

# EXCELLENCE IN EDUCATION VERSUS HIGH-STAKES STANDARDIZED TESTING

**Asa G. Hilliard, III**

Georgia State University

*The evidence is missing for high-stakes standardized testing of either intelligence or achievement as a reform tool, a tool for producing higher student achievement. The focus on invalid testing allows us to ignore the well-documented importance of the quality of teaching as a key factor in student achievement. There is no doubt that teachers can become powerful agents in raising the academic performance of children to high levels, regardless of common impediments such as poverty, bilingualism, and so forth. With the extreme and almost exclusive focus on high-stakes testing, we lose the opportunity to support valid staff development that would make all of our teachers powerful.*

For years I have been interested in a special group of educators. They are the ones who create powerful education environments. They are the ones who are not puzzled about how to raise the achievement levels of children from any background to levels of excellence. They are the ones who see the universal genius, spirit, and humanity in all children. Things like poverty, bilingual status, single-parent families, and even threatening neighborhood environments present no obstacle to the attainment of excellence for their students.

These are the educators who are lost in the aggregate, which includes those unsuccessful educators who are puzzled about how to teach for excellence. Most successful educators are “invisible,” unknown, and not featured. There are thousands of them in hundreds of schools. When we focus our attention on clusters of these educators, there is much to learn. In fact, I have often referred to the distinction between such educators and the more typical educators as the distinction between two conversations. Not only are these successful educators distinct in their behavior from the more typical ones, when they do take the time to reflect on what they do and speak about their work, their conversation is distinct as well. They talk about different

things in different ways than their less successful counterparts.

The proverbial “fly on the wall” in the faculty rooms in schools where children are not succeeding would very likely hear a conversation in which children, their families, and their communities are seen as the source of the problems of low performance. The children’s attendance, misbehavior, “deficient” language, aggressiveness, and so forth are highlighted as common topics. Similarly, parents are seen as not caring, as unable to be of help to children, and as unwilling to become involved in the school. Even whole communities and cultures are seen as dysfunctional. All of this is reflected in the conversations that many professionals have about children and their performances in the schools.

## ***What Can We Learn From Educators Whose Students Achieve Excellence?***

In my experience, a very different kind of conversation takes place among educators who succeed in spite of all impediments. In my experience, such educators not only talk about different things, but they have a very different worldview as well. They talk about their trials

---

and errors. They talk about lofty goals, goals of excellence, and a sense of mission and destiny. They talk about the school as a family. They talk about the community and their part in it. They view each child as a special gift, a small bundle of enormous potential. Their high hopes for the children are matched only by the sacrifices that they do not need to be urged to make because they are driven to ensure that no child is left behind. They are problem solvers.

There are indeed educators who are successful, regardless of circumstances, in raising the achievement of virtually all children to levels of excellence, not merely minimum competency. I have spoken often about programs such as Project SEED (Special Educational Enrichment for the Disadvantaged) (Mezzacappa, 1990) and Beginning School Mathematics (Miller & McKinnon, 1995). In both cases, children as young as 3 years of age and all the way into higher education are taught abstract, conceptually oriented mathematics in Socratic-type interactive classrooms, working not with textbooks but with manipulatives or simply with chalk and blackboard. Professor Everard Barrett (1994, 1998) is a powerful teacher and an excellent staff developer who has similar accomplishments to his credit. For example, he taught regular fifth-grade students in Bedford Stuyvesant public schools, most of whom passed the ninth-grade achievement examination for New York state. These were poor and minority children who are normally in the lowest quartile. Barrett's techniques are nationally recognized due to students achieving excellence based on high assessment ratings given by school districts located in major American cities. In each case, these educators bring many years of success in a variety of socioeconomic settings. The stark contrast between the results achieved by such teachers and what is typically expected from children in low socioeconomic circumstances is enormous.

Children are not screened specifically for these educators. It is only the luck of the draw that determines whether children will be served

by the powerful teachers or by the alternative. This simple fact has enormous consequences for how we should think and behave in the design of educational opportunity. Hundreds of individual schools such as the New Suncock School in Lovell, ME; the Sankofa African Centered Charter School in East Lansing, MI; Lincoln High School in Dallas, TX (under recently deceased principal Napoleon Lewis); and Garfield High School in the Los Angeles barrios during the time of Jaime Escalante demonstrate quality teaching. Individual teachers, whole schools, and, in some cases, whole districts defy the odds. In preschool, elementary, secondary, or higher education, good teaching matters. When good teaching occurs, all children achieve at levels that we can currently describe to be excellent.

