DOCUMENT RESUME

ED 340 779 TM 018 035

AUTHOR Noggle, Nelson L.
TITLE The Misuse of Educational Achievement Tests for Grades K-12: A Perspective.
SPONS AGENCY Congress of the U.S., Washington, D.C. Office of Technology Assessment.
PUB DATE Oct 91
NOTE 16p.; Contractor report prepared for the Office of Technology Assessment titled "Testing in American Schools: Asking the Right Questions." For related document, see TM 018 025.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Educational Assessment; Educational Change; Elementary Secondary Education; *Evaluation Methods; Norm Referenced Tests; *Standardized Tests; Student Evaluation; Test Bias; *Testing Problems; Test Results; Test Use
IDENTIFIERS *Reform Efforts

ABSTRACT

At the present time the American educational system seems caught in a squeeze play between the historical need for tests that are simple to administer and understand versus the current demand for tests that are respected as thorough measures of the kind of learning needed in a competitive and changing world. Many of the misuses of test data arise when conclusions are drawn from inappropriate evaluation methods. There is good evidence that concerns about testing practices may be resolved by focusing on how quality evaluations should be conducted. Testing often seems to miss the actual targets of education, and the situation is further complicated by tests that are not reliable or that cover the wrong content. Reforms must prevent covert agendas of any interest group and they must reduce discrimination in testing. The major source of misuse and abuse of tests is ignorance about the nature of important capabilities for today's students and how to measure those capabilities during and after their acquisition. If experts could agree on the capabilities needed for employment, postsecondary education, citizen responsibility, and personal and family satisfaction, they would have a common focus, one that consumers of test information could understand. Two attachments illustrate the discussion. (SLD)

THE MISUSE OF EDUCATIONAL ACHIEVEMENT TESTS FOR GRADES K-12: A PERSPECTIVE

by
Nelson L. Noggle
Centers for the Advancement of Educational Practices
Chandler, AZ

for
Office of Technology Assessment
U.S. Congress

October 1991

THE MISUSE OF EDUCATIONAL ACHIEVEMENT TESTS FOR GRADES K-12

A PERSPECTIVE

SOURCES OF MISUSE

Achievement testing in America's schools is in a state of chaos, partially due to societal, technological and instructional changes during the past 50 years, and partially due to the stress of the current education crisis. For every person advocating increased testing, we can find one advocating its reduction or elimination. For those asking for a "simple yardstick" to measure student achievement, we can find someone calling for "multiple measures". The stalwart of testing, the standardized, norm-referenced test, is both revered and despised. Test anxiety, test bias, and invalidity drive many of the criticisms. One of the major criticisms, however, is the misuse of tests and test results.

Increased criticism of standardized tests has led to new forms of testing, or to old forms of testing renewed. The criterion-referenced tests of the 1970s and 1980s are giving way to performance-based and wholistic assessments, including non-test indicators of achievement. Policy makers, however, still want the normative features of standardized tests which allow them to view achievement against other schools, states, or nations. Teachers, parents and students, however, are asking for measures more relevant to each individual student's situation.

The Nation's response to the educational crisis, "educational reform", is nearly a decade in the making. National, state and local groups have been formed to set goals for education; however, most have been stymied when trying to agree on performance indicators. Much of the difficulty rests with the lack of faith in tests or test users. A simple analogy puts this into perspective:

> During the 1950's, a person could go into almost any local hardware store and get a wooden yardstick. Depending on the parent, that yardstick might be used to (1) measure a child's physical height (status measure), (2) measure a child's growth (trend measure), (3) compare the height or growth to those of other children (normative measure), and (4) combine with other measures of physical growth (wholistic measure). Also, depending on the parent, that same yardstick might be used to (1) spank the child for misbehavior, (2) tell the child he/she is too tall or too short, or (3) tell the child he/she is growing too fast or too slowly.
>
> As the wooden yardstick was replaced by metal rules and other advanced measurement devices, the parent was able to obtain more precise (reliable) measures of height. The basic increments of height, feet and inches, did not change (validity). A person who is measured as 5'3" in 1991 is seen as having the same height as one measured 5'3" in 1955. Changes made in the yardstick, however, might have altered how the measurement itself was obtained (test administration), also how the yardstick was misused; e.g., spanking a child with a metal ruler, or comparing a child's height or growth against outdated charts (norms).

