DOCUMENT RESUME

ED 109 245 TM 004 709

AUTHOR Ebel, Robert L.
TITLE Current Issues in the Use of Tests in Education.
PUB DATE [May 75]
NOTE 18p.; Paper presented at the Annual Meeting of the Washington Educational Research Association (Seattle, May 23, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Affective Tests; Cognitive Tests; Criterion Referenced Tests; *Educational Testing; Intelligence Quotient; Intelligence Tests; Minority Groups; Norm Referenced Tests; Predictive Ability (Testing); Test Bias; Testing; *Testing Problems; *Tests; Test Validity

ABSTRACT

Six current issues in the use of tests are examined in this paper. Are the tests used in education valid? In general and to a reasonably satisfactory extent they are. Is the emphasis these tests place on cognitive achievements a serious limitation? Because of the basic and pervasive role of cognition in human affairs, and in the absence of any good alternative, it is not. Should we replace norm-referenced tests with criterion-referenced tests? In certain areas of learning where they are particularly appropriate, yes. In general, no. Are the tests biased against minorities? There seems to be little basis for the belief that they are. Is the I.Q. a myth? The notion of a latent-trait I.Q. that strongly influences rate of learning or amount that can be learned ultimately probably is a myth. In the absence of strong supporting evidence, it probably should be treated as a myth for the good of education and society. Finally, the competition engendered by testing seems distinctly more helpful than harmful. (Author)

Current Issues in the Use of Tests in Education*

Robert L. Ebel
Michigan State University

Among educators like us, among those directly and substantially involved in the processes of instruction, these are years of uncertainty and discontent, of reassessment and reform. Inevitably, this concern has extended to the means used to evaluate the effectiveness of our educational efforts, the tests used in education. Are they valid? Are they broad enough in scope? Should they be norm-referenced or criterion-referenced? Are they biased for or against particular groups in our population? Is the I.Q. really a myth? Does testing result in destructive competition?

Criticism of tests is not a new thing. It is not wholly a product of the contemporary unrest in education. It must have existed in ancient China, where positions of prestige and power were largely reserved to those who could pass rigorous tests of classical scholarship. It did indeed exist in the medieval universities, whose oral examiners did not always question their students relevantly nor judge their answers fairly. It existed in Massachusetts in 1845 where Horace Mann insisted that written tests be used to assess the effectiveness of instructional efforts. It existed throughout this country in the 1920's when objective tests began to gain favor. It surrounded the development of state testing programs and the current activity in assessment. And, clearly, it is with us today.

The reasons why tests are criticized and opposed are not difficult to find. The tests themselves are imperfect. Indeed, some are seriously flawed. They

*Prepared for presentation at a meeting of the Washington Educational Research Association in Seattle, May 23, 1975.

are sometimes used unwisely, misinterpreted, overinterpreted, or handled as weapons rather than as tools. They reflect particular perceptions of the goals of education, which not all educators share. But the main reason for opposition to tests and testing does not lie in the faults of the tests, or errors in their use, or differences in educational values. It grows out of the reluctance of educators to be judged and held to account for their stewardship. Tests provide a means for auditing the accounts of the educator. They can report profits and losses, indicate competence and incompetence, identify good management as well as mismanagement. Education is blessed with a great many capable and dedicated teachers and administrators. But the profession of education also has its share of mediocrity and of false messiahs. It is from the latter group that the loudest protests are heard against tests and testing.

Fred M. Heckinger, one-time education editor for the New York Times, commented on the objections of educators to educational tests in these words:

> There is today something like a last-ditch battle being waged which attacks everything that permits schools and colleges to be subjected to comparisons. In this battle, tests are not sniped at primarily because they are foolish tests (which some of them are) or because youngsters are scared (which fewer of them are than many of their teachers think); tests are attacked because they test the school, the community, and the state and thereby violate the fundamental freedom of American education--the local option to be as comfortable as the local school board, superintendent, or college dean think good. I hear time and again--and again last week at the meeting of the Council of Chief State School Officers--the warning that national tests and such commissions as that of the College Board on English and Mathematics "put American education in a straitjacket" and "dictate the curriculum". I would rather use a different kind of image--that of the importance of a skeleton, a brain, and a heart to make of American education more than a formless mass.[1]

---

[1]Heckinger, Fred M. "The All American Picnic is Over". New England Association Review, January, 1961.

Though they were written more than a decade ago, these words still provide an accurate diagnosis of one major source of opposition to the widespread use of educational tests.

