SESSION ONE: Evaluation of Educational Achievement of Underrepresented Minorities

Session Chair
Beatriz Chu Clewell
Principal Research Associate and Director
Evaluation Studies and Equity Research Program
Education Policy Center
The Urban Institute
And former Executive Director
Commission on the Advancement of
Women and Minorities in Science,
Engineering and Technology (CAWMSET)
National Science Foundation

Presenters
Gerunda B. Hughes
Assistant Professor
Howard University/School of Education
Center for Research on the Education
of Students Placed at Risk (CRESPAR)

Carlos Rodríguez
Principal Research Scientist
American Institutes for Research

Discussant
Jane Butler Kahle
Division Director
Elementary, Secondary and Informal Education
EHR/NSF

Guiding Question

- Several Federal agencies — for example, NSF, NASA, DOE and ED — support science and mathematics education reform. Evaluation of their efforts includes paying attention to student academic achievement. What issues surround the evaluation of science and mathematics achievement, especially with respect to underrepresented populations? The discussion should highlight the cultural context of this area of evaluation in light of relevant literature.

Discussion Highlights
Gerunda B. Hughes

Session chair, Dr. Beatriz Clewell, opened the meeting by highlighting the importance of collecting, analyzing, properly interpreting, and reporting achievement data. The results can help educators and policy makers set an agenda for educational reform and improvement, she said. The benefits have been seen on the international level through the Third International Mathematics and Science Study (TIMSS).

The relatively poor showing of U.S. students when compared with some of their peers from other nations drew negative public reaction.   The study revealed a gap in mathematics and science achievement the size of which few knew existed. The TIMSS results prompted many recent mathematics and science reform efforts in U.S. classrooms.   In evaluating these reforms, asserts Dr. Clewell, there is a need to measure not only achievement, but also the factors that may influence it.   This information fosters better understanding of student performance and facilitates the design of program and interventions modeled after "best practices." Such practices are most useful when they reflect the goals and objectives of American education, especially the goals to provide equal and equitable learning opportunities.

In light of the gap between white and underrepresented minority students in the area of mathematics and science achievement, we face an important question: Can we be any less diligent in implementing programs, policies, and practices to eliminate the national gap than in addressing the international gap? Dr. Clewell asserted the need to be even more diligent nationally because achievement by "American students" is inextricably tied to performance of all sub-populations, including "underrepresented minorities," who comprise an increasing proportion of the total student enrollment. NSF is a leader in support of projects and programs that improve mathematics and science achievement. But are NSF-funded education reform projects implementing effective evaluation designs that will capture data that is necessary to make sound decisions about project effectiveness and success? Regarding the assessment of achievement by underrepresented minorities, what issues, cultural and otherwise, must be considered?

Following Dr. Clewell, Dr. Carlos Rodríguez addressed three topics: contextual considerations, misconceptions and myths, and guiding principles. America is in transition in many ways, he said. Although minorities are quickly becoming the majority, critical masses of minority students remain scientifically and technologically illiterate. Dr. Rodríguez also noted that many people do not meet minimum levels of proficiency in literacy and quantitative capability. Most are Latinos, blacks, and poor whites. If these problems are ignored, many minorities will not receive the material rewards that follow competent performance, nor will they be able to participate fully in a democracy. According to Dr. Rodríguez, these circumstances pose serious implications for peace on the homefront and challenge the United States' position as a global leader in technology.

Dr. Rodríguez suggested that program evaluators bear much of the responsibility for bringing about effective reform. "What we do today in program evaluation is important because we are setting the pace, envisioning the future, changing the present," he asserted. "Part of what is setting the pace today is the call for high expectations (including high standards) and disciplined effort." Dr. Rodríguez added that, too often,

these noble goals are confused with high stakes testing, which is often found where students have not had an adequate opportunity to learn. "Are standards, frameworks and high stakes testing simply yet another excuse [for] telling minority students how deficient they are?", Dr. Rodríguez asked. If this is the tenor and context of current program evaluations, he continued, we have not progressed very far.

Dr. Rodríguez also warned that if our program evaluations produce still more "blaming the victim" litanies about minority students' bleak performance in mathematics and science without also examining their caretakers' willingness to improve, we do nothing but subscribe to myopic views. He said that the result is captured in the Spanish saying, "*El que adelante no va, atras se queda*" ("Who doesn't go forward, stays behind.") Dr. Rodríguez cited recent reports by the National Assessment of Education Progress that Hispanic eighth graders are more likely than non-Hispanic whites to take no science courses, while Hispanic and black students are more likely than any other group of students to take remedial mathematics and English classes. He said we continue to invest in schools and teachers like amateur gamblers, placing small bets but expecting huge winnings. It is no wonder, he said, that we have yet to realize either the ideals embodied in *A Nation at Risk* (1983) or the goals of *Goals 2000: Educate America Act* (1990).

