Journal of Negro Education

Human Diversity, Assessment in Education and the Achievement of Excellence and Equity

A. Wade Boykin   Howard University

*In recent years, large scale educational assessments coupled to more localized formative assessments have served as key drivers for educational reform. However, without other crucial considerations taken into account, current assessment practices can easily serve an exclusionary function for individuals whose experiences are construed as outside the mainstream of our society. This wastes crucial human talent. To counter this, more focus should be put on assessment **for** learning rather than assessment **of** learning; and on assessment of the learning context and not just assessment of students. Moreover, educational assessments should be coupled with a schooling purpose that emphasizes more human capacity building rather than sorting and selecting. The thrust here is that it is a societal good to foster extensive, high level knowledge, skills, and abilities in intellectual, technical, and civic participation domains, for successive cohorts of the American population. And in turn, assessments should function principally to help actualize such human capital production.*

INTRODUCTION

This article will address several topics in reaching its conclusion. Initially, it will be important to provide the background that assessment is a broad-ranging conception, which encompasses a host of differing practices, and has served several different functions in our society at large, and in the field of education in particular. Most strikingly today, we witness an especial stress in education placed on what is called *high-stakes testing*, a method of assessment that speaks to the role that large-scale educational assessments currently play in ascertaining the quality of student learning, and in turn, the quality of the education afforded to students. But most importantly, the obtained results have major implications for those tested and hold consequences for schools and their educational stakeholders as well.

It will then place the practice of assessment in historical context and show that, while assessment historically was conceived to provide greater opportunities for persons from diverse backgrounds, paradoxically, in terms of present day understanding of diversity in social backgrounds, assessment practices can easily serve an exclusionary purpose for individuals whose experiences are construed as outside the mainstream of our society.

The study documents what is largely known, that certain demographics and racial/ethnic minority groups fare relatively poorly on virtually all educational assessment measures, but that the obtained pattern of findings defy conventional explanations. Therefore the remedies that have typically been proffered for such low performance need to be thoughtfully reconsidered. In light of the need for explanatory reconsiderations, given that high-stakes testing is now used in education as a key driver for increasing school and student achievement, and more broadly for pursuing school reform, important questions should be posed and addressed. The discussion will then lead to raising and then answering the following questions: "Can we test our way to greater opportunities for diverse students and to promote school reform?" "Should we?" A case will also be made for focusing not just on assessment *of* learning but assessment *for* learning. Additionally, arguments will be made for expanding the reach of assessments to include not just assessments of students, but assessments of educational contexts; and a case will be made for expanding issues of validity to include matters of consequence and interpretation of assessment results. A case will also be made for converging educational assessments with a redirected purpose of formal education in our society to place greater

emphasis on human capacity-building rather than sorting and selecting. The article will conclude with a culminating summary of the arguments and issues that have been raised, and provide a framework for more proactively addressing issues of race, culture, excellence, equity and assessment in the American social order.

MULTIPLE FUNCTIONS OF ASSESSMENT AND THE CURRENT PREOCCUPATION IN EDUCATION WITH HIGH-STAKES TESTING

Over the course of the 20th century, and until the present day, assessment has played a major role in American schooling, and even in how our society has come to understand human ability and capacity. There is assessment for discerning competence or qualifications; for purposes of selection; for sorting or screening of candidates, and for certification (Gipps, 1999; Madaus, Raczek, & Clarke, 1997). Gipps argued that assessment has served the purpose of "controlling" examinees' access to further levels of education or professional positions (Gipps, 1999). Assessment also serves a diagnostic function, as well as an academic placement function. In recent years assessment has been a tool of accountability in educational settings, and has been used in the service of school reform (Linn, 2000; Shepard, 2000). In this regard, it has gotten increasing attention as an index of student learning. It is argued that assessments used in this way are controlling what is taught, and even how it is taught, and are placing accountability parameters on those professionals who are responsible for student learning. Related to this is the current emphasis in assessment with preparing students for college and the demands of the 21st century labor market. The latter point is now looming large as another major reason for the importance of national high-stakes assessments (Phillips & Wong, 2010). Assessment is now used as the basis for discerning comparative performances of students, classrooms, teachers, districts, states, and even nations (Madaus, Raczek, & Clarke, 1997; National Research Council, 2001). Shepard (2000) observed that high-stakes assessments have had an impact on routine classroom practices, which now emulate what transpires during standardized testing, and reflect the preferences of policymakers and the public at large for evidence of school and schooling quality to be based on standardization and quantification. Moreover, recent attention has been given not just to assessment of learning but assessment for learning as well (Stiggins, 2006). Assessment for learning involves classroom-based assessments in which greater attention should be given.

It may not be surprising that some would argue assessments have served a hegemonic function with regard to extant educational practice. As Connell (1993) has commented,

> In Western school systems, and Western-style school systems elsewhere in the world, a particular assessment regime is hegemonic. This means both that it is culturally dominant, connected with the society's central structure of power; and that it functions to maintain the social power and prestige of dominant groups (p. 75).

However, in more recent years, others asserted that assessment should play an equity function. For example, in 2001, the authors of a National Academy of Sciences report on student assessment asserted that large-scale assessments should principally serve the purpose of equity so that greater numbers of children will be more successful in school to learn and retain greater amounts of information and knowledge (NCR, 2001). Assessments they said, especially in standardized forms, must be better able to chronicle students' "accomplishments" and the strides they are making in terms of learning. Therefore, the influence of psychometrics is most apparent in the widespread use of standardized tests to monitor the progress of individuals and groups and to support consequential decisions about them (Heubert & Hauser, 1999). What is gained through such a process is prioritized over other forms of evidence (Porter, 1995). It has been observed that this increased preoccupation with assessment as a vehicle for gauging academic progress in our educational activities and institutions was initially fueled by the influential *A Nation at Risk* (1983) report. Presently, there is less concern about whether we should assess and more on what and how we should assess. As educational accountability has increased over the years and decades, so has the focus on testing and assessment. Therefore, an increased emphasis has been placed on the assessment of learning.

THE PARADOX OF THE HISTORICAL FUNCTIONS OF ASSESSMENT WITH PRESENT SOCIOCULTURAL REALITIES OF AMERICAN SOCIETY

Testing and assessment have a very long history in human societies. The Chinese developed a civil service examination as far back as the third century BC, so that "men of merit" could be selected for the work to be accomplished and minimize the use of patronage for job selection (Madaus, Raczek, & Clarke, 1997). In Europe, similar efforts emerged across the 17th to the 19th centuries for similar reasons and expressly to mitigate the influences of family history and birth, wealth, and corruption in gaining access to government positions and other middle class professions, including the medical profession. In America, in the 20th century, the College Board's Scholastic Aptitude Test (SAT) examination allowed for entry into prestigious colleges for students who did not come from privileged backgrounds or presumably favored ethnic groups. Consequently, such assessments over the course of time allowed for access to opportunities for individuals from diverse backgrounds (Gipps, 1999). From one angle, these assessments did pursue matters of equity across diverse social groups as a purpose for their administration in terms of potentially providing positive consequences for "diverse" test takers. Nevertheless, regardless of intentions, the exams still served sorting, selection, and certification functions.

Over the latter half of the 20th century, American society has been embroiled in a protracted reshaping ushered in by the struggle to protect and promote the social and civil rights of designated minority populations. Among the consequences has been a push for *affirmative action*, a push for the reduction, the elimination of discrimination based on social status, demography, ethnicity, race and gender, against the backdrop of the awareness that America is becoming an increasingly multiracial and multicultural society. In the wake of such developments, issues of equity and fairness in assessment have gained more pronounced importance. This then is a social justice approach. It is not seen as sufficient that groups or individuals from differing backgrounds have the opportunity to take a given assessment, but the results from these examinations should also reflect equity and fairness, in terms of the distribution of the obtained outcomes. When the obtained results are not evenly distributed, this for many raises the specter of discrimination inherent in the assessments themselves. The argument gains strength when the results yield relatively lower performance from certain groups who have been historically disenfranchised in the American social order. Therefore, whatever is the context for eventual participation, if for example, proportionately fewer people from certain identified groups are not selected for participation in those settings, then the tests are vulnerable to being seen as unfair. Equity is not obtained if the results do not lead to greater (or at least equal) access to the educational, professional, or status opportunities or to at least equal performance outcomes for socially marginalized groups within the realities of a multiethnic, multiracial, and multicultural society, where status and opportunities are arrayed along ethnic, racial, and cultural lines. It can be argued that wealth and social status of one's background is correlated with successful assessment outcomes. Such a reality certainly captures 21st century American society. The selection, sorting and certifying roles for assessments still remain a chief function today, even while all groups now, regardless of backgrounds, can and will be assessed under certain circumstances. Additionally, in some circles the differential outcomes for varying groups justify the lack of access to certain valued opportunities in our society, or the basis for less desirable placements.

