DOCUMENT RESUME

ED 175 936 TM 009 675

AUTHOR Turnbull, William W.
TITLE Intelligence Testing in the Year 2000.
PUB DATE Apr 79
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Measurement; Cognitive Processes; Educational Trends; Elementary Secondary Education; *Intelligence Factors; Intelligence Quotient; *Intelligence Tests; Labeling (of Persons); Native Speakers; Neurological Organization; Norms; *Prediction; Psychoeducational Processes; Skill Development; Testing Problems; Test Results; Test Validity

ABSTRACT

Reflections on past and future uses of intelligence tests are presented. Three current approaches to intelligence tests are described: (1) neural efficiency, which relates the speed or quality of functioning of the neural system to test results; (2) information processing--cognitive micro-processes used in solving test items; and (3) psychoeducational, which relates classroom phenomena and test content. In projecting the future of intelligence testing, suggestions are made to discard the term IQ; to use commonly understood words; and to label specific skills by the content being tested, such as "addition", rather than "quantitative skills." Predictions on the future of intelligence tests involve the reporting of separate scores, each of which reflects a distinct ability; use of a different standard of measurement, rather than the IQ ratio; and a reporting system which will have greater educational value. Measuring the development of separate skills rather than an aggregate is recommended, as well as constructing tests related to classroom instruction that are designed to measure specific skills as they develop over the years. Comparable tests of the same abilities which measure children's competencies in the dominant language, as well as in the language of instruction, are also recommended. (MH)

INTELLIGENCE TESTING IN THE YEAR 2000

William W. Turnbull

President

Educational Testing Service

Paper prepared for a symposium at the Annual Meeting of the American Educational Research Association, in San Francisco, California, on April 10, 1979.

INTELLIGENCE TESTING IN THE YEAR 2000

"The Year 2000" has been used as a surrogate for "the distant future" for so many years that we may not have noticed how close it is coming. It now lies as far ahead of us as the post-Sputnik year 1958 lies behind us. Taking the title of this symposium literally, then, and assuming that the past is one reasonable guide to the future, I was tempted, in writing this paper, to suggest that intelligence testing will probably change about as much in the next twenty-one years as it has in the twenty-one just past.

Since that line of thought did not help me a great deal, I decided instead to look back to the turn of the last century, to see if one could contrast the state of the art around 1900 with the probable or possible state of the art around the year 2000.

The approach of the 20th century found James McKeen Cattell experimenting with "mental tests" which he gave to college students. Anastasi[1] summarizes the tests used--"measures of muscular strength, speed of movement, sensitivity to pain, keenness of vision and of hearing, weight discrimination, reaction time, memory and the like," noting that "In his choice of tests, Cattell shared Galton's view that a measure of intellectual functions could be obtained through tests of sensory discrimination and reaction time. Cattell's preference

for such tests was also bolstered by the fact that simple functions could be measured with precision and accuracy, whereas the development of objective measures for the more complex functions seemed at that time a well-nigh hopeless task." And, of course, Cattell had illustrious company in his school of thought.

I invoke this bit of history for a purpose. We continue in 1979 to see the modern descendants of the psychological approaches of the turn of the last century, albeit their theory and their methodology in most cases are vastly improved. I'll say just a few words about this line of inquiry, which I shall call the "neural efficiency" approach, then consider briefly the second main strand today, sometimes called the "information processing" approach, and finally devote most of my remarks to the third line of attack, which might be termed the "psycho-educational" approach. This last set of procedures is the one toward which I would look for the greatest help in producing tests that are useful in schools and colleges for a few decades to come.

The "neural efficiency" approach, alive and well in laboratory settings, is of course the attempt to find physiological measures that tap directly into the speed or quality of functioning of the neural network. Examples are the evoked potential work of Ertl[2] in Canada and of the Hendricksons[3,4] in England, or the latter-day reaction time experiments of Jensen[5] on which he reported at last September's

meeting of APA. With the advances in theory and technology in this century, one may hope for some success in these efforts by the year 2000.

The discovery of more sophisticated neurological bases for Spearman's "g," or the development of simple behavioral measures of neural efficiency such as response time, may, if they prove valid, lead us to especially useful tests for very young children, as in screening for retardation or other handicap. As the child grows older, however, he or she experiences progressive differentiation in psychological functioning, in the range of tasks confronted, and in the broader environmental conditions in which development takes place. By the time the child is in school, where most intelligence testing is done, the differentiation has progressed sufficiently that behavior is more situation-specific. It is my thesis that for tests to be most useful in academic settings--the environment in which I propose to discuss them--they should reflect responses to the varied tasks posed in school, demanding performance of increasing depth of understanding and involving a progressively broader array of attributes that are called into play to produce a successful, adaptive response to tasks and environments that become more and more complex as the child grows older.