Dr. Rodríguez identified the following misconceptions and myths, related to the dynamics of race, culture and language, which affect student learning and the organization of learning and opportunities
to learn:

- Past inequities and inequalities have been addressed and no longer require attention.
- Merit can be defined by test scores.
- Fairness is best achieved through race-neutral policy.
- The goals of excellence and equity are irreconcilable.
- Test scores alone tell the whole story.

Dr. Rodríguez further noted that many variables can affect a student's test score, including:

- The quality of the student's education,
- The student's skill, ability or knowledge about a particular topic,
- Preparation for the test, or even
- What the student may have eaten for breakfast on the day of the test.

Dr. Rodríguez suggested that program evaluators and teachers need to spend time together to reflect on the diagnostic potential of tests and assessments to improve science and mathematics achievement among minority students. He also shared the five guiding principles of evaluation as adopted by the American Evaluation Association in 1994:

- Systemic inquiry,
- Competence,
- Integrity and honesty,
- Respect for people, and
- Responsibility for the general and public welfare.

"There are insidious notions in popularized ideas about the very ability of non-mainstream students to learn that appear ready to sabotage and derail any progress put forth by program evaluations, policy, research and practice," Dr. Rodr'guez warned. He said that political and ideological processes may essentially be determining the "improvement" of science and mathematics achievement among underrepresented minority students, and he added that program evaluators should not ignore or deny them.

Dr. Gerunda Hughes began her presentation by noting how often evaluation designs are very narrow in scope, focusing on the achievement of students as operationally defined by a test score. According to her, this often leaves many questions unanswered, especially when determining the effectiveness or success of a project that has been designed to improve science or mathematics achievement. Dr. Hughes noted that improved student achievement is a direct or indirect goal of many educational reform efforts but that academic achievement does not exist in a vacuum; it is correlated with factors that may or may not be within the control of the reform effort. To the extent that these factors are controllable, Dr. Hughes said, they may be planned for in the project design. If they are not controllable, they can at least be measured and statistically controlled in the analysis of achievement data. According to Dr. Hughes, school factors include: the quality of instruction; the opportunity to learn the material being assessed; and teacher characteristics such as efficacy, knowledge of content, teaching skills, attitudes and beliefs about the children's capacity to learn and years of experience. Personal factors include cultural orientation and socioeconomic status. The assessment of all of these factors (and others) can increase understanding of the academic achievement of minority and non-minority students, Dr. Hughes said.

Like Dr. Rodríguez, Dr. Hughes noted that the goal is neither to belittle the test score nor to ignore standards. Minority students must be held to the same high standards as other students, if they are going to be prepared to compete in the global marketplace, she said. However, she suggested that the test score does not tell the whole story. She said program evaluators must be sensitive to the widely divergent educational experiences, backgrounds and cultures of students and explore the ways in which those factors interact with the cultures of teaching, learning and assessment. With such sensitivity, Dr. Hughes said, evaluators and other measurement professionals are better able to identify and use traditional or innovative approaches to testing and assessment — methods that yield valuable information about what individuals know and can do.

In addition, Dr. Hughes said, it is possible that changes in the focus of curricular goals have a major impact on the kind of data that program evaluators collect and how it is collected. As an example, she noted that the NSF Working Group on Assessment in Calculus acknowledged that the mathematics community needs to change fundamental ways of assessment at all grade levels because of an increased understanding of what it means to think mathematically. The group suggested the use of paper-and-pencil tests as well as performance tasks, open-ended items, investigations and projects, observations, interviews, portfolios and self-assessment. Although there is some overlap in their purposes, each alternative assessment method taps into a slightly different aspect of learning. Dr. Hughes said that this has implications for the professional development of teachers in assessment, testing and measurement, especially as they work with program evaluators to provide useful information that can inform instruction and explain outcomes more fully. In closing, Dr. Hughes challenged all evaluators, and minority evaluators in particular, to go beyond the obvious and pay attention to factors that correlate with student achievement.

Dr. Jane Butler Kahle, the session discussant, stressed the need to look for achievement trends, rather than to try to established causality. She also reinforced the importance of disaggregated data, citing her recent study that showed different factors affected the science achievement of 8th grade African American children. She encouraged using multi-variant analyses, such as Hierarchial Linear Modeling, to analyze the complex factors surrounding minority achievement, as well as combining qualitative data with quantitative data. She stressed that until an adequate number of minority evaluators could be educated, efforts to educate majority evaluators in understanding minority issues, cultures, and contexts needed to expand.

Asked to identify the important points of session discussions, participants suggested the following guidelines:

- Utilize
  - quantitative and qualitative approaches in the evaluation design
  - multiple measures for project evaluation
  - multiple data sources
  - short-term, intermediate and long-term objectives
- Define "success" in multiple ways (e.g., not just in terms of test scores and standardized measures, but also in terms of increased attendance and decreased mobility rates; positive student-teacher interactions; increased parental involvement in the math or science education of the child; increased self-esteem among students; increased persistence rates; and improved attitudes about schooling).
- Disaggregate data according to race, gender and ethnicity (also along contextual lines, if appropriate).
- Include evaluators who are sensitive to issues of diversity (especially minority evaluators) and who will frame the right questions.

