

# A Critique of Test Standardization

by Judith Orasanu, Raymond P. McDermott, A. Wade Boykin, and the Laboratory of Comparative Human Cognition

In the course of going to school in America, most children face standardized tests designed to measure either their general mental ability or their achievement in the development of some skill. On the surface there are good reasons to employ both kinds of tests. The first principle of a good pedagogy is to start where a child is and to build on that foundation at a rate commensurate with the child's abilities. At its best, information from standardized tests can help us to construct a pedagogy that is sensitive to the capacities and skills of the individual. In this way all children receive equal educational opportunity; the democratic ideals of greater personal, social, and economic opportunity are thus facilitated by the formal testing movement.

Beneath this ideological surface, however, there is much grumbling that standardized tests play into the social stratification system, not only by reproducing generations of enfranchised and disenfranchised, but

ironically by making the hegemony of the successful look legitimate because of their superior performance on school tests. In this light, standardized tests appear to be one of the most powerful and efficient sorting mechanisms available in modern societies.

It is argued that instead of functioning in the service of pedagogy, tests work against pedagogical efforts. Clearly, there are dramatic differences among children in their abilities to advance when given similar curricula. Teachers are naturally concerned about such differences and are compelled to help the slower children catch up to their classmates. Of course, extra time for slow learners implies ignoring children who might excel if given that time. Standardized tests provide teachers with reasons why underachievers rank below their peers; i.e., they are less capable, less intelligent. Apart from the questionable logic of these rationalizations, they tend to foster a self-fulfilling

prophecy; namely, children who start off behind their classmates fall further behind each day they are in school as a result of the different kinds of attention they receive from their teachers (McDermott, 1977). In addition, children who obtain low tests scores may be tracked into slower programs, which usually means a sentence to inferior education for the duration of the child's school career, with its attendant credentialing problems (Mercer, 1973). This system is particularly rough on children who come into school with skills which are in some way different from what most teachers and test makers expect, and which are rewarded in student evaluations. It is in this light that the insensitivity of standardized tests to the potentials of minority and poor children is so reprehensible.

How are socially concerned educators and policy makers to evaluate these conflicting opinions concerning test use? We will offer a set of criteria which we feel tests must meet in order to fulfill their promise as tools of educational opportunity. Then we will assess the degree to which current testing practice meets those criteria.

In view of our stated concern with the relation between tests and teaching, it would seem that our focus should be exclusively on achievement tests. However, our scope will include general mental ability (IQ) tests

---

*JUDITH ORASANU is a cognitive psychologist and RAYMOND P. McDERMOTT is an anthropologist; both are at the Rockefeller University. A. WADE BOYKIN is assistant professor of psychology at Cornell University. This paper summarizes an unpublished research report arranged by the Laboratory of Comparative Human Cognition and the Institute for Comparative Human Development, under the direction of Michael Cole and William S. Hall, at the Rockefeller University. The authors wish to acknowledge that they have borrowed extensively from the insights of Michael Cole, Mary Cross, Anderson J. Franklin, William S. Hall, Michael Pratt, Sylvia Scribner, and Kenneth Traupmann. Hall, Scribner, and Traupmann also offered helpful comments on this paper. Finally, the support of the Carnegie Corporation and the Ford Foundation is acknowledged.*

---



as well, for two reasons. First, in practice IQ tests are frequently treated as achievement tests; the two are used interchangeably to determine the curriculum level prescribed for a child. Second, item analyses and the method of test construction indicate that the two types of tests are frequently indistinguishable (Levine, 1976; Schwartz, 1975). The issues we will address clearly pertain to all standardized instruments which are norm-referenced and which depend on standardized administration procedures.

### CRITERIA FOR ADEQUACY IN TESTING

Our criteria for the social and educational adequacy of standardized tests are quite simple:

- Testing for competence in some skill may be allowed only if there is some theoretical model of how a task (e.g., reading) is to be performed, and the skill tested for can be shown to be an essential component of that task.
- The test must provide relatively unambiguous information concerning presence or absence of the skill.
- Testing for competence in some component skill may be allowed only if there is a proven pedagogy for building that skill and eventually developing task competence in the child.