To be sure, the pursuit of sorting does not inevitably have to be attached to sinister intentions, or even lead to achievement-deflating results. Diagnostic assessment, at least in theory, can serve to determine the strengths and weaknesses in students' present ability to accomplish a task or display what they have learned. The resulting categorization allows educators to then shore up areas in need of improvement, or gives attention to the fortification of skill sets where the particular deficiencies currently exist. Such sorting could also lead to the provision of enrichment or acceleration activities if this seems called for. Of course, it is crucial that whatever prescriptions for students that ensue should be predicated on fidelity in the initial assessment results. That is, the results obtained for students should accurately reflect what quality of understanding or skill a student actually possesses.

In light of arguments presented elsewhere in this article, such fidelity is not always guaranteed. Moreover, when the sorting function leads to the relatively permanent placement of students into categories that are differentiated with respect to the quality and level of education that is afforded to them, then this is where sorting itself may not serve the interests of the schooling process to actualize equity and excellence in education for all students, especially for those from racially and culturally diverse backgrounds. To the extent that present-day diagnostic assessments do not take into account context, social-cultural considerations, and other experiential factors linked to students' diverse backgrounds, they may not be fine-grained enough, or suitably structured or conceptualized to address equity and excellence concerns. These matters require further exploration and expansion, and will be addressed more substantially later.

The focus of this article, however, is more squarely on educational *assessments* more broadly considered. The notion of assessment *per se* as a term in educational parlance has had a notable history of its own, apart from the notion of "testing." The *Oxford English Dictionary* reported that the origin of the word "assessment" dates back to 1626 and is used with reference to estimation or evaluation. The term came into educational parlance in 1956 and was offered as an alternative to "examination" (Madaus, Raczek, & Clarke, 1997).

It is also the case that the term assessment was coined originally in educational circles primarily to capture what occurred through widespread group administration. Unsurprisingly, assessments were geared toward being efficient, manageable, standardized, easily administered, objective, reliable, comparable, and inexpensive. In this way they mimicked the testing purposes and formats that arose earlier in the 20th century with respect to intelligence testing and "aptitude" testing such as the Scholastic Aptitude Test (now known simply as the SAT), which were arranged in multiple choice formats in the service of efficiency of administration, uniformity of design, and objectivity and reliability of scoring.

The term assessment came into wider use in the late 1960s with the advent of the National Assessment of Educational Progress (NAEP) program in the United States. The notion was tied to the triangular links between teaching, learning, and assessment (Madaus, Raczek, & Clarke, 1997). Assessment presumably was a means to discern how much has been learned or needs to be learned. Madaus and colleagues (1997) among others (Linn, 2001; Shepard, 2000), argued that those who "control" assessments have had power over the curriculum; what is taught, how it is taught, and how and what is learned.

Such a focus was also connected to the pervasive implementation of ability grouping beginning in the mid-1950s, which was seen as an innovation crucial to getting the most out of students with different levels of "talent," and coupled with the premium placed on objectivity and efficiency in testing, it was not surprising that the focus of testing was clearly on outcomes and not the processes that produced these outcomes (Clarke et al., 2000). This focus on testing and identifying talent was fueled by public dissatisfaction with educational quality in America during the late 1950s and early 1960s due to public consternation that the U.S. was falling behind the USSR in science and technology following the launch of Sputnik. Then in successive waves of public frustration with American schooling: there was the basic skills movement of the 1970s; and in the wake of the *A Nation at Risk* (1983) report in the 1980s and early 1990s, there were increasing calls for school reform, which was the centerpiece of the Goals 2000 initiative in the mid- to late-1990s; and now with present-day concerns about the various achievement gaps that plague American schooling and the specter of an unprepared future work force that resulted in the landmark legislation known as *No Child Left Behind Act of 2001* (2002). Through such developments over time, one has witnessed that high-stakes testing/assessment and concomitant accountability have been increasingly ushered into the shaping of schooling practices in American society.

In the current era of high-stakes testing, policymakers find assessment attractive as a vehicle for reform. Tests and assessments are comparatively cheaper than other approaches such as reducing class size, increasing teachers' salaries, increasing time for instruction, and greater professional development. It can be more readily externally mandated, especially when compared to altering what happens inside classrooms. It can be more easily and quickly implemented to fit within the political lifespan of elected officials. It is also more readily apparent to the public.

But one must ask in this era of increased accountability, how trustworthy is the assessment information? Moreover, how useful is the information gained in the service of equity and excellence combined? We must also consider that in the 1950s, prompted largely by the influential writings of James Conant, the search went out to find talented students. We should have "universal elementary education, comprehensive secondary education, and highly selective meritocratic higher education" (Cremin, 1989, p. 22).

Tests were then deployed to identify the gifted students for selective comprehensive high school programs, and selecting students for higher education. This approach reinforced and solidified a talent-sorting approach rather than talent development. This of course diverges greatly from the current preoccupation (putatively) with having all students reach high standards. Tracking remained the rule and not the exception well into the 21st century, and it can be argued remains as such in the U.S., and the quality of instruction is correlated with the level or type of one's track.

Although there is uniformity across assessments in chronicling the low performance of students from certain racial/ethnic groups in America, the patterns of findings are complex and defy easy straightforward explanations

Concerns about schooling in America very often have become squarely focused on the achievement gap(s) existing primarily between African American and Latino students and their White (and Asian) counterparts. Over the last several decades, there has been repeated documentation across a plethora of indexes for the existence of such gaps. Although there have been a host of concentrated efforts launched and the expenditure of substantial resources all aimed at the eradication of such disparities, by and large the gaps remain unabated.

These gaps have shown up in the disproportionate over-placement of Black and Latino students in programs for the learning and emotionally "disabled," and the under-placement of such students in programs for the talented and gifted. Evidence also abounds for greater disciplinary referrals and school suspensions for such students, and greater levels of school dropouts and failure to graduate from high school. But most prominently the achievement gap(s) between Black and Latino school children and youth, and their White (and Asian) counterparts has been understood in terms of performance on various measures of ability and achievement assessments. Achievement gaps have been obtained with children as young as three years of age (Burchinal et al., 2011), as indexed through a standardized test of "school readiness" skills. Gaps in math and reading achievement have been prominently obtained with other standardized educational assessment using the Woodcock Johnson Test Battery (Murnane et al., 2006). For example, gaps have been obtained in a longitudinal study using a national cohort sample when the children initially were four-year olds, and again for them in the first, third and fifth grades (Burchinal et al., 2011). These findings are comparable to those reported in a previous longitudinal study reported elsewhere using the Woodcock assessments (Murnane et al., 2006). Similarly, achievement gaps have been reported in investigations using the Early Childhood Longitudinal Study (ECLS-K) data set (Condron, 2009; Fryer & Levitt, 2004) as well, across essentially the same age range.

With comparatively older samples of students, marked achievement gaps in reading, writing and mathematics consistently have been produced for both the SAT and ACT "college admissions tests" (Luppescu et al., 2011; McKinsey et al., 2009). As a case in point, the SAT scores for the high school graduating class of 2011 showed marked disparities between the performance White and Asian American students on the one hand and those of African American and Latino students on the other. These same disparities were obtained for the reading, math, and writing subtests.

The assessed disparities have also showed up consistently across the K-12 spectrum with respect to statewide, "high-stakes" achievement tests (Luppescu et al., 2011). Results favoring the performance of White students have been consistently revealed over the last four decades from the administrations of the NAEP. NAEP has been popularly referred to as "the nation's report card," because it is the one test that has been consistently administered across the country since the late 1960s, and for fourth, eighth and twelfth graders. Although there was some narrowing in scores

among groups for a period in the 1980s, that then was reversed in the 1990s, for reading and math subtests, there have been consistent score disparities between White students and their Latino and African American peers that otherwise have not changed to the most recent test administration (Boykin & Noguera, 2011; Luppescu et al., 2011; McKinsey et al., 2009). These gaps continue to hold at the fourth-, eighth- and twelfth-grade levels.

While the focus has been on ethnic group differences in performance, data have been also parsed in other ways. One category of students that has received prominent attention in recent years has been English Language Learners (ELLs; Basterra, 2011). Work with such students has received attention in accounting for the existence of disparities in test performance between such students and their White, native English-speaking counterparts. For example, recent work in this regard has shed light on whether "mainstream" tests are culturally valid for ELLs.

Solano-Flores (2011) has asserted that differences in communication patterns, values, beliefs, and lived experiences may to a degree account for the comparatively lower test performance of ELLs. For example, better performance has been attained by ELLs with the same standardized achievement test items, when the items have been modified to reflect local dialect, or have been linguistically simplified, that is, using shorter sentences, active tense, concrete or abstract depictions, and so forth Similar findings have been obtained when the items are modified to be more experientially meaningful for such students (Solano-Flores, 2011).

Beyond this, closer analyses of recent NAEP data sets reinforce what is known at the national level, but also reveal issues worthy of further exploration. For example, results have been examined for Black and White fourth and eighth graders in several urban school districts in the country, and in categorized criterion levels. At both grade levels, White students do substantially better than their Black counterparts in districts such as Chicago, Cleveland, Los Angeles, New, York City, and Atlanta. Additionally, White students' scores across such districts are preponderantly located at the proficient and advanced levels of functioning, while Black students' scores placed them more frequently at the below basic and basic levels. But intriguingly, in the eighth grade, while White student scores are still substantially higher than those for Black students, the distribution of score levels for both Black and White students are markedly lower than in the fourth grade. Students in general are doing less well in middle school than in elementary school.

