DOCUMENT RESUME

ED 078 913 PS 006 527

AUTHOR	Shapiro, Edna
TITLE	Examining Criteria for Evaluating Educational Programs.
INSTITUTION	Bank Street Coll. of Education, New York, N.Y.
PUB DATE	Mar 73
NOTE	9p.; Paper presented at the biennial meeting of the Society for Research in Child Development (Philadelphia, Pennsylvania, March 29 - April 1, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Cognitive Tests; Compensatory Education Programs; *Evaluation Criteria; Primary Grades; *Program Evaluation; Psychological Testing; *Response Style (Tests); Speeches; Student Evaluation; *Testing

ABSTRACT

The negative results to an assessment of a Bank Street-sponsored Follow Through program raised questions about conventional ways of assessing the effects of educational programs. Children in a Follow Through program using the developmental-interaction approach and children in conventional classrooms were observed and tested for comparison. The classroom observation data were quite different for the two types of classrooms, but there were no significant testing differences attributable to the program. Test responses are affected by unique factors such as examiner variables. Certain aspects of the test situation receive little attention but are especially relevant to the assessment of educational programs. Test responses also assess an individual's ability to generalize and transfer from one situation to another. Children in a conventional classroom, with its emphasis on the teacher's dominant role, are better attuned to the testing situation, in which the examiner questions and the child responds. There is, as well, greater uniformity of experience in the conventional classroom. In the child's passive role in the testing situation, demonstration of cognitive ability is quite dependent on language usage. The approach that may be appropriate for assessing some kinds of functions and groups may lead to misevaluation of others. What is needed is a flexible, responsive, diversified use of formative evaluation and a delay in summative evaluation until a program has had a chance to take effect. (KM)

Examining Criteria for Evaluating Educational Programs*

Edna Shapiro
Bank Street College of Education

My purpose today is to take a critical look at the way in which we usually go about trying to demonstrate the effectiveness of educational programs. The impetus to do this came from a study which yielded what are customarily referred to as negative findings. Resisting the impulse to bury the report, I have been drawn to reexamine the basic assumptions of this kind of study, and consequently to reappraise conventional ways of assessing the effects of educational programs.

Let me briefly describe the study, so that we can refer to it in developing the argument. It was designed to assess aspects of a Bank Street-sponsored Follow Through program, a program which has a comprehensive approach with multiple goals for both children and teachers. Teachers are expected to embrace new ways of teaching, not merely introduce specific instructional methods or materials. We have tried to spell out the basic assumptions of this approach elsewhere, and have described it as a developmental-interaction approach to the education of young children.[1] Learning and development are viewed as a function of both intellective and emotional processes. Children's interaction with the teacher, other children and materials are actively encouraged. A major purpose of the study was to try out techniques that would be appropriate for the evaluation of programs of this kind.

The Study

First grade children in three schools in the Bank Street FT program and three schools not involved in the program were compared. The pairs of schools were

---

*Paper presented at the biennial meeting of the Society for Research in Child Development, Philadelphia, Pennsylvania, March 1973.

1. Edna Shapiro and Barbara Biber, "The education of young children: a developmental-interaction approach." Teachers College Record, 1972, 74, 55-79.

located in three geographic regions--two in northern metropolitan centers, the third in a southern semi-rural area. The sample consisted of 150 black children from low SES family backgrounds. The data included: observation of classroom activities in each of the six classrooms, an individual testing session with each child, the children's school records (when available), teacher interviews and family background information.

Classroom observation data were gathered by two observers who made three observations in each classroom. The children were tested in individual sessions as close to the end of the school year as possible by black testers whom we had selected and trained.

The techniques used were aimed at tapping attitudes and expressions of feeling about the self, about school, about learning, and aspects of cognitive functioning that do not depend only on information, but on the disposition to respond, measures of divergent rather than convergent thinking.

Six techniques provided a range of measures and gave the children some variety in task requirements: a set of general interview questions about the children's interests, activities, and feelings about school; sentence completion items; Draw-A-Person; a self-rating technique; and two techniques adapted from Wallach and Kogan, Instances of a Category (e.g., "Tell me all the things you can think of that are round"), and verbalization to Line Drawings (a set of ten lines or patterns, each drawn on a card; the child is asked to tell what each "looks like").[2]

The records and ratings of classroom activities showed striking differences between the FT and comparison classrooms in each of the three pairs. The FT rooms were characterized as lively, vibrant, with a diversity of curricular projects and

---

2. Michael A. Wallach and Nathan Kogan, Modes of thinking in young children. New York: Holt, Rinehart, & Winston, 1965.

children's products, and an atmosphere of busy cooperative endeavor. The non FT classrooms were characterized as relatively uneventful, with a narrow range of curriculum, uniform activity, a great deal of seat work; teachers as well as children were quieter and concerned with maintaining or submitting to discipline. In each of the three geographic regions, the programs and teaching methods of the comparison classrooms exemplified a traditional educational ideology, with its emphasis on meeting conventional standards of achievement, the prerogatives of adult authority, formal expectations of competence and concern for inducting the child into the adult culture.