As the analogy suggests, much can affect the use and misuse of a yardstick. In the case of achievement tests, much more has happened to them, over time including major variations in test content (validity) and test use (purpose). As a result, America has lost faith in its educational yardstick. Above all, tests did not provide information which was adequate enough to avert the current educational crisis.

Domino Effect of History

In the 1950's educational achievement was measured largely by standardized tests and teacher-made assessments. These standardized tests resulted from rigorous validity and reliability studies which supported four very important hypotheses:

1. Multiple-choice measures can compare favorably with other measures of achievement, such as performance observations, recitations, and essays.

2. Standardization of test-administration procedures ensures objectivity of the resulting achievement measures.

3. Content of standardized tests can adequately cover the broad range of knowledge and skills commonly emphasized in schools across the nation.

4. Standardized, multiple-choice tests are more cost effective than performance observations, recitations, and essays.

As a result, early standardized achievement tests were quite lengthy, requiring many hours of testing for a complete test battery. Students were sent to gymnasiums, cafeterias, or auditoriums in order to accommodate the test administration. The tests themselves were treated as "broad trait measures" (reading, math, language, etc.), rather than specific assessments of the local curriculum. When Sputnik entered us in the "space race", it only served to fuel the need for standardized trait measures, especially math and science.

During the 1960s and early 1970s, most of the technological advancements in testing were devoted to "milking" more precision from standardized multiple-choice tests. Schools were pressing for shorter subtests and shorter testing times to conform with their 40-45 minute class periods. Reliability estimates (correlations) for the typical subtest dropped from the high 0.90s to the low 0.90s or high 0.80s, as subtests shrank from 60-80 items to 30-50 items. In addition, the content of math tests changed dramatically as the schools experienced the "modern math" aftermath of Sputnik.

During the late 1960s and early 1970s three extremely important, interrelated movements occurred: (1) federal programs were funded for disadvantaged students, (2) "mastery learning" drove a wave of instructional development, and (3) "back-to-basics" called for a detailed, narrow specification of learning objectives. Standardized tests were used to select students for programs; and criterion-referenced tests measured mastery of each learning objective. Publishers responded by developing criterion-referenced tests, and revising standardized testing to look more criterion referenced.

As the "accountability period" of the late 1970s and early 1980s materialized, educators and policy makers were faced with a wide variety of significantly short, narrow standardized tests. In addition, publishers were encountering greater difficulty in obtaining nationally representative norms. Where the bulk of the students reside, in the Nation's 40 largest cities, the schools were inundated by research studies. Various test publishers had difficulty getting participation of these schools during their norming studies. One wonders if this was responsible for certain tests yielding significantly different results -- there seemed to be too much variation in content and norms across publishers.

During the 1970s and 1980s, test publishers seemed to be "steering in troubled waters". As test results of different ethnic groups and cultures were reported, publishers responded with studies of test bias as an integral part of their R&D efforts. As schools called for shorter and more locally relevant tests, publishers responded with shorter tests for each grade level, and better coverage of the curriculum in large population areas and in their major customer's schools. With the call for more diagnosis and remediation, publishers produced diagnostic and prescriptive score reports; and they linked their standardized tests to their criterion-referenced tests. As test publishers scrambled to keep up with the demands of schools, so did the textbook publishers. Many test publishers, however, continued to look to the best selling textbooks as a major source of what to place in their tests. As these textbooks began to vary, both in approach and content, test publishers were faced with the largest ambiguity of all -- what is the curriculum standard?

As we commence the 1990s, policy makers still want simplistic scores that they can compare across schools, states, and nations. In addition, they seem more open to the use of other measures which allow performance-based interpretations of a student's knowledge and skills. The educators, however, are concerned about the classroom relevance of any test; especially if increased testing drives certification and pay-for-performance decisions. The American education system seems caught in a "squeeze play" -- the historical need for tests which are simple to administer and understand, versus the current demand for tests which are respected as thorough measures of the kind of learning needed in a competitive and changing world.