Lest the reader think us unduly hard on the testing industry, consider the public outcry that would result if medical science attempted to ignore the types of criteria we are suggesting (Scribner, 1977). People would not subject themselves to medical tests unless they could anticipate a diagnosis and, if all is not lost, a treatment. Concern for possible negative side effects precludes administration of a medical treatment unless a need for it has been clearly demonstrated and its efficacy proven. Even if the negative effects are minimal, medical tests are only given when there is a symptom which cannot be unambiguously tied to an underlying cause. Moreover, wide-scale testing of the population is not undertaken to de-

termine the general health quotient (HQ) of individuals. The people would not stand for it, and we believe that similar warrants must be demanded of the tests put to our children.

Details of our criteria will be amplified. Following Cole, Hood, and McDermott (1977), we will argue that in order for a test to be useful in the description of what a child knows relative to what is to be learned, the test must offer well-defined tasks which are essential components of what must be done in the performance of some complex skilled behavior, such as reading. That is, we cannot give a child a reading test until we can show that the items on the test are well defined in the test taker's eyes and

---

### Testing for competence in some component skill may be allowed only if there is a proven pedagogy for building that skill and eventually developing task competence in the child.

---

that they relate to the skill we are trying to teach. This requirement presumes that a complete and adequate analysis of the target skill is available. Adequate task analysis describes what a person must do in order to perform successfully on the final task, e.g., read and comprehend a page of text; furthermore, it must identify subskills so that tests can be constructed which will monitor a child's progress on these components.

If a task analysis is adequate, the test taker's behavior must change systematically over a range of parametric variations in the task. It should be possible (1) to design items which are specifiably equivalent or different, based on the theory underlying the task, and (2) to predict which varia-

tions will be irrelevant and which will make a problem more complex (and therefore more difficult). Errors on certain types of items would indicate difficulty with a particular component skill which would then become the focus of teaching for that child.

### INADEQUACY OF STANDARDIZED TESTING PRACTICES

How well do current standardized tests meet our criteria for diagnostic adequacy? Our conclusion is that they fall far short for a variety of reasons. Paramount is the fact that practically without exception, standardized tests in use today are norm-referenced instruments. That means that their items were selected not because of their relation to a theory of learning or intelligence, but because a certain proportion of children at given ages answered the items correctly. The tests were designed to correlate with age and especially with school grades. The fact that older children get more items correct might mean that their intellect has developed; it might also mean that their vocabularies are better, their reading skills are better, their knowledge of the world is more like that of the test makers, or their test-taking skills are more sophisticated. It is clear that much more is involved in academic success and high test scores than "intelligence" or task-specific knowledge. However, because of the theoretical nature of test construction, the tests do little to inform us of what the essential school skills are, who has them, and how to give them to those who have not yet developed them. In short, we do not know what the norms are norms of.

Accordingly, how are differences or similarities in scores to be interpreted? In practice, norm-referenced tests, whether of general ability or of achievement, are interpreted as reflecting differences or similarities in some underlying entity (e.g., IQ, reading achievement, etc.) which is assumed to be measured by the test. Several critical assumptions are implicit in this attribution process, however.

On the one hand, items should be



more or less difficult because they tap, respectively, higher- or lower-level skills on the dimension measured. We noted above that test construction has not been based on task analysis. Thus, we have little reason to assume that such gradations in item difficulty are in any way reflective of a single underlying dimension.

On the other hand, standardized tests are assumed to measure the same ability in all individuals. However, we have little evidence that the same norm-referenced tests applied to different people truly measure the same capacity in each, although we call it by the same name. It is obvious that if a child doesn't speak fluent English or can't read, the final score on an intelligence test will be in part an English score or in part a reading score. Most cases are not so painfully obvious, but the important point is that in *no* case does a norm-referenced test provide an adequate description of a child's abilities so that a pedagogical prescription can be written. In principle, norm-referenced tests cannot satisfy our criteria because of the method by which they were constructed.

Research to date indicates that the criteria we have suggested for developing tests (which follow closely those that define criterion-referenced tests) will not be easy to meet. The difficulties are evident in research on two sources of variability in test results: namely, features of the test itself and the social interaction of the test taker and the test giver (Boykin, 1977; Orasanu, 1977). We will consider both in some detail, for until test makers deal with the kinds of problems we will outline, there can be little justification for the continued use of standardized testing in our schools. Much of our evidence will be based on careful data analyses by cognitive psychologists and sociologists. For the most part, the psychometric literature does not offer data appropriate to our analysis.