To recognize that some criticisms of tests and some objections to testing are motivated more by the desire to protect educators than to serve education is not to deny the value of critical commentary from fair-minded, well-informed observers. Neither our tests nor the uses to which they have been put are above reproach. The enterprise of educational evaluation, and of the whole process of education, stands to gain from honest assessment of its limitations as well as its contributions, of its shortcomings as well as its accomplishments. Let us return to the concerns about testing mentioned at the beginning of this paper.

Are educational tests valid? By validity, we mean the extent to which the test measures what needs to be measured. Thus, the validity of a test depends not only on what it does measure but also on what it ought to measure. What the test does measure is almost certain to be some aspect of cognitive development. Regardless of the title of the test, or of any claims made for it, if it is a paper and pencil test, it is a test of the examinee's knowledge--knowledge of some subject of study, knowledge of the world he lives in, or knowledge of himself. The task of answering the questions on a paper and pencil test is inevitably and inescapably a cognitive task.

Of course, there are tests of aptitudes as well as of achievement. There are tests which purport to measure motivation, creativity, and various dimensions of personality. There are tests of interests and attitudes. But each of these tests presents the person who takes it with a series of cognitive tasks. They ask the examinee in substance, "What do you know about yourself, and the world, and about how to answer a question like this?" They are all tests of knowledge.

Is this what our tests ought to measure? Surely cognitive development is an important aspect of human development. Perhaps it is the only important aspect of mental development. The peculiar excellence of human beings is their ability to produce and to use verbal knowledge. We invest heavily in research to discover knowledge. We write and speak to communicate it. We maintain libraries to preserve it. The focus of most of our educational efforts, from primary school to graduate school is on the transmission of knowledge.

This knowledge is more than information. It implies understanding and ability to use. It requires development of a coherent structure of concepts and relations. A mind stuffed with facts does not possess the kind of knowledge we here recommend.

What are the alternatives to knowledge as an objective of education? Is it the ability to think? Is it intellectual skills? Is it cognitive strategies? Each of these would find some support among educational psychologies. But successful exercise of each seems to depend on knowledge of how to proceed. To the extent that they have been developed, they exist as part of a person's structure of knowledge. We may separate knowledge from thinking verbally and conceptually, but operationally the two are inextricably intertwined. They do not constitute alternative goals of education in any real sense.

We should not, therefore, criticize the educational tests we use for lack of validity on the ground that the test cognitive development instead of testing something else that is more fundamentally important. But if the tests demand knowledge of trivia rather than of important comcepts, of fundamental principles, or of useful proc dures, they may be seriously lacking in validity. If they provide unreliable scores because the questions are too few, or too ambiguous, or too hard, or too easy, the tests also may lack validity. Teacher-made tests, incidentally, are far more likely to exhibit such shortcoming than are published tests. And these faults are remediable. They are not at all inherent in the test form.

Let us turn from the question of validity to consider a second issue in the use of tests in education. Are the tests we use broad enought in scope? Do they neglect the affective domain? Do they neglect non-cognitive aspects of human development like interests, attitudes, ideals, appreciations, self-concept, social concerns, and ethical responsibility? Have we dehumanized education in our concern for cognitive development?

I do not think so. It is surely true that feelings are important. Nor can one deny that all of us need to be sensitive to the feelings of others with whom we deal. This is especially true of teachers, I think. But none of this persuades me that schools should devote a substantial portion of their time and effort to something called affective education. It does not persuade me that our testing programs should include tests of affective development. Why not?

First, note that feelings are not subject to educational development by instructional processes. They are part, and a most valuable part, of our biological heritage. We are born with an extensive repertoire of them--hunger, sleepiness, love, rage, fear, pain, and so on. Maturation adds others. Experience attaches these feelings to a wide variety of objectives and situations. It also teaches us to cope with unpleasant feelings, and to curb the expression of certain of our feelings in certain situations. But feelings cannot be developed in the way that physical skills and cognitive abilities can be developed.

Much of the current concern for affective education stems from the work of a Committee of College Examiners which met under the chairmanship of Benjamin Bloom from 1949 and 1953. They divided all educational objectives into three domains: cognitive, affective, and psychomotor. Their outline of cognitive objectives was published in 1956,[2] and of affective objectives in 1964.[3]

---

[2]Bloom, B.s., and Krathwohl. D. R. Taxonomy of Educational Objectives: Cognitive Domain. New York: David McKay Co., Inc., 1956.