Finally, there is an impressive and growing body of literature supporting what I have said. *Cumulative and Residual Effects of Teachers on Future Students' Academic Achievement* (Saunders & Rivers, 1998), *Achieving Despite Diversity, The Study of High Achieving, High Poverty Schools in West Virginia* (Hughes, 1995), the national study of excellent schools by Schmoker (1996), reported on in his book, *Results*, and the Educational Trust documentation of gap-closing schools (Haycock, 1998) feature educators who provide an empirical foundation to show that they know what to do and do what they know. What is the source of this excellence? Is it the use of high-stakes standardized testing? Hardly! Few of these educators, if any, use standardized high-stakes testing as a tool to get high achievement.

Much more could be said about these two conversations, between those who produce and those who do not. However, one thing we might ask about in probing both of these conversations is how they are similar or different in the way that they view high-stakes standardized testing. In my experience, they differ markedly in their level of preoccupation with such tests. They differ markedly in their interpretation of the meaning of the tests.

---

### **The Legitimate Desire for Accountability and for the Highest Standards**

Children, parents, communities, and the public at large, including policy makers, have every right to raise the difficult question of whether schools achieve appropriate goals. More than that, they have the profound responsibility to ask. There can be no argument about the necessity for high standards. Therefore, there should be no confusion when high-stakes testing is criticized. Confusion makes some interpret the criticism of high-stakes standardized testing as opposition to high standards. Nothing could be further from the truth. It is very important that we understand what the real issues are in this debate and dialogue. High standards may not be connected to high-stakes standardized testing at all. In fact, an argument can be made that the content of some standardized high-stakes testing reflects low standards, (Zacharias, 1977; Schwartz, 1977a). The high standards mean little unless they are met. The real issue is to give children the teachers who can make them reach high standards. I call this power teaching.

Many years ago, Dr. Barbara Sizemore and I were cochairs of the 10-member Task Force on Black Academic and Cultural Excellence. (Hilliard & Sizemore, 1984). We coedited a publication for the National Alliance of Black School Educators titled *Saving the African-American Child*. In that publication, we called for the following high standards for all high school students. We did so as a task force because we were quite familiar with educators who know how to get virtually all of their students to reach those standards.

The Task Force notes that performance goals for public schools are usually stated in terms of normative scores (where students, classes, or schools rank as compared to other students). This is grossly inadequate for our purposes. We believe that a high level of academic achievement is within reach of virtually all African American children. Specifically, we believe that the following criterion levels of performance in basic subjects must be established as standards and must be reached by our children by the end of the twelfth grade.

1. Mathematics: Criterion performance, algebra in the sixth grade and calculus by the twelfth grade.
2. Economics: Criterion performance, understanding of the workings of the American economic and other economic systems.
3. Political Science: Criterion performance, an understanding of and the ability to discuss the workings of the American political system.
4. Computer Competence: Criterion performance, ability to write computer programs in one or more languages.
5. History: Criterion performance, an understanding of and the ability to discuss African American perspectives on standard historical topics commonly taught in schools.
6. Language Arts: Criterion performance, to be able to write a term research paper demonstrating the ability to use common English, appropriate documentation of ideas, and appropriate presentation of ideas.
7. Foreign Language: Criterion performance, a speaking, reading, and writing knowledge of at least one foreign language. The acquisition of competence in an African language should be available as an option.
8. Sciences: Criterion performance, a passing grade in a course equivalent to general chemistry. This assumes that the common practice is to require appropriate coursework in biological and physical sciences as a prerequisite.
9. Vocational: Criterion performance, typing, child care, work habits, employability.
10. African American History and Culture: Criterion performance, ability to tell the general story of African and African American people from earliest times to the present.

In addition to the criterion performances in academic areas cited above, certain general academic goals which transcend the disciplines mentioned above are attainable. Among these are the following:

- The demonstration of critical thinking
- The demonstration of creativity
- The demonstration of the acquisition of a systematic approach to problem solving and the demonstration of an understanding of the scientific method

Our thinking on the criterion levels of performance is in essential harmony with two recent education reports. The College Board Report, *Academic Preparation for College: What Students Need To Know and Be Able To Do* and *Paidiea Proposal* edited by Dr. Mortimer Adler.

These are not unreasonable goals for African American students, if fair, equitable, appropri-

---

ate, and high quality instructional support is offered to them. (Hilliard & Sizemore, 1984, pp. 35-36)

### ***The Language of High Standards and the Construction of Failure in Intelligence or Aptitude***

Publication of *Saving the African American Child: A Report of the National Alliance of Black School Educators, Inc.* (Hilliard & Sizemore, 1984) preceded the current emphasis on high standards. In fact, at the time that the National Alliance of Black School Educators called for these standards, the cry was for "minimum competencies" at the national level, as national policy! In our report, we chose to emphasize course completion and rigorous academic curricula as the centerpiece of evidence for meeting high standards. But even beyond traditional academic high standards, the think tank that produced that report spoke to the need for developing students who would become men and women of high character with the capacity to create joyful environments that would be fulfilling to their spiritual needs. Indeed, I know of no higher or more comprehensive standards called for then or now. Thus, meeting excellent standards of performance is not the issue for most of the educators who I know. If not, then what is the issue?