Ambiguity of Competing Perspectives

Level-to-Level Needs. Regardless of the type of tests being used, much of the misuse arises from the differing needs or priorities across each level of our educational system. Too often, a single type of test is called upon to serve too many of the following stakeholders who desire or demand achievement data:

1. Departments of Education (federal, state, local)

2. School Boards and Legislators (federal, state, local)

3. Advisory Groups (federal, state, local)

4. Media (newspapers, television, radio)

5. School Administrators (superintendents, principals, etc.)

6. School Teachers (Classrooms, subjects, specialists)

7. Classroom Helpers (trained aides, student teachers, parent helpers)

8. Students (grade level, level of achievement, level of at-risk)

9. Parents (economic status, education level, family situation)

10. General Public (faith in schools, attitude toward taxation, sense of community)

11. Business Leaders (standards of employment, technological advancements, economic competition)

12. Institutions of Higher Education (general studies requirements, program/major requirements, enrollment management)

Teachers argue that standardized test scores tend to underestimate the success story they observe in the classroom. The nation's leaders say that good measures would assure them that student achievement is equal to or greater than that of other countries like Germany and Japan. Teacher unions argue that testing practices must be fair especially as they are used to judge teaching performance. Business leaders point to the need for measures of communication and problem-solving skills among high-school graduates. Institutions of higher education want test scores which are better predictors of future academic performance, and which emphasize the higher-order thinking skills. The public wants test scores which reflect appropriate use of their tax dollars, and the quality of education their children are receiving. The school boards and media want information to alert the public about the success of schools or administrators.

Faith in Tests. As the history of testing has unfolded, the various misuses and abuses of tests has had a differential effect on those needing achievement data. For every school-board member, administrator, teacher, parent or student who respects or trusts testing, there is someone who has lost faith in testing, at least as it is currently done is our schools. Some of the lost faith is due to technical or practical issues surrounding the tests themselves, especially test relevance. Much of the problem, however, is caused by those who use the tests incorrectly, as they attempt to go beyond the purpose of the tests or try to over interpret the test data.

Accountability and the Fear of the Unknown. Many business leaders assert that educational accountability has "no teeth". Yet, teachers fear that test scores are being scrutinized somehow to devalue their teaching or hold them back from salary increases or promotions. With the exception of merit-pay and career-ladder programs, many educators do not know how or if accountability is being implemented. They tend, therefore, to be afraid of the unknown; but, they know that tests are somehow involved.

Teacher Training. The "rubber meets the road" in the classroom, where teaching and testing occur in tandem with one another. Too often, however, the acts of teaching and testing seem like "oil and water". The following analogy might shed light on this issue:

> When we drive a car, we are constantly looking left-to-right and front-to-back. We continually assess the flow and speed of traffic, watch for unusual events, and judge whether or not this route is getting us where we are going or getting us there on time. Do we call this testing or assessment? Or, do we call it driving a car?

Assessment should be integrated as a natural expansion of the art and science of teaching. Tests are but one type of assessment used by good teachers, who use many different kinds of measures of learning. They observe, ask for recitations, assign special projects, give essays, administer teacher-made tests, administer published tests, and use challenging exercises or problems. They look for "multiple lines of evidence" to judge student achievement. Unfortunately, too many teacher-training institutions minimize or neglect assessment courses, as they push a singular view on testing, typically traditional standardized, multiple-choice testing, or the notion that testing has too many faults to be taken seriously.

CIA Paradigm. Whether it is in the classroom or for an entire school, district or state, there is the issue of how assessment practices drive test use. Attachment A illustrates the CIA Paradigm, which simply depicts the obvious -- a good educational program finds an appropriate alignment of curriculum (C), instruction (I), and assessment (A). When the alignment comes from "partnership or ownership" models of stakeholder involvement, a program chooses its curriculum, promotes the type of instruction which best delivers the curriculum, and devises appropriate measures to find out if the curriculum has been learned. When the alignment is driven by "special interests", assessment can become the driving force, almost as follows:

> When assessment starts to drive educational programs, we make our programs susceptible to both "overt" and "covert" activities of special interest groups. It is no surprise that we find increased evidence of teaching to the test, narrowing of the curriculum, and lower levels of cognitive skill development.

