-black watery ink at my right as I sat with scratchy pen in hand. Oh, how I worked! To copy those two sentences in my best hand was a venture as difficult as crossing the Niagara on a high wire. After twenty minutes of pain I was through – in my best hand. But what I turned in to the teacher was no more a sample of my handwriting than it was of the man in the moon's. If my teachers wanted to know how I wrote, all they had to do was to look at what I was writing on any school day.

All too many achievement tests have similar characteristics: concurrent or predictive validity is too often a neglected consideration. We ought to be concerned not only with immediate effects, but with long-term effects. What the child will be like a year after the course is over is a far more telling question than how he behaves on the final exam.

In this sense it might be more reasonable to conceive of dependent variables as independent variables. After all, what we want to predict is

surely more than test performance. Test performance can be, and in my opinion ought to be, looked upon as a predictor of the future. Using the standard conception, the dependent measure tells you whether the treatment has been effective during the treatment period. *Now* the educational question emerges. Given that it has been effective during the treatment period, does it last? Is it used? Does it make a difference in how people function? There is no doubt in my mind that such questions will be difficult to answer, but there is also no doubt that these questions must be addressed if we are to know if schooling is more than a game.

A fifth characteristic of experimental research which filters into evaluation practices is the extreme brevity of the treatment that is provided. Making important and enduring differences in people requires either a great deal of time or a very powerful treatment, something in the form of a peak experience or one that is traumatic. Neither peak experiences nor traumas are typical of our experimental efforts; thus, time is required to bring about changes of a significant and enduring variety. Yet the average duration of experimental treatment time per subject in experimental studies reported in the last two volumes of the *AERA Journal* is about forty minutes – forty minutes to bring about a change that is to have educational significance! (Everyone, I am sure, realizes that such changes require at least an hour!)

Now there is a good reason for the brevity of experimental treatments. Short treatments increase control, and control reduces confounding. When confounding occurs, the ability to explain is reduced. Yet paradoxically, the more controlled the experiment the more difficult it might be to generalize it to classroom practice, for it is precisely the lack of tight control that characterizes most classrooms. It seems to me we need longer treatment periods as well as more sensitive instruments with which to evaluate the programs that are provided in schools.

Yet despite the caveats I have enumerated, there is interesting work taking place in the field of educational evaluation. The development of criterion-referenced testing is useful even though several of Professor Ebel's[9] reservations are well grounded. The aptitude-treatment-interaction work by Professors Cronbach and Snow[10] also holds promise even though consistent interactions have not been found. The idea of aptitude-treatment-interaction is persuasive and makes psychological sense. Daniel Stufflebeam's[11] context, input, process, and product model is a more comprehensive conception of the loci for evaluation than has been previously articulated. Michael

Scriven's[12] contributions are important new ideas in the evaluation field. In short, good work is being done. I would like now to contribute to that work by explicating the three ideas I identified at the beginning of this paper.

Many of you are familiar with prescriptions concerning the use of instructional objectives in curriculum planning and evaluation. The rationale for their use is straightforward: one must know what it is that a student is able to do in order to determine the effectiveness of the curriculum. This idea was developed in prototype by Franklin Bobbitt in 1924, refined by Henry Harap in the late twenties, rationalized by Ralph Tyler in his famous curriculum syllabus in the early fifties, exemplified in the cognitive domain by Benjamin Bloom in the middle fifties, extended into the affective domain by David Krathwohl, and given extremely precise meaning by Robert Mager in the early sixties. I will not reiterate the limitations that such a concept has as a comprehensive view of educational planning. Many readers are familiar with the views I have expressed in various journals and monographs on the subject.[13] My effort to conceptualize the expressive objective was intended to provide some balance to what I consider to be an extremely narrow vision of what education is and how planning for it should occur. The expressive objective is an outcome of an activity planned by the teacher or the student which is designed not to lead the student to a particular goal or form of behavior but, rather, to forms of thinking-feeling-acting that are his own making. The expressive curriculum activity is evocative rather than prescriptive and is intended to yield outcomes which, although educationally valuable, are not prescribed or defined beforehand. The task of the teacher is to look back, as it were, to evaluate what happened to the student rather than to ask whether the student achieved '90 percent mastery of a set of items placed before him during a forty-minute period.' The expressive activity is one in which the creative personalistic use of skills gained in instructional activities can be employed, developed, and refined. The expressive objective is the outcome of such activities.

