DOCUMENT RESUME

ED 478 486 TM 035 085

AUTHOR Hilliard, Asa G., III; Amankwatia, Baffour, II

TITLE Assessment Equity in a Multicultural Society: Assessment and Instructional Validity in a Culturally Plural World.

PUB DATE 2003-04-00

NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 22-24, 2003).

PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)

EDRS PRICE EDRS Price MF01/PC01 Plus Postage.

DESCRIPTORS *Cultural Relevance; *Culture Fair Tests; Educational Assessment; Language; Measurement Techniques; *Standardized Tests; Test Use

ABSTRACT

In the past there were no substantial challenges to the idea that standardized, mass produced assessment would be universally beneficial. Culture was ignored or minimized as a factor in creating testing routines or in interpreting testing and assessment data. In recent years, challenges to this idea have arisen, but the primary pressure for the consideration of cultural context in mental measurement has come through the courts rather than through the academy or the testing profession. The more linguists study the semantic and practical meaning conveyed by language, the less comfortable they become about the possibility of accurate measurement of tests that use language as a medium. It is beginning to be believed by many that the most critical measurement points, at least as far as language is concerned, are the ones least susceptible to quantification. Psychologists do not appear to be responding to these issues, as the acceptance of the reality of diversity will undermine the possibility for standardized, mass produced, universally applicable measurement instruments. It must be recognized that cultural pluralism is a reality, and not rhetoric. Cultural salience seems to be a taboo topic in testing, but it is one that cannot be ignored. Courageous psychologists must decide whether the profession will consider taboo topics and whether it will embrace beneficial professional practice. (Contains 41 references.) (SLD)

ASSESSMENT EQUITY IN A MULTICULTURAL SOCIETY [ASSESSMENT AND INSTRUCTIONAL VALIDITY IN A CULTURALLY PLURAL WORLD]

Asa G. Hilliard III-Baffour Amankwatia II
Georgia State University
Department of Educational Policy Studies and
Department of Educational Psychology/Special Education
American Educational Association Annual Meeting
National Council on Measurement in Education
Chicago, April, 2003

A physicist speaks about what is required for "measurement."

> "The main defect in both sides, or either side, of this argument is that the protagonists pay so little attention to the **quality of the data base**.
> ...The worst error in the whole business lies in attempting to put people, of whatever age or station, into a single ordered line of "intelligence" or "achievement" like numbers along a measuring tape: 86 comes after 85 and before 93. Everyone knows that people are complex—talented in some ways, clumsy in others; educated in some ways, ignorant in others, call, careful, persistent, and patient in some ways; impulsive, careless, or lazy in others. Not only are these characteristics different in different people, they also vary in any one person from time to time. To further complicate the problem, there is variety in the types of descriptions. The traits tall, handsome, and rich are not along the same sets of scales as affectionate, impetuous, or bossy.
> ...As an old professional measurer (by virtue of being an experimental physicist) I can say categorically **that it makes no sense to try to represent a multidimensional space with an array of numbers ranged along one line**. This does not mean it is impossible to cook up a scheme that tries to do it; it's just that the scheme won't make any sense. It's possible to make an average of a column of figures in a telephone directory, but one would never try to dial it. Telephone numbers at least represent the same kind of idea: they are all addresslike codes for the central office to respond to.
> ...Implicit in the process of averaging is the process of adding. To obtain an average, first add a number of quantitative measures, then divide by however many there are. This is all very simple. **provided the quantities can be added**, but for the most part, with disparate objects, they cannot be." pp. 69-70 (Jerrold R. Zacharias, 1977) (Emphasis mine.)

A socio-linguist speaks about using language to construct measurement devices.

> "Meanwhile it will do for us to examine some recent uses of quantitative language analyses from the perspective of the linguist. As noted earlier, linguists generally

feel more comfortable about using quantitative analyses to probe for patterned differences than to generalize for broad groupings. **Likewise, the more linguists study the semantic and pragmatic meaning conveyed by language, the less comfortable they become about the possibility of accurate measurement of tests which use language as a medium. It is beginning to be believed, in fact, that the most critical measurement points of all, at least as far as language is concerned, are the ones least susceptible to quantification**.

A basic problem is that the goal of getting responses that will be comparable across subjects or across testing times is often realized by forcing **one standard interpretation of a question (or stimulus) and answer (or response) that is, in fact, not uniquely interpretable but rather is vague and can be fully specified only with reference to specifics of the individual test-takers' background and the individual test-taking occasion**. (Shuy, 1975)

In preparing for this presentation, I recalled old conversations that I had with Jerrold Zacharias at MIT, and Roger Shuy at the Center for Applied Linguistics, and similar conversations with other scholars in related disciplines. The more I reflected, the more I knew that I had to modify the title that I began with, in order to have a title with a better fit to the main arguments that I want to make. My modified title is, "Assessment and Instructional Validity in a Culturally Plural World." I changed "equity" to "validity" I changed "multicultural" to "culturally plural." If tests and assessments are truly valid, then equity is assured. By calling for validity, we keep the matter of scientific adequacy evaluation before us. For many in psychometrics, equity does not challenge technical adequacy. I use cultural pluralism in recognition of the lived experiences of ethnic families. For some, multicultural refers to the amalgam of cultural material in the "mainstream," without reference to ethnic families at all.

