DOCUMENT RESUME

ED 242 770 TM 840 185

AUTHOR Bracey, Gerald W.
TITLE What Should Be Assessed in the Future: Theoretical Considerations from Research on Brain Function and from Research in Cognitive Science.
PUB DATE 7 Jun 83
NOTE 15p.; Paper presented at the National Assessment of Educational Progress Conference on Large Scale Assessment (Boulder, CO, June 6-9, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cerebral Dominance; *Cognitive Measurement; Cognitive Processes; *Cognitive Style; Elementary Secondary Education; Evaluation Methods; *Individual Testing; *Student Evaluation; *Test Construction; Testing; Testing Problems

ABSTRACT

Reasons why there are few incentives for test publishers to make significant innovations in what is tested or how it is tested are given. A brief discussion of research on growth spurts in the brain, hemispheric differences, and other neurological phenomena is followed by a discussion of some conclusions drawn from this work. While skepticism is expressed over the great inferential chasms one must leap to arrive at some conclusions, hope is expressed that the field will ultimately prove fruitful in permitting more sensitive assessment of individual children. Recent studies in cognitive psychology are discussed and hope is expressed that these areas, too, will lead to improved assessment although their current relevance to practice is not great. Finally, some areas of investigation that are currently being ignored are mentioned as being potential sources of useful evaluation. (Author)

What Should Be Assessed in The Future:

Theoretical Considerations

From Research on Brain Function

And

From Research in Cognitive Science

Presented to the Annual Conference
on Large Scale Assessment
Boulder, Colorado
June 7, 1983
by
Gerald W. Bracey, Ph.D.
Director of Research Evaluation and Testing
Virginia Department of Education

WHAT SHOULD BE TESTED IN THE FUTURE

Gerald W. Bracey, Ph.D.

ABSTRACT

Some reasons are given for why, at the current time, there are few incentives for test publishers to make significant innovations in what is tested or how it is tested. A brief discussion of research on growth spurts in the brain, hemispheric differences and other neurological phenomena is followed by discussion of some conclusions that have been drawn from this work. While scepticism is expressed over the great inferential chasms one must leap to arrive at some conclusions, hope is expressed that the field will ultimately prove fruitful in permitting more sensitive assessment of individual children. Recent studies in cognitive psychology are discussed and hope is expressed that these areas too, will lead to improved assessment although their current relevance to practice is not great. Finally, some areas of investigation that are currently being ignored are mentioned as being potential sources of useful evaluation.

Although I have had far less time than I had hoped to explore the libraries in preparation for this speech, I think I did uncover some things to be considered about what ought to be tested in the future and how it ought to be tested even though I will say those things with much less organization than I am comfortable with and much more tentatively than I would prefer.

Let me first discuss what I see as a major obstacle to any of the innovations in testing practice that I see as desirable. The obstacle is that for test publishers, there is little incentive to produce changes in tests other than perhaps broaden the scope of what is tested, in the sense of adding new curriculum areas and thus sell more tests. That in itself might not be a bad idea depending on how the tests were conceived and constructed: If it is true, as is so often alleged and as I tend to believe, that what is tested is, if not what is taught, what is emphasized, then developing more comprehensive batteries in science, fine arts, foreign language would help drive instruction, to use Jim Popham's phrase, in a healthy way. That would be fine as far as it goes, but unfortunately it would not go very far. The reason it would not go very far is, that when you expand the scope of testing to include more subjects and more objectives within a subject, the technology of testing soon falls short. For example, we are developing in Virginia an assessment component for a set of learner outcome objectives, k-twelve, all subject areas from science and math to fine arts and physical education. We are thus faced with assessing objectives such as

The student will gain insight into the culture and history of a people through the study of literature.

or

The student will describe physical, chemical, and nuclear changes using the law of conservation of matter and energy.

We have made a conscious decision that assessment strategies must be provided for all objectives, not merely those that lend themselves to easy assessment because of our concern that we not establish a hierarchy of importance within the objectives by selecting only some for assessment. We are thus faced with the development of some very creative assessment strategies. Believe me, we are not having an easy time of it.

Additionally, areas such as affective measures are badly ignored in assessment programs to the detriment of education in general. The State Board of Education has established nine goals of public education in the state of Virginia, but at this time, the assessment program address only two of them, those dealing with achievement. We are also moving in these areas, but the going is slow.