These ideas are not new. I have written about them before. Why then do I reiterate them here? Only to refresh your memory so that the distinction between the instructional and the expressive objective can be seen more clearly as I describe a third type of objective that I believe to be logically distinct from the latter two.

As I thought about instructional and expressive objectives, it occurred to me that neither of these types adequately fits the kind of tasks given to designers, architects, engineers, and commercial artists.

Product designers, for example, work for a client who generally has a problem – a specific problem – that he wants the designer to solve. He might say to the designer, 'I need a device that can be marketed for under fifty cents, which can be made on a vacuum press, and which ladies can use to carry cosmetics.' The problem that the designer has is to take the specifications that are provided by the client, specifications which define function but which do not provide a solution, and to invent an image that provides a solution within the parameters set by the client. In such a situation, the problem is highly delineated but the range of potential solutions is, in principle, infinite. Furthermore, there is generally little difficulty in determining the success of the solution. In this Type III objective – I do not have an appropriate name for it yet – the designer, or architect, or engineer must bring his imaginative resources to bear upon a highly specific problem but one that makes possible a wide variety of solutions. In distinction, the instructional objective provides the solution; what the student is to be able to do is specified in advance, and the objective and the student's behavior or product at the end of an instructional sequence are ideally isomorphic. In spelling or mathematics, for example, instructional objectives describe answers that are known in advance. In Type III objectives, although the problem is known the solutions are not. Ingenuity of solution, appraised on the basis of the parameters or specifications of the problem, is the ideal.

An example might make the use of Type III objectives clear in curriculum planning. Let us assume that a teacher or curriculum development group is working in the area of the social studies and is attempting to develop objectives and learning activities for children around 10 or 11 years of age. The topic being dealt with is the way in which the community handles the movement of people within its boundaries. The curriculum writers are interested in helping children understand that population density affects the type of controls imposed and that optimal solutions to movement need to be appraised by a variety of criteria, time, cost, aesthetic considerations, and so forth. To help children appreciate and understand the dimensions of such a problem, the curriculum writers decide to deal with problems of traffic control and traffic flow and formulate a Type III objective which asks children to improve the flow of traffic by modifying in some way the current traffic patterns near the school. These modifications should make auto traffic more efficient and time saving and should make it safer for students. No new streets can be built; changes must come about by modifying traffic flow on existing streets. A classroom

teacher using such an objective would, if he desired, set up teams of students to study this problem and to formulate potential solutions. Each team's solution will be presented and considered by the class as a whole. What the teacher looks for in evaluating achievement is not a preconceived fit between a known objective and a known solution but an appraisal, after an inquiry of the relative merit of solution to the objective formulated, in this case solving problems of traffic control.

With the expressive objective, neither the parameters nor the specifications are given. The student can define his own problem or task and create his own solution. Thus, of the three, the expressive is the most open, the Type III objective is less open, and the instructional objective is least open. But the distinctions between the three types of objectives are not, I believe, matters of degree but matters of kind. Type III objectives encourage the teacher to provide high degrees of structure in setting the problem but also encourage him to leave the avenues for potential solutions wide open. Within the specifications provided, anything that works well, works.

One can legitimately ask whether the distinctions between the three types of objectives I have described are simply an exercise in analysis or whether they have any practical utility for educational planning. I believe these concepts are heuristic, that is, they lead one to view curriculum decision making and evaluation in unconventional ways. They lead to fresh questions. With the three types of objectives, we can now examine a curriculum, one developed either nationally or by the classroom teacher, to determine the extent to which objectives of each type are provided and the degree of emphasis devoted to each. We can compare curricula in different subject matters to determine their use of such objectives.

Furthermore, we can now begin to examine evaluation tools such as standardized tests to determine the extent to which they provide for items or tasks related to these objectives, and if they do not, we can build instruments appropriate for such objectives. Finally, we can consciously begin to design learning activities within the parameters suggested by these types of objectives.