The discussion of high-stakes standardized testing must include reference to the politics of standardized testing. There is a well-documented history here, a history of abuse of minorities and the poor. We may like to think that such things, if they exist at all, are in the past. However, they are as present today, even though they may be presented in a more disguised form.

Few people have read *The Bell Curve: Intelligence and Class Structure in American Life* (Herrnstein & Murray, 1994). Fewer still know its history and context. It is a highly sophisticated rendering on popular old ideas in new covers. Skip past the pseudo science reflected in the many charts and graphs and go straight to the end of the interpretive and policy recommendation chapters (especially Chapters 17-22 and, more specifically, Chapters 21 and 22). The

real agenda for the book and for those who provide the support for the authors is contained in these seemingly innocent chapters. Here we have the real dangers of the misuse of one type of high-stakes standardized test exposed. The authors see IQ as *intelligence*, and they see intelligence as the causal factor for poverty and for affluence, or low- and high-class status. This means that the low class does not have the mental capacity to be improved, and therefore, charitably, they should be neglected.

Interestingly, Herrnstein and Murray (1994) see the underclass as a virtual aggressor and at least a threat. The following demonstrates their view.

#### **Facing Reality About the Underclass**

What new ways? There are many possibilities, but the central ones all involve the underclass. We fear that a new kind of conservatism is becoming the dominant ideology of the affluent—not in the social tradition of an Edmund Burke or in the economic tradition of an Adam Smith but "conservatism" along Latin American lines, where to be conservative has often meant doing whatever is necessary to preserve the mansions on the hills from the menace of the slums below. In the case of the United States, the threat comes from an underclass that has been with American society for some years but has been the subject of unrealistic analysis and ineffectual, often counterproductive policy. The new coalition is already afraid of the underclass. In the next few decades it is going to have a lot more to be afraid of. Now is the time to bring together from many chapters throughout the book the implications of cognitive stratification for the underclass. (pp. 518-519)

Herrnstein and Murray pronounce the final sentence on children with low intelligence. In doing so, they are unaware of or fail to take into account the thousands of powerful teachers who raise the performances of low-intelligence children to high levels. These teachers destroy the IQ correlation, because IQ is tied to poverty and to achievements, but not with good teachers in the mix (Fuller, 1977).

There will still be jobs for low skill labor, mostly with service businesses and private households, but the natural wage for those jobs will be low. Attempts to increase their wage artificially (by raising the minimum wage, for example, or mandating job benefits) may backfire by making alternatives to human labor more affordable and, in many cases, by making the jobs disappear altogether. People in the bottom quartile of intelligence are becoming not just increasingly expendable in economic terms, they will some-

---

time in the not-too-distant future become a net drag. In economic terms and barring a profound change in direction for our society, many people will be unable to perform that function so basic to human dignity: putting more into the world than they take out. (Herrnstein and Murray, 1994, pp. 519, 520)

What is the destiny of the poor? Herrnstein and Murray (1994) suggest a part of the answer:

In short, by custodial state, we have in mind a high-tech and more lavish version of the Indian reservation for some substantial minority of the nation's population, while the rest of America tries to go about its business. In its less benign forms, the solutions will become more and more totalitarian. Benign or otherwise, "going about its business" in the old sense will not be possible. It is difficult to imagine the United States preserving its heritage of individualism, equal rights before the law, free people running their own lives, once it is accepted that a significant part of the population must be made permanent wards of the state. (p. 526)

This custodial state is a life sentence with no possibility of parole for the poor because they are not merely regarded as poor but as beyond help. This is contrary to the record cited earlier on power teaching and power schools. Then Herrnstein and Murray raise the ante. Things will get even worse.

Herrnstein and Murray (1994) settle for finding a place for the poor in the new caste system:

#### The Goal and a Definition

The broadest goal is a society in which people throughout the functional range of intelligence can find and feel they have found a valued place for themselves. For "valued place," we offer a pragmatic definition: You occupy a valued place if other people would miss you if you were gone. (p. 535)

Thus, the authors pronounce with finality that low-achieving children are defective materials. They must find a "place" and be put in it, and they should stay in their place: "Government policy can do much to foster the vitality of neighborhoods by trying to do less for them" (p. 540).

Ultimately, the authors recommend abandonment of the poor, notwithstanding their nice way of saying it.