This simple title change captures the essence of my experiences in four decades of applied psychology in the schools, worldwide. Zacharias and Shuy quoted above are two of many scholars representing two of several related academic disciplines, whose messages have rarely if ever been heard by psychologists and educators in the field of "mental measurement." They offer glimpses of the porous foundation upon which the field of mental measurement is constructed, and upon which symptoms appear to be problems. For example "learning disabilities," the high incidence soft diagnostic category in special education, constitutes over half of all special education students. Since the category has serious construct validity problems, the large numbers are a symptom of the problem with validity, and are not the problem. Disproportionate placements of poor and minority ethnic groups in special education categories, and the political uses to which they are put, are symptoms of the problem of a lack of validity in assessment or instruction or both.

One more example of the avoidance of scientific information by psychometric experts and researchers who apply findings from mental "measurement." "Racial" comparisons

in "intelligence" are ubiquitous. From the beginning of the uses of IQ tests in education, until now, invidious comparisons are made between and among "racial" groups that do not exist. Regardless of how psychologists stand on matters of "race," it is a scientific flaw of the greatest proportion to refuse to engage findings that may threaten the validity of psychological findings. A physicist on measurement, a linguist on language use in measurement, and now, anthropologists on the appropriateness of using race in science are three examples of the gross failure in psychological test makers and producers. Consider the argument on "race," by Professor Alexander Alland Jr.

> ...The following letter, published in the Anthropology News for September 1998, from Jefferson M. Fish, a psychologist and professor at St. John's University in New Your City clearly illustrates his view of the danger of letting racist ideas professed by members of the academic community go unchallenged.
>
> > As a psychologist member of the AAA, I have been reassured by its handling of matters of "race."...
> > Anthropologists might be interested in a contrasting experience from the APA Monitor, the newsletter of the American Psychological Association. Its August 1977 edition carried a staff writer's front page story "When research is swept under the rug: Some of the best psychological research suffers for the sake of 'political correctness." This story portrayed Rushton (along with Arthur Jensen and others) as a serious scientist who has been victimized because his views are not 'politically correct." As a psychologist informant, I can confirm that the story accurately reflects the predominant view in academic psychology....
> > **Meanwhile the Monitor has turned down my request that they print the AAA's 1994 resolution on "Race" and Intelligence, the AAA's 1995 Statement on the Misuse of "Scientific Findings" to promote Bigotry and Racial and Ethnic Hatred and Discrimination, and the AAPA's (American Association of Physical Anthropologists) 1996 Statement on Biological Aspects of Race. I had suggested that, since APA is also in Washington, they cover the 1997 meeting-organized around the theme of race-and do a story contrasting anthropologists' understanding of race with that of psychologists. They chose not to because the Monitor covers psychology and not anthropology.**

Professor Fish extended his arguments concerning psychology's blindness to the false nature of race as a biological concept in an article published in the American Anthropologist. In this article, Professor Fish said the following.

Psychologists' ethnocentric assumption that "emic" [culture based] categories of "race" in the United States are biological realities appears particularly intractable,...My frustration at the imperviousness of American psychology, as a cultural group, to attempts to challenge this assumption with scientific data from anthropology has led me to a perverse respect for the power of ethnocentrism. It has also led me to write about "race"—including efforts like this article aimed at

encouraging anthropologists to help out....(Fish 2000,558) (Alland, 2002, pp. 170-171)

TABOO TOPICS.

I have spent the better part of my professional career trying to start discussions about the empirical study of context in testing and evaluation, especially the context of language variety. I first issued the challenge to testing advocates speak to these issues at an American Psychological Association Panel in San Francisco during the 1970s. The presidents of the Educational Testing Service and the Psychological Corporation were on the panel, along with six or so others, including the attorney for the APA. David Weschsler, author of the Weschsler Intelligence Tests, was in the audience. I raised the language issues there, asking merely that they be discussed. They were ignored. During the 1980s, I did the same thing with the senior professional staff of the American College Testing Program in Iowa. Good questions were raised by the attendees, with what appeared to be and understanding of the importance of the language issue. I have not heard of any follow-up. I also served on an invited panel at the Educational Testing Service, specifically on the topic of "bias" in testing, chaired by Anne Anastassi. I raised the same issues there. Again, I have not heard of any follow up. I served on the Committee on Testing and Assessment (CPTA) for the American Psychological Association, and was only minimally successful in getting a discussion there. To the best of my knowledge, the discussion never went beyond the committee. An adequate response, in my opinion, would be to invite the experts from the related disciplines to engage in a dialogue with psychometrics experts. If such meetings have occurred, I am not aware of them. Although, shortly after my presentation at the American Psychological Association annual meeting, the President of the Educational Testing Service became a member of the board of directors of the Center for Applied Linguistics, where I was also serving as a board member. Still, I know of no engagement of this branch of scholarship in any serious way in the work of test design and use.