[3]Krathwohl, D. R. et.al. Taxonomy of Educational Objectives: Affective Domain. New York: David McKay Co., Inc., 1964.

One who examines the outline of affective objectives will find that they have more to do with behavior than with feelings; with attention, acquiescence, response, acceptance, preference, commitment, conceptualization, organization, and characterization. What is the basis of that behavior? If it is rational behavior, it is cognitively based. If it is habitual behavior, it may be based on custom or imitation. Need we ask which of these, reason or habit formation, should be the principal concern of the schools?

Now it may well be true that schools should be more concerned with the pupil's cognitive development in the area of feelings, self-concepts, and interpersonal relations. This might mean more courses in psychology, perhaps, and more study of problems of behavior. This may be all that advocates of affective education are really asking. But if this is so, cognitive education is a means, not an alternative, to affective education. If this is true, tests of cognitive development in the understanding of feelings, self-concepts, and interpersonal relations will be quite adequate to assess a pupil's affective development.

One note of caution. Those who would de-emphasize cognitive goals in favor of affective goals are often quite vague about the precise nature of those goals, of the instructional procedures that lead pupils toward them, and of the assessments that will show whether or not they have been attained. Thus, professed concern for a somewhat intangible outcome can serve as an excuse for lack of achievement of the more tangible goals. Those concerned with effective utilization of our educational resources should not be put off so easily. To the extent that a proposed educational outcome is intangible (i.e., not clearly defined), it is worthless as a goal of instruction. We do not say, "If you can't measure it, it doesn't exist" We do say, "If you can't define it clearly, you cannot teach it purposefully or measure it validly."

To sum up the discussion of this issue, let us say that the tests we use probably are not too narrow in scope. If they lack ideal coverage, it is not

because they are largely limited to measuring cognitive achievements. It is because school programs may not include all of the most important areas of cognitive development. For the cultivation of cognitive development is the essential task of the schools.

Consider next the issue of test types. Should our tests be norm-referenced or criterion-referenced? A norm-referenced test is so called because it interprets the test score of a particular pupil in relation to norms established by testing other similar pupils. Most norm-referenced tests also sample the domain of a particular achievement diffusely. A criterion-referenced test reports which, or how many, of a set of specific goals for achievement a particular pupil has reached. Instead of sampling diffusely the multitude of elements of knowledge or skill included in a domain of achievement, the criterion-referenced test concentrates on a limited number of specifically defined goals, testing each of these repeatedly to make certain that the particular goal has actually been achieved.

There are advantages and limitations of each type of test. A criterion-referenced test can report more specific and detailed information on pupil achievement at the cost of a more extensive and cumbersome report form. A norm-referenced test provides a more concise summary of a less clearly defined area of achievement. Criterion-referenced tests emphasize mastery of specifics by all pupils, though there is some arbitrariness in the definition of mastery and in the choice of specifics to master. Norm-referenced tests encourage and reward individual excellence in achievement, through here too there are elements of chance in the selection of tasks on which excellence is to be demonstrated.

There is a difference between the two types of tests in the conceptions of learning they reflect. Criterion-referenced tests treat learning as if it were acquired by adding separate, discrete units to the collection of things learned, somewhat like adding beads to a string. Norm-referenced tests are consistent

with the notion that learning consists of building a structure of numerous relations between concepts, somewhat like a spider's web hung among twigs and branches. In some early stages of learning, such as the recognition of letters and numbers of memorizing the basic facts of addition or multiplication, the notion of accumulating separate and largely unrelated elements in sequence may be reasonably accurate. But even here relationships begin to appear, and as one progresses to more advanced learnings, the relations become more and more important and numerous. It becomes more and more difficult to identify a limited number of specifics to be learned to mastery.

The notion of mastery is appealing, but perhaps deceptively so. Complete mastery of any but the simplest ideas and skills is unattainable. Nor does there seem to be any good reason to insist that each pupil who completes a course should have mastered all that was set out to be learned in it, that he must continue to study it until he has learned all that any other student learned from it. Since none of us has unlimited time for living or for learning, there are serious practical obstacles to making time, rather than amount learned, the principal variable in the learning process.

Inherent in criterion-referenced testing, with its emphasis on separate specifics rather than on structures of understanding, there is a pedagogical danger. It is that teachers will teach too directly to the particular tasks used in testing. Pupils may be induced to learn particulars by rote instead of struggling to understand. One can teach badly by being too specific about the goals of learning, as well as by being not specific enough.