At one point, experts were giving educators the "green light" to invert the CIA Paradigm and place assessment at the top -- they called it "measurement-driven instruction (MDI)". One wonders if this is "putting the cart before the horse."

Search for Simplicity. Raising test scores sounds simplistic; everyone can understand that they went up/down or remained stable. Such simplicity is highly seductive. Unfortunately, learning is not simplistic; nor is its measurement. Even when we narrow down learning to mean "remembering important facts" or "recognizing correct answers", we find it extremely difficult to devise scores people will correctly interpret and use. Raw scores, percent correct, percent mastery, percentiles, grade-equivalent scores, normal-curve equivalents, stanines, and standard scores, tend to cause more confusion

rather than less. What will happen when we really begin to obtain multiple measures of achievement, e.g., across subjects, levels of reasoning, and types of application? Can we keep it simple? The following personal experience tends to keep things in perspective:

> I had been asked by a superintendent to attend his school-board meeting, primarily to give him technical back up as he presented test results. After about one hour of presenting technically accurate tables and graphs of test scores, the superintendent was faced with the following question by a prominent board member: "That's all well and good; but I don't have the answer to my basic question -- can my son read?"

High Scores vs Low Scores. Unfortunately, the poor seem to achieve lower test scores. Certain cultures and ethnic groups seem to achieve lower test scores. Despite programs for the disadvantaged, tests and test scores have become a major target of those who claim discrimination. In some instances, the tests were found to have culturally biased items. In other situations, the test scores were used inappropriately to "label" students as low achievers or as having low ability. Even when tests and test use are improved, standardized tests continue to separate the high scorers from low scorers -- this is one of their main purposes. Criterion-referenced and performance-based tests, which are designed to show what has or has not been learned, suffer from the same problem -- some students do not score as high or master as much as others. How to address this honestly without insensitivity to the rights of everyone involved still remains a major issue. When educators are able to prescribe what they will do to remediate low achievement, and consistently follow through successfully to raise it, perhaps tests will be viewed more favorably.

Test Publishers. The history of testing since the 1960s has been both a dream and a nightmare to test publishers. Test publishers tend to operate like the insurance industry. Once a school district or state adopts a brand-name testing program, it tends to stick with it for many years. Each year the same tests are re-administered and scored, generating a growing database depicting the school's achievement. As this database locks in the school, the publisher is guaranteed an income which reoccurs and possibly expands year-to-year. When criterion-referenced tests came on scene, test publishers were frightened that their "locked-in" customers would run to another publisher or create their own tests. To turn this potential nightmare into a dream, publishers added criterion-referenced tests to their catalog, revised their standardized tests to appear more criterion referenced, added prescriptive references between their tests and textbooks, went into the business of custom tailoring tests, and improvised score reports and customization by using the growing computer technology. The current wave of performance-based tests, (etc.), represents yet another possible "nightmare-to-dream" sequence for test publishers.

Sampling vs Aggregation. The Title I/Chapter 1 evaluation models created another publisher's dream. These models, as implemented, secured the comparison of pre-post gain relative to a norm-group's gain, and solidified the practice of aggregating test scores across classrooms, schools, districts, and states. Together, these practices tended to reduce the use of highly detailed diagnostic tests in the classroom, and devalued cost-effective methods of sampling and survey testing. Teachers were forced to use tests

which were not locally relevant, unless they practiced MDI (emphasized the content/items in the test). Policy makers and certain experts became afraid MDI was leading to exaggerated test-score gains. Likewise, there were growing concerns about test anxiety and too much testing.

Assessment versus Evaluation. Many of the misuses of test data arise when conclusions are drawn from inappropriate evaluation methods. Even when an achievement test is highly reliable and valid, there are other threats to the validity of interpretation, such as the timing of the test administration, the relevance of comparison groups, the representativeness of sampling procedures, the existence of confounding variables, the intentions of those taking the test, the objectivity of those analyzing the results, the correctness of statistical procedures, and the like. Classic pre-post and control-group evaluation designs are often inadequately implemented or improperly selected in the first place. Likewise, descriptive studies are frequently used to make inappropriate predictions about the future; or conclusions are extrapolated beyond the scope of the existing data. There is strong evidence that our current concern about testing practices may be resolved by focusing on the larger issue of how to conduct quality evaluations.