THE PURPOSES, PRACTICES AND THE UTILITY OF TESTS AND ASSESSMENT APPROACHES

I entered the field of education with training as a psychologist, working as a school psychologist, an educational psychologist, and as a teacher. At that time, psychological work in schools could be summed up as follows. We were responsible primarily for executing "valid" and reliable testing and assessment for the purposes of ranking, classification and placement of students, for treatment in special education and or school tracks. A small number of psychologists and special educators also claimed to be doing "diagnostic testing and assessment." Few psychologists or other educators raised fundamental questions about what we were doing, why we were doing it. More importantly, few questions were raised about whether there was evidence to show that students benefited from these practices. We were preoccupied simply with executing standardized procedures and routines or recipes flawlessly, as a part of the big education service machine. The value of education and psychological services, procedures and routines were taken for granted.

During these decades there have been fundamental conceptual changes in the science and academics of testing and assessment. One of the most interesting is the cognitive change approach. (Feruerstein, Rand and Hoffman 1979) (Lidz, 2000) Another set of changes has been the acceptance of the underdevelopment of the field of mental measurement by experts, and the development of a vision of possibilities. (Rowe, 1991) In spite of these conceptual changes for a significant number of scientists, there have been far fewer changes in professional practices among the overwhelming majority of professionals in psychology as applied to education.

TESTING AND ASSESSMENT ISSUES SPECIFIC TO ETHNIC MINORITIES

It will be seen that where the "equity problem" in testing and assessment in education for minority cultural groups is concerned, these groups are the canaries in the miners cap, signaling deep problems with the whole enterprise of mass produced standardized testing and assessment, a paradigm problem. The Zacharias and Shuy quotations above, show that we are in deep water with this "measurement" business. Cultural anthropologists and experts from other related academic disciplines can inform the work of psychometrics and the assessments that are linked to that work. When applied to minority ethnic groups in a cultural and political environment, simple problems can be magnified.

I was taught in my training that good assessment was reliable, meaning that repeated assessments of the same type would yield the same general results. We turned to such references as the Buros Mental Measurement Yearbook, with its independent test reviews, to evaluate evidence of how good or bad our testing instruments were. The higher the reliability, the happier we were, convinced that both our instruments and our use of them were appropriate.

I was taught also that good assessment had to be "valid." While validity was the most important thing to be sought, it was linked to reliability. With too little reliability, or consistent testing results, it was impossible to get validity. At the time, we had great confidence in our concepts of validity, or at least few of us seemed to be bothered much by our regular uses of these concepts.

The common types of validity that we talked about were "construct validity," "face" validity, "content" validity, "concurrent" validity," and "predictive" validity." By far, "predictive validity," or forecasting, was valued most, with "intelligence" and cognitive testing. In general, predictive validity rested on the assumption of fixed abilities, resulting in the expectation that students will maintain their approximate rank in a distribution of scores of "mental ability." These various validities were applied in different proportions to "cognitive" or "intelligence" testing and to "achievement" testing, collectively the primary forms of school testing. Of course, "personality," "interest" and other types of testing were sometimes used to a lesser degree.

For the most part in practices, testing was all that there was to "assessment." There are few rigorous and commonly accepted standards for "assessment" as there are with

testing. The romantic notion that many sources of valid data are combined with testing, and integrated in an instructionally valid way, simply has not been documented. (Donovan and Cross, 2002) The standardized routines that were executed were accepted, on faith, as important activities to support, implicitly and rarely explicitly, beneficial school services. Gradually, some psychologists and educators began to be concerned about alternative meanings and values of the testing/assessment enterprise in schools. A few took action to develop approaches based on a new paradigm. However, even today, testing and assessment activities tend to be compliance activities, not to be confused with activities that inform instructionally valid design.

When I entered the field, there was not even a hint of a substantial challenge to the accepted idea that standardized, mass produced, "one-size fits-all" assessment for all children, or that it would be beneficial universally for any students. Culture was ignored or minimized as a factor in creating testing routines, or in interpreting testing and assessment data. Student's economic opportunity and what we now call "opportunities to learn" were also ignored. In fact, I do not recall that culture was "on the radar screen" at all. Moreover, I recall no serious discussions in any classes or at any professional meetings of psychologists about culture, based on the work of appropriate experts. I recall no recognition by psychometricians of the academic disciplines that could inform cultural considerations, such as cultural anthropology and cultural linguistics. In general, most responses to the challenge of cultural deficiencies in testing and assessment tended to be political.

To be sure, a minority of critical voices, some strident, have been raised in challenges to "one-size-fits-all" testing and assessment, but without the robust academic and scientific basis for the challenges. Oddly, the situation is not much better today.

Some educators and psychologists began to argue for "equity" in testing and assessment, for "culture free," "culturally fair," and "culturally relevant" or "culturally salient" testing and assessment. There was and is also "non-discriminatory" testing and assessment which speaks to equity without specific mention of culture. These equity oriented and intuitive responses to the validity problem fell far short of what was needed, mainly because the problem was not really understood as a validity problem or as a matter of science. No scientifically based data were collected routinely on "opportunity-to-learn," "non-discrimination" and "cultural" factors to be included in the assessment. No standards were developed to demonstrate the scientific basis for such adjustments. Therefore by which standards were these culturally responsive and "opportunity-to-learn" data to be evaluated, and for what purpose were the procedures to be performed? Clearly, no serious scientific response the challenges have yet been made in general practice.