The children's responses to the techniques used in the individual sessions were analyzed in qualitative and quantitative terms. An analysis of variance was performed on all scores, with program, sex and geographic region as the main variables. There were no significant differences attributable to program, although there were differences between boys and girls and between the three regional groups.

How can we put together the dramatic differences in classroom behavior with the nul differences in test behavior?

Of course there were factors that may have obscured or mitigated against demonstrating differences between the FT and comparison groups in the test situation.[3] And of course we cannot be sure which of a host of confounding factors has been responsible, or whether all have contributed in some degree. But the differences examined in this study were pervasive and affected almost all aspects

---

3. It was not possible to control all variables, for example, the non FT children had had more previous school experience than the FT children. And in two of the comparison schools where children were grouped by ability, principals had selected classrooms which were top or second for the grade level. Also, the short duration of the program--it had been effectively operative for about four months when the children were tested--mitigated against demonstrating differences.

of the children's school experience.

The disparity between what happened in the classroom and what happened in the test situation raises questions about our basic assumptions. In this study, as in most, the child's responses in the test situation were considered critical. What children do in the classroom--the kinds of questions they ask, the kinds of activities they engage in--indicates not only what they are capable of doing, but what they are allowed to do. We cannot know whether the comparison group, given the same opportunities, would behave in similar ways. And we don't know whether, if the opportunity were removed, there would be any carry-over to a new classroom situation. Nor is it easy to separate the contribution of and effect upon individual children in the group. We had assumed that the internalized effects of different kinds of school experience could be inferred only from responses in test situations, and that the observation of teaching and learning in the classroom should be considered auxiliary information; its primary function was to document the differences in the children's group learning experiences.[4]

The rationale of the test, on the other hand, is that each child is removed from the classroom and treated equivalently; differences in response are presumed to indicate differences in what has been taken in, made one's own. But if we minimize the importance of the child's behavior in the classroom because it is influenced by situational variables, don't we have to apply the same logic to the child's responses in the test situation, which is also influenced by situational variables?

Tests, Testing and the Test Situation[5]

The individual's responses in the test situation have conventionally been

---

4. Patricia Minuchin, Barbara Biber, Edna Shapiro, and Herbert Zimiles, The psychological impact of school experience. New York: Basic Books, 1969.

5. I discuss some of the issues raised here more fully in an article entitled "Educational Evaluation: Rethinking the Criteria of Competence," to appear in the School Review, August 1973.

considered the primary means to truth about psychological functioning. Yet we often seem to forget that responses to test items are made in a unique interpersonal setting in which the rules of the game are carefully specified.

It is generally accepted that examiner variables (ethnic background, sex, manner and style) can have a powerful influence on responses in the testing situation. A number of studies have reported rather dramatic differences in obtained IQ as a result of optimizing testing conditions.

But certain aspects of the test situation which have received less attention are especially relevant to the assessment of educational programs, especially to comparing effects of different kinds of programs. Responses to tests also assess an individual's ability to transfer from one situation to another, the ability to generalize from information learned and attributes fostered in the classroom to the content and attitudes appropriate in the testing situation. There seems little question that the conventional schoolroom (and structured learning program) with its emphasis on the teacher's dominant role, on children's rather passive acceptance of what the teacher tells them and tells them to do, is much closer to the test situation than the more informal, open, program-centered classroom. Children are more tuned in to the teacher-question-child-answer kind of interchange, to the notion that there is a right answer and a right way to do things. In more open classrooms--certainly in those following the developmental-interaction approach--there is more exploration without specified outcome, more questioning, and more self-initiated activity. Different kinds of competences are fostered. The fact is that educational programs vary in their emphasis on teaching children to perform on demand, in the practice given in test-like activities and the value placed on the kinds of skills that are conducive to success in test-taking.

Furthermore, in conventional programs, there is much greater uniformity of experience in the classroom than there is in the more open programs. In the

comparison classrooms I observed, the children had much more homogeneous experiences in school than those in the FT classes where different children had different kinds of experiences. Susan Stodolsky also points out that when the children's experiences have been heterogeneous, one cannot consider the educational program a treatment, in the usual sense.[6]

While the situational constraints that operate in the testing situation apply both to adult and child, obviously the examiner is the freer agent and the one who determines the course of events. It is a situation of face-to-face interaction in which one party holds almost all the power. The major options open to the person being tested are to withhold, or give minimum or distorted responses. Usually what you get is the language of respect; the child tells you what he thinks you want to hear, in the terms he considers appropriate. In the test situation (as in the conventional schoolroom) the demonstration of cognitive ability is heavily dependent on language usage. The two kinds of competence are intimately connected. Yet in recent years, a wealth of data has shown that speech is extremely susceptible to situational influence.[7]

Dell Hymes'[8] concept of communicative competence is pertinent here--communicative competence requires being able to switch between parts of one's verbal repertoire, to be fluent and facile in many domains.

