

#### CHAPTER 1

# The Need to Move Beyond Standardized Methods

[A]ny fundamentally new approach to a scientific problem inevitably leads to new methods of investigation and analysis.

Vygotsky, Mind in Society

How can we discover what a child knows? Consider two different methods. The traditional approach is to investigate the child's thinking by employing tests or standard tasks. The practitioner administers IQ tests, language tests, achievement tests, and the like in a standardized fashion. All children receive the same set of questions or test items in essentially the same way. Responses are not difficult to score, and test results can easily be computed and compared. Similarly, the researcher gives all subjects the same series of tasks, each presented in a uniform fashion. The resulting behaviors can be coded with adequate reliability, and results can be quantified and analyzed with apparent precision.

The method of standardized testing is extremely influential, dominating the assessment and study of children's thinking. Indeed, this method has virtually come to define what we consider to be "scientific." Use of the standardized method is the dominant procedure in research, as can be confirmed by even a cursory examination of journals like *Child Development* or *Developmental Psychology*, and it is the typical procedure employed in the assessments conducted by practitioners, for whom various psychological tests are indispensable tools.

Consider, however, a radically different approach to understanding what a child knows – an approach that I will argue makes more



#### Entering the Child's Mind

sense than the traditional method. In a clinical interview, the examiner begins with some common questions but, in reaction to what the child says, modifies the original queries, asks follow-up questions, challenges the child's response, and asks how the child solved various problems and what was meant by a particular statement or response. The interviewer makes "clinical" judgments concerning the child's motivation and personality and uses these judgments to modify the questioning in appropriate ways, perhaps pressuring one child but treating another with kid gloves. The examiner is constantly engaged in interpreting the child's response so as to follow up on it in an incisive manner.

In employing the clinical interview, the examiner – practitioner or researcher – literally treats each child differently. Indeed, the clinical interview is *deliberately nonstandardized*, thus violating the central tenet of traditional testing. If the traditional method depends on uniformity, the clinical interview thrives on individual variation. In a collection of clinical interviews done for the purposes of either practice or research, it is possible that no two children are treated in an identical fashion and that no two children receive an identical collection of questions. The clinical interview is often considered to be preliminary, sloppy, imprecise, lacking in rigor, "unscientific" – in short, unsuitable for objective practice or rigorous research.

This book is an attempt to show that for certain key purposes the traditional methods of standardized testing are inadequate. Based on outmoded theory, standardized methods often fail to provide adequate insight into cognitive function; they are not effective techniques for understanding the processes of thought. By contrast, the clinical interview method offers a useful and seldom understood alternative to standardized testing. Although deliberately nonstandardized and difficult to employ, the clinical interview method can provide both researcher and practitioner with deep insights into children's thinking. Indeed, this nonstandardized, nontraditional, and, in some circles, controversial method is the "scientific" method of choice for the purpose of entering the child's mind.

Because the method of standardized administration has been so influential, it deserves a serious critique. This chapter therefore describes the method of standardized administration, explicates the as-



#### The Need to Move Beyond Standardized Methods

sumptions underlying it, and finally points out flaws which weaken the method's usefulness for examining conceptual frameworks and strategies. Although useful for some purposes, the method of standardized administration is not effective for others. It should not be allowed to provide the exclusive definition of what is "scientific" in research and practice.

# THE METHOD OF STANDARDIZED ADMINISTRATION

Consider first the basic procedures involved in the method of standardized administration, and second its underlying rationale.

### The Basics of the Method

Suppose that I want to gain insight into such aspects of the child's mind as changes in moral judgment from childhood to adolescence, a student's conceptual abilities in mathematics, the differences between boys and girls in logical reasoning, or the comprehension abilities of a child diagnosed as suffering from "learning disability." To answer questions like these, the traditional method is to employ a *task*,<sup>1</sup> or a series of tasks, which I will call a *test*, with or without norms. For example, if as a practitioner I am interested in examining processes of logical reasoning in an individual child, or if as a researcher I want to study age differences in logical reasoning in groups of 4- and 6-year-olds, the traditional procedure would be to