The material from *The Bell Curve* above must be placed in context. Herrnstein acknowledges that he and Murray were at the American Enterprise Institute and at Harvard. In addition, support for the book came from the Bradley Foun-

dation in Chicago. Moreover, Murray appears to be a fixture at the Heritage Foundation, as well as having been a policy adviser to conservative presidents. Moreover, the policy recommendations in *The Bell Curve* are closely aligned with the policy proposals that come out of the conservative think tanks that have housed Herrnstein and Murray. Blaming the poor for poverty, asserting that nothing can be done because of their innate lack of capacity, and calling for neglect, individualism, and a withdrawal of fiscal support, along with many other such policies, are really the widely supported and increasingly implemented conservative agenda. Perhaps this explains the popularity of *The Bell Curve*, a *New York Times* best-seller, but not among professional psychologists. It takes the general public to make a best-seller. Furthermore, *The Bell Curve* was published without peer review. My friend, Herb Kohl, gave me an 800 number of a foundation that invited calls from higher education institutions, offering \$20,000 to any college or university that would host Charles Murray in a debate on the book. This sounds like propaganda, not scholarship. "Bell curve thinking" is more typical than many will admit (Snyderman & Rothman, 1990).

The point here is that we cannot take the high-stakes standardized testing movement as an innocent school improvement activity. There are those who like children—even other people's children and poor and minority children—and who are also attracted to the mystique of standardized testing, some believing that beneficial use can be made of it. In some cases, beneficial use can be made of achievement tests. However, following the conservative agenda and practices, making final admissions and placement decisions solely on the basis of these tests (high-stakes decisions) is unscientific, unprofessional, and bad public policy.

Bell curve thinking is the enemy of the millions of children who have unrecognized genius. As documented above, savage inequalities in school services are eliminated by giving students equal access to good teachers. The power of good teaching now has overwhelming support in the literature. The high-stakes standardized testing movement advocates cannot

---

ignore the connection of such uses to the pessimists who are, in fact, enemies of the poor.

We cannot escape the fact that low expectations for a large segment of our population are endemic in America. Not only are low expectations endemic, but there is also an alienation, even a dislike for poor children and for certain minority ethnic groups in the United States. A visceral reaction, often very negative, lies just below the surface of polite discourse about low-achieving students. The best recent example was when the Oakland Public School "Ebonics" caused the hysterical, irrational and uninformed, overwhelmingly negative and dehumanizing reactions of so many educators, politicians, bureaucrats, civil rights leaders, and various media sources opposing the children, teachers, administrators, families, and the African culture in general (Perry & Delpit, 1995). No one can read Perry and Delpit and feel anything but shame and apprehension about the negative feelings expressed in the crisis toward the children and this culture. The question is, how many more issues pertaining to low performance would bring these latent negative feelings back to the surface (Adger, Christian, & Taylor, 1999)?

### **WHAT IS HIGH-STAKES STANDARDIZED TESTING?**

There is a fundamental disconnect between high-stakes standardized testing and the movement toward excellence in education. First, there is the issue of test validity. Simply put, it is rare to find high-stakes standardized tests that have any meaningful validity. Second, there is the question of appropriate standards. In few districts do the lofty goals adopted by boards of education match the standards reflected in high-stakes standardized testing. In some cases, neither the school goals nor the high-stakes standardized tests reflect high standards.

There is very little debate but that standardized tests, if they measure anything of value at all, measure a very narrow range of behaviors, many of them at a very low level of thinking. Much more could be said about the validity and appropriateness of high-stakes standardized

testing. We may summarize the criticisms in the following way: High-stakes standardized testing tends to fall into two categories. On one hand, there are tests that purport to tell us what a student's capacity is, sometimes called "aptitude tests," "IQ tests," and/or "intelligence tests." In the second category, there are achievement tests that purport to tell us whether a student has mastered the curriculum.

The capacity, IQ, and/or intelligence tests are normally evaluated primarily by whether the scores on those tests are associated or correlated with later achievement test scores by students. That association or correlation is called "predictive validity." The goal of predictive validity is for educators to make good guesses about education outcomes. Sometimes, those who argue for such tests claim that they can be used as "diagnostic" devices, devices that explain what performance is, how it came about, or the presence or source of malfunctions. Implicit in the practice of diagnosis is the idea that there is an approach to teaching, based on this diagnostic information, that would improve academic achievement outcomes for children. Of course, no documentation of valid test/teaching connections appears in the technical manuals that accompany such tests. The technical manuals merely record the research that documents association or correlation between test scores and achievement.