In addition to these tasks, we can enquire whether there is an appropriate rhythm among the types of learning activities implied by the three types of objectives and whether principles can be formulated that teachers might use to decide when to emphasize each type. In short, the distinctions I have drawn indicate more acute directions that can be taken in the construction and evaluation of educational programs. When we recognize that we do not need, indeed cannot

successfully have, a single, monolithic conception of educational objectives, we are in a position to generate alternatives in curriculum development and educational evaluation that a single view will not permit. Type III objectives, for want of a better name, can, I believe, provide a wider scope for such enquiries.

A second idea that I would like to discuss deals with an image of the types of outcomes that it seems reasonable to assume are the products of teaching. The dominant, if not exclusive, orientation toward evaluating the effects of instruction is one which is aimed at determining the extent to which objectives are attained. Objectives in turn are usually couched within some subject matter field, especially when it comes to the evaluation of academic achievement. Such a vision or model of evaluation fails, I believe, to attend to other, perhaps equally important, consequences of instruction. For example, it is part of educational lore that a teacher not only teaches a subject matter, but he also teaches himself. Those of us who have had the good fortune to have studied under great teachers know this in acute terms, but even lesser teachers teach themselves. How teachers attack a problem, what their standards of excellence are, their sense of excitement or boredom when they encounter a new idea, their expectations for deportment, their tolerance for ambiguity, their need for precision: these are all teachable characteristics that teachers inevitably convey to students during the course of their work. These effects one might call teacher-specific outcomes. Outcomes dealing with subject matter achievement are content specific.

Teacher-specific and content-specific outcomes are not the only ones that arise in the course of instruction. The student makes his own outcomes. As a result of his previous life history, his particular interests, his turn of mind, the angle at which he comes at things, the student, like all of us, makes his own meaning. Although a substantial portion of the meanings made during a course will be common to virtually all students in that course, each student will develop meanings that are unique. Each is likely to construct from that course ideas which are peculiarly his own. These outcomes are student specific. Seen in the image of a triadically divided circle (see fig. 1), one-third of that circle represents *content-specific outcomes*, a second third represents *teacher-specific outcomes*, and a final third represents *student-specific outcomes*. Content-specific and teacher-specific outcomes are likely to be homogeneous in character across students. That is, the characteristics and values that teachers teach by virtue of what they are, are in large measure common to most, if not all, students in class. Especially in

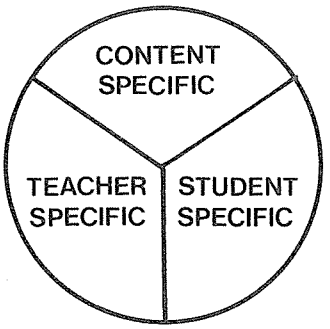

Figure 1

mastery learning are such common outcomes desired.[14] But along with the common inevitably comes the unique. The way a student personalizes meanings – the ideas he creates that are spin-offs from the content of the course or from the musings of the teacher – is also important. Indeed, in the long run they might be among the most important contributions of schooling. This dimension as well as the contributions that the teacher makes to students because he is a particular type of human being have been neglected aspects of educational evaluation. Yet if we are to understand the effects of the programs that are provided, surely these outcomes too must be examined. Thus, this triadic image of outcomes, bounded by a circle representing their unity, discloses the second idea of the three that I mentioned earlier.

Finally, I want to suggest a set of methods that I consider promising as a complement to the quantitative procedures now used so widely for educational evaluation. That set is the procedures and techniques of art criticism. The criticism of art is the use of methods designed to heighten one's perception of the qualities that constitute the work. The end of criticism as Dewey observed is the re-education of the perception of the work of art.[15] To achieve this end, the critic must bring two kinds of skills to his work. First, he must have developed highly refined visual sensibilities; that is, he must be able to see the elements that constitute a whole and their interplay. Second, he must be capable of rendering his perceptions into a language that makes it possible for others less perceptive than he to see qualities and aspects of the work that they would otherwise overlook. The critic, like a good teacher or book, directs attention to the subtle, he points out and articulates, he vivifies perception.