Criterion-referenced testing has the appeal of novelty and innvoation. It may seem to offer more meaningful measures of achievement, as well as escape from some of the problems inherent in norm-referenced measurements. But it creates special problems. Incidentally, the presumed necessity of redoing test

theory to accommodate criterion-referenced testing is probably based on mistaken assumptions. Even if test scores with little or no variability should be attained under the mastery learning model, which is seldom if ever the case, we can still test the effectiveness of the test, and of the items, by administering the test before and after instruction. The job of any achievement test, whether criterion-referenced or norm-referenced, is to differentiate those who have from those who have not learned. Classical measures of test reliability and of item discrimination indicate how well the test has done that job.

No doubt there are many situations, especially in the early grades and in program of technical training, where criterion-referenced tests are more suitable than norm-referenced tests. But the notion that criterion-referenced tests provide a generally superior method for measuring educational achievement is probably not true.

One final caveat concerning criterion-referenced testing. Some proponents suggest that appropriate objectives should be selected, and satisfactory levels of mastery be determined by local schools, by individual teachers, or by each pupil. If this suggestion is adopted the possibility of wide scale testing disappears. With it goes the possibility of meaningful comparisons of achievement among schools, among classes, and among pupils. What a convenient escape from evaluation for all concerned! If each can set his own goals and make his own assessments of how well they have been achieved, who stands to lose anything? It is like the caucus race in Alice in Wonderland, in which everyone wins and everyone gets a prize. Not all criterion referenced testing involves this kind of psychometric sleight of band. But some does, and if it does honest educators will join in crying "beware".

Consider next the issue of bias in tests. What does it mean to say that a test is biased for or against a particular group of examinees? To some it

means only that the group makes scores that are lower, on the average, than scores made by others. By this definition of bias, a spelling test might be said to be biased against poor spellers. To others, bias means that the test systematically underpredicts the performance on the job of members of the particular group. But underprediction is not an infallible indication of bias. A test which underpredicts may be a perfectly unbiased measure of what it does measure. If it does not measure all components of job productivity, such as willingness to work, energy, strength, etc., and if the particular group excells in these qualities, the perfectly unbiased test will underpredict when used as the sole predictor.

There are other definitions of test bias, such as those proposed by Thorndike[4] and by Cole.[5] This is not the place nor is there time now for a close look at the justifications and consequences of these two definitions. Let it suffice to say that by these definitions, a test is biased against anyone who scores below the mean on the test, whether or not he is a member of a particular cultural, racial, or ethnic group. Nor can this kind of bias be eliminated so long as our tests are less than perfectly reliable.

Claims of test bias can be sustained or refuted by at least two lines of argument. The first is empirical. The second is rational. Those who contend that our educational tests have a middle-class bias which makes them unfair to minorities often simply assert that the bias is "well known", without citing specific evidence. Such specific evidence of bias against minorities seems quite hard to find. On the other hand, substantial evidence of test bias in favor or minority group members has been reported. Stanley and Porter[6] in a

---

[4]Thorndike, R. L. "Concepts of Culture Fairness." Journal of Educational Measurement, 1971, 8 (2) 63-70.

[5]Cole, N. S. "Bias in Selection". Journal of Educational Measurement, 1973, 10 237-55.

[6]Stanley, Julian C. and A. C. Porter. "Correlation of Scholastic Aptitude Test Scores with College Grades for Negroes Versus Whites," Journal of Educational Measurement, 1967, 4, 199-218.

study of the prediction of success in college found that Scholastic Aptitude Test scores tend to overpredict the performance of Black students. Other investigators[7,8,9] report similar findings.

The rational argument focuses on the content validity of the test. If the items in the test sample representatively the areas of knowledge and skill that constitute competence in the ability being tested, then the test possesses content validity for all examinees regardless of cultural, ethnic, or racial origins. Language or experience differences that handicap the minority examinee in his attempt to demonstrate knowledge and skill on the test are likely to handicap him also in his attempt to utilize that knowledge and skill in other situations. When this happens, the test cannot be said to be biased against the members of the minority. Rather it reflects quite accurately the usuable competence he possesses.