### ***False Prophets: Predictive Validity Without Regard for Access to High-Quality Teaching***

The easiest way to challenge such claims of predictive validity is to look at the examples of powerful educators and powerful schools such as those mentioned above. Over and over again, children whose initial scores would cause test users to predict the children's failure not only succeed, but they succeed at the highest levels of excellence when offered appropriate instruction. The high correlation or association between tests of aptitude, intelligence, capacity, and so forth can be accounted for by the simple fact that both the capacity tests and the achievement tests, if measures of anything, are

mainly measures of opportunity to learn testlike material. They are pure and simple measures of exposure. Moreover, the tests of intelligence, ability, and aptitude cannot be regarded as diagnostic, nor is there a body of literature to show connections between such tests, instructional "prescriptions," and beneficial outcomes for children. The ritual or routine of testing is quite widespread and prominent. The benefits of such testing are nonexistent (Heller, Holtzman, & Messick, 1982; Hilliard, 1994; Skyrtic, 1991).

Diagnostic assessments that are linked to powerful instruction that results in powerful student learning outcomes do exist. However, such approaches should be characterized as assessments rather than high-stakes standardized tests of capacity (Feuerstein, 1979; Feuerstein & Rand, 1997). The only real purpose of standardized capacity tests is ranking/classification for segregating students into nonbeneficial instructions. The outcome of such practices places poorly served children, the victims of "savage inequalities" (Kozol, 1991), at a severe disadvantage for opportunities to learn.

Standardized intelligence tests are the primary basis for most special education placements, a truly high-stakes decision. A cursory look at the makeup of the special education population in the United States should raise questions immediately. As Table 1 shows, more than one half of the special education placements are of students regarded as "learning disabled," a very soft category, meaning that there is great ambiguity and uncertainty about this label. Look at the hard diagnostic categories such as auditory and visual impairments. Clearly, the more ambiguous the assessment, primarily IQ based, the higher the frequencies. (Heller et al., 1982; Skyrtic, 1991).

When we add the lack of evidence of efficacy of treatment in special classes to the problem of questionable validity of assessment (Skyrtic, 1991), it is easy to see that the consequences of being given a label, classified into a category, and given special services are a high-stakes assessment problem of tremendous proportions. Stigma and isolation are but two aspects of the problem.

TABLE 1: Types of Disabilities Served

Disability	IDEA, Part B	
	Number	Percentage
Specific learning disabilities	2,513,977	51.1
Speech or language impairments	1,023,655	20.8
Mental retardation	580,855	11.6
Serious emotional disturbance <sup>a</sup>	428,168	8.7
Multiple disabilities	89,646	1.8
Hearing impairments	65,568	1.3
Orthopedic impairments	60,604	1.2
Other health impairments	106,509	2.1
Visual impairments	24,877	0.5
Deafness-blindness	1,331	0.0002
Autism	22,780	0.4
Traumatic brain injury	7,188	0.0014
All disabilities	4,915,168	100

NOTE: From U.S. Department of Education (1996).

a. IDEA now also calls this category "emotional disturbance," dropping "serious" in some but not all of its provisions.

Some high-stakes standardized testing, such as admissions testing to colleges and universities, rely on the presumed predictive validity of tests. Yet, over and over, we see evidence that challenges this validity (Smith, 1984; Slack & Porter, 1980; Cole, 1991; Messick, 1980). Scores can be raised by coaching. If so, which score is the real score? How do we deal with differential access to good coaching?

Even when we rely on standardized tests for prediction, many who do so are not aware of an important difficulty. Whatever the strength of the relationship between a distribution of scores on an aptitude test and the predictive validity coefficient (a later measure of achievement), the strongest part of the relationship is at the middle of the distribution. It is at the highest or lowest score levels where the greatest errors in prediction occur.

## ACHIEVEMENT TESTING

In theory, standardized achievement tests may offer some benefit. What is being claimed by achievement test makers and users is that standardized achievement testing is a fair measure of the curriculum that is offered by the school. All achievement test makers claim implicitly or explicitly that tests have content validity, that is, that they are a fair match to the

---

**school curriculum and objectives.** The popular term for this criterion is *test/curriculum alignment*. No school district would or should knowingly buy and use a standardized achievement test that obviously did not match its curriculum. We need not take much time to demonstrate the falsity of that claim of alignment. Some years ago at the Institute for Research on Teaching, a report contained the summary of empirical observation. This observation was about the comparison of textbook content (i.e., academic topics covered in textbooks) and standardized achievement tests content (i.e., academic topics covered by achievement tests). Do they match? The content area was fifth-grade mathematics. Are they aligned?

In fifth-grade mathematics, where one might expect to find fairly high agreement regarding content, the reality is that nationally standardized tests show a very poor record of alignment with nationally standardized textbooks. In fact, it appears that the average overlap between standardized textbook content and standardized test content hovers somewhere around 50%. The situation is even worse in other areas of the curriculum, where there is less agreement on content. This means that children who sit for such standardized high-stakes exams, if they are nationally standardized tests, often must attempt items on topics half of which they have not been taught! We can only imagine the gross nature of the mismatch between texts and tests in curricula areas where there is low consensus about content.