In spite of the creative work I see happening within the Department and a few other places, I think that my assertion concerning the disincentives for test innovations hold. I draw my conclusion in part not only from the limited technology of testing but from John Goodlad's recent conclusion that "in the how and the what of what is taught, a school is a school is a school." Bob Glaser had made a similar observation somewhat earlier. Goodlad noted that the schools were characterized by a low level of cognitive demand and cognitive response. Glaser mentioned the inflexibility of curricula treatments. Until there is a change in what Sorotnik, one of Goodlad's co-invesitgators, called the persistency, consistency and mediocrity of life in classrooms, there seems little value in, little incentive for test manufacturers developing better disgnostic, prescriptive instruments unless, again- those instruments can be used to drive instruction. That is a big if, because I recall last year Eva Baker here wondering aloud if curriculum makers and test builders would ever talk to one another.

I want to talk now about the implications of two areas of research: Testing considerations drawn from research tying brain function to learning and research stemming from investigations in cognitive psychology.

Implications from the studies of the brain.

This is an area of study where I had been hearing a lot of second hand talk about what exciting things were going on but when I went to the library cupboard, I found not that it was bare but that it wasn't very amply stocked. In 1978, the NSSE Yearbook was entitled the Brain and Education and it contains many articles by many of the major brain researchers in which many of them make some rather interesting conjectures of the relationship of their work to education but that's what they are -conjectures. Most of the work revolves around MacLean's notion of a Triune brain, Sperry's work involving lateralization, and Epsteins work on growth spurts. Herman Epstein who has postulated theories of learning growth based on the studies of spurts in brain growth largely using brains of dead children asserts that Head Start was bound to fail because it did not occur during the period of a growth spurt. Someone, whose name escapes me now, has also postulated that since the growth spurt at age 11 seems to be twice as great for girls as for boys that perhaps we should consider sex segregated schooling during the middle school years. Many of the conjectures appear contradictory and after three hundred and seventy odd pages of this, editors Allan Mirsky and Jeanne Chall are left to say are in a chapter entitled implications for education, "Gee ain't this interesting but what does it all mean?"

That does not stop Mirsky and Chall however, from posing the following futuristic scenario:

The test battery of the twenty-first century would be the responsibility of a team of specialists including the educational neuroscientist. It would encompass behavioral and photographic analyses designed to identify motor patterns, cerebral dominance and related psycho- and physiomotor capacities; it might also include electrographic and sensory tests that would provide data about the relative maturity and efficiency of processing information in all relevant sensory modalities. Attentional capacities would be assessed by both behavioral and electrophysiological means, and the sources of attentional difficulties (if any) categorized and identified with respect to inter- as opposed to extra-cerebral causes. Brain size, maturity, and relative degree of myelinization in key areas would be assessed by means of noninjurious neuroradiological techniques. Oxygen utilization in various brain regions at rest and during a variety of mental activities would be assessed by means of dynamic energy utilization techniques. Such methods currently exist and need only to be refined further. Brain neurohumoral balance and maturity would be assessed by means of biochemical assays performed on a few drops of blood and urine. Computer-assisted analyses of these data would enable the educational neuroscientist to perform accurate assessment of the child's developmental stage, his particular strengths and weaknesses, the instructional materials he would best be able to handle, and the problem areas that would most likely be encountered during his educational career.

Mirsky and Chall close the volume by saying:

As exciting as this utopian aid to education may be in the twenty-first century or a few years earlier, it must be realized that it cannot be applied in the absence of that most effective and essential of all educational forces -- able, patient, and caring teachers.

Work such as that just quoted, led Barbara Hutson of Virginia Tech, in a piece entitled "Brain Based Curricula - Salvation or Snake Oil ?" to note that

"MacLean's work is based on surgical experimentation with green lizards and squirrel monkeys, the work of Sperry and others is based on human split brain studies and noninvasive analogues, Epstein's work on human cadavers. The MacLean/Hart position credits us with three brains and three minds; the Sperry/Samples position credits us with two hemispheres and two minds; the Epstein/Toepfer position credits us with one mind which works -- sometimes".

She notes as well that many of the assertions require as yet to be demonstrated brain structures, as yet to be demonstrated brain functions and/or as yet to be demonstrated linakges between the first two and psychological processes.

Despite the great inferential chasms one must leap from research on brain function to assessment, there is a general implication here which I found in all of my research

whether I started with brain based research, developmental psychology, cognitive psychology, policy studies, or what: The research all implies sizable differences in learning styles, growth rates, perceptual preferences,information processing etc., which imply that both instruction and assessment should be more tailored to the individual than they currently are or probably can be in the immediate future.