Papers/Presentations

Evaluation of Educational Achievement of Underrepresented Minorities: Assessing Correlates of Student Academic Achievement
Gerunda B. Hughes

The evaluation of the educational achievement of children involves more than analyzing and reporting test results. It carries the responsibility of providing important information to stakeholders that facilitates both internal and external decision-making about a project. This additional charge is reflected in the definition of evaluation as stated by the Joint Committee on Standards for Educational Evaluation (1981). In its report, the committee defined evaluation as "the systematic investigation of the worth or merit of an object." Similarly, Webster's International New Dictionary defines evaluation as "the examination of the worth, quality, significance, amount, degree or condition of [an object]." The target of most educational evaluation is a project, program or product that has as one of its primary or secondary goals, improved student achievement. The evaluation of achievement usually comes near the end of the project, for example, and is part of the summative evaluation. Inferences and conclusions about the success or failure of a project are drawn from data collected during the project. Sometimes these evaluations involve the use of controlled or matched groups and sometimes comparisons of student achievements are made on the basis of sex, race or ethnicity. In the latter case, what is often reported in the research literature and what characterizes many evaluation reports about minority students' educational achievement is limited to "the amount and degree of achievement." This is best illustrated by the first sentence of the first chapter in the book, *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. It states: "African Americans currently score lower than European Americans on vocabulary, reading, and mathematics tests, as well as on tests that claim to measure scholastic aptitude and intelligence (p. 1)." This statement may be true. But is it the whole story? Are there factors that may explain differences in the academic achievement of African Americans and European Americans? Can project staff plan effectively for the influence of such factors on project outcomes? If so, how? One way is by systematically assessing those factors known to correlate with student achievement. The additional information can provide greater insight into why certain outcomes, such as gaps in educational achievements, exist.

Goal of Educational Reform

Improved student achievement is, directly or indirectly, a major goal of educational reform efforts. Yet, the academic achievement of children does not exist within a vacuum. It is influenced by and correlated with a variety of school and personal background factors. These operate in ways to facilitate or inhibit the academic achievement of children in different contexts. School factors include how children are taught, how they are assessed, teacher expectations and opportunity to learn. Personal background factors include cultural orientation and socioeconomic status. Assessment of these factors in evaluation studies can lead to a more comprehensive understanding of the academic achievement of minority and non-minority students.

Researchers in the field of testing, measurement and assessment have noted that the systematic assessment of African Americans, Native Americans, Hispanics, women and persons of low socioeconomic status is as appropriate, as desirable and as necessary as it is for any other group (Davis, 1948; Johnson, 1979; Gordon, 1996). Additionally, there is no argument against the logic that individuals within these groups must develop the same body of skills and expertise that standards require. What is argued, however, is that a single test score does not reflect all of reality, that it should not be used as the sole basis for making inferences about individuals or groups. Rather, it is indeed necessary to look beyond the test score to the widely divergent educational experiences, backgrounds and cultures of the test taker and explore how these interact with the cultures of teaching, learning and assessment (Johnson, 1979; Ladson-Billings, 1995). By doing so, evaluators and other assessment professionals are in a better position to identify more effective approaches to testing and assessment that yield valuable information about what individuals know and can do. These approaches to testing and assessment very often take into account the multidimensional variables associated with educating people of color —especially how they are taught and how they learn.

Culture and Cognition

Issues about culture have always played an important part in schools. Though these issues were not always directly addressed, children were, nonetheless, judged and evaluated in terms of how much of the mainstream culture they espoused. In essence there is a positive correlation between the amount of mainstream acculturation one has acquired and achievement on traditional mainstream tests of cognition. In fact, not so long ago children who came from impoverished environments and who generally scored lower on tests of achievement and cognition were described as "culturally deprived."

Fortunately, being "cultured" is no longer associated with having high test scores. In its most basic form, culture entails the way a particular group of individuals translates reality. Many of these translations are evident in how children respond to test items. Boykin (1994) and Giroux and McLaren (1986) suggest that culture embodies a set of practices and ideologies from which different groups draw to make sense of the world. It embodies belief systems and ways of knowing and valuing. Sometimes the culture of schooling and the assessment of what goes on in schools matches that which the children bring to the classroom, and sometimes it does not.

Delpit (1988) argues that a culture of power exists in American classrooms. That culture of power reflects the practices of those in power and consequently reflects mainstream culture. Some children come to school with the rules of the culture already understood. They come to school inclined to be receptive, if not endowed with the culture because of their previous experiences with it. Other children, however, do not have or have not been sufficiently exposed to the "rules of the culture" and hence, are directly or indirectly penalized for not knowing them. In other words, some children possess the requisite "cultural capital" while others do not. The latter are among the "culturally deprived."