---

6. Susan S. Stodolsky, "Defining treatment and outcome in early childhood education," in Herbert J. Walberg and Andrew T. Kopan (Eds.), Rethinking Urban Education. San Francisco: Jossey-Bass, 1972.

7. Courtney B. Cazden, "The situation: a neglected source of social class differences in language use." J. soc. Issues, 1970, 26, 35-60; and William Labov, "The logic of non-standard English," in Frederick Williams (Ed.), Language and Poverty. Chicago: Markham Publishing, 1971.

8. Dell Hymes, "On linguistic theory, communicative competence, and the education of disadvantaged children," in M. L. Wax, E. S. Diamond, and F. O Gearing (Eds.), Anthropological Perspectives on Education. New York: Basic Books, 1971.

The study observers and the Bank Street program representatives reported that the children in the FT classrooms were enthusiastic, open and communicative. However (and especially in the southern sample), the free and easy verbal interchange quickly disappeared in a one-to-one interview or test situation, even when the interviewer was someone who was familiar to them. Both the comparison and the FT children were able to respond adequately to the questions and tasks. Their responses, however, were so similar that no group differences could be discerned. While there was some variation, the general impression was of well socialized six and seven year olds, rather passive and conforming, who gave superficial, often cliché responses, and who seemed to think that their task was to say what they thought the adult wanted them to say (In school we..."work," said 55 percent).

It is not that testing the ability to make transitions is irrelevant. Cognitive competence, like communicative competence, requires effective functioning in different domains, the ability to respond to the requirements of different situations, flexibility in dealing with different kinds of content, in different modalities. But when we are assessing the ability to switch, we should know that that is what we are doing. It makes little sense to assess cognitive (or any other) competence in one domain by setting up demands and expectations appropriate to another. And even less sense to assume that competence, or lack of it, in one domain means equivalent competence or lack of it in others.

Can it be assumed that everyone is motivated to do his best in a testing situation? And that best means the same for all? There is evidence that many children of minority background are not as achievement oriented as middle-class (white American) children. Nor do we know how age and developmental maturity influence a child's ability to adapt his responses and mobilize his resources in different situations. It is likely that as children mature they become more adept at reading situational requirements, but there is no evidence that this kind of learning

is an inevitable consequence of growing older. Rather, it seems likely that the range of situations that have been experienced and the kinds of face-to-face interactions that have been encouraged will affect differential responsiveness. But that is another way of saying that the test situation does not provide a uniform situation for children with different educational experiences, for children of different ages, or of different SES backgrounds.

What may be an appropriate situation for assessing some kinds of functions and some groups may well lead to misevaluation of others. The discrepancy between the test situation and other life experiences is greater for poor children from minority groups than for middle-class children. The standard test situation has built in lines of continuity with middle-class experience, as well as with conventional and teacher-centered structured learning programs.

Furthermore, when we ask a child a simple factual question, to which--let us assume--he knows the answer, many variables influence the speed and efficiency with which he will respond. His understanding of the question as asked, his desire to please, to show off his ability, the importance he places on being correct, his anxiety about making a mistake, his confusion about why you ask such a question... But now we all know that assessing cognitive proficiency is not enough. We want to know not only how much the child has learned or even how well he can apply his know-how to a new problem. We want to know how he feels about it; we want to know his image of himself, his sense of competence, his feeling of power and control over events; how much he likes school, his teacher, his mother; what his hopes and aspirations are... It seems likely that for such issues, the test situation is even more constricting than for the assessment of cognitive performance.

It wouldn't matter so much were it not that when we evaluate children's performance in test situations, we are almost invariably making inferences about

their capacities. Yet the experimental literature is rich in instances where variation in conditions of testing and specific task demands leads to differences in performance and consequently in assumed underlying capacity.[9]

* * * * * * * * *

I realize that I have skimmed the surface of a critique of the standard evaluation format. Many of these criticisms are not new. But evaluation goes on. We nod, and go out to construct more and cleverer tests. Perhaps spend a little more time selecting and training the testers. It seems to me that it is time to stop perpetuating the misevaluation of children and programs, time to give more than lip service to criticisms of testing programs.

This, then, is a plea for a more imaginative approach to educational evaluation--less rigidly psychometric, more flexible, with a more diversified use of different kinds of situations, a more fine-grained analysis of what goes on in classrooms and of the relation between type of program and of measurement devices. Formative evaluation, with a delay of summative evaluation until we can legitimately expect that a program has had a chance to demonstrate its effectiveness, and that children's responses to the rather special demands of the test situation have psychological and educational significance.

---

9. Morton Bortner and Herbert G. Birch, "Cognitive capacity and cognitive competence." American Journal of Mental Deficiency, 1970, 74, 735-744.