Although some of the challenges to standardized testing, especially IQ tests, and assessment practices were made in a variety of states, California by itself gives a good example of these challenges with three landmark court cases. In 1970, Dianna vs. California State Board of Education dealt with the failure to take Mexican culture and language into account in testing. The Mexican children actually got and average of 15

point gains on IQ tests, when tested in their native language. In 1972 Lau vs. Nichols dealt with the failure to take Asian language into account in assessment for access to school services. In 1972, Larry P. vs. Wilson Riles dealt with the failure to take African culture into account because of the use of biased IQ tests. Similar cases were tried throughout the nation, Hobsen vs. Hansen on IQ and tracking in the District of Columbia in 1967, PASE [Parents in action for Special Education] vs. Hannon in Chicago in 1980 on IQ bias and special education, and Mattie T. vs. Holiday in Mississippi on IQ and special education in the 1980s.

So the primary pressure for consideration of cultural context in mental measurement came through the courts, rather than through the academy or the profession, which had turned its back on these questions. That is still true.

Of all the cultural challenges, the most transparent one is language diversity. Why would families and communities have to go to court to get psychologists and educators to understand the existence of and the meaning of language diversity, even in California, one of the most diverse states linguistically? California was also the state where the "ebonics" controversy flared up a few years ago, the source of almost hysterical opposition to the recognition of the cultural uniqueness of the language spoken by many African students. (Adger, Christian and Taylor, 1999) (Delpit and Perry, 1998) (Delpit and Dowdy,) (Crawford, 2001) (Jones, 1995) The Center for Applied Linguistics and the Linguistic Society of America joined in virtual unanimous support of the substance of the approach that the Oakland Unified School District was taking. Clearly, the level of hysteria was exceeded only by the level of the ignorance that propelled it.

A reading of a bit more of linguist Roger Shuy's (1979) analysis of the construct of intelligence, and the possibility of its measurement, using language data, is a powerful example of the validity of cultural linguistic criticism of IQ testing, in particular, and standardized tests and assessments in general. This is not trivial criticism. I must quote at length from Shuy's brilliant presentation.

> "...Now, if such a basic principle is overlooked by the schools, it's also overlooked by those who measure things in schools. It's common practice to measure that which can be seen and that which can be counted. In the area of foreign-language instruction, what is most frequently assumed in the vocabulary, phonology, and grammar is that what you can measure most often is that. We need to be cognizant **that ability to use language to get things accomplished is difficult to measure, not very physical and virtually impossible to count**. Naturally, it is seldom tested.
>
> **"It's small wonder then that if those who measure intelligence are not cognizant of the interference of surface forms on deep structure**, they're not alone. Most everybody else does the same thing." (Shuy, 1979. p. 2)
>
> "At this point, it has been asserted that subjective judgments of all sorts, including judgments about intelligence, are made by teachers, employers, researchers and

the general public regardless of the languages or settings in which they occur. **Rather compelling evidence rejects every claim made by those who attempt to show linguistic variations as a deficit. Most arguments put forth to support this claim misrepresent linguistic theory and reveal naive methodology through lack of cultural understanding**.

"One basic contribution of linguistics to this question is **that no one language or dialect, standard or nonstandard, is known to be significantly more complex than another in its basic grammatical or semantic characteristics**. The Cakchiquel Indians, for example, in Guatemala were said to be primitive by the Spanish people, but they had over a hundred times more verb forms than Spanish does; and whatever complexity means, it certainly isn't that.

"Well, linguists have not found any speech community with a native language that can be said to be logically or conceptually primitive. Likewise the so-called nonstandard dialects of English spoken by lower-class families in the inner-cities of this country are fully formed logical languages with only superficial differences in the means of expression from standard English - sometimes superior.

... "There's probably no real greater issue in linguistics today than the issue of what constitutes language. Our real task is this: **Assuming that there can be some agreement about what intelligence is, how can we find primary data about it**? Often it is said to be reflected in language use. p. 3

"**My point today is that such a representation is already several steps away from real data, for language use is influenced by a multitude of developmental, cultural, stereotypical and representational variances. This brings us a long way from a happy feeling about measurement of any sort**.

In my opinion, the distance is too great to be of any real significance in that **if ever the construct of intelligence can be shown to exist, the attempt to reflect it in language is far too distant to be of any real help**." .p. 6

THIS IS PARADIGM BREAKING STUFF! Psychologists cannot simply leave such profound professional opinions and documentation hanging with no response. I know of no meaningful responses from the psychometric discipline to this seminal linguistic challenge, or to other equally powerful challenges like it in other academic disciplines. I suspect that psychologists are incapable of responding because the consequences of doing so are enormous. The acceptance of the reality of diversity is to undermine the possibility for standardized, mass produced, universally applicable measurement instruments. Agree or disagree, it is a fundamental scientific flaw to ignore this to ignore this particular challenge. It seems that there is a code of silence here among many professionals.

The doubts about the validity of testing and assessment are not new. Some of the early pioneers were clear about the limitations of standardized testing.

> Existing instruments (for measuring intellect) represent enormous improvements over what was available twenty years ago, but three fundamental defects remain. Just what they measure is not known, how far it is proper to add, substract, multiply, divide, compute ratios with the measures obtained is not known; just what the measures obtained signify concerning intellect is not known. We may refer to these defects in order as ambiguity in content, arbitrariness in units, and ambiguity in significance."
> Edward L. Thorndike (Cited in Houts, 1977, p. 23 by Sheldon White)

Unfortunately, the testing movement seemed to be propelled by inertia. It was not able to hear and to respond appropriately to the other scientific information that would have a major bearing on the work of psychologists.