- develop a test involving a series of logical-reasoning problems or tasks ("Jane rides a bicycle. Bicycle riders are athletic. Is Jane athletic?");
- develop instructions which make clear what needs to be done ("I want you to tell me whether the answer to the question is yes or no");
- use procedures for establishing rapport and motivating the children ("We're going to have a lot of fun playing some games today. Listen carefully to the stories I tell you.");
- present the problems in as attractive a way as possible (perhaps use a colorful picture of Jane on a bicycle as part of the procedure);



#### Entering the Child's Mind

- make sure that the physical surroundings are comfortable and conducive to diligent work;
- use a checking procedure to determine whether the subject pays attention and seems to understand the basic information in the problems ("What does Jane do?");
- administer the entire procedure in a standardized fashion, perhaps with several such problems or tasks given in the same order to all subjects;
- limit the children to responses that can easily be scored in an objective fashion, such as simple "yes" or "no" responses (so that independent judges would easily agree on the scoring);
- if the responses must be longer (as in the response to a moraljudgment dilemma), develop clear criteria for coding the responses;
- carefully record the child's behavior, responses, and overall performance.

This example of logical thinking illustrates several basic features of traditional methods. First, it involves a *task* – a situation contrived by the psychologist to elicit behavior (activities, verbalizations, etc.) that will provide information concerning the "dependent variable" of interest, in this case, the child's logical thinking. In general, the task is of limited scope, focused on a particular topic, like the problem concerning Jane and the bicycle. The task elicits a fairly limited response, like the yes or no answer to the specific question concerning Jane's athletic prowess. The task is not modified regardless of what happens in the study; once decided upon, the task does not change.

Second, the task or series of tasks (the test) is administered in a uniform fashion to all subjects. This kind of *control* and *standardization*<sup>2</sup> is the essence of the traditional methodology. Both researcher and practitioner attempt to hold constant the test "stimuli," to make sure that all subjects receive the same problems in the same way. The controlled conditions must involve precise and uniform directions for administering the test, including use of the same questions, materials, tasks, instructions, and even the same tone of voice and pace in talking to different subjects. The testing environment should be the same for all subjects, including "lighting, ventilation, seating facilities, and



#### The Need to Move Beyond Standardized Methods

working space" (Anastasi, 1988, p. 34), although this is difficult to achieve in practice (as all of us who have had to test children in the school nurse's office know full well).

Third, standardized administration often contains devices for promoting rapport, attention, and motivation. Obviously, both researcher and practitioner want children to take the test seriously, cooperate with the examiner, work hard, concentrate, and in general do as well as possible. As a result, we normally stress having the children attend to the test, we minimize distractions, and we encourage the children to work hard. Sometimes we build into the test techniques for checking our assumptions about rapport, attention, and motivation.

Fourth, tasks and tests may or may not be based on norms. I may develop the logical-thinking problems myself, perhaps relying only on informal pilot work. Or I may use already existing tasks about which various amounts of information may be available. If I borrow the problems from an existing research study, I will at least know how that researcher's subjects performed on the test (and other subjects in similar studies). If I use a test with extensive norms concerning age, ethnicity, social class, etc., then I know a good deal about how children "typically" respond to the test. In research, we tend to use tests which do not involve extensive norms; in assessment practice, we tend to use normed tests. But whether the tests are original or borrowed, normed or not normed, their essence is standardized administration as described above.

### Two Rationales for Standardized Administration

Standardized administration can be justified on both scientific and ethical grounds.

#### SCIENCE

The basic scientific justification for standardized administration originated in the 19th century, when experimental psychologists, concerned mainly with sensation and psychophysics, required rigorous control over the conditions of testing. Suppose that you as an experimental psychologist want to investigate the ability to discriminate



#### Entering the Child's Mind

different colors. You want to know, for example, whether people can see the difference between red and green, or between one reddish color and another not so red. To find out, you need to make sure that all subjects are presented with the same two colors – in the first case, the same shades of red and green, or in the second, the same "reddish" color and the same "not so red" color. (It would help tremendously if you could define these colors in precise, physical terms, rather than just describing them as "reddish" and "not so red.") You would not obtain useful information were some subjects to see two very similar colors and other subjects two extremely different ones.