### SITUATIONAL VARIABILITY IN TESTING

#### Textual Factors

Task analyses of problems posed to subjects on tests or in experiments often break down with slight changes

in how they are presented to the subjects. Tasks that are defined by psychologists as isomorphic frequently seem to be taken as different, not only by different people, but by the same person over time. Changes in the wording, the colors of the stimulus materials, the presence or absence of the stimulus materials, and many other incidentals have all been shown to make a difference when they are not expected to.

A powerful example of the effect of changing the wording used to present a problem is available in Hayes and Simon's (1977) research on problem solving. After a thorough task

---

**What is assumed to be irrelevant from the adult point of view is frequently an overlearned habitual component of the task which does not require active attention from the adult. Failure to identify these task-relevant factors frequently leads to the conclusion that young children lack a particular ability.**

---

analysis, they constructed two isomorphic problems, each involving three different-sized monsters holding three different-sized globes. The task was to alter the situation so that the size relationships that held between monsters matched the size relationships of the globes. One of the variations in the wording of the problem concerned the kind of changes that could be made; in one variation, the globes were to be transferred, and in another they were to be shrunk or enlarged. Even though the number of steps needed to solve the two problems was the same, the slight differences in wording produced differ-

ences in the problem-solving strategies subjects used and the time it took them to finish the task. However, according to Hayes and Simon's task analysis, the wording differences should not have made a difference. Thus, their task analysis is inadequate, for it cannot account for the variable performances on isomorphic tasks. This raises the question of how we might come to a conclusion about the differential performance of children unless a task is sufficiently analyzed.

When inadequate, task analyses may fail to show the importance of certain features for children at particular ages. What is assumed to be irrelevant from the adult point of view is frequently an overlearned habitual component of the task which does not require active attention from the adult. Failure to identify these task-relevant factors frequently leads to the conclusion that young children lack a particular ability. This point is important since most norm-referenced tests are age graded.

For example, Turgeon and Hill (1977) showed that children could perform "like adults" when the verbal concepts used in a two-phase learning task were easily accessible to the children. Rapid learning of the second component required application of a verbal concept learned in the first phase. Abundant research has shown that older children and adults apply the concept, while younger children do not. However, Turgeon and Hill pretested children to identify verbal concepts which were readily available and unavailable to each age group (ranging from four to 18 years). When the two-phase task involved learning of readily available concepts, all age groups applied the concept in the second phase. When relatively unavailable concepts were used, none of the children, even the oldest, applied the concepts. Obviously, the original task analysis did not take into account availability of the concepts for subjects of various ages.

Conversely, inadequate task analysis can lead to an overestimate of adult abilities. For example, on a word association task commonly used



to study the development of verbal concepts and relations, adults "normally" reply with words from the same grammatical class as the cue word, e.g., red-black. Young children, on the other hand, reply with words which could be used in a sentence with the cue word, e.g., red-ball. However, when Stoltz and Tiffany (1972) presented adults with cue words which were unfamiliar and infrequently used, their replies were "childlike." The task was not supposed to be a vocabulary test, but it is clear that relative familiarity of the items to children of various ages could contribute to the "developmental" trend normally obtained. Likewise, differences between ethnic or social-class groups could be mediated by differential exposure to the items on the test (Cole, 1975; W.S. Hall, personal communication).

Children taking tests frequently complain that they know the answer but didn't understand the question. A case in point is a study by Blank (1975) who found that three year olds failed to provide "correct" answers when she asked them to describe the basis for their choice after learning to pick one of two geometric objects (say, a circle or a square). When she asked her young subjects, "Why did you pick that one?" replies were frequently of the form, "Because I wanted to." They referred to an internal motivational state rather than to the name of the appropriate object. It occurred to Blank that the children interpreted the question much as an adult would interpret the question, "Why did you sit down?" "Because I was tired" would be a perfectly appropriate reply. However, in the experimental situation, the adult would probably reply to "Why did you pick that one?" with "Because the circle is correct," the type of reply Blank was hoping to elicit from the youngsters.

Blank modified the task only slightly. Before asking the question, she hid the stimulus array so the children could not simply point to the correct item. She changed the form of the question to "Which one did you pick?" Under these modifications a substantial number of three year olds

provided the name of the correct item. Clearly, the important aspect of this task was understanding not the literal meaning of the question, but its practical import and what could constitute an appropriate reply. This is a provocative finding given recent sociolinguistic research on speech styles and interpretation of communicative intent which suggests that the form of test questions can be a potential source of difficulty especially for test takers who do not speak standard English (Wolfram, 1976; Hall & Freedle, 1975).