Validity, meaning to take such variables as culture and opportunity-to-learn into account. These things matter greatly in testing and assessment, and in the delivery of instructional services to students. Context includes culture and socio-economic status. The research community, belatedly within the past decade, recognizes contextual variables as influential. These variables pose great threats to standardized mass-produced testing and assessment validity, especially in the absence of controls for context variation in validation studies. They also pose a great challenge to profits from mass production. It must be noted that typical validation studies of IQ tests are correlational and not experimental. How can the linkages between powerful instruction and assessment be demonstrated without some use of experimental studies and, controlling for critical contextual variables? These are scientific matters, not political ones. When will we ever learn?

The issue for ethnic minorities is the reality of their existence. That reality cannot be denied and must be taken into account when professional tools and practices impinge upon their lives.

VALIDITY MORE THAN EQUITY: INSTRUCTIONAL VALIDITY MORE THAN OTHERS

Typically, the foundation academic disciplines for the support of education have been psychology, sociology, and anthropology to a lesser extent. When we move beyond day to day school instructional services, disciplines such as economics, business, and other areas contribute to policy studies. What is interesting about all of these disciplines, and especially about psychology [testing and assessment], is that until very recently, there was no explicit benefits criterion for the evaluation of instruction and the testing and assessment practices that inform instruction. In other words, for example, psychology, the primary testing and assessment discipline, was not evaluated in terms of its contribution to beneficial instruction, only in terms of the faithful execution of psychological procedures or routines and recipes. Under current job descriptions, school psychologists take no responsibility for powerful instruction, only forecasting student performance.

In other words, the traditional validity criteria for "mental measurement" did not include empirically determined "instructional validity," the linking of testing and assessment to the improvement of instructional outcomes. Some psychometric experts refer to this as "consequential validity." To name "instructional validity" is to set in motion activities to provide empirical documentation to determine if it exists. In fact, that is precisely what happened, when the first National Research Council of the National Academy of Sciences committee on disproportionate placement of black males [later changed to "minorities"] in special education sought to determine if testing and assessment and special education services produced benefits, not merely predictive validity. The report was entitled Placing Children in Special Education: A Strategy for Equity. (Heller, Holtzman and Messick, 1982) A second National Academy of Sciences report during the same year, Ability Tests, Uses and Consequences echoed the instructional validity call. (Wigdor and Garner, 1982) Both reports begin with the clear statements.

> "Our ultimate message is a strikingly simple one. **The purpose of the entire process, from referral for assessment to eventual placement in special education, is to improve instruction for children. The focus on educational benefits for children became our underlying theme**, cutting across disciplinary boundaries and sharply divergent points of view.
> . . . These two themes the validity of assessment and the quality of instruction are the subjects of this report. **Valid assessment, in our view, is marked by its relevance to and usefulness for instruction**. (pgs. x, xi) (Heller, Holtzman and Messick, 1982)
>
> . . . **The basic principle underlying the Committee's discussion of testing in the schools is that the classification of pupils is warranted only when the decision rules, whether based on tests or not, have instructional validity**. No school child should be relegated to a program of instruction that is not expected to enhance performance. (pg. 5) (Wigdor & Garner, 1982)

These clear statements introduce a "benefits" criterion to determine the "**instructional validity**," of professional practices and services for the first time, to my knowledge, in major scientific or professional publications. This is a seismic shift in educational and psychological paradigms, from a custodial to a remedial paradigm

Included in the final report, Placing Children in Special Education: A Strategy for Equity, (Heller, Holtzman and Messick, 1982}, and in subsequent years, empirical research was done to determine **if the linking of testing, assessment and school services produced student achievement benefits**. The work showed essentially that there were few benefits to negative benefits. This work began at the instigation of the Office of Civil Rights. The initial target for investigation was African male extreme disproportion in the high incidence categories. This was changed to consider disproportion in general. Twenty years later, the second National Academy of Sciences report on minority disproportion, Minority Students in Special and Gifted Education (Donovan and Cross, 2001) found virtually the same results as the first, with the review of even more studies seeking to document benefits. The Donovan and Cross report confirmed the substance of

the two 1982 reports, and extended their findings. In other words, a cost benefit analysis of the current popular use of testing and assessment in the schools would result in a finding of virtually all cost and few to no benefits, in the high incidence special education categories. In fact, the lack of evidence for "instructional validity" and utility caused the second National Academy of Sciences committee to **provide the basis for the argument for the elimination of IQ tests from school use!** The following excerpts from the National Academy of Sciences Report add rationale for this thinking. (Donovan and Cross, 2001)