Boykin (1994) notes that the school culture and, I might add, the present mainstream "assessment culture," reward individual possession of specific intellectual and social attributes. By emphasizing competition among individuals and devising reward systems to honor such accomplishments, schools and the testing professions reinforce what they value. But what if tests and assessments were designed to be aligned with the cultural integrity of the children and at the same time maintain the content validity they were designed to possess? It may be possible to demonstrate that the much talked about gap between black and white children is not as large as one might think — or at least there may be a way to explain the variance in the context of culture.

Snow and Lohman (1989) note that "cognitive study of the targeted aptitudes, achievements and content domains for which educational measures are to be built might suggest alternative measurement strategies and refinements for existing instruments." With this perspective, they help provide a rationale for seeking alternative ways of assessing what different subgroups of the populations can do on tests and assessments. Furthermore, they note that much of cognitive research on the nature and development of ability suggests that learner experience, which includes out-of-school experience and the cultural context of the home, is an important determiner of what attributes are measured by any test and how many different attributes are measured.

Teachers and Teaching

The role that teachers play in the academic achievement of minority children cannot be overemphasized or understated. Teachers' personal and cultural attributes as well as their attitudes, beliefs and behaviors are important. They influence self-concept and attitudes of students as well. Irvine (1990) notes that students identify teachers as significant others in their lives, and how a child feels about himself or herself may be — to a large extent — based on how the child perceives the teacher feels about him or her. Many children who believe their teacher does not like them, in turn, do not like themselves or school and eventually fail academically. This effect is exaggerated for low-income and minority students because they are more teacher-dependent and are more likely to hold the teacher in high esteem.

Teacher expectations of students' performance are also related to students' academic achievement and are mediated by factors such as the characteristics of the teacher and the students. In content areas, and especially in mathematics and science, achievement among minority students can be greatly inhibited by teachers' low expectations simply because of the small numbers of minorities who are currently in the those fields.

Using Multiple Means of Assessment in Mathematics

In the Assessment Standards for School Mathematics, the National Council of Teachers of Mathematics (NCTM) (1995) encourages the use of multiple forms of assessment for both short-term and long-term instructional planning. Multiple forms of assessment provide evidence about student learning that may be difficult to capture by administering singular formats. Observations and questioning, for example, offer opportunities for understanding the influences of students' unique prior experience. Furthermore, making valid inferences about students' learning requires familiarity with every student's response in a variety of modes, such as talking, writing, graphing or illustrating in a variety of contexts. The NCTM also notes that "culture considerations

are also important; however, care should be taken not to make assumptions based on cultural stereotypes, because each student has unique responses to experiences in and out of school" (p. 52). The recognition by the mathematics community of alternative ways to assess what students know and can do, is indeed a step toward accommodating cultural differences, where appropriate. Thus, in addition to the use of paper-and-pencil assessment tasks, the NCTM recommends the use of open-ended items, student-constructed tests, performance tasks, investigations and projects, interviews, portfolios and self-assessment in mathematics instruction and assessment. Finally, the expanded use of alternative assessments, including the use of performance assessments in the classroom, must be accompanied by appropriate training and professional development for classroom teachers and university faculty (Johnson et al. 1998).

Role of Evaluators in Mathematics and Science Projects

Clearly, the charge for minority evaluators is to go beyond the obvious. While there may be gaps in the academic achievement of minority and non-minority students, it is not enough to leave such gaps unexplained. By paying attention to the factors that are correlated with minority student achievement, and designing and implementing evaluation models that measure these factors, evaluators can inform the instructional and assessment practices that aim to improve student achievement in general, and minority student achievement in particular.

References

- Boykin, A. W. (1994). Afrocultural expressions and its implications for schooling. In E. R. Hollins, J. E. King, & W. C. Hayman (Eds.), <u>Teaching diverse populations: Formulating a knowledge base</u> (pp. 225-273). Albany, NY: State University of New York.

- Davis, A. (1948). <u>Social class influences upon learning</u>. Boston: Harvard University Press.

- Delpit, L. (1988). The silenced dialogue: Power and pedagogy in educating other people's children. <u>Harvard Educational Review</u>, <u>58</u>, 280-298.

- Giroux H. & McLaren, P. (1986). Teacher education and the politics of engagement: The case for democratic schooling. <u>Harvard Educational Review</u>, <u>56</u>, 213-238.

- Gordon, E. W. (1996). Toward an equitable system of educational assessment. <u>Journal of Negro Education</u>, <u>64(3)</u>, 360-372.

- Irvine, J. (1990). <u>Black students and school failure: Polices, practices, and prescriptions.</u> New York: Praeger.