The common element in the preceding examples is that young children were assumed to lack the ability tested for; but in all cases they were able to demonstrate the skill, given certain modifications in the procedure. If teaching were based on test results, the children would be taught something

---

**The assumption is false that if a person has an ability or skill, it will be manifest, regardless of the specifics of the test used to measure it.**

---

they already know how to do—certainly not the most motivating of situations.

This point is not restricted to inferences about children's abilities. Research conducted by Cole, Gay, Glick, and Sharp (1971) among nonliterate adults from the Kpelle tribe in Liberia yields the same conclusion. Kpelle men showed very poor performance on a typical oral free-recall test. After five trials only 11 items out of 20 were recalled from a list composed of five common examples of four taxonomic categories (clothes, food, etc.). Then Cole altered the procedure slightly. Instead of merely holding up each item to be remembered and naming it, he held it over one of four chairs. Recall skyrocketed, especially when all the items from one category were associated with the same chair. Why associa-

tion with the chairs should facilitate recall is still a mystery, but it clearly indicates that Kpelle memory is much better than would have been thought based on the first study.

These examples together yield two general observations: (1) intuitive task analyses were not adequate in any of these cases, despite the fact that these were tightly controlled psychological experiments; and (2) the assumption is false that if a person has an ability or skill, it will be manifest, regardless of the specifics of the test used to measure it. The assumption that, if it is "there" one should be able to view it no matter how the question is asked or what specific items are used, derives from the conception of intelligence as a general factor, *g* (Spearman, 1904). Variability in a single individual's performance in the examples we have presented clearly contradicts this view.

#### Contextual Factors

Even if we conduct an adequate task analysis, it may not be possible to assume that tasks can be presented in an identical manner to different children. Factors beyond the text come into play in every presentation of the task, and subtle variations in procedure grossly diminish any assumption of test standardization across persons or successive presentations.

Perhaps the most interesting research along this line centers on the issue of rapport or friendliness, a vague interactional quality which remains unspecified in most analyses of its effects on test performance. Some behavioral accounts of rapport, however, indicate that interactions reportedly high in rapport are marked by intense mutual attention by participants to each other, given some shared task at hand (Charney, 1965; Schefflen, 1973). In a less specifiable way, testers have long talked about the importance of rapport for interpreting a child's performance, and many have a rule of thumb that children can gain an average of 10 IQ points if they are allowed to acclimate to the tester and the test-taking conditions. For the most part, such improvements in performance are understood in terms of a child's increased motivation, and the bulk of



the research has centered on locating the motivational factors which could make the difference. In addition, an intriguing line of observational research by sociologists suggests that optimizing conditions simply function by directing a child's attention to the task to be performed. Unfortunately, we will be able to present only enough of both materials to make our point.

Two lines of research have successfully shown that motivating or optimizing conditions function to increase children's test performance. Zigler and Butterfield (1968) and Zigler, Abelson, and Seitz (1973) have shown marked increases in the performance of minority children when they were encouraged to do their best on every item. And Thomas et al. (1971) have given us a closer look at how such factors can make a difference. Their subjects were working-class Puerto Rican children from six to 16 years of age. Two examiners were used, both bilingual Puerto Rican females and experienced testers. A particularly dramatic set of findings concerned children who were tested twice on the same test, once by an experimenter who was described as "encouraging" on the basis of prior observation, and once by an experimenter who was described as "impersonal." The two administrations of the test were separated by a little over a year, and the order in which the two examiners tested the children was appropriately counterbalanced. On the average the encouraging experimenter obtained scores 17 IQ points higher than the impersonal experimenter.

How are such striking results accomplished? Thomas et al. do not offer a strongly specified answer, but their sense of how it happened is summed up in the descriptive reports on the children by the two testers. The encouraging tester described the children as pleasant, warm, and relaxed. And the impersonal experimenter was more likely to describe the same children as rigid, aloof, shy, and hostile. It is easy to imagine how in the presence of the first tester, the children would try harder to find the most appropriate answer to the tester's

question. It might be considerably easier for the child to focus on the intellectual task at hand in the presence of the comforting experimenter. Provocative as these reports are, we need some more detailed accounts of how rapport helps a child to perform on tests.