The second approach I mentioned, well represented on this panel, is the "information processing" approach, which Sternberg[6] has explored impressively especially in relation to verbal analogies. This work may, I believe, be of enormous value not

only to the understanding of intelligence but also to the development of tests that draw in the most efficient combinations upon the "components," as Sternberg calls them--the cognitive micro-processes--that the individual employs in solving the tasks comprising the test, or rather the test item. Information processing approaches have made notable contributions to theoretical psychology. In relation to academic testing, I see them as likely to contribute to the more precise formulation of test questions in the familiar psychometric or educational formats. I do not believe it likely that they will give us separately measured components that stand themselves as the variables of utility in the schools.

The third approach to intelligence testing, the "psycho-educational" approach, is of course the basis for most of today's standardized testing. It is conceptually imprecise and it deals in phenomena--especially classroom-related phenomena--that are extraordinarily hard to analyze with precision, although Scandura[7] and others are opening up new fields of analysis. To a degree, the present tests work because they mirror in themselves the complexities of the classroom behaviors that constitute the criteria of interest.

As we all know, the early breakthrough in testing useful in relation to schooling came just after the turn of the century in Binet's work. He was successful precisely because he accepted the difficulties of measuring complex functions, and instead of concentrating on simple responses presumed to

yield indices of efficient neural functioning, created tasks that simulated real-world problems or posed classroom-relevant questions to be answered. My belief is, in short, that at least for the rest of this century the most promising avenues for the development and improvement of "intelligence" tests that are to find utility in the classroom will owe more to Binet than to Galton.

What will intelligence tests be like in the year 2000? Instead of an I.Q. they will yield scores that reflect separately the various aspects of ability that are of interest, will express those abilities in a more manageable metric, and will report them in terms of greater educational utility. The learning tasks confronting the child will be multi-faceted, and will yield scores that reflect separately the various pertinent aspects of developed ability.

In recent years we have moved away from the original ratio metric from which the I.Q. was derived--the ratio of "mental age" to chronological age--and toward the substitution of a standard score as the measure of intelligence. This is a step very much in the right direction. Roger Lennon[8], speaking about the I.Q. at the annual meeting of AERA and the National Council on Measurement in Education in 1978, said "A persuasive case can be made for elimination of this term [I.Q.] on the grounds that it now carries, in professional and lay minds alike, an insupportable freighting of emotional and otherwise

irrelevant connotations." In this regard, I would agree with him completely. Lennon goes on, however, to say "But it is sensible to wonder whether it can that easily now be exorcised from the language, or whether the terms invented to replace it will be more accurately interpreted." As usual, Lennon has a point.

I would suggest that in the attempt to rid our society of the term "I.Q.," exorcism is unlikely to be effective but that "benign neglect" may at last be a term for which an appropriate use could be found. I believe we should stop using I.Q. as an appellation. When we have just succeeded in substituting a standard score for a ratio, it seems counterproductive to continue using the term "quotient" or its abbreviation in describing intelligence. In response to Roger Lennon's query as to what terms might be used to replace "I.Q.," I would suggest that we could do worse than to use a list supplied earlier in his same paper. He said "To be sure, the content of most intelligence tests, from Binet to the present, has been drawn heavily from about a dozen types of tasks: vocabulary, general information, analogical reasoning, series or sequence manipulation, perceptual acuity, spatial abilities, quantitative skills, classification, syllogistic reasoning." Moving toward use of those terms, and toward measurement of the child's development with regard to those skills separately rather than in the aggregate, would mark a considerable advance in intelligence testing.

Even better than using the language of factor analysis to describe test content would be the further demystification of the tests by substituting, for the psychological terms we tend to use, words that are more common to the classroom and to the home. If a test is in fact a test of addition, for example, or more broadly of arithmetic, it doesn't help the teacher or the parent to call it a test of "quantitative skills." To have a child who can't add is one thing--regrettable but understandable and, one may hope, remediable. To have a child who is "deficient in quantitative skills" is, to most people who deal with children on a daily basis, only marginally intelligible but distinctly ominous, as if all hope for a complete child must be abandoned.