However, this is not the worst of it. The institute report deals with "ideal" circumstances. What is the real curriculum that is offered to children? First, what are the actual topics that are taught to children in the classroom? These topics may be very different from what appears in the textbook. Therefore, there is even less overlap between what is actually taught and the content of standardized high-stakes testing. There is one further issue: How do we account for the variety of competencies among teachers? What is the quality of instruction to which students are exposed, even when the goals and content are common? How confident can we be in the equitable exposure of children to high-

quality teaching of all topics in the textbook or curriculum outlines and in a nationally standardized test that is aligned with the student's textbook or curriculum outline? This is not a trivial matter. The source of children's low or high performance may well be due to the poor quality of instruction that they receive:

A June analysis of middle school students' poor performance on international math and science tests found that the low scores can't be explained by the wider diversity of U.S. students taking the test, a frequent scapegoat. The real culprit? A weak curriculum.

In January, the American Association for the Advancement of Science rated as "satisfactory" only four of 12 common middle school math textbooks, and those weren't the four texts appearing in most classrooms.

Why else would only about a fifth of U.S. eighth-graders take algebra, while in countries ranking highest on international math tests, that figure rises to nearly 100%?

And why do so many middle school math teachers lack college training in the subject they teach? An estimated 36% of all math teachers lack a major or minor degree in the subject; most of those fall in the middle school grades of seven and eight. ("Education's Weakest Link," 1999, p. 3)

Now let us look at one district where there is documentation about how teachers are distributed. If this is accurate, then high-stakes standardized testing is not the appropriate response.

Not only does New York City have a disproportionate number of uncertified teachers, but those who do pass the state certification exams tend to score much lower than their counterparts in the suburbs and upstate, according to state data that will be placed into evidence today in a landmark school-financing case.

The data shows that 31 percent of teachers in the city's public schools failed the main certification exam—the liberal arts and sciences tests—at least once, compared with only 4.7 percent of teachers elsewhere in the state. About 47 percent of city teachers who took the state certification test for math failed at least once, compared with about 21 percent of teachers elsewhere. And 26.9 percent of city teachers who took the state test that measures elementary teaching skills failed, compared with 3 percent elsewhere.

About 12 percent, or 10,000, of the city's teachers are uncertified. Because senior teachers can often choose which schools they teach in, *many of the uncertified teachers work in the schools with the poorest children and the lowest student performance* [italics added].

---

("City Teachers' Low Scores Cited in School Aid Suit," 1999, p. 1)

Even the minimal amount of information from the two examples from the Institute for Research on Teaching report shows the complicated problems for assessments in a nation that values local control of the education process and that abides savage inequalities in school services. The bottom line is that there is no national curriculum. Consequently, there can be no nationally standardized test that is valid for measuring this curriculum that does not exist. The professionals in testing refer to this issue as a matter of content validity. The bad news is that the situation is even worse than I have described so far, for the simple reason that we must assess whether either textbook or standardized achievement tests or both together actually constitute an excellence-oriented curriculum. We are very likely to find minimum competencies emphasized, not high standards.

Given the successful experiences of children with good educators, such as those who I have shown above, we would have to say that the entire process of high-stakes standardized testing of a high-stakes nonstandardized curriculum gets failing marks (Barrett, 1994, 1998; Haycock, 1998; Saunders & Rivers, 1998). In other writing, I have argued that high-stakes standardized achievement tests have been used more as decoys than as measures of standards of quality (Hilliard, 1999). Policy makers are often timid in approaching the problem of what to do about supporting schools and demanding excellence from them. Policy makers have sought a quick fix, a simple fix, a cheap fix, and an uncomplicated fix. This is the easiest thing for a policy maker to do, to appear to make a valid effort to solve the problem of school improvement, sometimes called "school reform." When the money and the resolve are not there, the easy thing to do is to manipulate the requirements for high-stakes standardized achievement tests and call it reform.

Generally speaking, there is very little drain on the budget when using test requirements as the hammer to attempt to create high achievers, as evidence of commitment to school improvement. On the other hand, real school

improvement—improvements that guarantee high achievement on comprehensive high-level goals—would require appropriate resources. For example, it is a fact that many teachers who teach mathematics and science course work are teaching outside their field of preparation. It is also known that the distribution of unprepared teachers often follows the patterns associated with minority status and poverty, with the best-prepared teachers going to the most affluent and preferred groups. I know of no accountability systems for schools and students that face this horrible fact or mobilize to address it.

The weight is being taken by the children for the failures of adults, either on the policy level or in the classroom. It is because of these conditions reflecting the maldistribution of good services that when we settle on a rigid use of high-stakes standardized academic achievement tests, we support invalid and unfair practices, practices that penalize neglected children.