Beyond this, when examining the pattern of findings across the various assessments of interest, the explanations to account for the disparities defy easy and straightforward explanations. Hughes (2003) for example found that achievement test performance differences are found between third-grade Black and White students who live in a uniformly wealthy community and attend schools in a notably affluent, highly resourced school district, with pervasive, highly educated school personnel. Elsewhere, Nettles (2000) has reported differences between 100-150 scale points for Black and White students' SAT scores, differences that held up across all income levels.

Fine-grained scrutiny of the data from the NAEP administration (NAEP, 2009) provides further interpretation complications in accounting for the achievement score gaps. For these comparisons, average NAEP mathematics and reading scores were examined for eighth graders classified in terms of their race and their parents' highest level of education. It was found that the scale point disparity in the scores for Black and White students whose parents were college graduates was approximately twice as large as that obtained between Black and White students whose parents did not even finish high school. A similar picture was obtained for the score gaps for Latino students, except that the gap was approximately four times greater for students whose parents were college graduates relative to those whose parents were not high school graduates. At the twelfth-grade level the pattern of disparities was essentially the same, except the disparities were even greater between the scores of students whose parents were college graduates, relative to those for students whose parents did not finish high school. In all, these findings presented cannot simply be explained away in terms of differences in socioeconomic-status (SES) levels between White students and their Black and Latino counterparts. They cannot be readily explained away in terms of differential access to resources in the homes and in the communities. We must search for more nuanced, or process-based answers, or for answers that may be tied to what goes on inside the schools and classrooms differentially where White, Black and Latino students matriculate.

©The Journal of Negro Education, 2014, Vol. 83, No. 4

Some credence for looking in such directions is gained from further findings. The achievement gap in terms of test scores actually grows larger across the years in school (Fryer & Levitt, 2004; McKinsey et al., 2009; Murnane et al., 2006). In fact, this widening holds across the entire K-12 spectrum (Burchinal et al., 2011; McKinsey et al., 2009). Moreover, additional evidence points to the gap getting larger across a given academic school year (Barbarin, 2002; Condron, 2009). These matters will be taken up more substantially in subsequent sections.

Can we test our way to greater opportunities for diverse students and to promote school reform and greater achievement levels? Should we?

The answer to the first question is yes and no. The answer to the second question is that testing plays a part but not necessarily the only or not even the most important part for enhancing school and student achievement.

It was the case a decade ago and may be even more the case today that in the realm of educational assessments in the United States, much more importance and virtue is attached to large-scale assessments of students and educational programs for enhanced student and school achievement than to ones focused more so on classroom considerations (National Research Council, 2001). More financial resources, labor and time are devoted to research, development, and implementation of large-scale assessments than to fostering enhancements and the implementation of assessments geared directly to classroom-based factors. The trend has certainly been in that direction over the years.

We have gone from district-wide in the 1960s to statewide in the 1970s to nationwide in the 1980s and 1990s to international in the 1990s and early 21st century, with increasing numbers of grades tested during these years at the different levels as well (Stiggins 2002). An emphasis on large-scale assessments has merit. By definition the testing is widespread, therefore offering insight into the bigger picture of education. Such testing is cost- and time-efficient, provides for common yardsticks and for gleaning different grain sizes of information from the international, national, state, district, school, classroom, and even individual unit of analysis. The increased intensification of focus has been animated by avowedly good intentions: to increase achievement levels for all students regardless of their backgrounds. There is considerable merit in the pursuit of *equity* with respect to increasing achievement for all students, and to hold students and educators to high standards, thereby promoting *excellence* for all as well. It is notable also to ground this pursuit in systematic, empirical data, using indices characterized by psychometric soundness in order to provide arguably objective criteria for imposing education accountability to attain criterion performance levels across the various layers of the schooling enterprise. But time has proven that this tactic has shortcomings that must be rectified if in the future one will truly attain equity and excellence in education in this society.

Nonetheless, Madaus and colleagues (1997) argued, "while [large-scale] assessment can assist in reform efforts, the nation cannot assess, test, nor examine its way out of its educational problems" (p. 5). This observation still rings true today. At least, perhaps, one cannot do so with the current disproportionate preoccupation with large-scale educational assessments and with the almost exclusive linkage of the results from such assessments to the consequences of the high-stakes accountability agenda.

There are three problematic issues attendant to the current rendition of high-stakes, large-scale assessments. First, the accountability mechanism is flawed as a lever for positive school and student change. Second, there is a synchronization challenge associated with when the assessment results are made available and if they can substantially impact the changing of educational practices. Third, there is difficulty in knowing why the particular results occurred, what they mean, and what should be done as a consequence.

Regarding the first point, over the last several decades, three crucial elements in the nation's assessment agenda have emerged. They are standards, assessments, and accountability. The three have been inextricably linked, but not necessarily in the most effective ways. The focus has been on high academic standards, and therefore what academic success is defined as, which in turn must be

reached as indexed by assessment; and we hold educational stakeholders accountable for seeing that the assessment results demonstrate that the standards have been met or are at least are being increasingly approximated. Schools are "rewarded or sanctioned" based on how "successful" they are in this enterprise. All told, this set of processes speaks to *assessment of learning*. As Stiggins (2002) stated, "These tests tell us how much students have learned, whether the standards are being met, and whether educators have done their job they were hired to do" (p. 2).

However, it can also be argued that such assessments are being used in such manner for purposes of behavioral control, and for coercion to learn. It can be further argued that such purposes can undermine both the informational value of assessments, as well as the quality of student engagement in learning.

As Stiggins (2002) has posited, policymakers, test designers, and politicians are joined in their conviction that the way to improve schools and student achievement is through a process of coercion and pressure, aimed at impelling schools and teachers and even school systems to provide better instructional services to students. Moreover, test publishers are all too willing to oblige, because there is profit to be made in offering a product to serve such ends; accountability is important for them. Having consequential educational decisions based exclusively on high-stakes tests largely leaves unattended matters concerning the provision of daily and "moment-to-moment" feedback that students need in the course of learning at school. There is also relative silence on what opportunities should be implemented to better prepare stakeholders to meet the accountability expectations. Consequently, prevailing emphases on large-scale assessment conceivably are doing more harm than good when it comes to student learning, in spite of purportedly good intentions. Toward the goal of student and school improvement in the context of the need for accountability, we must have a more equitable balance between large-scale standardized assessments and classroom based, more instruction relevant assessments (Stiggins, 2002).

This conclusion leads naturally to the third critique posed. Large-scale assessments provide little insight into what the results actually mean, why the assessment results occurred, and what educational practices should ensue. Well-constructed large-scale assessments will have clear achievement targets. However, they will likely have ambiguity about what a test score means, and reveal little in the way of hypotheses of what the next instructional steps should be to optimize learning outcomes for the students (Stiggins, 2002). In this regard, Stigler and Hiebert's quote is telling; "a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve" (Black & Williams, 2010, p. 81). The increased focus on standards in assessment and teaching and learning certainly has its merits. Moreover, raising the bar for already disengaged and even disenfranchised students, and then testing to these raised standards exacerbates the achievement problem, rather than solving it. There is virtue in questions posed by Stiggins (2002): **"**how can we use assessment to help all of our students *want* to learn? How can we help them feel *able* to learn?" (p. 1).

One crucial recommendation of the National Research Council (2001) report cited is its call to focus less on assessment *of* learning and more on assessment *for* learning. This call has been greatly pursued over the years by Stiggins and his associates (Stiggins et al., 2004). They argued that assessment of learning is typically used to confirm student competence, sort students based on ability/achievement status; for grading/evaluation purposes; and extrinsic motivation is to be the engine that impels it (threat of punishment, promise of rewards). Assessment for learning is to be done on a daily basis. According to Stiggins (2002, 2006, Stiggins et al., 2004) assessment for learning entails teachers:

- setting achievement targets in advance of instruction (displaying good and bad examples of pertinent work);
- making these known in understandable ways to students; becoming adept at assessment so that accurate student portrayals are obtained;
- providing students with informative, nonjudgmental feedback (focusing on strengths as well as weaknesses); adjusting instruction accordingly;

- having students themselves become fluent in self-assessment (e.g., what are my goals, where am I now, what must I do to improve, am I making progress?), or as Stigler and Hiebert (2009) have labeled it assessment *while l*earning; and
- having students come to share with others the status of their achievement strivings.