In effect, in designing tests we have used psychological constructs to provide the basic architecture, but in their development we have drawn heavily upon classroom behaviors that have been found valid in relation to those constructs. We have then named the test scores for the constructs rather than for the behaviors. To stick with my example of "quantitative skills," or "numerical ability," or "N" we have used this construct or cluster of constructs to help lay out what we want the test to include. We have then asked what manifestation of that factor one can expect to observe in children at a certain grade level, and concluded that problems involving addition and subtraction would be appropriate. The children taking the test have duly added and subtracted. But then we have called the

test score not a measure of addition and subtraction but of quantitative skills--a respectable part of intelligence.

To the psychological or educational research person, that escalation of vocabulary conveys a broader and richer sense of the variable in question, provided the facet of behavior or attainment tested is indeed an adequate basis for generalizing to the construct. But in the escalating process, we tend to lose altogether the people who are trying to make sensible decisions on the basis of the scores, and who would have a chance of doing so if the names placed on them were close to the operations that generated them. This process finds its apotheosis in the terms intelligence and I.Q.

One of the most unfortunate side effects of our use of the terms for constructs rather than for classroom-observable behaviors is the apparent justification it provides for basing long-term judgments and decision on test scores. In the case of young children, especially, ascribing long-term implications to the scores derived from tests as they are today is hazardous business. The tests measure abilities and skills that are learned by children at a period of dynamic development, in the first place, and under highly differentiated conditions of exposure to opportunities to learn the skills being measured in the second. The scores can be very helpful indeed in indicating how well a child can perform the specific operations required by the test, and consequently what that child is ready to do next. This kind of interpretation is encouraged by test titles

and score reports that sound like the language of the classroom. By contrast, the generic construct titles inevitably suggest enduring characteristics of the individual. That suggestion in turn produces a temptation to extrapolate present performance into the future, to classify students rather than teach them.

Whereas the use of test scores for short-term assignments to new learning tasks is eminently sensible, the tendency toward long-term assumptions or predictions may be the most pervasive negative aspect of testing in the schools today.

There must be, I think, a law that says the validity of the inference drawn from a test score varies inversely with the remoteness of the criterion. This law holds not only for test scores but for any information about people, and especially about young people. The point to be made here is that we encourage long-term prediction when we escalate the terminology we apply to abilities we have measured: from classroom skills to factors and from factors to eternal verities like I.Q., presumed in our society to encompass much of the person's permanent intellectual equipment if not total worth.

With tests that are geared recognizably to the kinds of questions to be resolved in the classroom, some long-standing issues may be less vexing. The more situation-specific the test, the less one is tempted to expect that performance will be invariant over time, ascribable to heredity, and generalizable to a large domain. The limitation on generalizability must be recognized as a loss, but the trade-off for the virtue of

greater validity for the decision at hand is likely to be highly worthwhile. Tests of this kind are likely to be very much the product of the learning environment, and it is difficult to see that the nature-nurture issue will burn as brightly as it has in the past in relation to single-score tests often of more abstract content.

The problem of labels has, of course, been of particular concern in relation to people whose cultural and linguistic backgrounds are not those of the mainstream. Two principal approaches have been proposed to the problems involved in interpreting the scores of pupils whose dominant language is other than English. The two approaches are separate norms or separate tests.

The solution through separate norms is probably a transitional step. Separate norms may be seen as useful if one is interpreting scores in terms of factors rather than in terms of the specific operations required by the test itself. The question being asked, in this case, is "What is this child's verbal ability?" The observation is that the child has made a low score on a test of reading passages in English. Immediately someone will point out that this child began speaking a different language at home and has had less exposure to the English language than have others in the norms group, and therefore his or her verbal ability cannot be judged in relation to the performance of the others in the group. Ergo, we need norms based on other children who have had limited opportunity to

learn English so that we can infer this child's verbal ability in relation to pupils of similar background.

The picture changes entirely if you ask, not "What is this child's verbal ability?" but "How well can this child read English?" If the child has a low score, and if the learning task to be predicted or assigned in the classroom is reading English, the teacher's main problem after seeing the test results is to make sure that the pupil is next given reading exercises at the proper (easy) level of difficulty. No assumption about the child's generic verbal ability is involved. It may be of some interest and even of some value to know that most other children from non-English backgrounds have equal, more, or less difficulty with the material, but such a discovery is largely immaterial to the classroom decision to be made about the child in question.

