DOCUMENT RESUME

ED 057 112 TM 000 985

AUTHOR Dyer, Henry S.
TITLE Testing Little Children -- Some Old Problems in New Settings.
PUB DATE Oct 71
NOTE 16p.; Paper presented at the National Leadership Institute in Early Childhood Education, Washington, D. C., October 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests; Children; Cognitive Tests; Criterion Referenced Tests; Diagnostic Tests; *Early Childhood Education; *Educational Testing; Instructional Improvement; Intellectual Development; Intelligence Tests; *Measurement Goals; Observation; Preschool Tests; Screening Tests; *Testing Problems; *Test Interpretation

ABSTRACT

Problems concerning the misuse of tests by teachers and administrators are discussed. Two common misconceptions of the basic function of tests are: (1) the test as an incentive for study; and (2) the test as a learning experience. Rather, a test should be conceived of as a measuring instrument. More appropriate conceptions of tests are: (1) tests for selection; (2) tests for pupil guidance; (3) tests for evaluating instruction. It is necessary to disassociate the selective function of testing from its diagnostic function. Special problems are involved in intelligence testing since few people realize that intelligence can be taught. (MS)

TESTING LITTLE CHILDREN -- SOME OLD PROBLEMS IN NEW SETTINGS*

Henry S. Dyer
Educational Testing Service
Princeton, New Jersey

The subject of this talk -- "New Tests for Little Children" -- is, as you may imagine, a very large subject indeed. Its magnitude is attested by the fact that the Head Start Test Collection, which resides in the library at ETS, and is a cooperative venture with OCD, now contains 908 different tests, with more coming in all the time. And the Collection does not include an additional 2199 research instruments that are standing in the wings hoping to qualify for a regular part in the Collection itself.

Going literally from A to Z, the titles in the Head Start Test Collection range from the "ABC Inventory," which is a school readiness measure, to the "Zip Test" which is another readiness measure for migrant children with Spanish speaking background. Most of the tests in the Collection are quite recent, like the "Thomas Self-Concept Values Test" (1969), but some are old standbys, like the "Vineland Social Maturity Scale" (1936). The tests cover a lot of territory in both the cognitive and affective domain. Some of them are psychometrically respectable; some are trying to become respectable; and some are innocent of any known psychometric properties whatever.

Well, having paid my respects to the subject on which I was expected to talk, I shall now follow the usual practice of changing

*Talk to National Leadership Institute in Early Childhood Education, Washington, D.C. October 7, 1971

the subject to fit my own concerns. What I want to talk about is "Some Old Problems in New Settings" in the testing of young children. This is not such a big switch as you might think, because the spate of new tests, especially for young children, is itself part of two old problems -- namely, the problem of trying to keep up with all the new tests that are forever coming on the scene and the much tougher problem of trying to figure out which ones yield some dependable information about children and which ones are mostly flights of fancy, or reincarnations of old and sometimes unsatisfactory gimmicks. But there are some deeper problems as well that I think need attending to, and I want to discuss some of them.

I

For a short spell, about 25 years ago, I taught a course called "Educational Measurement 1." It was a required one-semester course for people who were bucking for a master's degree in education. Most of the students were already in teaching or administrative jobs, and a lot of them were working in elementary schools, or headed in that direction.

The over-all objectives of the course, in my mind at least, were (1) to get the students to acquire a decent respect for data, (2) to distinguish between dependable data and untrustworthy data, and (3) to secure some practice in generating good data for their own classroom use, particularly test data, and I mean "tests" in a broad sense to include not just paper and pencil devices, but any of various ways of observing and sizing up pupil performance.(1).

I learned a lot from teaching that course -- probably more than the students learned. One of the things I learned was that many of the teachers -- especially the teachers of young children -- became traumitized by anything that smacked of statistics -- even such simple statistics as means, standard deviations, and standard errors of measurement, which in my view are concepts that should and could be mastered routinely by every child before the age of 12. I'll give the kids until age 15 to get a firm grip on things like correlation and the fundamentals of regression analysis. And I'm only half joking.

Another thing I learned from the course was that most of those teachers -- and the administrators too -- tended to be very sloppy in the terms they used for characterizing pupils and their learning, for describing explicitly what goes on in the teaching-learning process, and for specifying the goals they thought they and their pupils should be reaching for. In short, they seemed to be quite unacquainted with rigor in thought and talk about educational matters, and indeed seemed to deplore the exercise of rigorous thinking altogether, as though it might conceivably interfere with the exercise of compassion in their dealings with the young. It took me the whole semester to get some of them to see that the rigors required by measurement are not incompatible with good teaching, in fact, that careful, dispassionate observation of how the children are getting along is a necessary ingredient of careful compassionate teaching.