- Jenck, C. & Phillips M. (1998). <u>The Black-White test score gap</u>. Washington, D.C.: Brookings Institution.

- Johnson, S. (1979). <u>The measurement mystique: Issues in selection for professional schools and employment</u>. Washington, D.C.: Howard University, Institute for the Study of Educational Policy.

- Johnson, S., Thompson, S., Wallace, M., Hughes, G., Manswell-Butty, J. (1998). How teachers and university faculty perceive the need for and importance of professional development in performanced-based assessment. <u>Journal of Negro Education</u>, <u>67(3)</u>, 197-210.

- Joint Committee on Standards for Educational Evaluation. (1981). <u>Standards for Evaluation of Educational Programs, Projects, and Materials</u>. New York: NY: McGraw-Hill.

- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. <u>American Educational Research Journal</u>, <u>35</u>, 465-491.

- National Council of Teachers of Mathematics (1995). <u>Assessment Standards for School Mathematics</u>. Reston, VA: Author.

- Snow, R. E. & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), <u>Educational measurement</u> (pp. 263-331). Phoenix, AZ: The Oryx Press and American Council on Education.

Presentation

Assessing Underrepresented Science and Mathematics Students: Issues and Myths
Carlos Rodríguez

Good Morning. Thank you for the opportunity to be here. I have prepared my remarks to focus on three areas related to the question we are to address today. What are the issues surrounding the evaluation of science and mathematics achievement, especially the academic assessment of underrepresented populations? I understand that the relevancy of this question to you and the purpose of this meeting are in the context of science and mathematics program evaluation and the training of science and mathematics program evaluators. First, I will offer you some contextual considerations; secondly I will dispel some common myths about assessments and minority students; and thirdly, I will articulate some principles that should guide evaluation practices that I hope you take with you as tools in the important work you do.

Contextual Issues

"Many in our society and its educational institutions seem to have lost sight of the basic purposes of schooling, and of the high expectations and disciplined effort needed to attain them." High standards, high stakes testing and even program evaluations rarely talk about high expectations and disciplined effort.

What I have just presented to you is paraphrased from the introduction to *A Nation at Risk* written 17 years ago in 1983. Almost two decades have passed and these words are as true today as they were then. If minorities are quickly becoming the majority, does this mean that we should expect even greater failure rates, even less success? We answer this question with a resounding "NO."

Success, access and opportunity, however, do not happen accidentally for most people. They happen deliberately, and unfortunately, slowly. What we do today in program evaluation is that much more important because we are setting the pace, envisioning the future, changing the present. There is a saying in Spanish: "*El que adelante no va, atraz se queda*" ("Who doesn't go forward, stays behind.")

Many people in the United States today do not possess the minimum levels of know-how in literacy and numeracy, and training essential to this emerging millennium— and they are mostly Latinos, blacks, and poor whites. In fact, with Latinos this is true for almost half of the total U.S.–born population, and slightly less so for African Americans. Many then, of these groups, our brothers and sisters, will effectively be left out, not simply from the material rewards that accompany competent performance, but also from the chance to participate fully in our democratic society. These levels of educational achievement are inadequate when 80% of all new jobs this century will require at least some post-secondary education or training (Carnavale, 1999).

Without enumerating all the data, you should know as well as I do that, with critical masses of our minority students, we are raising a new generation of Americans of color that is scientifically and technologically illiterate. There is a real world of techno haves and have-nots. With program evaluations, we cannot be content to simply continue to represent the gap.

Program evaluators have the opportunity to close the gap by identifying the missing pieces in program interventions — the missing links, if you will. Program evaluators have the opportunity to build the bridges among policy, practice and research.

Schools may be emphasizing such basics as reading and computation at the expense of other essential skills such as comprehension, communication, analysis, problem solving and drawing conclusions. Do you know what the standards-based curriculum really contains in your state, school district, schools and classrooms? Do you know how well your testing programs are aligned with these standards and the enacted curriculum (what's really going on behind closed doors)? How well do you know if the standards are really changing how the poor and minorities are taught?

If you are not engaged in answering these kinds of questions in your evaluation efforts, you have a lot of work to do. Or have standards and frameworks and high stakes testing simply become yet another excuse for telling minority students how deficient they are? The training of program evaluators in science and mathematics must include deep understanding of contextual issues, and curriculum, and pedagogy, and assessments — both the promises and the limitations of each of these things. No mean task. But remember, comprehensive problems require comprehensive solutions. Comprehensive solutions can be built from comprehensive program evaluation approaches. Piecemeal solutions and Band-Aid approaches yield piecemeal and Band-Aid solutions.

If, for example, as we evaluate the loss of our black and Latino students from science and mathematics programs funded by NSF, NASA, the JPL, and the like, we stay at the level that reiterates the deleterious effects of resource-poor educational opportunities, without condemning the insistence of the state or district to perpetuate same, we relegate our greatest resource, our students, to shame. If our program evaluations produce yet more blame-the-victim litanies about minority students' bleak performance in mathematics and science, without also examining the willingness of their interveners to correct same, we do nothing more than subscribe to myopic views.