It is crucial to distance assessment for learning from the very similar activities of benchmark assessment and formative assessment (Stiggins, 2002). In the former case, benchmark assessment is not substantially distinct from outcomes assessment insofar as the results from both are typically available at some time past when the actual relevant material was covered in the class; even though the time lapse is certainly greater for outcomes assessment. Thus, "in-time" feedback is not provided. It is also the case that evidence supporting the efficacy of benchmark assessments for raising achievement is not strong (Heritage, 2011). In the latter case, formative assessment typically is conducted for student learning diagnostic purposes *per se*. It is conceived to provide teachers with knowledge of the areas or skills for which students are in need of improvement. But the focus in such tests is usually on what students do not know or cannot do. This then leads to a remediation tactic for addressing the problems uncovered, and this typically entails re-teaching, or otherwise devoting more time to the student so that the student can do better. However, such an assessment approach, while well intentioned, typically fails to shed light on why the student has deficiencies in the area of focus, and especially not on what instructional changes the teacher should make to ensure that teaching will be more effective than before.

The takeaway from this section is that assessment for learning should be coupled with assessment while learning, and then linked explicitly to making classroom learning processes more transparent; and in turn these classroom-based practices should be synchronized with large-scale assessments of learning. Such an amalgam would position one better to test the way to greater schooling outcomes.

Nonetheless, additional factors must be entertained if we are to functionally connect such considerations to matters of diversity and greater excellence and equity. In this regard, we must also consider assessments of learning contexts. We must broaden the scope with respect to key forms of assessment validity, and we must have a more nuanced, robust, and textured understanding of how assessment should be done with due consideration to issues of diversity and culture, and how such assessment relates to the functions that schooling should serve in the American social order. These topics will be addressed in the remaining sections.

Greater emphasis should be given to assessment, not of students per se, *but of (academic) contexts*

Accountability is important in terms of decision-making and policy, but to what extent is the schooling process itself measured. As Oakes (1989) proclaimed in her seminal piece on this topic, we need to see if enabling conditions are present inside classrooms that "promote high-quality teaching and learning" (p. 186) for all students. This was an "underdeveloped measurement technology" when her paper was first published. It still is an underdeveloped technology. One simple way to capture the issue of assessment of context would be whether the material was actually covered in class. Polikoff (2010) has referred to this phenomenon as "instructional sensitivity." But more germane to the present concerns would be questions such as do students from diverse backgrounds actually have legitimate opportunities to learn? How would or do we know? As a case in point, if assessments are designed to test critical thinking skills, are such skills being taught and taught successfully so that discernible and pervasive increases in critical thinking skills are now evident inside classrooms? Mathews and Lowe (2011) picked up on this issue with respect to critical thinking. They defined critical thinking as 'the act of challenging ideas and considering alternatives based on developing valid and plausible premises through sound logic and reasoning" (p. 60). Critical thinking would, for example, be manifest when students look for similarities among things that seem different and differences among things that seem similar. They posited that critical thinking is undermined by, for example, stifling students' questions, creating authoritarian classrooms, prioritizing the reproduction of information as a desirable learning outcome, and presenting information that is

disconnected from students' lives outside of school. And beyond this, the press for assessment would not be just to come up with a checklist of "effective" teaching strategies to see if they are present or absent inside classrooms. It is also crucial to discern not only whether certain information was covered, but also if it was covered well, broadly, and if it was accurately learned, at sufficient dose level, with suitable cognitive load, and in ways that foster long-term retention, application, and knowledge transfer.

The examination of classroom environments has had a several decades-long history in the field of educational psychology. Such work dates back at least to the late 1960s and early 1970s (Moos, 1974; Wahlberg & Anderson, 1968), leading to the development of the Learning Environments Inventory (Wahlberg & Anderson, 1968) and the Social Climate Scales (Moos, 1974). Beginning in the 1980s, work in the domains of classroom environment assessment increased dramatically, owing to the influential collaborations of Barry Fraser and Daryl Fisher, and their colleagues (Fisher & Fraser, 1981; Fraser & Fisher, 1983; Fraser & Wahlberg, 1981) as additional measures were developed across the K-16 spectrum. Such assessments included the My Class Inventory, Classroom Environment Scale, Individualized Classroom Environment Questionnaire, and the Constructivist Learning Environment Survey, among others (Dorman, Aldridge, & Fraser, 2006; Fraser, 1998; Taylor, Fraser & White, 1994). The measures have taken different formats, including direct classroom observations, and teacher and student surveys. These measures presently are in widespread use around the world.

Over the years, this important and influential line of scholarly pursuit has not been substantially linked to matters of school reform in the United States, or to addressing the achievement gap; nor has it been put to widespread use to inform professional development for teachers and administrators aimed at the improvement of instruction. The earnest and sustained attempt to make such connections is perhaps well overdue.

It is crucial to appreciate that for a classroom context to be truly enabling, it must be one that fosters authentic, high-quality learning opportunities. If the classroom context is a toxic or disabling one, then the opportunities for effective learning will not be manifested. An enabling classroom context does not guarantee that a student will learn well and ultimately score highly on an outcomes assessment. But it is submitted that examples of "*assessed false negatives*" will be minimized. The notion of assessed false negatives is that a given student, who is fully capable of learning to high standards and, in turn, scoring highly on the assessment of achievement, will fall short in his or her assessment outcome and then be judged inappropriately as an inadequate learner.

Beyond this, an enabling classroom also should be gauged by the quality of transpersonal dynamics and affective tone of that classroom context. In this regard, it should be a "safe place" for students to learn (Heritage, 2011). Those who understand learning processes appreciate that, by definition, learning exposes a student's intellectual vulnerabilities. It requires them, sometimes publicly, to admit what they do not know or cannot presently do or do well. Teachers or peers cannot use this vulnerability as a weapon against a student. "It should not be cause to embarrass, ridicule, or demean a student's current understanding, or intellectual ability. But instead it should be appreciated even encouraged to be proclaimed, as sources for new learning" (Heritage, 2011, p. 19). A sense of mutual trust and encouragement must be fostered in this learning environment so that students feel comfortable asking for help, raising questions, announcing to others what they do not yet know. An enabling classroom is certainly not a site that functions principally to publicly separate the good from the bad learners, the smart from the not so smart students. It should not be a site for such sorting functions. It *should* be a site where it is widely understood that students' academic capacities will be built up, will be further developed, as increasing levels of learning for all students are encouraged, fostered, and celebrated (Boykin, 2000; Kelly & Turner, 2009; Strambler & Weinstein, 2010). If the environment is enabling, it would be supportive for effective learning and one should expect that greater, genuine student involvement in learning would transpire.

In recent years, involvement in classroom work has been increasingly captured in both the research literature and discourse on teacher practice, the concept of *engagement*. There is mounting evidence that student engagement in classroom learning activities is a key, even an essential ingredient in capturing the quality of classroom teaching, and can serve as an important proximal

gauge for quality of student learning, and a crucial marker for eventual performance level on standardized outcomes assessments (Bodovski & Farkas, 2007; Li & Lerner, 2011; Strambler & Weinstein, 2010). It is crucial to mention that engagement does not simply boil down to student time on task. Authentic engagement should be captured as progressive involvement with a given lesson that is leading toward increased learning and understanding of the lesson content (Engle & Conant, 2010). It entails active, positive, non-superficial, and progressive involvement in the learning process. In the influential work of Fredericks, Blumenfeld, and Paris (2004), engagement is conceived as a three dimensional phenomenon. There is behavioral engagement, which speaks to the level of student attention, persistence, and effort; affective engagement, which entails positive emotional absorption, value, and interest; and cognitive engagement, which denotes deep processing of information and critical/higher-order thinking. The more authentically and pervasively engaged are students, the more educationally enabling is the classroom environment (Borman & Overman, 2004; Kelly & Turner, 2009). It is strongly recommended that an indispensable aspect of the assessment of context should be the surveyed perceptions of students and teachers with regard to student engagement, and the real-time, third-party observational indexing of such engagement as well.

THE EMERGING IMPORTANCE OF CONSEQUENTIAL AND INTERPRETIVE VALIDITY

The arguments and stances presented in this article need to also be amplified through consideration of two forms of validity that receive little or no attention in the discussion of educational assessments. These are *consequential* and *interpretive validity.*

Consequential validity concerns the individual, social, or societal ramifications of the deployment of an assessment, which are deemed either harmful or beneficial. What is harmful or beneficial can likely be complex in its construal. To be sure, assessment for sorting purposes does not necessarily have to be problematic. For example, when it is done, it may be beneficial to yield accurate diagnosis of needed remedial support for students. But in the main, it is argued presently that when an assessment serves to sort students, or select the "best" students, this would likely be beneficial for those sorted into a favorable category, or selected for a favorable result. But this assessment process would likely be seen as producing a less favorable result for those weeded "out" or not selected into the valued grouping. This would be a negative consequence for such students at the individual or even social group level. However, by selecting the "best qualified" as indexed by the assessment, the societal consequence would likely seem to be beneficial to some observers. This line of argument could be taken to imply that there is little in the way of educational benefits. There certainly could or would be benefits, but the benefits could be more widely distributed.