The commitment to such invalid high-stakes standardized testing practices changes the very nature of the dialogue on school improvement. For example, during the implementation phase of Goals 2000, the process was nearly halfway completed before it was acknowledged, and possibly before it was realized, that some vital aspects of a whole curriculum had been left out of the goals altogether. Art and music, for example, had not been included. It was only with a supreme lobbying effort that they were added to the goals in the later stages of the process. What else is left out of the curriculum? In what kind of dialogue would we engage to see the big picture, make judicious choices about the content of school curriculum, and appropriate assessment of the results of our efforts?

It is a real pity. Over the course of my more than 40 years of experience in education, it appears that we are in the same place where we have been all along. It is a shame to be in this place, given our greater knowledge of the technical inadequacies and the political abuses of high-stakes standardized testing and greater knowledge of the inequities that are created by policies that are touted as being aimed at producing benefits for children.

---

One of psychology's most defining assaults on others was our early misuse of one of our dearest pursuits, namely, the definition and measurement of intelligence. Along with the biologists, anthropologists of their time, American psychologists assumed the intellectual superiority of the white race. (Strickland, 2000, p. 332)

The critiques of standardized testing are numerous and devastating. The documentation of savage inequalities in the provision of student opportunities in school is equally damaging. The challenge of a different type of curriculum and different types of standards has been made numerous times by visionary educators who paint the big picture, a big picture that reflects a deep and abiding faith in the humanness and in the human capacities possessed by virtually all of our students. We seem to have descended to the lowest common denominator even as we talk about high standards.

### ***Where Do We Go From Here?***

I have often said that the matter of school improvement is hardly rocket science. To eavesdrop on the conversation that goes on among powerful educators who do succeed is to hear that hard work, rather than new or esoteric methods, matters. Furthermore, among those educators, there is little to no support for high-stakes standardized academic achievement tests. The great educators such as Abdulalim Shabazz and Jaime Escalante produced students whose achievements were excellent, far surpassing the mediocre goals reflected in standardized tests (Hilliard, 1991). The achievements were not just academic but social and cultural as well. The problem is not that the high-stakes standardized tests are too difficult but that they do not represent the schools of the real world. Rather, they represent enormous profits for the corporations that make them. It may not even be their fault that their tests fail to match the real-world curriculum. More at fault may be those policy makers among us who are unable to produce a consensus on what constitutes the outcomes of edu-

tion. Such policy makers give the assessment community an impossible task.

The high-stakes standardized movement actually interferes with and impedes the opportunity for educators and communities to frame a human future for our children. The current dialogue is bogged down in trivia. Our children are special gifts with awesome potential. They may never get the opportunity to be challenged by the types of educators that they deserve, partly because our pathological preoccupation with high-stakes standardized testing deflects attention from the need to build greater teaching capacity among our educators.

In closing, I must say that appropriate high-quality structures can release the genius of teachers, genius that is currently bound up by the imposition of hastily conceived, oversimplified schemes for school improvement. Many teachers whom I see have become depressed and terrorized by the mindless demands for inappropriate standardization not only in testing but in teaching as well. Their humanity has been challenged in much the same way that the system challenges the humanity of our children. Many are being turned into robots that lead drill-and-kill sessions to prepare for low-level high-stakes achievement tests. For many, the joy is gone, the hope is gone, and their pride is crushed. The climate in many schools that I see is more like a factory than anything else.

The way out of the mess that we are in is to operate in much the same way that sports figures do in the Olympics. To run the race, to be at the center of things, one must run and win qualifying heats. To be in the decision-making position in the design of the educational futures for children, the price for participation must be a track record of success. Master teachers and master educational leaders can design teaching and learning processes, school structures, and, yes, assessment approaches that exhibit both standards of excellence and valid relationships to excellent educational achievement. They can also rescue most of the low-performing teachers and help them to become powerful agents of student change in achievement.

---

In the final analysis, the real questions are the following: What do we really want, and what do we have the will to achieve? Is the value of every child such that we will leave no stone unturned in guaranteeing for them a future that brings out their full potential? Will we keep "weighing the elephant" to make it grow; will we feed and nurture it? We need high-stakes services for our children. Then, any testing will be a matter of small concern.

## Bibliography

The following is a listing of supplemental publications regarding the preceding information.