At present, since tests and scores carry factorial labels rather than operational ones, and since those who interpret and make decisions on the basis of scores are caught up in the escalation of inference to higher levels of abstraction, we probably need separate norms. The need for differential norms will tend to fade as we label and interpret tests more modestly. Whether or not the need will have disappeared by the year 2000 is a moot point.

Another approach, of course, is to provide comparable tests in a variety of languages--tests that are as nearly parallel as is possible. This is a costly procedure but one that is, of

course, technically feasible. With parallel tests available, children can be tested in their dominant language, achieving scores that more nearly reflect their abilities assuming that their opportunities to develop their competencies in the language of the test have been about equal. This solution has some utility if the question being asked is ability in the factor: a child who reads well in Portuguese and who is tested in Portuguese can demonstrate skill in reading and hence verbal ability, or "V." If the ensuing instruction is to be in Portuguese, the finding also is relevant to the academic decision to be made.

The situation is different if the ensuing instruction is to be in English. In such an instance, a high score on a reading test given in Portuguese tells you nothing about where the child is ready to begin the program of teaching and learning in English. In order to make that decision, you need a reading test in English, although of course it would be folly to interpret the latter score as indicating the child's standing on "the verbal factor." The score on the Portuguese-language version might not be without utility, however. If you had two children of comparable background, both with low reading scores in English, but one with high scores in Portuguese and the other with low scores, you might infer a greater developed reading skill in the former that could transfer to the learning of English. But the proof of the pudding would still be how

well each of the two children did indeed acquire the English-language reading skill.

Conclusion

My view, then, is that over the next twenty years or so we are likely to see evolutionary rather than quantum changes in intelligence tests, at least as they are used in academic settings. We are likely to see tests that provide separate scores on a variety of abilities. They are likely to be standard scores. The ratio defining the I.Q. may by then have been abandoned everywhere and the term I.Q. may have disappeared into psychological and educational history.

The new terms to replace I.Q. may well be drawn from factor theory at first but increasingly may refer rather to the skills required daily of the children in the classroom. The testing itself may likely be seen to draw its relevance and hence its utility more from the tasks of teaching and learning than from psychological theory, although the development of the test may draw importantly on psychological as well as educational theory.

With the movement toward rather concrete tasks embedded in the flow of learning, it is likely that those who interpret the scores will be more inclined to use them to make near-term decisions about the next problems to give the child and to refrain from assumptions about his or her long-term potential. Perhaps our most severe problems of test score misuse come from

decisions that cannot be modified or reversed in the near future on the basis of further evidence. Hence the development of a mode of test use that ties the scores to decisions with proximate consequences, if it comes about, will be of inestimable value. The same new emphasis on a variety of test scores as part of a dynamic system of instruction is likely to resolve the heredity-environment issue, for these tests in these circumstances, in the direction of environment.

Since the schools will still be dealing with pupils whose backgrounds of language and culture have provided differential opportunities to learn the tasks that make up their academic environment, we will need differential norms as long as people persist in relating the scores to psychological constructs rather than to classroom tasks. A more satisfactory solution will be at hand when comparable tests of the same abilities are available to describe the child's competencies in both the dominant language and the language of instruction.

If all these changes come to pass by the year 2000--as I believe they will--three questions remain. Will we then call these tests intelligence tests? If not, will we need still other measures to call intelligence tests? If the answer to either of those questions is "yes," why?

Bibliography

[1]Anastasi, A. Psychological Testing (4th ed.). New York: Macmillan, 1976. p.9.

[2]Ertl, J. P. Fourier analysis of evoked potentials and human intelligence. Nature, 1971, 230, 525-526.

[3]Hendrickson, A. The biological theory of intelligence. Paper presented at the meeting of the International Congress of Applied Psychology, Munich, 1978.

[4]Hendrickson, E. The physiological measurements of intelligence. Paper presented at the meeting of the International Congress of Applied Psychology, Munich, 1978.

[5]Jensen, A. R. g-Outmoded theory or unconquered frontier? Paper presented at the meeting of the American Psychological Association, Toronto, 1978.

[6]Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. New York: Wiley, 1977.

[7]Scandura, J. M. Problem solving: A structural/process approach with instructional implications. New York: Academic Press, 1977.

[8]Lennon, R. T. Perspective on intelligence testing. NCME Measurement in Education, 1978, 9 (2), 1-8.