Now suppose the goal that decision makers hold for an assessment is that it should lead to improvement in student or school achievement for all students. Also, suppose in the wake of the assessment implementation at targeted schools only a few students do improve, and overall the schools' scores do not meet expectations. Moreover, if the students who do not improve come most often from certain social (cultural, racial, low economic status) groups, then at this social level and the level of the individuals who belong to such groups, the consequences are particularly unfavorable as well. For example, the results from the assessment deployment could lead to lowering standards or expectations for such students. It could also be argued that students from certain social groups do not improve on the assessments because they are incapable of doing better or even as well as the assessment-favored group. The extension of this logic would often be that the assessment itself is not problematic, but teaching these assessment unfavored groups will not yield more positive results because of their innate deficiencies or because they are not sufficiently prepared to benefit from instruction due to their deficient out-of-school experiences. Otherwise, it could be concluded that the school itself is doing a poor job of educating its students, and the school will be subject to negative consequences, even its possible reconstitution. In any case, any of these scenarios represent negative consequences resulting from the assessment's deployment at the individual, social group, school and even societal levels. The "unintended consequences" of NCLB would seem to be in line with the various possibilities outlined.

But alternatively, some decision makers conceivably could subscribe to the sorting function for schooling and still deem that it is unjust to have unequal distribution of assessed "excellence" across social groups. In such cases, the assessments themselves could be judged unfair or at least ineffective, and the push is to allocate the distribution of assessed excellence more equally across groups. But this could lead to a backlash of sentiment that the revised assessments will have to compromise on attaining true excellence, and be unfair to those who deserve to be labeled excellent by virtue of their true measured ability level.

There is a growing chorus of scholars and educators who promote the position that schooling's purpose should more so be construed as promoting human capacity building, of human capital production, as indexed through pervasive improvement and attainment of an excellent criterion level for all students. For this aim, it is advocated that a system (not an individual assessment) of coordinated assessments, collectively serving multiple purposes, should be devised that places diversity in the forefront of the educating and testing nexus. This system would seek to discern what instructional vehicles can be deployed and what insights about diverse lived experiences and social identities can be brought to bear on fulfilling the pervasive human capacity-building agenda for schools. The aim here would be that assessments and schooling both must be changed, and in doing so, what is to be minimized would the negative consequential validity of the assessment enterprise.

Seen from this angle, a preponderance of current assessments work at cross purposes with such a capacity-building function for schooling, in that they likely serve to cloud, misrepresent or even fail to detect the ability, competence, or potential of many students from diverse backgrounds. Such would be *consequentially* problematic, and interpretationally insufficient or even inaccurate.

It can be further argued that such a consequence arises in large measure, from a lack of *interpretive validity* of these assessments. That is, what is inferred from the results of many conventional assessments may be faulty, incorrect, or misleading. This may arise for several reasons. The following quote should be kept in mind: ". . . by its very nature . . ., assessment is imprecise to some degree. Assessment results are only estimates of what a person knows and can do" (National Research Council, 2001, p. 2). Therefore, when one obtains an assessment score, it can prove difficult to discern if it is an estimate of one's ability, learning achievement potential, or teaching quality to which one has been exposed (Hickey & Zucker, 2005). Moreover, they argued that motivation to do well on tests is a major compromiser to how much results reflect competence, knowledge, or ability. Hickey & Zucker (2005) even posited that "knowledge ultimately resides in the context of its use" (p. 278). Furthermore, they offer that if one is not meaningfully participating in a given "knowledge community" (e.g., the classroom) it is ambiguous as to whether one is less competent or has less potential. It may be overly simplistic to infer that knowledge resides solely inside the individual's head.

Additionally, a given test may be one of low stakes, and certain non-mainstream students do not put forth the necessary efforts to do well on them; or they may be high-stakes tests in which many minority students feel overly anxious because they are fearful that if they perform poorly it will reflect badly on their social group as a whole (Steele & Aronson, 1995). Or it may be that the tests are insensitive to the instruction actually covered prior to the testing (Polikoff, 2010), or the classroom coverage was different, inadequate or ineffective. In all, there can be non-trivial latitude in, for example, what a typical standardized achievement test score tells us. Is it about learning, ability, misinterpretation, and test wise-ness, exposure? The score by itself cannot provide a sufficient answer. One must take into account contextual considerations.

Issues of interpretations enter also into this discussion in a different way. What about the interpretations that students make about themselves as students or learners, about their intellectual capabilities as individuals or as members of a given group? These self-referential insights are rich information sources for getting at how to understand a student's assessed performance. But in taking into account such data sources, we cannot rely only on the methods linked to classical psychometrics (Gipps, 1999). Qualitative data would be garnered, and through ethnographic methods linked to students' own lived, every day realities and phenomenological frames of reference. From such one can glean students' intentions, buy-ins, definitions, understandings, values, and even interpretations of the task/assessment demands at hand. When it comes to testing, one can glean whether students

are "on the same page" with the testing's "official" purpose. The information gathered will be subject to interpretation. Indeed, the pursuit of systematic interpretational inquiry, or *hermeneutics*, should be integral to this assessment in context enterprise (Cole, 1996; Moss, 2003).

An essential of such inquiry is to understand better what it is about the same "objective' situation that it holds multiple and differential meanings for different persons in that same setting, or why objectively different settings hold the same meanings but for divergent people.

This will require getting inside the black box of classroom teaching and learning dynamics (Black & Williams, 2010) to enrich formative assessments and assessments for learning, as one captures the phenomenology of students' lived classroom experiences and as they are linked to their lived realities outside of school, or their own experiential perspectives. Even when students are too young to provide rich explanatory insights, one can extract such insights from the various ways they are responsive to varying classroom contextual conditions and circumstances.

The assessments done should be poly-angulated to include the perspectives of students, teachers, and third-party observers. All this must be done with new assessment methods that reduce the time and labor intensity of such efforts, and ensuring the availability of the results to be packaged for reasonably efficient digestion by teachers as real time, informative feedback.

Herein may be where recent and future technological advancements in educational technology may play a vital role (Pellegrino & Quellmalz, 2010). Therefore, advancing technology may be especially useful with respect to processing more complex information, and real-time customized feedback to students, educators, and future assessment specialists, in the melding of information-rich quantitative and qualitative data from multiple sources. It is also conceivable that technology can aid in the alignment of formative and classroom-embedded assessments with outcomes assessments at the district, state, national, and international levels; all in the service of supporting the actualization of high-quality, capacity-building, teaching and learning.

It would seem to follow that greater, keener, more "accurate" insights are needed, through assessments, on

- why differing students perform as they do,
- how these differences should be adaptively addressed in the service of human capacity building, and ultimately, and
- how educational stakeholders are to be prepared to be more effective in these regards.

While the last point is beyond the scope of this article (although it deserves great attention elsewhere), the first two concerns will be discussed.

The upshot of this line of reasoning is that when we move away from schools functioning to sort for, select, certify, and even anoint the very best students, and more so to the purpose of pervasive human capacity building, then the focus should be on crafting opportunities for all students to learn at high levels. As mentioned previously, this pursuit would necessarily raise and require answers to several pertinent questions: Do students have genuine opportunities to learn? Opportunities to learn what? Opportunities to learn how? Opportunities to learn when? Opportunities to learn where?

The matter of opportunities to learn *what* can be expanded to the consideration of what should be in the official cannon of the curriculum and what would be the sought-after learning outcomes. This would seem to encompass actualizing potential, reaching high standards, and with respect to knowledge, skills, abilities, and civic participation. Moreover, Darling-Hammond and Falk (1997) captured some of this territory when they stated that we must support "schooling that will encourage all students to construct, integrate, and apply their knowledge, to think critically, and invent solutions to problems; and to respond creatively to the unforeseeable issues that will confront them in the complex world of tomorrow" (p. 51). Moreover, it would seem that the principal purpose of (educational) assessments would be largely to gauge the presence and consequences of opportunities to learn; and to learn in order to be prepared for the rigors, realities, and responsibilities of the 21st century.

In all, the push in educational assessments for the 21st century should be for greater support for a human capacity-building approach to schooling, where greater opportunities to learn for all students is greatly enacted, and where we deploy a poly-angulated system of assessments that leads to increased student engagement, while discerning the psychosocial integrity in the lived experiences of extant students from diverse backgrounds that schools and school districts are intended, indeed required, to serve.

CULTURE, RACE, EXCELLENCE, EQUITY AND ASSESSMENT

A growing literature points to engagement as particularly linked to favorable learning outcomes in minority students, who have been placed at risk for academic failure (Borman & Overman, 2004; Tucker et al., 2002; Wenglinsky, 2004). The data typically show that classroom engagement is among the most potent predictors of student outcomes. In reviewing research on the learning and performance of African American and Latino students, the evidence strongly suggests, to the extent that engagement processes are positively manifested, that this can dictate greater success for these populations (e.g., Bodovski & Farkas, 2007; Borman & Overman, 2004; Li & Lerner, 2011; Lutz, Guthrie & Davis, 2006; Skinner & Belmont, 1993). Findings in support of this claim have been documented across the full K-12 spectrum (Balfanz & Byrnes, 2006; Borman & Overman, 2004). For example, in a study by Balfanz and Byrnes (2006), self-reported effort and engagement emerged as a significant predictor in Black and Latino middle school students from an "urban background" of whether the students' gains in math performance exceeded what would have been expected by average yearly grade-equivalent increments. According to the authors, this outcome suggests by implication that the prediction is of gap-closing math performance.