> "In addition to the limitations of IQ tests from the perspectives of cultural psychology, it is **questionable whether the costs of IQ tests are worth the benefits** in special education eligibility determination."
> ... "The use of IQ tests and IQ-based disability determination does not promote the achievement of those critical goals**; therefore, IQ should be abandoned, even if that action complicates the work of other agencies**."
> ... "Moreover, test authors and test publishers all acknowledge that IQ tests are measures of what individuals have learned – that is, it is useful to think of them as tests of general achievement, reflecting broad culturally rooted ways of thinking and problem solving. The tests are only indirect measures of success with the school curriculum and imperfect predictors of school achievement."
> "**Perhaps the most convincing of the arguments against IQ tests is that the results are largely unrelated to the design, implementation, and evaluation of interventions designed to overcome learning and behavioral problems in school settings. For example, IQ is not a good predictor either of the kind of reading problem that a student exhibits or of the student's response to treatments designed to overcome that reading problem**"
> ... "The same general interventions appear to work with basic skills problems regardless of whether the student is classified with mild mental retardation (MMR), learning disability (LD), or emotional disturbance (ED)"
> ... "**The differentiation between LD and MMR that is done primarily with IQ test results does not lead to unique treatments or to more effective treatments**."
> ..."**No contemporary test author or publisher endorses the notion that IQ tests are direct measures of innate ability. Yet misconceptions that the tests reflect genetically determined, innate ability that is fixed throughout the life span remain prominent with the public, many educators, and some social scientists**."
> ... "The most frequent use of IQ tests today is in determining whether a 'severe discrepancy between achievement and intellectual ability' exists as per the federal criteria for LD (34 C.F.R.300.541) and state LD classification criteria. Several problems exist with this procedure. First and most fundamental, there is no 'bright line' in performance that can be used to determine the appropriate size of the discrepancy; the size required is arbitrary."
> ..."**The present study suggests that the concept of discrepancy operationalized using IQ scores does not produce a unique subgroup of children with reading disabilities when a chronological age design is used;**

> **rather, it simply provides an arbitrary subdivision of the reading-IQ distribution that is fraught with statistical and other interpretative problems"...Poor readers who make up 70 to 80 percent of the current LD population seem to have the same needs and the same cognitive processing profiles, and they respond to the same treatments regardless of their IQ status** (it should be noted that children with IQs less than 80 generally were excluded from the NICHD studies). **Therefore, arbitrarily dividing poor readers into subgroups with higher IQs (those who meet the current LD criteria) and those with IQs similar to their reading achievement levels is invalid. With regard to reading-related characteristics, these subgroups are much more similar than different, calling into serious question the current LD diagnostic practices**.
> "... **Importantly for future policy development, the IQ test results and whether or not a child shows a discrepancy between IQ and reading achievement have little significance for understanding or treating a reading disability**." (See especially pages 283-91)

Why did it take so long to discover the lack of value, or even the presence of harm in professional practice, invalid professional practice? "Blinders" were built into the professional conceptualizations and routines. Professional habits, really fixations, rarely critiqued, are self perpetuating.

I must hasten to add here that instructionally valid beneficial testing and assessment approaches have been available for more than half a century.(Feuerstein, 1979 and Lidz, 2000) However, it comes from and entirely different paradigm, a marginalized paradigm. This alternative paradigm, rooted in the cognitive psychology of Binet and Piaget, assumes malleable intelligence, and demonstrates the instructional validity of diagnostic cognitive assessment and remedial teaching or mediation. The diagnostic assessment is called "dynamic assessment" for structural cognitive modifiability. The remedial teaching is called "instrumental enrichment" for changing cognitive structures in an enduring way. **The remedial paradigm changes the assumptions, the questions, the goals, the roles of professionals, and many other things**. This approach, although marginalized, does show powerful achievements benefits for students.

A brief word is in order here about achievement testing. The primary concern about achievement testing is that tests be "content valid." It is not rocket science to say that tests should measure what schools promise to teach. Of course, in the real world, students rarely are exposed to a common curriculum with a uniformly high quality of professional service. Rarely do valid tests exist of the actual curriculum that is offered, the actual opportunities to learn. Time and time again, empirical work demonstrates the lack of content validity of popularly used standardized tests. Empirical documentation of the "Savage Inequalities" (Kozol, 1991) in opportunities-to-learn in school services offered to students is abundant.

Of course, since few curricula are culturally salient, few tests are either. Cultural salience is of great importance, since students should have the opportunity to demonstrate what

they know and can do; in perceptions, operations, functions and structures, using cultural material that is familiar to them. In addition, the human experience is an enormously diverse experience. The truth of the human experience must be reflected in the diversity of the school world. Any non-diverse curriculum is an untruthful one, be that a curriculum in mathematics, science, art, music, history, etc.

Mass produced standardized achievement tests have a poor content validity track record. That will remain a problem because savage inequalities in opportunities-to-learn shows no signs of abating. I do not argue that standardized mass produced tests should be eliminated. The argument is that they are not measurement when applied in the world of diversity. This means that they must be takes for what they are, rough instruments for data gathering as a point of departure for assessment. They may give some indication of the results of exposure to instruction. Some of these data may be useful, potentially even beneficial. However, the benefits must be demonstrated empirically.

CULTURAL PLURALISM IS A REALITY, NOT RHETORIC

It is a simple fact, empirically demonstrable, that there is no universal human culture. Testing and assessment are dependent upon communication. Systems which are tied to language. Language is culturally embedded and is a core part of the communication process. Cultural linguists are scientists who have developed a body of knowledge and models of meaning that are essential to the work of those who use language in making measurement devices, and in the interpretation of the meaning of mediation and responses. (Shuy, 1979) Standardization in testing and assessment is of high value to entrepreneurs, who make money, if they can mass produce products, such as tests. The best of all possible worlds for them is "one size fits all." Unfortunately for their needs, humanity made up of a plurality of cultures.