Similarly, you need to make sure that the lighting conditions are the same for all subjects. It would be a mistake for you to arrange conditions so that illumination was "bright" for some subjects and "dim" for others. (Here again, precise, physical measurement is helpful: What are "bright" and "dim"?) You also need to use the same instructions for all subjects. Thus it would be a mistake if some subjects were told to indicate when they "thought" one color was different from the other, whereas the other subjects were required to say when they were "sure" that the colors differed.

This then was the model of scientific procedure as developed by 19th-century experimental psychology. And it makes a good deal of sense: if anything is to be learned from investigations like these, the experimenter must have control over the stimuli and must be able to ensure that procedures are fully standardized.

In an effort to provide "scientific" measurement, this logic of control and standardization was then applied to the psychological testing developed in the late 19th and early 20th centuries. The argument was as follows: The goal of testing is to measure variation in some trait across individuals. We want to know whether A is more intelligent, aggressive, or depressed than B. We may want this information in order to compare groups of individuals (are boys more aggressive than girls?), to compare traits within individuals (is intelligence related to lower amounts of depression?), or to compare the effects of independent variables (does increasing stress heighten aggression?). In all of these cases – the comparative study of groups, the study of individual differences, or experimental research – the aim is to get reliable and accurate measures of psychological traits.



#### The Need to Move Beyond Standardized Methods

How to do this? The basic rationale is again control and standardization. According to Anastasi (1988), who has brought a good deal of sound judgment to bear on the subject of psychological testing: "If the scores obtained by different persons are to be comparable, testing conditions must obviously be the same for all" (p. 25). All subjects must receive the same test stimuli – whether this be a list of nonsense syllables or a paragraph to read or an IQ test – in the same manner. To do otherwise, the argument goes, would be to render the data uninterpretable. If subjects received different test stimuli, then the tester could not confidently attribute subsequent variations in performance to individual differences in a particular trait.<sup>3</sup> Consequently, strict uniformity of administration is required. "Such a requirement is only a special application of the need for controlled conditions in all scientific observations" (Anastasi, 1988, p. 25). This scientific argument then leads to the whole apparatus of standardized administration.

#### ETHICS

Standardized administration can also be justified on ethical grounds, particularly in terms of a specific kind of fairness. As far back as 1845, Horace Mann offered several reasons for introducing standardized testing to the schools (Wainer, 1992):

They are impartial.

They are just to the pupils.

They prevent the officious interference of the teacher.

They take away all possibility of favoritism.

They make the information obtained available to all.

They enable all to appraise the ease or difficulty of the questions.

(p. 15)

In one way or another, these reasons revolve around *fairness* or impartiality. One justification is that standard tests prevent the teacher from favoring some children over others (perhaps by giving some children easier questions than others) or from interfering with the process of testing. Another justification is that the tests make the process public, so that an outside observer can judge whether the



#### Entering the Child's Mind

questions are too hard or too easy. In this way, testing can be fair and "just."

The originators of psychological testing were also vitally concerned with fairness. In the France of the early 1900s, "alienists" (assessment specialists) used subjective methods to perform diagnoses of mental retardation in schoolchildren. Those children identified as retarded were pulled out of regular classes and assigned to special instruction. Although the alienists' goal was laudable – to identify children "unable to profit . . . from the instruction given in ordinary schools" (Binet & Simon, 1916, p. 9) – the accuracy of their diagnoses was by no means guaranteed, with the result that children were often mislabeled and then denied a mainstream education. Binet and Simon (1916) were particularly critical of the unstandardized nature of the alienists' testing, which seemed to involve

haphazard decisions according to impressions which are subjective, and consequently uncontrolled. Such impressions . . . have at all times too much the nature of the arbitrary, of caprice, of indifference. Such a condition is quite unfortunate because the interests of the child demand a more careful method. To be a member of a special class can never be a mark of distinction, and such as do not merit it should be spared the record. . . . [T]he precision and exactness of science should be introduced into our practice whenever possible. (pp. 9–10)<sup>4</sup>

Binet and Simon felt that fairness demanded standardized, uniform administration. Their goal in developing the standardized intelligence test was to be *fair* to the child, whose "best interests" demand a precise and exact diagnosis. Haphazard, subjective assessment can result in the child's being labeled as retarded, which is never a "mark of distinction" and which can cause the child to miss out on normal instruction and be consigned to the special classroom. It is a great irony that Binet and Simon's motivation underlying the creation of the intelligence test, which is today so severely criticized for its bias and lack of fairness, was to avoid inaccurate and unreliable diagnoses of schoolchildren.