Moreover, engagement that is informed by certain learning attitudes or guiding functions, and by certain classroom dynamics, themes, arrangements and processes that have been referred to as *asset-focused factors* are particularly instrumental in promoting the requisite levels of engagement that lead to the desired outcomes, or even impact these outcomes in a more direct fashion (Boykin & Noguera, 2011). Prominent among the guiding functions or learning attitudes are *self-efficacy, self-regulated learning,* and *incremental beliefs* about intelligence or ability (Bandura, 1977; Schunk & Zimmerman, 2007; Yeager & Walton, 2011). Self-efficacy entails the confidence that one has that she or he can do what it takes to accomplish the desired outcomes. Self-regulated learning entails the propensity for planning, monitoring, and assessing one's own learning. Incremental beliefs (as opposed to fixed beliefs) involve the belief that one's smartness or competence is malleable and thus potentially incremental. Recent evidence indicates positive outcomes accrue for struggling minority group learners when self-efficacy (Borman & Overman, 2004; Byrnes & Miller, 2007; Kitsantas, Cheema, & Ware, 2011); self-regulated learning (Horner & O'Connor, 2007; Mason, 2004; Schunk, 2005; Zito et al., 2007), and malleable beliefs about ability (Blackwell, Trzesniewski, & Dweck, 2007; Yeager & Walton, 2011) are manifested. As one example, a study by Mason (2004) speaks to the potency of self-regulated learning. Participants in this investigation were fifth-grade students with low reading achievement test scores (falling in the 10th to 40th percentile range). Over 90 percent of the sample was African American children from low-income backgrounds. One-half of the sample received explicit training in self-regulated learning to aid in reading comprehension. The other half received standard reading instruction. Reading performance was superior among those in the intervention group who were taught to deploy goal-setting, self-monitoring, and self-reinforcement (for making progress) while engaged in reading. Elsewhere, research evidence shows that African American students are more likely to blame their own shortcomings for academic difficulties or failures (Good, Aronson, & Inzlicht, 2003). That is, they make internal, stable attributions for negative outcomes. Nevertheless, when they come to believe in an incremental or malleable theory, they are much less likely to blame their own fixed intellectual shortcomings for their current struggles. Instead, they come to view mistakes as an indicator that they did not try hard enough or did not appropriately approach the problem. Having African American students come to believe that competence and ability are changeable and not fixed can have positive impact on their achievement

strivings, in that they become more task-engaged and perform better on tests of achievement (Blackwell, Trzesniewski, & Dweck, 2007).

Research also indicates that the deployment of asset-based strategies have a direct positive impact on student classroom engagement and gap-closing achievement, raising academic outcomes (Boykin & Noguera, 2011). As the term implies, this approach seeks to build on the assets that children and youth from diverse backgrounds bring with them into learning settings, provide classroom conditions that encourage the expression of these assets, and help create academically relevant assets for students in school settings (Boykin, 2012; Boykin & Noguera, 2011). Among the most empirically verified, factors are:

- *teacher-student relationship quality*, which entails the provision of a socially and emotionally supportive yet demanding and high-expectations classroom learning environment;
- *collaborative learning,* which entails collaborative intellectual exchanges among students and ensures that all classroom participants are actively involved in the learning process;
- *meaningful learning*, which conveys building on students' past experience and prior knowledge and making connections in school to significant events in students' lives;
- *cultural resources,* captured as building proactively on the cultural, family and community assets, values and practices students bring with them to the classroom; and
- *information processing quality,* which involves directly teaching students problem-solving and learning strategies, and promoting higher-order thinking and critical understanding with respect to subject matter.

There is ample evidence in the extant research literature that asset-focused factors, when incorporated into classroom teaching and learning can be particularly beneficial for ethnic minority students across the K-12 spectrum. The research evidence reveals that while teachers generally hold lower expectations for Black and Latino students than their White counterparts, even when controlling for SES, positive teacher-student relationship quality has achievement-enhancing, gap-closing consequences for low-income ethnic minority students (Hamre & Pianta, 2005; Stewart, 2006; Tenenbaum & Ruck, 2007). The promoting of mastery goal structures leads to enhanced engagement, achievement, and positive climate for ethnic minority students (Kaplan & Maehr, 1999). A meta-analysis indicated that collaborative learning more positively impacts achievement outcomes for Black and Latino than White students (Rohrbeck et al., 2003). Meaningful learning has gap-closing consequences at both the elementary (Anand & Ross, 1987) and secondary level (Cohen et al., 2006). The deployment of culturally relevant strategies in the form of promoting such themes as communalism (Hurley, Allen & Boykin, 2009) or incorporating popular culture (Crumpton & Gregory, 2011; Lee, 2001) has enhancing impact on achievement outcomes for many African American students. Moreover, directly teaching ethnic minority students efficient and effective information-processing skills have positive achievement consequences as well (Jitendra et al., 2007; Mason, 2004; Ramani & Siegler, 2011; Williams et al., 2005).

As one case in point, the potential gap-closing potency of learning through collaboration has been documented in a meta-analytic review conducted by Rohrbeck and colleagues (2003). All studies reviewed were at the elementary school level. In general, peer-assisted learning contexts yielded greater math outcomes than did contexts marked by individualistic or competitive learning. Yet more particularly, the differences favoring peer-assisted learning were greater in urban over rural and suburban settings, low-SES over mid- and high-SES levels, and for minority status (Black and Latino) compared to majority (White) status. In fact, this ethnic minority status difference favoring collaborative learning was the largest of all comparisons. Subsequently, the benefits of collaborative learning for Black and Latino students as compared to White students, was the largest difference among all the comparisons.

In spite of such evidence from the research literature, such findings have not substantially found their way into educational practice. All too many racial minority students are on "trajectories of marginalization" in our nation's schools. This has led to task disengagement, further leading to subject matter disengagement, to school disengagement, and eventually to societal disenfranchisement. Increasing engagement (thereby decreasing disengagement) at any or all points

along this trajectory can lead to greater opportunities to learn. As stressed in this article, assessments of engagement/disengagement are needed, as are assessments of the instructional and curricular activities, and the enabling or disabling conditions inside classrooms and other learning settings. One marker to be mindful of is captured in the following:

> The gate keeping role that [subject matter] plays in students' access to educational and economic opportunities is not limited to ways of knowing associated with participation in the practices of different communities. Instead it also includes difficulties that students experience in reconciling their views of themselves and who they want to become with the identities they are invited to construct in the . . . classroom. (Cobb & Hodge, 2002, p. 249)

When students, are encouraged and supported by their teachers and peers to reach high levels of performance, they can respond favorably in school. When students see their lived social and cultural experiences meaningfully incorporated into classroom learning activities, they can respond favorably in school. When students are not penalized for not doing or knowing what to do academically, rather than being penalized and blamed for this, they are provided with intellectual tools to handle the lesson material more effectively and efficiently, and they can respond favorably in school. Incorporating such factors into the schooling process can likely lead to positive academic outcomes, and the assessment of the effective presence of such factors inside classrooms should become a measurement priority in the 21st century. These indices should be both quantitative and qualitative. They should incorporate surveys, direct observations and performance appraisals. Data should be gleaned from several stakeholders. Educational personnel must be prepared to provide such enabling learning conditions inside classrooms and, for that matter, other learning venues as well. All of this is done to foster an authentic human capacity-building function for our students, in our schools, in this century.

In light of the issues raised, outcomes assessments could expand their scope to include broader notions of academic understanding, or intellectually valued skills, such as the creative use of information under conditions of limited resources; the ability to thrive in the midst of social contradictions or in the midst of adults who do not convey that they have your best interests at heart; or problem-solving in real life social settings, where multiple, even seemingly incompatible social agendas have to be simultaneously and successfully negotiated.

To add more density to the understanding of the issues at hand, it is crucial to distinguish between vertical and horizontal diversity, in regard to matters of schooling. Vertical diversity speaks to a hierarchical ordering of performance, skill, competence, and so forth, where this ordering can be imminently measured in quantitative terms. Vertical diversity often implies that the experiences or attributes of lower ranked individuals or groups account for the lower levels of performance. Horizontal diversity speaks to differences among individuals or groups, which are more qualitatively captured and for which relative ranking would not be appropriate. The differences are just differences, and no particular ranking value are attached to what experiences or attributes are better or worse per se. At issue is that even when students from one racial or cultural grouping perform more poorly than others in school or on assessments, we should not necessarily presume that their social or cultural experiences or their intellectual proclivities are also of lesser quality, or that they explain away the ranking differences. Yes, certain groups do perform more poorly on assessments. But it should not be automatically concluded that their lived experiences and the manifestations of these experiences are of a lesser quality. Consequently, from a horizontal diversity lens, we should seek the possible equally potentiating value of these experiences and inclinations, and proffer assessments that go beyond current indices for academic understanding to include ones that represent the diversity of experiences of the students inside classrooms, and that manifest value for these experiences and for the social, intellectual and cultural capital that they generate.