CURTAILING MISUSE

As one acknowledges the possible reasons for the misuse of tests, some of which have been discussed above, it becomes clear that there are too many critical variables, too many purposes, too little training, too many fragmented historical events associated with testing, and too much emphasis on precision rather than relevance. As Kerlinger, one of our premier statisticians, said after he retired:

> "If I had it to do over, I'd spend far less time on how to turn 0.01 into 0.001 statistically. I would spend my time on educational significance."

The Target -- Knowledge versus Skills versus Capabilities?

One of the subtle difficulties associated with testing is that for most people it seems to "miss the target". If one adopts a multiple-choice approach to testing, in order to obtain highly objective, cost-effective test scores that can be aggregated and subjected to statistical analysis, then one is caught in a major dilemma. If the test length must remain short, one cannot probe very effectively for higher-order thinking and is trapped into testing primarily knowledge and recognition. Multiple-choice tests with less severe length requirements, coupled with the latest advancements in test construction, can creatively probe more deeply into broader-range concepts and applications. Unfortunately, the historical trend toward shortened tests and the "anti-testing" mentality among educators would stand in the way of expanding test length. It is interesting, however, that teachers seem more willing to invest increased testing time in performance-based tests. Is this a short-lived phenomenon; one which subsides when these tests are misused/abused? Perhaps the issue is that the various stakeholders have not agreed on what is to be learned. The following example illustrates the problem:

Suppose a school board decided to assess how many students have the capability to balance a checkbook. None of the existing standardized tests probe important capabilities. From existing tests, the school board could learn how many students have the enabling skills needed to balance a checkbook (reading, writing, adding, subtracting, keeping decimals lined up, etc.); and they could learn how many students have the appropriate knowledge (definitions of deposit, withdrawal, balance, and familiarity with the layouts of the checkbook and bank statement). The school board would need to allocate significant funding to use a combined program of norm- and criterion-referenced tests to assess the enabling knowledge and skills. In fact, they might be able to verify that 70% of students have mastered the enabling knowledge and skills, and that the student body is above the national average in reading and math. But if an interested researcher probed to see if the students could balance a checkbook, the school board might be stunned when only a handful could do it.

In this example, the problem is one of AIM (appreciation, interest, and motivation -- Noggle, 1989). The reason so many students do not have important capabilities is because they do not AIM to have them. Instead, they tend to AIM for high test scores. If they have any track record of achieving high scores, they become primarily interested in test scores and grades, rather than learning. If they cannot get high enough scores, they become disinterested altogether. Attachment B illustrates two barometers of learning. Our testing programs too often emphasize the bottom one-third of only one barometer -- trivial knowledge. Students may be AIMing at what seems to them to be a non-relevant target.

Eliminating Bad Tests

The test publishers have responded well to the challenges of history. They have attempted to meet the varying needs of consumers, and they have used advancing technology in so doing. As a result, each publisher has a variety of "different" tests in terms of their ability to meet some combination of competing consumer needs. The standardized tests of today may be less reliable because of shortness, far less comprehensive in content, and less representative of population norms. Are they bad tests? No, in the sense that the test publishers have done the best they could do while staying in business. Yes, in the sense that after all that investment in meeting everyone's needs, the consumer has little faith in the resulting product.

The states and local school districts have responded with tests of their own. Criterion-referenced tests first, and now performance-based tests (etc.). The quality of these tests are directly proportional to the resources allocated to acquiring testing expertise. Test publishers have been very supportive of these ventures, as many of them offer custom-tailored testing options.

The institutions of higher education, however, have varied widely in their impact on quality testing practices. The continued focus on increased precision, rather than validity, was probably one of the two most devastating aspects of their involvement. The second aspect, has been the poor training of teachers, both preservice and inservice. Teachers are too often unprepared to use norm- and criterion-referenced tests; and they are frequently

unprepared to develop good tests on their own. Those teachers who depend on unreliable and invalid tests to assign grades may be guilty of inadvertently but permanently damaging their students' self esteem.