Earlier today you heard Dick Schutz say "forget about Aptitude-Treatment-Interaction". I am tempted to say "Forget Dick Schutz on ATI". Everything that I read implies that it ought to be there, although we might wish to call it "Learning Style Treatment Interaction" or "Lateralization Treatment Interaction". There is a real question as to how much of any variance the differences in style or lateralization would account for as opposed to the communalities shared by all humans, but I would be loathe to dismiss ATI at this time.

While I poked fun at the Mirsky-Chall description of a twenty-first century test battery, I don't think it preposterous to consider using the emerging technology to provide much more sensitive instrument for each child. And I don't just mean in terms of difficulty level. Gagne, Sternberg and others have posed that children must learn certain skills to a certain level of automaticity. If this be true then it doesn't seem like a tremendous technological problem to have a computer program with a realtime clock that can assess reaction time or even the amount of time spent on different parts of the problem. In fact, Sternberg has obliquely recommended such latency measures and last week I saw a program built for an 8K computer that contained two real-time clocks. If you embed two realtime clocks in an 8K machine, the technological problems in assessing latency and other temporal aspects of problem solving do not seem great at all.

There are other studies, derived from cognitive science, cognitive psychology that may also require the assistance of microprocessor technology and it is to these that I now turn. Some years ago, Anne Anastasi noted that increasing specialization had led to a concentration upon techniques for test construction without sufficient consideration of psychological research for the interpretation of test scores. There is no real "theory" underlying Item Response Theory except a theory of test construction which becomes, in the end, a theory of technique not of substance. More recently, Bob Glaser has affirmed that theories of learning have been ignored by test developers, but has also noted that until recently, theories of learning were based on experiments contrived to fit the convenience of the experimenter - and, I would add, they continue to fit the reward structure of the universities, and that because of this the ap-lication of learning theory to real life, long term learning and the development of competence or even expertise has been relatively minimal.

The current Zeitgeist with its emphasis on competence, excellence, and expertise and higher order skills such as analysis, synthesis, and problem solving has turned our attention to the assessment of such skills and, as with the research on brain and learning, I find on closer examination that we don't know as much as we ought to about how to assess such skills or even how to describe them.

Some of the research in cognitive science derives from what are called "expert systems" usually in medicine and mathematics. Such systems are developed as computer programs to make explicit the rules for problem solving that are implicit or tough to see when humans do them. Such systems can be used for diagnosis in medicine and in the instance of physics for the development of sets of rules to solve problems.

One system developed by Gordon Novak found that problems in physics textbooks which appear to call for the use of a couple of equations actually called for ten or twelve and that the laws of physics needed to solve the problems and explanation of how to solve the problems presented in the textbook7 was totally inadequate or even missing

altogether. Thus, in order to solve the problems sutdents had to use hidden laws and equations. One reason, it is alleged that students are "bad" at science is that the instructional materials they use never provide them with the materials necessary to understand well the fundamental concepts or to solve the problems presented them. The implication is clear from some of the work using expert systems, that we will need to assess not just what a child knows in a summative way, but assess what he knows about a concept and how this knowledge may show fundamental misconceptions. Tom Romberg will expand on this in his presentation.

A second set of researches has looked at the differences in how experts approach a problem and how a novice does. Novice is defined in different ways depending on the study. In one instance for example, a novice was someone who had completed a college course in physics while an expert was a graduate student in physics. Presented with a set of physics problems, novices tended to react to superficial qualities of the problem (e.g., these all involve inclined planes) while experts tended to invokve the underlying principles (e.g., these all can be solved using Newton's second law). Clearly, what needs to be really assessed here, what is really important here, is not only how does a novice approach a problem in contrast to an expert or a less competent person as opposed to a more competent person, , but how to we assess the process from novice to expert, how do we assess the development of competence. If this sounds a little like it requires a merger of instruction and assessment, that is no accident. The schism between instruction and assessment has certainly been as damaging as that between education and psychology and we should begin to look at instruction/assessment as integral parts of a single process.

The studies of expert-novice behavior suggest, in part, that the difference between more and less expert people is in their ability to organize information in long term

memory in such a way that makes it readily accessible for a variety of purposes. Thus we need to assess there kinds of storage and retrieval and generalization capabilities. A problem of course, is that most assessments that take place in classrooms are so sar removed from any definable behavior as to make their analysis into information processing components virtually impossible.