In order to achieve some sense of the role rapport might play in the child's test performance, we turn next to a small literature by sociologists concerned with the interaction between children and experimenters in actual test-taking situations (Cicourel, Jennings, Jennings, Leiter, MacKay, Mehan, & Roth, 1976;

---

**Research strongly suggests that social interaction subtly structures a test situation so that children spend more or less time attending to the task as defined by the experimenter and engaging in activities that are more or less relevant to success at that task.**

---

MacKay, 1973; Mehan, 1973). Descriptions accomplished to date offer the conclusion that a child's performance on a given test is necessarily a social performance, and that the child's score is invariably and necessarily achieved in interaction with the test giver. A standardized procedure for giving a test assumes that an experimenter proceeds with all people in a similar way (or at least assumes that error with respect to these behaviors is randomly distributed over children). At first, this does not appear to be an outrageous assumption (although it does involve a curious rejection of the individual in a field which purports to study individual differences). In addition, the assump-

tion of standardized social interaction is simply unwarranted. Ideally a person is given an instruction or question by the experimenter, the person produces an answer, and the experimenter has only to score the answer as correct or incorrect. How much could go wrong between the standard instruction and the answer? How is it possible to fashion a claim that an intellectual performance on such a test is interactionally managed and that the specifics of the interactional managing make a difference to the child's final score?

Evidence on this point is provided in analyses by Mehan (1973) and Roth (1976), who videotaped administrations of the Peabody Picture Vocabulary Test. Their main point is made clear in a test analysis offered by Roth. A child is presented with a word and asked to point to one of four pictures which best represents the stimulus word. Between the time when the experimenter starts by offering a single word and the child finally picks an answer, there are more than 25 separate turns to talk taken by the experimenter and the subject. Roth goes on to show that in the course of their interaction, children and testers clarify and even change the instructions on specific items. The tester often provides detailed interpretations for the child in order to make the instructions clear. The mutual attention which marks interactional rapport appears to help the child and tester to organize themselves for performance of the task at hand. It requires a good deal of social "work" on the part of both participants to define the task clearly.

While we cannot be certain, because the directly relevant data are not at hand, the analyses of test performance provided by Roth and Mehan, when combined with the results reported in studies such as those by Thomas et al., strongly suggest that social interaction subtly structures a test situation so that children spend more or less time attending to the task as defined by the experimenter and engaging in activities that are more or less relevant to success at that task.

Thus, the possibility of reaching a firm conclusion about what a child



knows how to do on the basis of a test which does not take these contextual factors into account appears to be minimal.

## CONCLUSIONS AND RECOMMENDATIONS

In view of the fact that current standardized tests are not diagnostic in the sense we have described, our conclusion is that the negative social consequences outweigh the positive educational contributions these tests can make. Until tests and teaching systems which meet our criteria are designed, we see no reason to continue standardized testing and accumulation of tests records that are powerless to do anything other than haunt children as they struggle through school.

On a more constructive note, we can offer general guidelines that might establish a positive educational role for tests. Before any new diagnostic-prescriptive tests are constructed, decisions must first be made concerning what skills are to be taught in our schools. Once the target skills have been identified, task analyses must be carried out. Ideally, these will form the basis for diagnostic tests and will serve as a guide for pedagogy (Traupmann & Cole, 1977).

If we do not want tests to serve as instruments for maintaining social inequality, their function in the educational system must change (Hall & Pratt, 1977). Teaching must be organized for mastery of basic skills by all children. Tests capable of diagnosing a child's entry skills and prescribing progress toward mastery are desirable. But tests which rely on and emphasize differences between children will have no place in such a system. In order to achieve equality in educational opportunity, we must remain open to the potentials of the great variety of children who enter into our schools every year. ■