Yes, there are bad tests! There are unreliable tests at all levels; and there are tests covering the wrong content at all levels. While this was never the intention of those who prepare tests, we have not adhered to our standards, plain and simple.

Preventing Covert Agendas

When test scores become the main or only criteria for "high-stakes" decisions, everyone's focus is on achieving test scores. Establishing the curriculum and finding the best instruction to deliver the curriculum, as the CIA Paradigm suggests, remain on the surface as the overt agenda of school boards, school administrators, curriculum/instruction committees, teachers, and students. However, the covert agenda for each group, is to raise test scores. Is it any wonder why "certain tests" are approved or developed, why "certain exercises" are encouraged or followed in the classroom, or why "certain acts" are observed of students or parents prior to or during testing. It is not that people are unethical; it is just that they "know the game that needs to be played." We need to decide how to help kids learn, not play games.

Reducing Blatant Discrimination

When test scores rather than learning become the product of schooling, test scores become the language of discrimination. High test scores have been said to represent specific traits: e.g., ability, achievement, academic potential, opportunity to succeed, contributing member of society, etc.. Low scores, correspondingly, suggest that students lack such traits. This too often occurs when testing acts as a status report, or a "snap-shot-in-time" representation of achievement.

> This drives interpretations of learning as a "state of being" rather than a "process of becoming".

Yes, there are those who misuse or abuse test scores out of hate, bias, fear, anger, or intolerance. The most blatant form of discrimination, unfortunately, is ignorance. If the ignorant were educated, the few who purposely discriminate would be controlled. If a common understanding of the purpose of testing existed, if proper tests were used, and if proper interpretations of scores were made, then discrimination would be drastically curtailed.

Converging Competing Purposes

Business leaders may be right; education has a product-definition problem. Can it be defined by test scores, graduation rates, employer satisfaction, and post-secondary education participation? Can it be defined by improving tests and test scores to become better "trait" measures? Can it be defined by judging the mastery of independent examples of knowledge or skill? Can it be defined by improved definitions and alignments of the curriculum and tests in terms of the most important capabilities needed for further schooling, employment, citizenry, and parenting?

If the purpose of achievement testing is to "somehow help students learn", then tests can be properly integrated into our educational methods as the "evidence of becoming" rather than the "statement of being". All other purposes bring discontinuity to the ongoing good of education -- the self growth, improvement, and adaptability needed in a fast-changing world.

Creating a Language of Success

The consumers of testing are confused. The jargon of the testing experts, while giving definition to the science of test construction and test-score interpretation, has become "a part of the problem rather than the solution". Parents, educators, and employers desperately want and need to know if children and youth have the pre-requisite capabilities for what will face them next. To the degree testing formats and test scores give them believable information about agreed-upon capabilities, those needs and desires will be met. The major source of misuse and abuse is ignorance -- a lack of understanding -- about the nature of important capabilities and how to measure them during and after their acquisition. There are too many testing experts, curriculum experts, instruction experts, and policy makers going in different directions. If we were able to agree on the major capabilities needed for: (1) employment, (2) post-secondary education, (3) citizen responsibility, and (4) personal and family satisfaction, the experts would have a common focus -- they could avoid the subtle seduction of the CIA Paradigm.