Another issue for consideration is what is meant by "education." More specifically, some consider education to be a thing, a commodity that students acquire or receive. The views and arguments presented construe education more so as a process. To view education more so as a process than a thing or goods that one acquires over the course of his or her academic matriculation would entail the viewing of education as a family of processes in which one engages with teachers and other

514   ©*The Journal of Negro Education, 2014, Vol. 83, No. 4*

students (and likely even others outside of the formal schooling setting), and which takes the forms of participatory transactions, discourses, dialogues, and continuous skill and knowledge acquisitions, as well as varying vectors of effort and improvement.

In all, on matters of *culture, race, excellence, equity and assessment*, the fundamental argument is in the pursuit of a pervasive human capacity building agenda that undergirds student learning, students from diverse cultural and racial backgrounds must be given high quality, legitimate opportunities to learn. These opportunities to learn should be assessed in real time by an interlocking set of classroom-based assessments that have a broader conception of what should be learned; and this broader conception should be incorporated in the formulation of more long-term outcomes assessments.

In this view, student learning task engagement is key. It is the linchpin for both indexing the quality of learning and assessed achievement outcomes. Learning should be understood as not simply doing well on some test, but must be construed as gaining understanding, insight, skill and knowledge of the curriculum at hand. However, the notion of *curriculum* itself must be unpacked. Learning should take place with respect to the *intended* (official) curriculum, which may or may not be the same as the *taught* curriculum, and certainly could be at varying odds with the *learned* curriculum, as it manifests through individual and group phenomena. According to these analyses, students in the classroom may not necessarily be learning the overtly taught curriculum. The taught curriculum also includes the *hidden* curriculum vis-à-vis socialization messages conveyed by teachers and other schooling personnel, regarding who (what group) has the power to define what a good student should look like, act like, think like; what is educationally important to know; what is to be valued educationally; what is the prevailing ideology about what constitutes learning; and who is most capable of doing so. Therefore, students from diverse racial and cultural backgrounds seemingly could be disengaged from the formal lesson, but deeply responsive to the messages contained in the hidden lessons transpiring in the classroom. Or, they may be engaged in activities that are well-intentioned and personally meaningful, just not what the teacher was formally attempting to convey. All of these complexities also have profound implications for the interpretive and consequential validity of the *assessed* curriculum. The ultimate goal, in the pursuit of both equity and excellence is to increase the number of students, and those from diverse racial and ethnic backgrounds, who are three-dimensionally engaged in authentic learning activities that foster individual, (racial or ethnic) group, and school goals simultaneously. There must be engaged participation, calibrated by classroom-based capacity-building assessments, rather than disengaged participation in the better synchronized intended and taught curriculum, so that the learned curriculum of the students can be well-manifested on the ensuing assessments of learning that could be administered periodically, but certainly at the end of the academic year.

IMPLICATIONS

The converging matters of assessment, diversity, equity, and excellence can plausibly be approached from two (at the least) distinct yet interrelated stances. There is the social justice angle. There is also the human capital production position.

From the social justice standpoint, the argument would be that it is morally appropriate to critique the inherent fairness of assessment in terms of how negatively certain subpopulations within society seem to consistently fare poorly on various assessment measures. Many advocates of this stance also argue that assessments are even conceived to serve a hegemonic function, so that certain mainstream groups are to be privileged and prioritized. This is socially unjust, and such injustice must be eradicated. In all, then, the social justice argument is that the bias, problematics, or illegitimacy of such measures must be exposed and rectified, since, in their present forms, these assessments do not allow the abilities and propensities of certain groups to be accurately or adequately ascertained. From this stance, it is also argued that we need to determine how such assessments can more truly discern the performance capabilities of certain groups without undermining the psychometric quality of such assessments; without unduly penalizing those who fare well on such measures in their present forms; and without inappropriately undermining the pursuit of excellence in the standards of the outcomes

sought. On this last point, certain critics of the social justice stance argue that its pursuit permits access for less qualified people to opportunities that they have not earned or do not deserve. Stating this last point differently, the claim is that we should not diminish the pursuit of determining excellence in order to satisfy the principle of equity. Furthermore another danger that critics raise is that the seeking of social justice obscures the actual deficiencies either inherent or experiential of the diverse or non-mainstream groups by the preoccupation with inclusiveness in the pursuit of fairness.

CONCLUSION

The social justice conception can be contrasted with the human-capital production or capacity-building stance. While the former would claim that the pursuit of social justice in assessment is a societal good, the latter would claim that it is a societal good to foster extensive, high-level knowledge, skills and abilities in intellectual, technical and civic-participation domains, for successive cohorts of the American population. In turn, assessments should function principally to help actualize such human-capital production. Such a focus would shift away from the preoccupation with outcomes assessments, and with assessments as predictors of subsequent competence, to increased emphasis on assessments of a more formative kind. Such assessments would be more localized, and even situated to a good degree inside classrooms or other (putatively) capacity-building settings. The argument would also be that assessment is done, to a large but not exclusive extent, to guide the preparation of individuals, and from diverse backgrounds, to reach and even exceed the high performance expectations that society will require of them in the years and decades ahead. Viewed in this way, outcomes assessments will serve more so as confirmation or disconfirmation that the *preparation process* has been successfully executed. Moreover, the range of knowledge, skills and abilities examined in outcomes assessments should reflect the complexities of competencies needed by people for full participation in the increasingly diverse social fabric of American society, and in a globally interdependent world, and as they confront the rigors, roles and responsibilities in the 21st century. This would require rethinking some of the conventions of assessment, and the criteria for what high-quality assessments look like. The phenomenon captured in this line of reasoning is what should be assessed. To do the latter would necessitate movement away from seeing people with presently "non-mainstream" social, racial, and cultural backgrounds exclusively in pejorative terms, but as encompassed in a new, more inclusive notion of humanity. Otherwise, it would be argued, America will not meet the demanding manpower needs of the 21st century. From this stance, achieving equity is essentially a byproduct of the quest for pervasively high-level human capacity building; that is for excellence.

From a human capital production stance, the onus is on assessment to illuminate the potentialities of or suitable preparation processes for people from diverse backgrounds. In these ways, their identities and the experiences associated with these backgrounds either are construed as important, as a source of assets that can be capitalized on in the preparation process, or seen as making clearer the diversity in outcomes that assessments must measure and measure well. On this latter point, such would thereby broaden the range of possibilities for which excellence is (can be) ascertained. It is also plausible that the successful pursuit of a human capital production agenda will address successfully and relatively more permanently the intended goals of the social justice agenda.

In all, the evidence, arguments, and reasoning advanced attempts to make the case for the primacy of the human capacity-building approach for future assessment endeavors; or at the very least for assessment endeavors in the realm of formal education. It is also advanced that such a primary function for the assessment enterprise would only make sense if we also reimagine or re-envision what should be the function and purpose of formal education in the American social order in the years and decades ahead. For far too long, education has served the purposes of sorting, classifying, and selecting students, and of identifying existing student abilities. Instead, it would seem that the schooling enterprise should allot at least equal (if not more) weight to the cultivation and advancement of student potential, and the actualization of pervasively high levels of student intellectual, technical, and civic participation competence.

REFERENCES

Anand, P. & Ross, S.M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79*, 72-78.

Balfanz, R. & Byrnes, V. (2006). Closing the mathematics achievement gap in high-poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed At Risk*, *11*, 143-159.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychology Review,* 84, 191-215.

Barbarin, O. A. (2002). The Black–White achievement gap in early reading skills: Familial and sociocultural context. *Love to read: Essays in developing and enhancing early literacy skills of African American children.* Washington, DC: National Black Child Development Institute, Inc.

Basterra, M. (2011). Cognition, culture, language, and assessment: How to select culturally valid assessments in the classroom. In M. Basterra, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 72-95). New York: Routledge.

Black, P., & Williams, D. (2010, September). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*, 81-90.

Blackwell, L. A., Trzesniewski, K. H., & Dweck, C. S. (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. *Child Development, 78,* 246-263.

Bodovski, K., & Farkas, G. (2007). Do instructional practices contribute to inequality in achievement? The case of mathematics instruction in kindergarten. *The Journal of Early Childhood Research*, *5*, 301-322.

Borman, G., & Overman, L. (2004). Academic resilience among poor and minority students. *Elementary School Journal, 104*, 177-195.

Boykin, A. W. (2000). The talent development model of schooling: Placing students at promise for academic success. *Journal of Education for Students Placed At Risk*, 5, 3-25.

Boykin, A. W. (2012). Talent Development Model of schooling. In Banks, J. A. (Ed.), *Encyclopedia of diversity in education* (Vol. 4, pp. 2111-2115). Thousand Oaks, CA: Sage.

Boykin, A. W., & Noguera, P. (2011). *Creating the opportunity to learn: Moving from research to practice to close the achievement gap.* Alexandria, VA: ASCD Press.

Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., & McLoyd, V., & NICHD Early Child Care Research Network (2011). Examining the Black–White achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development, 82*, 1404-1420.

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity–propensity analysis. *Contemporary Educational Psychology*, *32*(4), 599-629.