This vivification of perception which it is the critic's office to further is carried out by a particular use of language. It is quite clear that

our discourse is not as differentiated as our sensibilities. We experience more than we can describe.[16] Thus, what the critic must do is not to attempt to replicate the visual, dramatic, or musical work verbally, but to provide a rendering of them through the use of poetic language. The vehicles the critic employs are suggestion, simile, and metaphor. These poetic vehicles carry the viewer to a heightened perception of the phenomena.

An example of the use of such linguistic resources can be found in art critic Max Kozloff's description of a painting by Robert Motherwell:

> As an example, let me take a 1962 canvas, 'Chi Ama, Crede' (Who Loves, Trusts), in which a recurring flaw, a disproportion of the generalized over the particular, is held at bay. It is a twelve-foot frieze of wandering tan zones, surrounding two utterly eccentric, squirming turpentine blots of cool rusts, all this laid on in very close values. Basically the picture posits a contrast between restful, opaque fields that hold the surface and uneven strains that, with their shifting shadows, open up a translucent space and suggest a watery, organic agitation. But these rather hormonal blots are hemmed or even locked in by the ground at every stop of their fading perimeters. Here the artist reveals an overloaded liquidity that had dried up and been absorbed, and a mat, diffident facade that discloses an unsuspected strength. But suddenly, at one point, he withholds the paint tissue and, in an irregular glimpse of white canvas, flicks a whip of splatters that are almost electric under the murky circumstances. The whole thing glows as a vicarious pageant of his psyche.[17]

What Kozloff is doing here is using the connotative aspects of language to disclose the 'ineffable' content of visual-emotional experience. Kozloff writes of his efforts as a critic:

> For this, the most appropriate devices at my disposal have been innuendo, nuance, and hypothesis, because what is peripheral to direct statement in language is often central to a pictorial encounter or its memory. The more willingly this condition is acknowledged, the more readily is it possible to avoid the imputation of fact to something which is not 'factual,' while remaining faithful to that catalyst of our aesthetic life – credulity.[18]

Much of what goes on in schools can be illuminated by the tools of criticism. As a generic method, criticism is especially suited to articulating the unique and the personalistic outcomes that are so highly prized by those who complain of the school's impersonality. The reason criticism is so suited is because it does not depend upon the conventional application of class concepts for description and because it does not restrict itself to the primary surface of a situation; the secondary surface, that is, the situation's expressive and underlying qualities, is also a candidate for description and interpretation. Such a mode of evaluation has not, as far as I know, been employed in its full-blown form (although one of my students is using such a method to examine teaching as an art form). There are in the literature examples that approximate such an approach to evaluation:

> I see again in mind my rickety raftered rocky prefab that split the melting frost in the spring. With Sammy Snail wandering down upon us from the rafters, the sun thick tangible bars across the rising dust from the bare floor boards, the loud ever-moving, ever-talking life of the New Race, from corner to corner, from wall to wall, both on the floor and upon the desks. Tall towers rocking precariously, fantastic shapes in colour leaping from the ten-child easel, Little Ones in eddying figures dancing, the clay-births, the sand turning into a graveyard under passionate brown fingers, the water through with one-pint building wharves, bombers zooming on the blackboards, outrageous statements in funnily spelt words on the low wall blackboards, children singing, quarrelling magnificently, laughing for nothing, infectiously, crying for nothing infectiously, Waiwini's Little Brother wailing to me that somebodies they broked his castle for notheen. Bleeding Heart laughing his head off, the Tamatis' dog snuggling about for a cuddle, Pussy insinuating herself fastidiously, the Ginger Rooster scratching about ambitiously for culture, pictures of the meeting house and pa and the Ghost and of big-footed people kissing and words like shearing shed and beer and graveyard and wild piggy and lollies, tongues patrolling Maori lips over intensely personal writing, voices raised in exuberance, in argument, in reading, laughter, singing and crying and How-do-you-spell-Nanny. And our *floor*! You should see our floor! Round about the ten-child easel where the colour drips, it's prettier than the face of the countryside itself. You'd think Autumn himself had

> passed this way with his careless brush; slinging his paint about in his extravagant way. And noise . . . noise! And the whole show rocking like an overcrowded dinghy on high seas.[19]

Although autobiographical, such an account of classroom life gives the reader a vivid picture of its qualities and of Sylvia Ashton-Warner's attitude toward it.