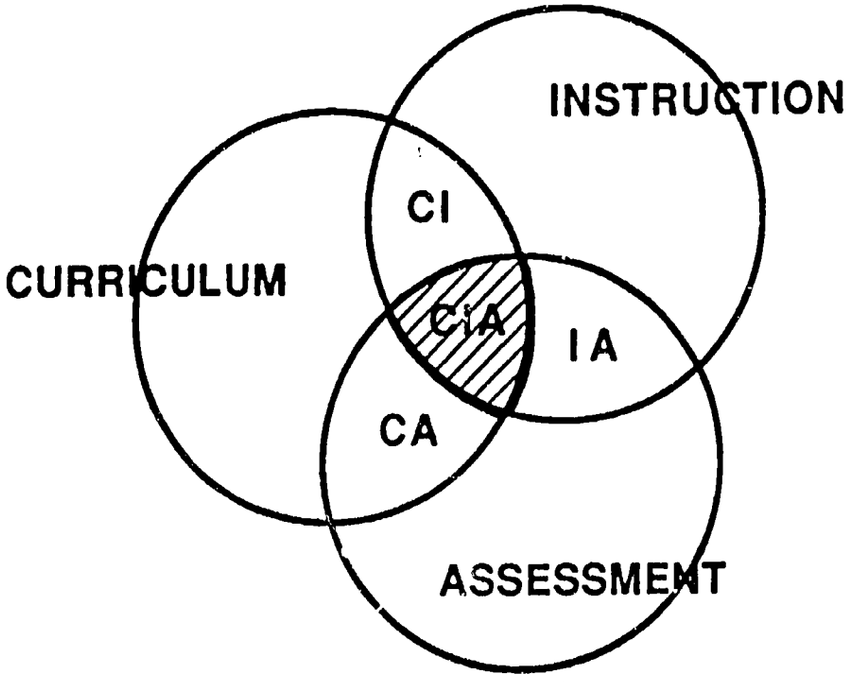
At the time of graduation from high school, I envision youth who are affirmed by knowing what they can do and what they still want to learn. There should be graduation assessments which help graduating students know their capabilities. Prior to graduation, I envision children and youth who are affirmed by knowing their growing list of capabilities and their progress toward others. There should be ongoing assessments which help students at all grade levels. I envision parents, educators, employers, and policy makers who are assured by knowing what children and youth can do, both in terms of progress toward and acquisition of important capabilities. There should be comprehensive evaluations which help parents, educators, employers, and policy makers judge progress and acquisition against specified capability statements, levels of investment, and changing needs. I envision a nation of people freel investing their tax dollars in the education of its children and youth. There success of our educational programs, as well as those areas needing improvement.

> If properly developed and implemented, assessment and evaluation would continually evolve a language which causes people to "celebrate and fix rather than identify, blame, and punish."

Nelson L. Noggle, Ph.D.
Centers for the Advancement of Educational Practices
3217 North Margate Place
Chandler, AZ 85224

(602) 345-0368

EDUCATION'S CIA PARADIGM



STUDENT ABILITY & MOTIVATION
TEACHER ABILITY & MOTIVATION
TIME AVAILABLE FOR INSTRUCTION
INSTRUCTIONAL MATL'S & SUPPORT

ATTACHMENT A

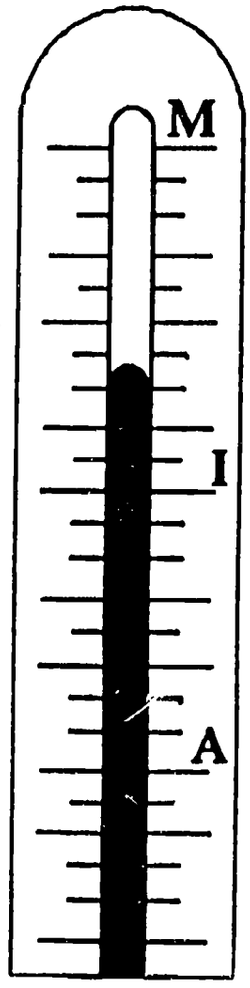
C-1.3

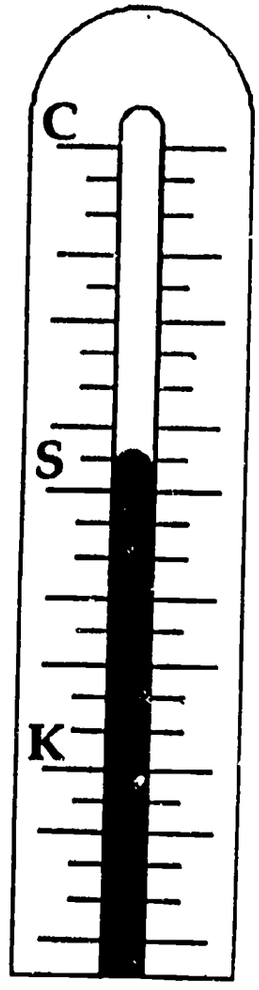
BAROMETERS OF LEARNING

MOTIVATION

INTEREST

APPRECIATION





CAPABILITY

SKILLS

KNOWLEDGE