The argument that educational tests are biased against minorities because they reflect middle-class values has a superficial appeal but is probably fallacious. Recall that these tests present essentially cognitive tasks. The items in them were selected because they represent important aspects of competence, not because they reflect the values of a particular culture. They reflect what is true and important in our common culture. That common culture needs to be maintained and extended. In a coherent society, the values which are shared

---

[7]Kallingal, A. The Prediction of Grades for Black and White Students at Michigan State University", Journal of Educational Measurement, 1971, 8: 263-65.

[8]Cleary, T.A. "Test bias: Prediction of Grades of Negro and White Student in Integrated Colleges." Journal of Educational Measurement, 1968, 5: 115-124.

[9]Pfeifer, C.M. Jr., and Sedlacek W.E. "The Validity of Academic Predictors for Black and White Students at a Predominately White University". Journal of Educational Measurement. 1971, 8: 253-61.

must be more numerous and more basic than those on which people differ. Cultural apartheid ought not to be encouraged in this society. We have suffered enough from it already in our schools.

The tests we use in education ought to be as free of bias as we can make them. But the extent and seriousness of bias in our current educational tests can be, and probably has been, exaggereated. The "well known" bias of tests against minority group members o[illegible] more fanciful than factual.

Consider next the I.Q. as an educational issue. Is it really a myth? The recent television special presided over by Dan Rather[10] argued that it was. Whether you agree depends in part on what you understand the I.Q. to be. If you define it in simple operational terms as a derived score on a particular kind of test, then there is nothing mythical about it. But if you define it as a measure of some latent trait that determines how rapidly a person can learn, and that places a limit on how much the person can learn, its real existance becomes much more debatable. It was this second conception of the I.Q. that Dan Rather sought to expose as a myth.

Let it be noted at the outset that there is no conclusive or overwhelming evidence on either side of this issue. Those who believe that the I.Q. is not a myth point out that studies of twins suggest a genetic basis for whatever is being measured by an I.Q. test. They note that blood types and some metabolic characteristics seem to be genetically determined, and that physical characteristics in general seem subject to genetic influences. Hence, they argue, the anatomical and physiological basis for learning is likely also to be subject to genetic influence. Individuals are likely to differ with respect to the quality of their basic equipment for learning. If they do, differences in rates of learning and in ceilings on learning seem plausible.

---

[10]Ravitch, Diane. "The I.Q. Myth-Criticisms, Complexities, Contradictions." The New York Times, Sunday, April 20, 1975. Section D, page 29.

Those who regard latent trait I.Q.'s as essentially mythical note that no anatomic or physiological differences correlated with I.Q. scores have been discovered among normal human beings. They acknowledge the possibility of such differences, but doubt that they can have much influence on how fast or how much a person learns. Human beings seldom try to learn as fast as they can, or as much as they can. Seldom do they extend themselves to the limit of their capabilities.

It is clear to those on both sides of this issue that I.Q. tests cannot test native ability directly. There is no task on any I.Q. test that any person was born knowing how to do. All require the examinee to show what he has learned to do. In this light, an I.Q. test is a test of learned achievement. Indeed, the tasks on some group tests of intelligence are indistinguishable from tasks on a general achievement test. Thus, it is quite reasonable to attribute at least some of the differences in I.Q. scores to differences in opportunities to learn, or to accidents of success or frustration in learning.

Since we cannot measure latent trait I.Q.'s directly, and since our indirect measurements are distorted by uncontrolled influences, it seems clear that we are unlikely to get highly reliable or highly valid measures of the I.Q. It also seems clear that such measures are not really necessary. All learning builds on prior learning. If we know from relevant tests of past achievement how firm the foundation for continued learning is, we can forecast with reasonable accuracy how well future learning is likely to proceed.

Surely it is true that different individuals of the same age have different advantages or handicaps in learning. Whether these were built-in genetically or developed by experience makes little difference in their effect on the individual who possesses them. But to those of us who operate schools, to those whose mission and responsibility is to foster educational development, it makes a considerable

difference. Until faced with strong evidence to the contrary, evidence that does not now exist and that gives no promise of appearing soon, we should probably behave as if there were no significant genetic limitations to the learning potential of any normal human being. It is to our advantage, and to the advantage of our society, to regard latent trait I.Q.'s as mythological.

Consider finally the issue of competition versus cooperation. Those who oppose the use of tests to obtain accurate measures of achievement, and of grades to report those measures, often deplore the emphasis that tests and grades put on competition. What modern societies need, they argue, is more cooperation and less competition. If we do not limit and discredit competition, they say, we will ultimately destroy the finer human qualities of compassion and social responsibility.