Let me turn to mathematics and science, then, for a few more minutes. We know that the vast majority of elementary school teachers are not prepared to teach mathematics and science. In fact, we can observe that the elementary curriculum is primarily organized around the language arts. At the middle school level, mathematics and science courses introduce students to inductive and deductive reasoning and provide them with experience in the practice of science. Minority students who do not have full access to these courses are unlikely to do well in the high school mathematics and science courses that are prerequisites for college entry. Hispanic eighth graders are more likely than non-Hispanic whites to be taking no science courses, and Hispanic and black students are more likely than any other group of students to be taking remedial mathematics or English courses. Most Hispanic high school students lag substantially (four years) behind non-Hispanic whites in science and mathematics proficiency, and black students are not far behind. Hispanic and black students are often denied access, that is, "tracked out" of regular science and mathematics courses. Minority students also are generally "tracked into" non-college preparatory courses. Today, black and Hispanic students are receiving, on average, more vocational course credits than academic credits, and are less likely to take algebra, geometry, and science courses — the minimum requirements for college admission. Did you know that only six states require algebra 1 of all students for high school graduation and none require

geometry? We need to get serious about high standards and make school boards and state departments of education accountable to provide the level of resources required for high standards of learning. We continue to invest in schools and teachers like gambling — we place small bets and expect huge winnings. Yet, we seem not to have any difficulty in finding enough money to barrage students with tests. Can you guess which groups of students are getting the better results with educational reforms based on high standards and high stakes testing?

So we see a rather easy-to-identify trajectory for the poor performance of most minority students in mathematics and science through high school. Even in high school, we know we have a cadre of influential and entrenched mathematics and science teachers who do not believe that the ability to learn mathematics and science is ubiquitous, but is really reserved for a select few. We organize learning in school as if intelligence is not normally distributed.

Dispelling Misconceptions and Myths

Program evaluators of science and mathematics often need to integrate some measurement data of academic performance, i.e., assessment and test scores.

So let's get to the misconceptions and myths about testing and assessment. This information should be axiomatic to the competent program evaluator of minority student academic achievement in science and mathematics.

Misconceptions (Change et al., 1999) that relate to the dynamics of race, culture and language affect student learning and the organization of learning and opportunities to learn:

Misconception One

Past inequalities in access and opportunities that racial and ethnic groups have suffered have been sufficiently addressed and no longer require attention. If this were true, most of this meeting would be unnecessary and the millions of dollars NSF invests in URM (Underrepresented Minority) programs a fraud. The fact is that low-income and minority children have significantly poorer access to quality schooling experiences, are concentrated in resource-poor schools and, due to persistent tracking and ability grouping modes, are usually found in the lowest groups.

Misconception Two

Merit can be defined by test scores. There appears to be a fairly common public notion that there are ways of measuring merit that are fairly precise and scientific, which may be true psychometrically. However, while tests may be shown to be statistically sound, policies based on such narrow definitions of merit inevitably exclude some students. The factors that determine merit and capacity for success — a mixture of ability, performance, talent and motivation — are not measured by standardized tests. The misuse of test scores beyond which they have been validated has had a systematic adverse effect on minority students.

Misconception Three

Fairness is best achieved through race neutral policy. This misconception contends that all individuals, regardless of race or ethnicity, should be judged on the same established criteria of competence, which are considered objective. Using the same standards, however, to judge individuals from majority and minority groups is unfair because differences in opportunities to learn prevent groups from having equal opportunity.

Let me turn to some common myths (Coleman, 1999) that also should be axiomatic to program evaluators when making the case about minority student academic achievement in science and mathematics:

Myth One: The Goals of Excellence and Equity Are Irreconcilable

Excellence and equity are not irreconcilable. They are neither intrinsically nor necessarily competing theories. Only when excellence is defined as flowing from some idealized notions of level playing fields, objectified neutral knowledge and meritocracy are excellence and equity irreconcilable. The educational foundations that guide policies promoting excellence can be — and should be — fully aligned with the promotion of equal opportunity for all students.

Myth Two: Test Scores, Alone, Tell the Whole Story

The value that test results can provide when making educational decisions about students does not mean that test scores should, as a matter of good educational practice, trump the need for thoughtful educational decision-making. A test's value as an educational tool is dependent upon its design, the context in which the test is administered and the ultimate uses of the test. Even when a test is used for purposes consistent with its design, a test is one tool among many. Just as tests are not perfect barometers of learning, conclusions based on those test results are not always error free. Many variables can affect a student's test performance, including: the quality of the student's education; the student's skill, ability or knowledge about a particular topic; preparation for the test; or what the student ate for breakfast on the day the test was administered. Does this mean that we should do away with tests? No. What it does suggest is precisely what test measurement standards affirm: the importance of considering multiple and educationally appropriate measures when making life-defining decisions about students. In 1985, the American Psychological Association (APA) Standards for Educational and Psychological Testing reminded us, for instance: "In elementary and secondary education, a decisionÉthat will have a major impact on a test taker should not automatically be made on the basis of a single test score" (APA Standard 8.12).