To my great regret, I do not know how much effect the course may have had in reducing ignorance about educational measurement among the students who took the course, nor do I know how much it may have increased the rigor with which they handled their educational problems. What I do know is that that kind of course, which was aimed, I remind you, at the general practitioner, was a rarity then, and I suspect from what I read and hear and see that it is even more of a rarity today. I think this is too bad. In fact, I think it is dangerous, for if there is a flight from the discipline of measurement in the training of teachers and other educational personnel, the schools are in for even more trouble than they are now experiencing. Which is another way of saying that I think a good part of the cause of what Silberman calls the "Crisis in the Classroom"(2) is that too many educators have too little respect for the dignity of data.

II

Let me get a bit more specific by taking a look at some of the confused thinking that seems to accompany the testing of children. As I see it there are two common misconceptions of the basic functions of tests -- that is, uses to which they get put but which have nothing to do with their function as measuring instruments. Probably the most common misconception is the one that conceives of a test as an incentive to study; that is, as a club to coerce children into doing what it is supposed they would not otherwise do. Most teachers

have been taught, but too soon forget, that saying to children, "You must learn this or that because you are going to be tested on it" undermines the whole educational effort. It is wrong in two ways. It is wrong educationally because it gets the kids to thinking, from kindergarten on up, that the main reason for going to school is to learn to pass tests rather than to learn to make some sense out of the crazy world they are going to inherit.

It is wrong psychometrically, too, because if a child studies with the sole purpose of getting a good mark on a test, the chances are the results will be inflated, or maybe deflated, and therefore will not tell you what you really need to know to help him grow and develop. Furthermore, it produces a mind-set in young children that stays with them the rest of their lives and that makes almost impossible any effort to get, by means of testing, any valid indication of how the child feels about himself or his fellow pupils or his teachers, or of any of those other things we assign to the "affective domain."

A second common misconception of the function of testing is that the main purpose of a test is to provide a learning experience -- a series of exercises for the child to perform because they might help him develop his intellect, or his personality, or his character, or whatever. The idea is that tests are "good for you," like spinach.

This wrong-headed notion about tests probably does not have such serious consequences as the wrong-headed notion that tests are to be equated with prods and clubs. The child is not likely to be seriously

harmed by it. In fact, it can be argued that good tests usually contain good learning exercises. But to use them as such as to destroy their use as measurement devices, and I think it is important for teachers and other school people to come to realize this lest the measurement function of tests be lost in the shuffle.

For instance, one college teacher, writing on the uses of course examinations worries in print that the measurement function of tests will be overstressed by both pupils and teachers. "As a result," he says, "it may get in the way of the pedagogic function. This happens, for instance, when in our concern to set examinations that can be graded accurately and uniformly, we test only the more measureable academic capacities."(3)

It seems to me that this misses the main point of testing altogether. It is saying that accurate measurement of a child's learning should not get in the way of helping him to learn. It is saying, "Don't sacrifice good teaching to good testing." But this line of thinking creates a dilemma where, in my view, there ought never to be a dilemma. For good testing, that is, accurate measurement of pupil performance, is, I submit, absolutely indispensable to good teaching. You never know whether your teaching is any good, or more importantly, how to make it better, unless you have some reasonably dependable way of observing what effects it is having on the child's learning -- and by "learning" I don't mean just academic learning in the conventional sense, but learning to cope

with the whole broad array of developmental tasks that have been described by scholars like Erik Erikson[4] and Robert Havighurst.[5] The primary and most important product of any test is the kind of information that tells you how better to help the child learn. If, as a byproduct, the tasks in a test happen to provide useful learning exercises for the child, so much the better. But if the quality of the test as a meauring instrument is neglected, then the main point of the testing is lost.

III

Having disposed, I hope, of the two principal misconceptions of the functions of testing as clubs and character builders, let's look at some more appropriate conceptions of what tests are supposed to do, and the problems even these can generate. There are three functions that stand out: tests for selection, tests for pupil guidance, and tests for evaluating instruction.