Clarke, M. M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies, 32*, 159-181.

Cobb, P., & Hodge, L. L. (2002). A relational perspective on issues of cultural diversity and equity as they play out in the mathematics classroom. *Mathematical Thinking and Learning, 4*, 249-284.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science, 313*, 1307-1310.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Harvard University Press.

Condron, D. J. (2009). Social class, school and non-school environments, and Black/White inequalities in children's learning. *American Sociological Review, 74*, 683-708.

Connell, R. W. (1992). Social justice in education. In *Schools and Social Justice* (pp. 11-19). Toronto: Our Schools/Our Selves Education Foundation.

Cremin, L. A. (1989). *Popular education and its discontents*. New York: Harper and Row.

Crumpton, H. E., & Gregory, A. (2011). "I'm not learning": The role of academic relevancy for low-achieving students. *Journal of Educational Research, 104*, 42-53.

Darling-Hammond, L., & Falk, B. (1997). Supporting teaching and learning for all students: Policies for authentic assessment systems. In Goodwin, A. (Ed.), *Assessment for equity and inclusion: Embracing all of our children* (pp. 51-76). London: Routledge.

Dorman, J. P., Aldridge, J. M., & Fraser, B. J. (2006). Using students' assessment of classroom environment to develop a typology of secondary school classrooms. *International Education Journal, 7*, 906-915.

Engle, R. A., & Conant, F. R. (2010). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community-of-learners classroom. *Cognition and Instruction, 20*, 399-483.

Fisher, S. L., & Fraser, B. J. (1981). Validity and use of the My Class Inventory. *Science Education, 65*, 145-156.

Fraser, B. J. (1998). Classroom environment instruments: Development, validity and applications. *Learning Environments Research, 1*, 7-33.

Fraser, B. J., & Fisher, S. L. (1983). Development and validation of short forms of some instruments measuring student perceptions of actual and preferred classroom learning environment. *Science Education, 67*, 115-131.

Fraser, B. J., & Wahlberg, H. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Education, 8*, 67-92.

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of evidence. *Review of Educational Research, 74*, 59-109.

Fryer, R., & Levitt, S. (2004). Understanding the Black-White test score gap in the first two years of school. *The Review of Economics and Statistics, 86*, 447-464.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, *24*, 355-392.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*, 645-662.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949-967.

Heritage, M. (2011). Formative assessment: An enabler of learning. *Better: Evidence- Based Education*, *3*, 12-13.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

Hickey, D. T., & Zucker, S. J. (2005). Engaged participation: A sociocultural model of motivation with implications for educational assessment. *Educational Assessment, 10*, 277-305.

Horner, S. L., & O'Connor, E. A. (2007). Developing self-regulation within the reading recovery context. In S. L. Horner & E. A. O'Connor (Eds.), Helping beginning and struggling readers and writers develop self-regulated strategies. *Special Issue of Reading and Writing Quarterly, 23*, 97-106.

Hughes, S. (2003). An early gap in Black-White mathematics achievement: Holding school and home accountable in an affluent city school district. *The Urban Review, 35*, 297-322.

Hurley, E. A., Allen, B. A., & Boykin A. W., (2009). Culture and the interaction of student ethnicity with reward structure in group learning. *Cognition & Instruction. 27*, 121-146.

Jitendra, A. K., Griffin, C., Haria, P., Leh, J., Adams, A., & Kaduvetoor, A. (2007). A comparison of single and multiple strategy instruction on third grade students' mathematical problem solving. *Journal of Educational Psychology, 99,* 115-127.

Kaplan, A., & Maehr, M. L. (1999). Achievement goals and student well-being. *Contemporary Educational Psychology, 24,* 330-358.

Kelly, S., & Turner, J. (2009). Rethinking the effects of classroom activity structure on the engagement of low-achieving students. *Teachers College Record, 111,* 1665-1692.

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics*, *22*, 310-339.

Lee, C. (2001) Is October Brown Chinese? A cultural modeling activity system for underachieving students, *American Educational Research Journal, 38,* 97-141.

Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology, 47*, 233-247.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*, 4-16.

Linn, R. L. (2001). Validation of the uses and interpretations of results of state assessment and accountability systems. In J. Tindal & T. Haladyna (Eds.), *Large-scale assessment programs for all students: Development, implementation, and analysis* (pp. 27-48). Mahwah, NJ: Erlbaum.

Luppescu, S., Allensworth, E., Moore, P., de la Torre, M., Murphy, J., & Jagesic, S. (2011). Trends in Chicago's schools across three eras of reform. *University of Chicago Consortium on School Research*.

Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *Journal of Educational Research*, *100*, 3-20.

Madaus, G., Raczek, A., & Clarke, M. (1997). Historical and policy foundations of the assessment movement. In A. Goodwin (Ed.), *Assessment for Equity and Inclusion: Embracing All Our Children* (pp.1-34). London: Routledge.

Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology, 96*, 283-196.

Mathews, S. R., & Lowe, K. (2011). Classroom environments that foster a disposition for critical thinking. *Learning Environments Research, 14*, 59-73.

McKinsey & Company (2009, April). T*he economic impact of the achievement gap in America's schools*. Washington, DC: Social Sector Office.

Moos, R. H. (1974). *The social climate scales: An overview*. Palo Alto, CA: Consulting Psychologists Press.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*, 13-25.

Murnane, R., Willett, J., Bub, K., & McCartney, K. (2006). Understanding trends in the Black White achievement gap during the first years of school. *Brookings Wharton Papers on Urban Affairs, 7,* 97-135.

National Assessment of Education Progress (NAEP). (2009). Washington, DC: U.S. Department of Education.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and the Secretary of Education, United States Department of Education*. Washington, DC: The Commission.

National Research Council, Pellegrino, J., Chudowsky, N., & Glaser, R., Board on Testing and Assessment, Center for Education (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

Nettles, M. (2000, October). *The status and consequences of admissions test performance for the nation's demographically diverse population of aspiring students*. Paper presented at the Fulbright's Educational Experts Seminar, Ann Arbor, MI.

No Child Left Behind Act of 2001, Pub. L. 107-118 (2002).

Oakes, J. (1989). What educational indicators? The case of assessing the school context. *Educational Evaluation and Policy Analysis*, *11*, 181-199.

Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education*, *43*, 119-134.

Phillips, V., & Wong, C. (2010). Tying together the common core of standards, instruction, and assessments. *Phi Delta Kappan, 91*, 37-42.

Polikoff, M. S. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice, 29*, 3-14.

Porter, T. M. (1995) *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low-and middle income preschoolers. *Journal of Applied Developmental Psychology, 32,* 146-159.

Rohrbeck, C. A., Fantuzzo, J. W., Ginsburg-Block, M. D., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*, 240-257.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40*, 85-94.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading and Writing Quarterly*, *23,* 7-25.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*, 4-14.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement. *Journal of Educational Psychology*, *85*, 571.

Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: An introduction. In M. Basterra, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 3-21). New York: Routledge.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Stewart, E. (2006). Family and individual-level predictors of academic success for African-American students: a longitudinal path analysis using national data. *Journal of Black Studies, 36*, 597-621.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*, 758-765.

Stiggins, R. J. (2006). Assessment for learning: A key to student motivation and learning. *Phi Delta Kappan EDGE*, *2*, 19, 3-19

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment* for *student learning: Doing it right—using it well.* Portland, OR: Pearson Assessment Training Institute.

Stigler, J. W. & Hiebert, J. (2009). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Strambler, M. & Weinstein, R. (2010). Psychological disengagement in elementary school among ethnic minority students. *Journal of Applied Developmental Psychology, 31*, 155-165.

Taylor, P. C., Fraser, B. J., & White, L. R. (1994, March). *A classroom environment questionnaire for science educators interested in the constructivist reform of school science*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Anaheim, CA.

Tenenbaum, H. R. & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253-273.

Tucker, C. M., Zayco, R. A., Herman, K. C., Reinke, W. M., Trujillo, M., Carraway, K., & Ivery, P. D. (2002). Teacher and child variables as predictors of academic engagement among low-income African American children. *Psychology in the Schools*, *39*, 477-488.

Wahlberg, H. J., &Anderson, G. J. (1968). Classroom climate and individual learning. *Journal of Educational Psychology*, *59*, 414-419.

Wenglinsky, H. (2004). Closing the racial achievement gap: the role of reforming instructional practices. *Educational Policy Analysis Archives, 12*, 1-24.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & deCani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, *97*, 538-550.

Yeager, D. S., & Walton, G. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81,* 267-301.

Zito, J. R., Adkins, M., Gavins, M., Harris, K. R., & Graham, S. (2007). Self-regulated strategy development: relationship to the social-cognitive perspective and the development of self-regulation. *Reading & Writing Quarterly*, *23*, 77-95.

AUTHOR

A. WADE BOYKIN is Director of the Graduate Program in the Department of Psychology at Howard University. He also serves as the Executive Director of the Capstone Institute.

All comments and queries regarding this article should be addressed to aboykin@howard.edu and jlee@capstoneinstitute.org