Conclusion to Misconceptions and Myths

Ultimately, good educational practices highlight the importance of considering objective measures such as tests in appropriate ways when making decisions about students. Not all assessments and tests are created equal, and tests should be used in ways that are valid for the particular purpose for which they are used. We must guard against cookie-cutter approaches to learning and assessment. Human beings are much

too complex to be reduced to learning and assessment approaches that assume that everyone, for example, learns how to read and solve problems and perform calculations in the same way.

Let me also sum up the challenges to school reform, including the reform of assessment: As states, districts, and schools use elements of standards-based pedagogical and curricular reform as well as standards-based assessment reform to enhance education capacity, they will face several continuing challenges.

The most critical challenge is to place learning at the center of all reform, assessment and program evaluation efforts — not just improved learning for students, but also for the system as a whole and for those who work in it. If the improvement of the learning experience for URMs is not a primary goal of the assessment practices and program evaluations in science and mathematics, they are not worth doing. If this improvement is not the tangible, realized goal of assessment and evaluation data, what end does it serve? For if the test givers — i.e., funders, teachers, administrators — are not themselves learning how to improve learning through assessment and program evaluation, and if the system does not continually learn from practice, then there appears to be little hope of significantly improving opportunities for all our youth to achieve to the new standards (CPRE 1995). This is what should be the grist of effective program evaluation — what the evaluator can say about the observed linkages between input and outcome with due consideration for the contextual considerations, while capturing all the nuances of unanticipated outcomes.

In my opinion, program evaluations need to contribute to the development of coherent and strategic approaches to capacity-building that take into account the needs and goals of the individual learner, school, district and state, not just for the immediate initiative, but for the long term. Resources are obviously a critical aspect of organizational capacity. A key target in addressing resource needs will be expanding available time to school personnel — time for teachers to collaborate in planning and assessing their instruction; time for teachers and administrators to participate in learning opportunities outside the school; and time for reforms to mature without falling prey to policymakers' readiness to halt reform if student test scores do not rise immediately. Allowing schools and districts to reconfigure schedules to provide time for teacher collaboration and learning is possibly the most cost-effective means of providing at least some of the additional time required. Finally, teachers and program evaluators need time together to reflect on the diagnostic capacity of assessments and the program evaluation itself to improve science and mathematics achievement among minority students.

Principles to Guide Evaluation Practices

In 1994, the American Evaluation Association adopted the following principles. The order of these principles does not imply priority among them; priority will vary by situation and evaluator role.

Systematic Inquiry: Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

Competence: Evaluators provide competent performance to stakeholders.

Integrity/Honesty: Evaluators ensure the honesty and integrity of the entire evaluation process.

<u>Respect for People</u>: Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients and other stakeholders with whom they interact.

<u>Responsibilities for General and Public Welfare</u>: Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Let me warn you, however, about certain seductions. Evaluators and educators must be very careful not to assume or be seduced into thinking that if program evaluations, policy and research and practice are perfectly linked, school improvement and advancement of minority students in science, mathematics or any subject will result as logical occurrences. There are insidious notions in popularized ideas about the very ability of non-mainstream students to learn that appear ready to sabotage and de-rail any progress that program evaluations, policy, research and practice put forth.

All are susceptible to manipulation. This is the "Who benefits?" question. Allington and Woodside-Jiron in the November 1999 *Educational Researcher* provide compelling and disturbing evidence, for example, for major policy manipulation at the state level that is currently on-going and is targeted at implementing a more "code-oriented" or phonics emphasis curriculum framework for reading instruction in the states, indeed nationally. In this analysis, they uncover the masquerades of research in the form of "expert opinion" to promote a one-size-fits-all approach to developmental reading instruction. They cite Benviste's work, The Politics of Expertise. In this work, Benviste indicates that through the political use of expertise, policy advocates consolidate a monopolistic position by promoting the appearance of an external professional consensus on a policy issue, often achieved by using highly selective research teams whose advice may not be easily dismissed.

Another example may be given drawing on *Issues in Education Research* (1999) edited by Ellen Condliffe Lagemann and Lee Shulman. Theodore Mitchell and Analee Haro in their chapter "Poles Apart: Reconciling the Dichotomies in Education Research"acknowledge that scholarly efforts "to put research knowledge into practice, working collaboratively and in mediated fashion, have been extraordinarily powerful, engaging teachers, researchers, and parents in discussions of educational aims and means and developing a sense of common purpose around the task of educating children." They warn, however, that "effective practices" often become cookie-cutter formulas for success, rather than tools for continued improvement. They posit this as a consumption problem — of a national appetite for solutions to the frustrating and complex problems facing education. A hunger that includes voters, parents, some school professionals, university presidents and, increasingly, funders of education inquiry and practice. Thus are fed the reductionist approaches that seek to take carefully crafted interventions and wholesale them. This is very appealing and seemingly cost-effective, and seductive.