There are a few researches that are more immediately cogent to elementary and secondary education. John Seely Brown and his colleages at Xerox' PARC have developed a computer program that analyzes "bugs" in kids arithmetic. That is, it systematically looks at patterns of errors, not just counts rights and wrongs. Building on this work, someone whose reference has been lost to me has pointed out that it is much more efficient, even necessary to point out to the child the "bug" in his problem solving strategy, not just note that he got it wrong. This pointing out of errors only tends to be more of what happens in schools or on standardized tests.

One intriguing study looked at the errors made in writing passages of students who made a large number of errors in writing. Most teachers would probably, given the enormous number of errors, be able only to send the child for some kinds of remediation which might or might not match the kids' problems, because a systematic analysis of errors in writing as in mathematics shows that there are patterns of mistakes which can be placed into categories while remediation would probably consist of a predetermined set of procedures adopted by the school system.

I'm certain that a computer program for such analyses is some ways off and one will probably be required as the analyses of error patterns is a rather laborious task. An intriguing aspect to these researches: When the kids were allowed to read, aloud their own badly written passages, they often spontaneously corrected most of the errors. The error producing problem appears to be something other than that the child doesn't understand the structure of the language.

Again, the utility of approaches to error pattern analysis is a ways down the road, but I think it bears watching. It really goes back to the old dictum derived from Piaget that a child's errors, if adequately examined, may tell you much more about what his cognitive level is than his correct responses.

There is one area of research that I have not yet had a chance to look at. Much activity in Russia, so I understood has focused on the old concept of Lev Vygotsky of the "zone of proximal development" which is now usually called the "zone of potential development". The thrust of these researches is to find ways of not only assessing where a child is in cognitive development but what his potential level is. Knowing these two facts would allow a tailoring of instruction that would not be too far above the level of present development, but aimed toward the potential. I don't recall Piaget using the phrase but in his terms, I think we could call this the "zone of potential accomodation".

Although this may be a tangent on what we think of in assessment, I think we need to begin a much more comprehensive assessment of the instructional materials in classes as well as kids. Most textbooks that I have seen are awful. Ditto most software. And when I have queried software developers on how they know that something is good, they often offer the retort that "You don't ask that of textbooks". No, we haven't but we damn well ought to. An intriguing line of research has been opened up by Thomas Malone in his analysis of what makes videogames intrinsically motivating. His findings--the challenge, the curiosity, and even the ambiguity of the rules initially are part of the motivationg aspect. Similar findings were reported at a conference on videogames and their implication for education held at the Harvard School of Education. Even allowing for the fact that the conference was funded by Atari, the researchers seemed to find much to gain from videogames for instructional materials. I am convinced,

from my own experiences with my Apple, that some of the games require such intense concentration and sustained attention that those abilities may in fact be facilitated. I, for one, am tired of reports like the Bell Commission report which resulted in calls to get tough on kids which usually means get boring and get punitive. Why not get challenging, get exciting? In any case, it will benefit us little to know all there is about information processing, brain functioning, problem solving if the materials in classrooms do little or nothing to evoke or enhance these processes.

Finally, there is one area of investigation that I see no one working on right now. It is an essential of adaptive intelligence to be able to deal with uncertainty in a situation. But I don't know of any investigators who are working in the area of finding out what does a person do when he doesn't know what to do ? Given a novel situation, how does a person go about deciding what to do next ? How does he decide how to gather information and which information to gather ? These kinds of skills, at least on the surface, seem more important in an information society than in previous ones. That is because, as I painfully discovered enroute to this paper, the information explposion has produced a concomitant ignorance explosion - somehow the electronic marvels I an others posess lead to more and more papers on more and more topics and I find that I know less and less about more and more. I understand that there are some 6000 to 7000 scientific reports published every day and certainly at least three or four of them are worth reading. Thus we need to asses how a child learns a) to separate the informational wheat from the informational chaff and b) how a child copes with a novel situation. Given the Goodlad conclusion that a school is a school is a school, however, the child almost never finds himself in a novel, information sorting situation there. Sternberg appears to have done some preliminary work in this area in what he calls executive processes that organize, plan and monitor behavior, but his work is relatively primitive and it appears to me at this time that little work has been done since the

topic was addressed by Miller, Galanter and Pribram in their 1962 book Plans and the Structure of Behavior.

Some of the topics I have covered seem remote from immediate practice and some of them are. I hope you will give them some consideration however, because, to me at least, they imply a much richer formulation of testing practice than is currently available.