Tests for selection have a long history. Such tests, like paper, printing, and gundpowder, were invented by the Chinese some 4000 years ago. As early as 2200 B.C. the emperor of China was using achievement tests to decide which of his officials should be promoted and which should be fired.[6] Later on they were real performance tests, too, covering the five basic arts: music, archery, horsemanship, writing, and arithmetic -- which I suppose made them reasonably appropriate "criterion referenced measures" of performance in government office. In any case, the use of achievement tests

or aptitude tests (which in my view are one and the same thing) for selecting people and sorting them out has persisted down through the ages and is unquestionably their most prevalent use throughout American society today -- even of course right down to the preschool and kindergarten level where "readiness tests" have become so predominant in delaying the entrance of children to kindergarten and the first grade.

This pervasive use of tests for selection is why tests in general are so frequently perceived as a personal threat, why cramming for tests is so widely tolerated if not condoned, why cheating on tests is so common, why minority groups complain, not without good reason, that tests tend to be biased against them and keep them from making it into the mainstream.

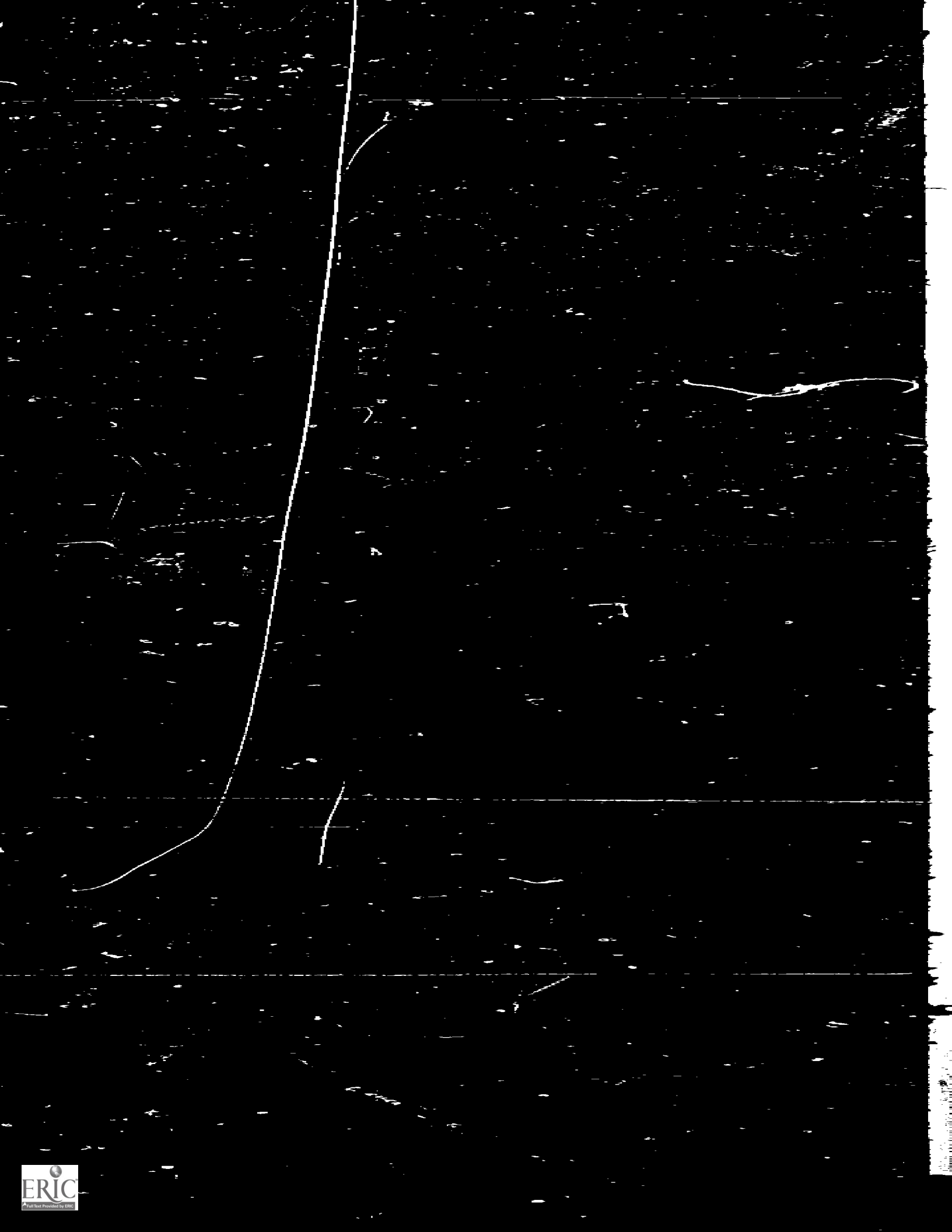
But let's face it. This widespread use of tests for purposes of selection, for deciding from kindergarten on up who will pass and who will fail, who will be winners and who will be losers, is not likely to go away in a hurry. For, whether we like it or not, it has become indigenous to the kind of competitive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

So, it seems to me, we are faced with an extraordinarily difficult problem. If we propose to use tests primarily for evaluating and improving instructional processes, rather than for slapping down children and teachers, how are we going to convince the victims that

we really mean what we say? Somehow we have got to separate the functions of testing in fact as well as in the minds of the children and the teachers affected. Otherwise the use of tests for selection purposes will tend to destroy their usefulness for guidance and for providing the feedback both the pupils and the teachers must have to make things go better in the classroom, or in the "open corridor," or wherever. For if children are learning from the very start to beat the tests or are learning to be turned off by tests -- as I'm convinced many children in the ghetto are -- the kinds of information the tests yield are not likely to be very helpful in pointing the way to teaching that is best adapted to the individual needs of children.

I have no easy answers to this problem, but I think it is one that should be squarely confronted and placed high on the agenda for training teachers and school administrators, and, I hasten to add, test specialists and psychometricians as well, since, as I read them, most of the developments of psychometric theory tend to center almost exclusively around the selection model.

Although the task of getting out of this bind will be difficult, it can be simply stated: it is one of getting past the usual rhetoric that is spouted in the name of educational evaluation and getting down to the hard business of finding practical and believable ways of dissociating the selective function of testing from its diagnostic or feedback function -- and it is here that I think the so-called



criterion-referenced tests, once we understand better what we mean by them, can play a part. At present the theory on which they rest is still somewhat confused and filled with controversy.(7)

IV

Now let's go back in history and look at another aspect of the problem of testing little children.

In the 67 years since Binet and Simon produced their first intelligence scale, we have learned a great deal that can and ought to be put to practical use in the testing of young children, provided educators can become sufficiently cautious in exploiting their possibilities.

It should be remembered, first of all, that tests in the Binet tradition have from the very beginning been primarily concerned with the testing of young children. The Binet-Simon scales of 1905 and 1908 were aimed at children ranging in age from 3 to 11. In 1912 Kuhlman's revision extended the scale down to the age of three months. Terman's Stanford revision of 1916 started with exercises for two-year olds, and most recently, the Wechsler Preschool and Primary Scale of 1967 focuses on children from four to six-and-a-half. The point I am trying to make is that, taken all together, the tests in the Binet tradition represent a vast amount of good empirical work in cooking up a great variety of developmental tasks for observing the mental functioning of children at early ages. And tests in the factor analytic tradition, culminating in the work of people like Guilford, has enriched the item bank still further.

It is useful to recall that Binet's original genius consisted chiefly in two fundamental contributions to the measurement of children's cognitive functioning: (1) the invention and development of a set of psychologically complex exercises that differentiated children who were doing well in school from those who were doing not so well and (2) the invention of a normative scale, the mental age scale, for ordering the children in accordance with their performance on a variety of these tasks.

The notion of a normative scale in its time was a real breakthrough. The trouble is that, as it has become more and more widely used in the schools, its basic nature and its limitations have not been well understood by the people who use it. In my experience, most teachers and other school people can't seem to get it through their heads that the use of years and months to measure a child's cognitive behavior -- or for that matter any kind of behavior -- is a fundamentally different operation from the use of feet and inches to measure a child's height. Which is to say that they have a hard time getting the idea that the units in a normative scale are not additive, that you cannot in logic say, for instance, that the cognitive functioning of a child with a mental age of eight is twice that of a child with a mental age of four. And of course when you move over to normative scales based on grades-in-school, you are in even worse trouble because grades-in-school are simply arbitrary inventions for organizing schools and slicing up curriculums

in any of a thousand different ways. People don't seem to realize that adding or subtracting or averaging grade-equivalent scores is only a little less absurd than it would be to add or subtract or average a set of social security numbers.

It was to Binet's credit that having devised the mental age scale, he stopped there. He did not invent that rich source of useless psychological controversy, the notorious IQ, which has distracted too many good minds from more fruitful educational endeavors and still does. It seems to me that the whole nature-nurture controversy, centering around the IQ, has become a disaster from the standpoint of trying to find better ways of teaching the young. Its widespread use in the schools has encouraged and reinforced the attitude among teachers and others that once you have figured a child's IQ at, say the age of five, and have put him in the "right slot," the whole academic establishment can rest on its oars and let the inevitable stream of academic routine do its work. The IQ tends to give teachers the notion that they can hide behind the numbers and thereby be relieved of any pressing obligation to keep studying their pupils to try to determine the manifold ways in which their behavior is changing and developing as a consequence of their school experiences.

The reversal of this attitude has been slow in coming, which is to say that it has taken a terribly long time for Jean Piaget's ideas to begin to percolate into the thinking of American education

to the point where teachers may, hopefully, become perceptive child-watchers and not just textbook watchers or, worse still, IQ watchers. Piaget's lectures on the language and thought of the child, which were delivered in 1922, represented a radical departure from the Binet tradition in the testing of children. Piaget began his investigations, you remember, by simply watching young children and observing the patterns of behavior that emerged as they interacted with the phenomena of their world. He was concerned with getting behind the numbers to try to infer what was actually going on in the minds of the children. This was another kind of breakthrough. In his book on The Psychology of Intelligence, Piaget himself has summed up the contrast between his approach and the Binet approach in two cogent sentences:

> It is indisputable, [he says] that these tests of mental age have on the whole lived up to what was expected of them: a rapid and convenient estimation of an individual's general level. But it is less obvious that they simply measure a "yield" without reaching constructive operations themselves.(8)

It is of course these "constructive operations" -- the different ways children perceive and think about and interact with their world that must be observed and understood as precisely as possible if teaching is to meet the child where he is and shape the school experience so as to maximize his development. But it is not until quite recently that any systematic attempt has been made to incorporate Piaget's ideas into testing procedures that could be used by teachers

to enhance children's learning. As far as I know, the first such attempt was the project that ETS undertook in some of the elementary schools in New York City back around 1964. It carried the title Let's Look at First Graders,(9) and its purpose, essentially,was to help first grade teachers observe and assess the specific aspects of the intellectual development of each pupil as accurately and objectively as possible so that they could teach them better.

But this sort of thing has been an uphill struggle, and as far as I can see, it has not yet made much of a dent. I remember talking a few years ago to a group of teacher-trainers about this project and whipping up quite a lot of enthusiasm about it. After I got through, one of the people there said, "You know, this is great stuff, and I've been trying myself to get my own student-teachers on top of it. But they just can't seem to grasp the Piagetan concepts or to apply them in their own classwork."

In short, it appears terribly difficult to make most school people believe that intelligence has many facets which can be observed and assessed if one pays close enough attention to the behavior of the pupils. The reason for this hang-up is that too few people realize that intelligence can be taught and that the primary business of schools and school teachers is to teach it -- and not just hide behind the numbers generated by paper and pencil tests concocted by somebody else. Such numbers, if properly used,

can be helpful, but I submit that the heart of the process of educational measurement ought to be thought of as the disciplined observation by teachers of the behavior of their pupils, and until we can get this idea across, most of the teaching, not to mention the testing, of little children will continue to be intolerably blind in its operations and dubious in its effects.

I think we can do better. I think it is up to you to see that we do.

References for Anyone Who Wants to Go a Little Deeper

(1) Dyer, H. S. "Educational Measurement--Its Nature and Its Problems" in Berg, H. D. (Ed.) Evaluation in Social Studies Washington, D.C.: National Council for the Social Studies, 1965. pp. 21-46.

(2) Silberman, C. E. Crisis in the Classroom: The Remaking of of American Education. New York: Random House, 1970.

(3) Beer, S. H. "Examinations as Instruments of Teaching" in Bramson, L. (Ed.) A Collection of Essays by Members of the Harvard Faculty. Cambridge, Mass.: Faculty of Arts and Sciences, Harvard University, 1963, pp. 71-87.

(4) Erikson, E. H. Identity: Youth and Crisis. New York: W. W. Norton, 1968, pp. 91-141.

(5) Havighurst, R. J. Human Development and Education. New York: Longmans, Green, 1953.

(6) Du Bois, P. H. "A Test Dominated Society: China 1115 B.C.-1905 A.D." in Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1964, pp. 3-11.

(7) Jackson, R. Developing Criterion-Referenced Tests. Princeton, N.J.: ERIC Clearinghouse on Tests, Measurements and Evaluation, 1970.

(8) Piaget, J. The Psychology of Intelligence (Piercy, M. and Berlyne, D.E., translators). Paterson, N.J.: Littlefield, Adams, 1963. p. 153.

(9) Educational Testing Service. Let's Look at First Graders: A Guide to Fostering and Understanding Intellectual Development in Young Children. New York: Board of Education, 1965.