Similarly, modern testers are also concerned with the issue of fairness, particularly in the case of minority students. Anastasi (1988) argues that when "prejudice may distort interpersonal evaluations,



#### The Need to Move Beyond Standardized Methods

tests provide a safeguard against favoritism and arbitrary or capricious decisions" (p. 67). She maintains that tests provide an opportunity for bright students of minority background to reveal their abilities in a forum not subject to the bias of teachers. The tests are a level playing ground on which all may compete equally. She argues that if standardized tests were eliminated, we would have to fall back on subjective judgment, which is subject to "unreliability, subjectivity, and potential bias" (p. 68), much as Binet and Simon had earlier maintained. In her typically wise summation, Anastasi claims that "tests can indeed be misused in testing cultural minorities – as in testing anyone else. When properly used, however, they serve an important function in preventing irrelevant and unfair discrimination" (p. 68).

I have tried to show that standardized testing springs from the noble motive of ensuring fairness. As articulated by Mann, by Binet and Simon, and by Anastasi, fairness has the goal of eliminating bias, discrimination against minorities, stereotyping, haphazard judgments, "officious" (or other) interference, and distortions of judgment.

The method for achieving this kind of fairness is to treat all people alike, to give everyone the same conditions for running the race. From this point of view, no one should be given easier questions than anyone else; no person should be judged on a different basis from another. A wrong answer should be a wrong answer for all. The same solution should not be judged incorrect if the child is African-American and correct if she is White. This kind of fairness is *colorblind:* It treats all children in the same way and ignores irrelevant attributes.<sup>5</sup>

The arguments in favor of the method of standardized administration are both scientific and ethical. The method aims at accurate measurement, unconfounded by variation in test administration, and it aims at fairness in the sense of impartiality.

#### Some Flaws in the Method

I claim that for certain purposes, the method of standardized administration (which I will sometimes designate as "traditional") often



## Entering the Child's Mind

falls short on both ethical and scientific grounds. There is a sense in which the method is often distinctly unfair and can provide us with only limited information concerning children's thinking. The method of standardized administration suffers from several basic flaws, each of which will be described in some detail below:

- Despite standardized administration, subjects nevertheless interpret tests in idiosyncratic ways, in ways not intended by the examiner.
- It is often not clear what cognitive activities are actually measured by standardized tests.
- Standardized achievement tests are usually based on outmoded assumptions about cognition.
- Standardized methods are not suitable instruments for studying complex thinking.
- Standardized methods are not suitable instruments for studying dynamic change.
- Standardized procedures cannot effectively motivate all children.
- Traditional methods are often not suitable for tapping the competence of those who are in some way "different" by virtue of culture, ethnicity, health status (e.g., the hearing impaired), or other reasons.
- Traditional methods are inadequate for the kind of exploration often required in research and practice.

# CHILDREN MAY NOT INTERPRET THE SAME TESTS IN THE SAME WAYS

As we have seen, traditional psychological research uses the method of standardized administration for a critical reason: to hold test stimuli constant; to ensure that all subjects receive the "same" problems. Although this method may work for the psychophysical experiments in which it originated, it often fails when other situations, including cognitive problems, are involved.

It fails because different people interpret the same *objective* stimuli in different *personal* ways. Given the "same" stimulus, we see different things in it. In Piagetian terms, we always *assimilate* objective reality into our own schemas. We interpret the world, and we respond more to the subjective interpretation than we do to the objective reality. As Shweder (1991) puts it: "'stimuli' are not external to or