Certainly it can be hypothesized that program evaluations, policy, practice and research can be aligned, and examples given. But we must be diligent at unmasking and respecting the discreet elements within each domain that may really be at play. The political underpinnings of program evaluation, policy, research and practice are real, yet not often made explicit or articulated. Political underpinnings refer to values, belief systems, power arrangements and divisions of labor.

Margaret Barrego Brainard reminds us that the word assessment is derived from a Latin verb *assidere*, which mean to sit beside. Evaluate comes from the Latin valoram, which means to place a value on something or to ascertain the value of something. Neither assess nor evaluate means to sit on top of or to hold back or to judge. They suggest that in order to reveal what a student really knows, it is necessary to be close to them, perhaps even moving alongside them on a path of learning. It means that we are challenged to see all our students, the teachers who teach them and those who organize their learning environment, the faculty that research them, and, yes, especially our minority students — we are challenged to see them in new ways, especially in ways that stop telling us that they are unteachable; in ways that convince us that each of us can contribute to increasing the degree to which our schools and our evaluations of the programs to enhance student learning can serve the full range of diversity that our students represent.

Where is the voice today of program evaluation surrounding the evaluation of science/ mathematics achievement, especially the academic assessment of underrepresented populations, outside this conference? Outside this conference, it is in the voice of high standards and high stakes testing in the public discourse. Which, (at the risk of sounding repetitive) if they do not provide diagnostic information upon which to improve student's learning, then they are but another face of blaming the victim. Blaming the victims because they attribute the causes of failure to deficiencies in the social, cultural and linguistic experiences of the students' themselves, and tangentially, if at all, to the organization of learning and the resources committed to it. In short, political and ideological processes, more often than not, may be determining the improvement of science and mathematics achievement among underrepresented minority students, and should not be ignored or denied by program evaluators.

Recommendations

The following questions and answers I recommend as suggestions for the thematic content of this conference:

- How do we determine if adequate integration exists among program evaluators, administrators, teachers and faculty and students on issues affecting SCIENCE AND MATHEMATICS learners?

  – Early and often.

  – Through collaborative inquiry between education evaluators and practitioners. By engaging teachers as researchers and evaluators in action research/program evaluation projects, for example.

  – By taking the long view. By establishing short-term accomplishable goals and placing them in a long-term continuum.

  – By remaining aware that programmatic and policy interventions almost always have political implications. At the site or grantee level, program evaluation may often require political strategies and political thinking.

- How is integration manifested?

  - Through deliberate discourse building.

  - Getting evaluators, researchers, practitioners and policymakers in the same room and talking through the process and building consensus on the desired outcomes.

  - When people know they are doing it.

- What barriers exist to integration among program evaluation, practice, research and policy?

  - Contradictory political values.

  - The lack of willingness to engage in comprehensive approaches.

  - Ignorance.

- How can diverse stakeholders advocate for minority students in SCIENCE AND MATHEMATICS in ways that will advance improved achievement?

  - Early and often.

  - Change the discourse.

  - Influence the alignment of assessments to what is actually taught. Require the use of multiple measures of students' abilities to inform instructional decisions.

  - Dispel the myth of a single American "mainstream." As my colleague at ETS, Tony Carnavale, so succinctly states: The notion of a single American culture is inconsistent with American history and ignores the realities of the development of individual and group identities in the modern world.

  - Integrate into program evaluations the rich natural resources of linguistic and cultural diversity in our communities.

  - Maintain comprehensive, holistic, ways of seeing the effects of program interventions on minority student achievement in science and mathematics.

Summary

The promotion of program evaluation of science and mathematics performance among underrepresented minorities to advance academic achievement provides one of the few tools we currently have to promote inclusionary values that foster learning success in schools and education writ large. Inclusionary values honor, respect and give dignity to the innate wonder and beauty and promise of each and every child, of each and every student. There is nothing more worth doing.

References

- American Psychological Association (1985). *Standards for Educational and Psychological Testing.* Washington, DC: APA

- Carnavale, Anthony P. (1999). *Education = Success: Empowering Hispanic Youth and Adults.* Washington, DC: ETC/HACU.

- Change, Mitchell; Witt-Sands, Daria; Jones, James; and Hakuta, Kenji (Eds.) (1999). *The Dynamics of Race in Higher Education.* Stanford, CA: AERA, Center for Comparative Studies in Race and Ethnicity.

- Coleman, Arthur E. (1999). Public briefing.

- CPRE (1995). *Policy Brief: Building Capacity for Education Reform.*