Until now, it is rare to see any academic/scientific expertise in culture, its contents and principles, reflected in mainstream psychology. The reality of cultural diversity and its meaning for science is often seen by measurement experts as a "minority concern," a "fairness concern," a "civil rights concern," an "equity concern," but not a scientific concern. This means that many scholars, often noting the concerns above, seem to believe that cultural diversity matters are more political than scientific. Moreover, there is real resistance among many traditional psychologists to engage in the required scientific study and dialogue about these cultural matters. Their cultural naiveté is almost legendary. The field of psychology is littered with the wreckage of attempts to function professionally in ignorance of cultural realities, and in ignorance of the cultural sciences that document and interpret those realities.

POLITICS, TESTING AND ASSESSMENT: MENTAL MEASUREMENT AND HEGEMONY

This section deserves extensive treatment. However, it will be brief. Yet, I cannot conclude this paper without raising the issue of the contamination of the professions, including psychology and other behavioral sciences, by the more than 400 year tradition

of white supremacy hegemony. Many of the invalid and bad practices in testing and assessment, in particular, stem from the well documented partnership between many powerful people in our field and the forces of slavery, colonialism, segregation/apartheid and white supremacy ideology.

It would take many books to treat this subject well. I will merely provide a few bibliographical references to support the charge. The most recent example is the widely respected and supported psychologist, Arthur Jensen ["How much can we boost IQ, 1979 and Bias in Mental Testing, 1980], to Richard Herrenstein and Charles Murray [The bell curve; 1994], and finally to the internationalization of the bell curve racist ideology, Richard Lynn and Tatu Vanhanen [IQ and the wealth of nations. 2002]. Other classic works document the continuing contamination of professional practice with white supremacy thinking. (Gould, 1996) (Guthrie, 1976) (Kamin, 1974) (Tucker, 1994) Few courses in history and systems offer the documentation and the discussion time that is due to the white supremacist use of the discipline. This was and remains a non-trivial matter.

The political nature of the mental measurement matter can be seen in its pseudo-science application, the comparasion of "races."

This is truly a taboo topic. White supremacy ideology is a contextual variable, subject to empirical confirmation. It is rarely investigated. It must be apart of validity considerations. The culture and equity testing and assessment investigation of validity cannot be complete, in the absence of an examination of this continuing phenomenon.

CONCLUSION

The professional remedies for the problems that I have tried to describe are worthy of careful consideration, though there is little time for that discussion here. I have merely tried to open up or re-open neglected topics. Cultural salience appears to be among the taboo topics, such as including the sciences of cultural linguistics and cultural anthropology in the psychometric endeavor. Even to enter into discussion about them is to threaten the derailment of the mass produced testing and assessment train, as it is now constituted. Perhaps that explains the grand silence on these matters.

As I see it, the main problems here are political and economic, not academic and professional at all. Professionally, scientifically and academically, the way has already been prepared for appropriate responses to instructional validity and cultural salience in testing and assessment and teaching linkage. Vested interests will seek stasis. Courageous psychologists will determine whether we consider the taboo topics, and whether we embrace beneficial professional practice.

The students are waiting.

Selected References and Bibliography

Adger, Carolyn Temple; Christian, Donna and Taylor, Orlando (1999) (Eds.) Making the connection: Language and academic achievement among African American students. McHenry, Il.:Center for Applied Linguistics and Delta Systems

Alland, Alexander Jr. (2002) Race in Mind: Race, IQ and other Racisms. New York :Palgrave Macmillian

Crawford, Clinton (2001) (Ed.) Ebonics and language education of African ancestry students. New York: Sankofa World Publishers.

Dandy, Evelyn B. (1991) Black communications: breaking down the barriers. Chicago: African American Images

Delpit, Lisa and Perry, Thresa (1998) The real ebonics debate: power, language and the education of African American children. Boston:Beacon Press

Delpit, Lisa and Dowdy, Jo Ann (Eds.)( ) The skin we speak.

Donovan, M. Suzanne and Cross, Christopher T. (2001) Minority Students in Special and Gifted Education, Commission on Minority Representation in Special and Gifted Education, Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press

Feuerstein, Reuven; Rand, Ya'acov and Hoffman, Mildred (1979) The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments and techniques. Baltimore: University Park Press

Fuller, Renee (1977) In search of the IQ correlation: a scientific whodunit. Stony Brook, N. Y.: Ball-Stick-Bird Publishing

Gould, Stephen J. (1996) The mismeasure of man. New York: W.W, Norton

Guthrie, Robert V. (1976) Even the rat was white: a historical view of psychology. N.Y.: Harper and Row Publishers

Hehir, Thomas and Latus, Thomas (1992) Special education at century's end: evaluation of theory and practice since 1970. Harvard Educational Review Reprint Series Number 23

Heller, Kirby A., Holtzman, Wayne H., and Messick, Samuel (Eds.) Placing Children in Special Education: A Strategy for Equity. Washington, D.C.: National Academy Press, 1982