- Gould, S. (1981). *The mismeasure of man*. New York: Norton.
- Guthrie, R. (1976). *Even the rat was white*. New York: Harper & Row.
- Hehir, T. & Latus, T. Eds. (1992). *Special education at the century's end: Evolution of theory and practice*. Cambridge: Harvard Education Review, Reprint Series No. 23. (See especially the seminal work by Thomas Skrtic, "The special education paradox: equity as the way to excellence.")
- Hilliard, A. G., III (1982). *The sociopolitical implications of minimum competency testing*. In John H. Neel and Shirley W. Goldwasser (Eds.), *Minimum competency education: Issues, methodology and policy for local school systems* (pp. 86-106). Atlanta: Georgia State University.
- Hilliard, A. G., III (1996) Either a paradigm shift or no mental measurement: The non-science and nonsense of the Bell curve. *Cultural Diversity and Mental Health Journal*, 2(1), 1-20.
- Schwartz, J. L. (1977b). Math tests. In P. L. Houts (ed.), *The myth of measurability*. New York: Hart.
- Schwartz, B. N. & Disch, R. (1970). *White racism: Its history, pathology, and practice*. New York: Dell.
- REFERENCES**
- Adger, C. T., Christian, D., & Taylor, O. (1999). *Making the connection: Language and academic achievement among African American students*. Washington, DC: Center for Applied Linguistics and ERIC Clearinghouse on Languages and Linguistics.
- Barrett, E. (1994). *Mathematics power learning for children*. Duluth, GA: Professor B Enterprises.
- Barrett, E. (1998). *Professor B Math: A video supplement to mathematic power learning. Book 1*. Duluth, GA: Professor B Enterprises.
- Cole, B. (1991). College admission and coaching. In A. G. Hilliard, III (Ed.), *Testing African-American students* (pp. 97-110). Chicago: Third World Press.
- Delpit, L. (1995). *Other people's children: Culture conflict in the classroom*. New York: New Press..
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore, MD: University Park.
- Feuerstein, R., & Rand, Y. (1997). *Don't accept me as I am: Helping retarded performers excel*. Arlington Heights, IL: Skylight.
- Fuller, R. (1977). *In search of the I.Q. correlation*. Stonybrook, NY: Ball-Stick-Bird.
- Haycock, K. (1998). Good teaching matters . . . a lot. *The Education Trust*, 3(2), 3-14.
- Heller, K., Holtzman, W., & Messick, S. (Eds.) (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hilliard, A. G., III. (1991). Do we have the will to educate all children? *Educational Leadership*, 49(1), 31-36.
- Hilliard, A. G., III. (1994). What good is this thing called intelligence and why bother to measure it? *Journal of Black Psychology*, 20(4), 430-444.
- Hilliard, A. G., III. (1999). The standards movement: Quality control or decoy? *Rethinking Schools*, 12(4), 4-6.
- Hilliard, A. G., III, & Sizemore, B. A. (1984). *Saving the African American child: A report of the National Alliance of Black School Educators, Inc.* Washington, DC: National.
- Hughes, M. F. (1995). *Achieving despite diversity: Why are some schools successful in spite of the obstacles they face?* Charleston: West Virginia Education Fund.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York: Crown.
- Messick, S. (1980). *The effectiveness of coaching for the S.A.T.* Princeton, NJ: Educational Testing Service.
- Mezzacappa, D. (1990, February 11) Calculated excitement in the pursuit of algebra. *The Philadelphia Inquirer*.
- Perry, T., & Delpit, L. (1997) The real ebonics debate: Language, power and the education of African-American children. *Rethinking Schools and Boston Beacon Press*, 12(1), 3-7.
- Sanders, W. I., & Rivers, J. C. (1998). *Cumulative and residual effects of teachers on future students' academic achievement*. (Unpublished report).
- Schmoker, M. (1996). *Results: The key to continuous school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Schwartz, J. L. (1977a). The illogic of IQ tests. In P. L. Houts (Ed.), *The myth of measurability*. New York: Hart.
- Schwartz, J. L. (1977b). Math tests. In P. L. Houts (Ed.), *The myth of measurability*. New York: Hart.
- Skyrtic, T. M. (1991). The special education paradox: Equity as the way to excellence. *Harvard Educational Review*, 61(2), 148-206.
- Slack, W. V., & Porter, D. (1980). The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, 50(2), 154-175.

- 
- Smith, G. P. (1984). The critical issues of excellence and equality in competency testing. *Journal of Teacher Education*, 35(2), 6-9.
- Snyderman, M., & Rothman, S. (1990). *The I.Q. controversy: The media and public policy*. New Brunswick, NJ: Transaction.
- Strickland, B. R. (2000). Misassumptions, misadventures and the misuse of psychology. *American Psychologist*, 55(3), 331-338.
- U.S. Department of Education. (1996). *To assure the free appropriate public education of all children with disabilities Eighteenth Annual Report to Congress on the implementation of the Individuals With Disabilities Education Act*. Washington, DC: Author
- Zacharias, J. R. (1977). The trouble with tests. In P. L. Houts (Ed.), *The myth of measurability*. New York: Hart.

**Asa G. Hilliard, III** is the Fuller E. Callaway professor of urban education in the Department of Educational Policy Studies and the Department of Educational Psychology/Special Education at Georgia State University, Atlanta. She is an educational psychologist with specializations in assessment, power teaching, and African history and cultural studies.