The record of discussions of this issue suggests that it is easy to jump quickly to quite superficial conclusions about the relation between competition and cooperation; about the evils of the one and the virtue of the other. Let us try to examine this issue more closely, to think about it as clearly as we can.

Are competition and cooperation mutually exclusive alternatives? Consider the members of a basketball squad. During practice they compete with each other in being cooperative with other potential members of the team. During a game the cooperate with each other in order to compete effectively against their opponents.

Are cooperation and competition opposite in meaning? If cooperation means working with others toward a common goal, then the most direct opposite must be to work alone toward an individual goal. Is cooperation always preferable to individual effort? Surely the answer must depend on the particular circumstances. Further, if competition means working to excel others in doing

something, then the most direct opposite must be to avoid working to excel anyone at doing anything. Viewed in their light, does competition look so bad? What would be bad is uncontrolled, cut-throat competition. There is no reason for a society to tolerate that kind of competition, nor for decent men and women to engage in it. The possibility that it might occur is not great enough to justify rejection of the whole idea of competition. Nor is there any good reason to think that people will automatically learn the values and techniques of cooperation if competition is banned.

To say, as has been said, that cooperation involves working <u>with</u> others, whereas competition involves working <u>against</u> others is only partly true. Consider a runner, a jumper, a weight lifter, an archer, a golfer, a bowler, or an ice dancer. Each tries to excel the other competitors. But if any of them does anything <u>against</u> the other competitors, he is subject to penalty or disqualification. When we compete it is always with others. Only in some cases is it against others.

Do test scores and course marks stimulate competition? Surely they can. Whether they do depends on the inclinations of those who receive them. Scores and marks make comparisons possible. They do not make competition inevitable. My friends and I compare our monthly telephone bills, our property tax assessments, the number of miles our cars run on a gallon of gasoline, and the sizes of the classes we teach, with little or no interest in competing with each other. We are more interested in sympathy than in superiority.

Those who pursue excellence in education or in any other endeavor can hardly avoid comparisons, for excellence is a comparative term. These comparisons may involve some overtones of competition; of competition <u>with</u> rather than competition <u>against</u>. And if it is competition, what then? Can the pursuit of excellence be discredited by calling it competition? Would we be better

off, individually and as a society if there were no competition in business or politics, in scholarship or invention, in surgical skill or social service? When competition stimulates efforts to achieve and leads to achievement it must be beneficial. It seems inherently more productive than destructive. When it results in hostility or dishonesty, shall we blame the competition or the competitors? What made Sammy run so destructively to such a tragic end was not mainly the competitive society in which he lived. It was the amorality of Sammy.

Cooperation and competition are not mutually exclusive alternatives. They are not direct opposites in meaning. Neither is all good or all bad. Each has its place in a good society. Our pupils need to be taught the values and techniques of both. To discredit tests and grades on the ground that they encourage competition is to make an argument that requires much stronger rational support than it has received thus far.

Six current issues in the use of tests have been examined in this paper. Are the tests used in education valid? In general and to a reasonably satisfactory extent they are. Is the emphasis these tests place on cognitive achievements a serious limitation? Because of the basic and pervasive role of cognition in human affairs, and in the absence of any good alternative, it is not. Should we replace norm-referenced tests with criterion-referenced tests? In certain areas of learning where they are particulary appropriate, yes. In general, no. Are the tests biased against minorities? There seems to be little basis for the belief that they are. Is the I.Q. a myth? The notion of a latent-trait I.Q. that strongly influences rate of learning or amount that can be learned ultimately probably is a myth. In the absence of strong supporting evidence, it probably should be treated as a myth for the good of education and society. Finally, the competition engendered by testing seems distinctly more helpful than harmful.

Can these conclusions be dismissed by those who may disagree with some of them as "merely one man's opinions"? Perhaps they can, though they are based on almost a half century of observation and reflection and study. Do we need more research on these issues? Good research is always helpful, but in areas of human characteristics and behavior it is often difficult to do good research that will produce sound and generalizable findings. Research can tell us with some degree of accuracy what is so. It can not tell us what we ought to do about it. Above all, we need to guard against the notion that research is a substitute for reflective thought and judicious deliberation. It can provide data for us to think about. It can enable us to base our decisions more soundly. But research is by no means an all-purpose, problem-solving technique. Human thought is. To the extent that it is clear, straight, and well informed it will give us the best answers we are likely to get.