## REFERENCES

- Blank, M. 1975. "Eliciting Verbalization from Young Children in Experimental Tasks: A Methodological Note." *Child Development* 46, 1: 244-251.
- Boykin, A.W. 1977. "Contextual Factors and Test Performance." Working Paper #2, Laboratory of Comparative Human Cognition and Institute for Comparative Human Development, The Rockefeller University.
- Charney, J. 1965. "Psychosomatic Manifestations of Rapport in Psychotherapy." *Psychosomatic Medicine* 28: 305-315.
- Cicourel, A.V.; Jennings, K.H.; Jennings, S.H.M.; Leiter, K.; MacKay, R.; Mehan, H.; and Roth, D.R., eds. 1976. *Language Use and School Performance*. New York: Academic Press.
- Cole, M. 1975. "Culture, Cognition and I.Q. Testing." *The National Elementary Principal* 54: 49-52.
- Cole, M.; Gay, J.; Glick, J.A.; and Sharp, D.W. 1971. *The Cultural Context of Learning and Thinking*. New York: Basic Books.
- Cole, M.; Hood, L.; and McDermott, R. 1977. "Inferences About Cognitive Abilities in Non-test Environments." Paper presented at the American Association for the Advancement of Science Meeting, Denver (February).
- Hall, W.S., and Freedle, R.O. 1975. *Culture and Language: The Black American Experience*. New York: Halsted.
- Hall, W.S., and Pratt, M. 1977. "Testing for Competence: Changing the Criterion." Working Paper #4, Laboratory of Comparative Human Cognition and Institute for Comparative Human Development, The Rockefeller University.
- Hayes, J.R., and Simon, H.A. 1977. "Psychological Differences Among Problem Isomorphs." In *Cognitive Theory*, Vol. II, ed. N.J. Castellan; D.B. Pisoni; and G.R. Potts. Hillsdale, N.J.: Erlbaum and Associates.
- Levine, M. 1976. "The Academic Achievement Test." *American Psychologist* 31: 228-238.
- MacKay, R. 1973. "Conceptions of Children and Models of Socialization." *Recent Sociology* 5: 27-43.
- McDermott, R.P. 1977. "Social Relations as Contexts for School Learning." *Harvard Educational Review* 47: in press.
- Mehan, H. 1973. "Assessing Children's School Performance." *Recent Sociology* 5: 213-239.
- Mercer, J.R. 1973. *Labelling the Mentally Retarded*. Berkeley, CA: University of California Press.
- Orasanu, J. 1977. "Some Constraints of Text and Setting on Measurement of Mental Ability." Working Paper #3, Laboratory of Comparative Human Cognition and Institute for Comparative Human Development, The Rockefeller University.
- Roth, D.R. 1976. "Intelligence Testing as a Social Activity." In *Language Use and School Performance*, ed. A.V. Cicourel, K.H. Jennings, S.H.M. Jennings, K. Leiter, R. MacKay, H. Mehan, and D.R. Roth. New York: Academic Press.
- Schefflen, A.E. 1973. *Communicational Structure*. Bloomington, IN: Indiana University Press.
- Schwartz, J.L. 1975. "A Is to B as C Is to Anything at All." *The National Elementary Principal* 54: 38-41.
- Scribner, S. 1977. "Intelligence Tests: A Comparative Perspective." Working Paper #1, Laboratory of Comparative Human Cognition and Institute for Comparative Human Development, The Rockefeller University.
- Spearman, C. 1904. "General Intelligence Objectively Determined and Measured." *American Journal of Psychology* 15: 201-293.
- Stoltz, W. and Tiffany, J. 1972. "The Production of Child-like Word Associations by Adults to Unfamiliar Adjectives." *Journal of Verbal Learning and Verbal Behavior* 2: 38-46.
- Thomas, A.; Hertzog, N.; Dryman, L.; and Fernandez, D. 1971. "Examiner Effects on IQ Testing of Puerto Rican Working Class Children." *American Journal of Orthopsychiatry* 41: 809-821.
- Traupmann, K.L. and Cole, M. 1977. "Can Task Analysis Inform the Development of Standardized Tests?" Working Paper #5, Laboratory of Comparative Human Cognition and Institute for Comparative Human Development, The Rockefeller University.
- Turgeon, V.F. and Hill, S.D. 1977. "A Developmental Analysis of the Formation and Use of Conceptual Categories." *Journal of Experimental Child Psychology* 23: 108-116.
- Wolfram, W. 1976. "Levels of Sociolinguistic Bias in Testing." In *Black English*, ed. D.S. Harris and T. Trabasso. Hillsdale, N.J.: Erlbaum and Associates.
- Zigler, E., and Butterfield, E. 1968. "Motivational Aspects of Changes in IQ Test Performance of Culturally Deprived Nursery School Children." *Child Development* 39: 1-14.
- Zigler, E.; Abelson, W.; and Seitz, V. 1973. "Motivational Factors in the Performance of Economically Disadvantaged Children on the Peabody Picture Vocabulary Test." *Child Development* 44: 294-303.