But one might ask, 'Isn't such a method merely subjective?' Can nonquantified description using poetic devices be anything more than the expression of taste and liable to the grossest forms of unreliability? Not necessarily. In a very important sense, criticism is an empirical method. The adequacy of criticism is tested on the work itself. If what the critic describes cannot be seen in the work, his criticism fails to perform its function. In short, what he points out must be capable of being seen. Such a test is easier to apply to non-ephemeral works such as visual art and music than to the qualities that constitute classroom life, but such qualities are surely not so fugitive that their existence lasts only for a moment. Much of what is important in teaching and in learning is recurrent and regular. Criticism as a set of methods for analysis and disclosure can, I believe, make them vivid.

As a *complement* to the quantitative procedures we now use, such methods hold much promise. Their realization will require the creation of programs designed to prepare individuals with such skills. In a venture of this kind, departments of art, English, drama, and anthropology might be called upon for assistance. The promise of such procedures for dealing incisively with educational programs that might in the future become much more individualized than they are now is persuasive.

Notes

1. This paper was originally prepared for presentation at the Distinguished Visiting Scholars Program. Department of Educational Psychology. Michigan State University. I wish to express my gratitude to the students and faculty of that institution for their incisive and stimulating comments and critique of this paper.
2. SCHWAB. J.J. (1961) 'The Teaching of Science as Enquiry,' in *The Teaching of Science* Cambridge, Mass., Harvard University Press.
3. It is estimated that the National Science Foundation and the U.S. Office of Education have allocated well over $100 million for teacher training and curriculum development in sciences and in mathematics over the past ten-year period.

4. See JACKSON, P.W. (1968) *Life in Classrooms*, New York, Holt, Rinehart and Winston.
5. For a lucid account of the psychometric work during the First World War, see JONCICH, G. (1968) *The Sane Positivist: A Biography of Edward L. Thorndike* Middletown, Conn., Wesleyan University Press, pp. 356–82.
6. CREMIN, L.A. (1961) *The Transformation of the School* New York, Alfred A. Knopf. Inc., passim.
7. The notion that scientific enquiry is value neutral has been disputed by numerous students of science. For two cogent accounts of this problem, see BRONOWSKI, J. (1959) *Science and Human Values* New York, Harper and Row; and KUHN, T. (1962) *The Structure of Scientific Revolutions* Chicago, University of Chicago Press.
8. SHULMAN, L. (1970) 'Reconstruction of Educational Research,' in *Review of Educational Research* 40, no. 3, June, pp. 374–75.
9. EBEL, R. (1971) 'Criterion-referenced Measurements: Limitations,' in *School Review* 79, no. 2, February, pp. 282–88.
10. CRONBACH, L.J. and SNOW, R. (1969) *Individual Differences in Learning Ability* Washington, U.S. Office of Education.
11. STUFFLEBEAM, D. (1967) 'The Use and Abuse of Evaluation in Title III,' in *Theory into Practice* 6, no. 3 June, pp. 126–33.
12. SCRIVEN, M. 'Education for Survival.' mimeographed, available from author, 57 pp.
13. EISNER, E.W. (1967) 'Educational Objectives: Help or Hindrance?' *School Review* 75, no. 3, Autumn; pp. 250–60; and 'Instructional and Expressive Objectives: Their Formulation and Use in Curriculum', in POPHAM, W.J. et al. (ed.) (1969) *Instructional Objectives* American Educational Research Association Monograph no. 3, Chicago, Rand McNally and Co, pp. 1–18.
14. BLOOM, B. 'Mastery Learning and Its Implications for Curriculum Development,' in EISNER, E.W. (Ed.) (1971) *Confronting Curriculum Reform*, Boston Little, Brown and Co., pp. 17–49.
15. DEWEY, J. (1934) *Art as Experience* New York, Minton, Balch and Co., p. 324.
16. For a brilliant discussion of the relationship of experience to discourse, see POLANYI, M. (1966) *The Tacit Dimension* New York, Doubleday and Co.
17. KOZLOFF, M. (1969) *Renderings* New York, Simon and Schuster, pp. 169–70.
18. *Ibid.*, p. 10.
19. ASHTON-WARNER, S. (1963) *Teacher* New York, Simon and Schuster, pp. 220–21.