Herrenstein, Richard and Murray, Charles (1994) The bell curve: intelligence and class structure in American life. New York: The Free Press

Hilliard, Asa G. III (1990) "Back to Binet: The case against the use of IQ tests in the schools." Contemporary education. 61, 4, 184-9

Hilliard, Asa G. III (1995) Either a paradigm shift or no mental measurement: The non-science and non-sense of the Bell Curve. Psych Discourse. 76, 10, 620

Hilliard, Asa G. III (1984) "IQ testing as the emperor's new clothes: a critique of Bias in Mental Testing" In C. Reynolds (Ed.) Perspectives on Bias in Mental Testing. New York: Plenum

Hilliard, Asa G. III (1988) Misunderstanding and testing intelligence." In John Goodlad and Pamela Keating (Eds.) Access to knowledge. New York: The College Board, 145-157

Hilliard, Asa G. III (1987) Testing African American students. Special Issue of the Negro Education Review. 38, numbers 2 and 3 (Republished 1995, by Chicago: Third World Press)

Hilliard, Asa G. III (1975) "The strengths and weaknesses of cognitive tests for young children." in J. D Andrews (Ed.) One child indivisible. Washingrton: D.C.: National Education for the Education of Young Children

Hilliard, Asa G. III (1983). " factors associated with language in the education of the African-American child." Journal of Negro Education, 52(1), 24-34.

Hilliard, Asa G. III (1994) "What good is this thing called intelligence and why bother to measure it? Journal of black psychology. 20, 4, 430-444

Houts, Paul (Ed.)(1977) The myth of measurability: IQ tests, standardized tests. New York: Hart Publishing Company, Inc.

Jacoby, Russell and Guberman, Naomi (1995) The bell curve debate: history, documents, opinion. New York: Times Books, Random House

Jones, J. Arthur (1987). Look at math teachers, not 'Black English'. Essays and Policy Studies. Washington, D.C.: Institute for Independent Education.

Kamin, Leon (1974) The Science and politics of IQ, New York: John Wiley and Sons

Kincheloe, Joe L., Sternberg, Shirley N. Gresson, Aron D. M. (1997) (Eds.) Measured lies: the bell curve examined. New York: St. Martins Press

Kozol, Jonathan (1991) Savage inequalities: children in America's schools. New York: Crown

Lidz, Carol and Elliott, Julian G. (2000) (Eds.) Dynamic assessment: theory, models, and applications. New York: Pantheon Books

Lidz, Carol and Elliott, Julian G. (2000) Dynamic assessment: prevailing models and applications. New York: JAI, An Imprint of Elsevier Science

Lynn, Richard. and Vanhanen, Tatu (2002) IQ and the wealth of nations. Westport, CT: Praeger

Rowe, Helga A.H. (1991) Intelligence: reconceptualization and measurement. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers

Shuy, Roger, (1979) "Is the Construct Intelligence a Twentieth-Century Myth?" (Transcript of Symposium Presentation) American Psychological Association Annual Convention, New York, 1979. (Other Symposium Participants: David Elkind, Chairman; John Horn, Renee Fuller, Asa Hilliard, Presenters; and J. McVicker Hunt, Discussant.)

Shuy, Roger (1975) "Quantitative linguistic data: a case for and some warnings against." Unpublished manuscript. (Shuy was at the Center For Applied Linguistics in Washington, D.C.)

Skrtic, Thomas M. (1992) "The special education paradox: equity as the way to excellence" in Thomas Hehir and Thomas Latus (Eds.) Special education at century's end: evaluation of theory and practice since 1970. Harvard Education Review Reprint Series No. 23

Smitherman, Geneva (Ed.) (1981). Black English and the education of black children and youth: Proceedings of the National Invitational Symposium on the King Decision, Detroit, Harlo Press

Tucker, William H. (1994). The science and politics of racial research. Chicago: University of Illinois Press.

Van Keulen, G., Weddington, G. And DeBose, C. (1998). Speech, Language, Learning and the African-American Child. Allyn Bacon.

Wigdor, Alexandra K. and Garner, William R. (Eds.) Ability Testing: Uses, Consequences, and Controversies, Part I. Committee on Ability Testing, Assembly of Behavioral and Social Sciences National Research Council. Washington, D.C.: National Academy Press, 1982.

Zacharias, Jerrold R. (1977) "The trouble with tests." In Houts, Paul (Ed.) The myth of measurability. New York: Hart Publishing Company, Inc.

Alland, Alexander Jr. (2002) Race in Mind: Race, IQ and other Racisms. New York :Palgrave Macmillian

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM035085

I. DOCUMENT IDENTIFICATION:

Title: Assessment Equity in a Multicultural Society [Assessment and Instruction Validity in a Culturally Plural World]

Author(s): Asa G. Hilliard III - Baffour Amankwatia II

Corporate Source: Georgia State University

Publication Date: April 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ↑

☐

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ↑

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ↑

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature:	Printed Name/Position/Title: Asa G. Hilliard
Organization/Address: Georgia State University, Educational Policy Studies, University Plaza, Atlanta, GA 30303	Telephone: 404-651-1270 FAX: 404-651-1009
	E-Mail Address: ahilliard@gsu.edu Date: 6/23/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions