Translating Standards into Assessments:

The Opportunities and Challenges of a Common Core

Presented to:

Brookings Institution

by:

Gregory J. Cizek

Professor of Educational Measurement and Evaluation

University of North Carolina at Chapel Hill

October 28, 2010

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Translating Standards into Assessments:

The Opportunities and Challenges of a Common Core

The development of assessments based on common core standards represents the logical next step in a fairly traditional process–a process that has been implemented hundreds of times in the past decade in the context of state student achievement testing programs. Typically, the process begins with the articulation of *content standards*; that is, statements about the knowledge and skills that comprise the important learning outcomes for students in specific grades and subjects. The content standards in this case are those embodied by the Common Core State Standards (CCSS) in English language arts (CCSSO & NGA, 2010a) and mathematics (CCSSO & NGA, 2010b).

Once content standards for grades and subjects have been established, mechanisms for measuring achievement vis a vis the content standards are undertaken. This aspect is often referred to as the assessment development process. It typically begins with explicit articulation of the purposes, types, and frequency of asessments. It includes specification of the breadth, depth, and weighting of the content standards to be assessed, the length, and formats to be used for the assessments, and other practical considerations such as the role of *performance standards*; that is, the required levels of achievement relative to the content standards.

Several opportunities and challenges lie ahead for the processes and policies that will guide the development of common core assessments (CCAs). At a general level, there is the opportunity that assessments developed under Race to the Top (RT) initiative have the potential to continue some of the positive aspects of the No Child Left Behind (NCLB, 2001) legislation.

For one thing, CCAs are likely to sustain a clear national education policy focus on instruction and achievement. The large number of states participating in consortia developing CCAs and the likely high public and professional visibility of the content standards upon which the assessments are based contribute to this potential. For another, CCAs will maintain and likely expand a desirable homogeneity of educational experience for American students. Under NCLB, students in a given state could expect, within grade and subject, a reasonable degree of standardization with respect to the content standards that would be the focus of instruction and assessment across the classrooms, schools, and districts in that state. Where student populations are increasingly diverse and mobile, such standardization extended across state boundaries can be seen as a positive influence on teaching and learning.

Assessment Challenges (AC)

There are other opportunities and anticipated benefits that are likely to accrue from common standards and assessments; drivers of the CCSS process have well-articulated the promising aspects of this reform initiative. Many specific assessment challenges also exist in the development of CCAs, however, and the challenges have received comparatively less attention. Thus, in this balance of this paper, I will not further catalogue the promising aspects, but will instead highlight seven specific assessment challenges and associated policy implications.

AC1 - Performance Standards

Increased rigor in common content standards and corresponding assessments for U.S. students is an oft-stated and desirable goal. The opportunity to develop assessments that move toward that goal is present, but the accomplishment of the goal depends on several factors. One

such factor is the establishment of performance standards.

In an assessment development process, it is possible to develop content standards that are more or less comprehensive, more or less demanding, more or less specific. It is also possible to develop a range of items or tasks that tap any single content standard. Items/tasks may be developed that are harder or easier, more or less complex, and so on. It is not yet clear how the CCSS compare on these criteria to the patchwork of individual state content standards, nor are there clear pictures of what "test forms" of items and tasks aligned to the CCSS will look like. However, let us assume that both the content standards and any items and tasks will be worth teaching to and represent increased rigor.

Even if those assumptions are satisfied, they could be undercut when the *performance standards* for the CCAs are established. For example, a large proportion of the readers of this paper would likely "pass" a test of pediatric ophthalmology administered in Iraqi/Arabic language if it consisted of 20 four-option, multiple-choice questions and the cut score was set at 4 or 5 items correct, despite the fact that no reader of this paper is likely to be a pediatric ophthalmologist or fluent in Iraqi Arabic–and almost certainly not both!

The example illustrates that, ultimately, rigor is a characteristic not only of the content to be assessed (and the level at which the content is assessed), but also of the performance standards(s) established on the assessment. An oft-cited deficiency of the current patchwork, individual state content and performance standards is that they preclude meaningful comparisons. To redress that deficiency and permit the kinds of meaningful comparisons that, common scale reporting mechanisms will be required. The concern here is that at least one identified weakness of NCLB will also be repeated. That is, it was possible for a state to establish less demanding performance standards even when it had adopted highly rigorous content standards.

At minimum then, it would seem necessary to establish one or more common performance standards for any CCAs. Expenditures of the amount contemplated for the development of CCAs (approximately $330 billion) would hardly seem justified if members within a consortium of state participants could set differing performance standards. And, as I have noted elsewhere "the word 'common' also means 'not distinguished by superior or noteworthy characteristics; of mediocre or inferior quality; second-rate'" (Cizek, 2010a). As was seen–unfortunately–in the context of NCLB, the requirement to derive a broadly-acceptable outcome can lead to performance standards that fail to reflect truly rigorous targets. It will be of great importance then, that current CCA development efforts resist the dual threats of either individual state performance standards, or common performance standards that do not truly reflect aspirational levels of achievement.

To be specific, a number of questions have yet to be addressed regarding performance standards, including: Who will be involved in setting the performance standards? What process and procedures will be used to set them? Will the performance standards be truly common? How many performance levels will be appropriate? How will we know if the are truly challenging? What kinds of consequences (both positive and negative) will be associated with meeting (or failing to meet) the performance standards?

Almost certainly, consortium members will need to be involved in the development–or, at least, approval–of common performance standards, and they are likely to face strident critique regardless of which course is pursued. For example, if truly rigorous performance standards are adopted, some of the same criticisms are likely to arise as were voiced regarding the aspriational achievement levels for the National Assessment of Educational Progress (NAEP) being set "unreasonably high" (Shepard, Glaser, Linn, & Bohrnstedt, 1993, p. xxii). If performance

standards are adopted that are considered to be insufficiently rigorous, some of the same criticisms are likely to arise as have been voiced regarding achievement levels for some NCLB-mandated state assessments being set too low. In either case, the rationale for the common core standards initiative would be undermined and the credibility for the entire enterprise would be at risk. In the end, it may well seem in retrospect as if it had been easy to gain support for the sets of common content standards compared to the challenge of winning support for common, rigorous performance standards.

AC2 - Inclusion

One of the noble goals of the NCLB legislation–a reaffirmation of what was set forth in the *Improving America's Schools Act* [IASA] (1994)–was to expand access and participation in assessment programs to all students. In the past, some students with limited English proficiency (LEP) or students with disabilities (SWD), were routinely excluded from full access and participation and often were not exposed to the same kinds of instructional programs as their peers. The mandates of IASA and NCLB meant that such students would "count" and it was required that they be provided with relevant, individualized instructional programs and included in assessment programs. As I have noted previously, "necessity was the mother of intervention" (Cizek, 2001).

A psychometric challenge and policy consideration in the development of CCAs will be the extent to which versions of the tests will be produced for students with limited English proficiency and those with special needs. The CCA proposals of both comprehensive (i.e., elementary through high school) consortia (Partnership for Assessment of Readiness of College and Careers [PARCC], 2010; Smarter Balanced Assessment Consortium [SBAC], 2010) contain

similar, promising language about inclusion. For example, the SBAC proposal indicates that: "This balanced system is designed to provide accurate assessment of all students--including students with disabilities (SWDs), English learners (ELs), and low- and high-performing students..." (p. 19). The PARCC proposal indicates that: "The Partnership will design each component in a manner that allows these groups of students, who have a wide range of cognitive abilities, content knowledge, and English proficiency, to demonstrate what they know and can do" (p. 74). However, both documents appear to propose only single test forms designed for the full population of students; mechanisms for accomplishing full inclusion are not described in detail, and inclusion strategies are largely limited to expert review and input and attention to features of universal design and various common administration accommodations.

If any lessons from the NCLB requirement to include all students in an assessment system are clear, they are that: 1) one assessment does not fit all; 2) the creation of different test versions appropriate for ELs and SWDs requires substantial additional time and resources; and 3) the modifications necessary to create such versions in terms of content coverage, format, level, administration modes, and other factors result in students' scores across versions that are not comparable in meaning. Beyond the psychometric concerns, clear challenges lie ahead as regards reaching consensus among states regarding common, acceptable, and valid accommodations. Additionally, decisions will need to be made regarding whether or how to assess students for whom the common acceptable accommodations are insufficient to permit sound measurement of their knowledge, skills and abilities.

AC3 - Instructional Alignment

It is not uncommon for changes to be made to the content standards upon which

assessments are developed. The changes may be implemented as a result of new theoretical ideas within a content area or age level, research findings, public, professional, or political preferences and other factors. Many problems arise, however, when content standards are changed too frequently. One concern is that dissemination regarding the changes, professional development, and–most importantly–instruction lag behind. Depending on the timing of implementation of the changes into the assessment program, the lag can create concerns about students' opportunity to learn the new content. Another unintended consequence related to frequent changes in content standards can be the feeling, expressed by many educators, that "the rules are always changing." This can lead to instructional inertia–a "wait and see" perspective on the part of teachers, curriculum planners, administrators and others, and a fall-back focus on the knowledge and skills deemed important by individual teachers, as opposed to those articulated in a state's content standards.

From a psychometric perspective, when content standards are revised–even sometimes in subtle ways–a shift in the construct measured by an assessment can be introduced. Because it would introduce confusion and inaccurate interpretations to report scores on different constructs as if they were the same, it is typically appropriate to establish a new scale for such a test, often using totally different scale values (e.g., adopting a new a scale ranging from 200 to 400 in place of a previous scale ranging from 50 to 80) and resetting the performance standards (i.e., cut scores). Even with appropriate linking or equating methods, construct shift introduced by changes in the content standards can reduce the usefulness of the test results, introduce confusion in score interpretation, and limit the ability to monitor growth/achievement over time.

As submitted, neither of the funded assessment consortia proposals (PARCC, 2010; SBAC, 2010) nor U.S. Department of Education guidelines appear to address this concern. A

clear policy challenge that lies ahead will be the need to develop guidelines that limit tinkering with the CCSS. A promising model is that of the National Assessment Governing Board (NAGB) for development of National Assessment of Educational Progress (NAEP) frameworks. Although the CCSS process has already deviated from the NAEP-like attributes of deliberate and democratic development of content standards, it is not too late to emulate the NAEP model as regards stability of frameworks over time and in the achievement levels setting process that will be required for CCAs.

AC4 - Innovation in Assessment

Funds allocated to the CCA program have the potential to stimulate innovation in student assessment. However, it is equally likely–or more so–that innovation in assessment will actually be reduced as an unintended consequence of the federal assessment grant program.

A review of the two comprehensive assessment proposals reveals two things. First, there are only two of them. Instead of spawning many diverse, innovative proposals, the competition resulted in only two large clumpings of individual states that associated in support of common proposals. Second, regardless of one's opinion about NCLB on other criteria, across the states, fairly substantial variation in approaches and incremental innovation efforts had marked the state assessment landscape. A common aspect of the two CCA proposals–that of computer-based testing–had already been incorporated into several state testing programs in a variety of ways. Admittedly, the incorporation of computer-based testing into large-scale state programs has not been unequivocally successful–a technical challenge that has plagued assessment programs at the state level and is as yet an unresolved challenge for the even larger-scale consortia proposals.

The bottom line here is that, in terms of diversity and innovation in

practice–encompassing aspects such as content standards covered, proportions of constructed-response items, innovative item formats, language versions, computer delivery, assessments tailored to special needs students, and so on–student assessment will almost certainly be more homogeneous after implementation of the CCAs than had been the case previously.

It might be asserted that the federal guideline permitting 15% of something to be added to the something will still allow some degree of innovation or uniqueness in state assessment programs. However, there are at least two concerns with that guideline; a non-subtle hint at the first concern is the intentional ambiguity and repetition of word choice–"something." Among the many perplexities that have plagued previous and present policy is the one found in "the 15% guideline." A reasonable test developer would appear to have fairly wide latitude for customizing a state test while remaining faithful to that guideline. Consider, for example, the following three instantiations of the guideline with italics added to emphasize an important difference among them:

> 1) "States are allowed to include up to 15 percent *additional items* to the Common Core State Standards." (Iowa Department of Education, nd, p. 2)
>
> 2) "The State Board of Education reserves the right to add up to 15 percent *additional standards* to the Common Core State Standards as appropriate." [Oklahoma Administrative Code, Title 210:15-4-3(b)(1)]
>
> 3) "States are allowed to augment the standards with an *additional 15% of content* that a state feels is imperative." (Achieve, 2010, p.1)

Depending on which of the three interpretations of the guideline is taken, effective test specifications could vary widely. And, on only a mildly technical note, all interpretations of the

guideline might assume equal and unit weighting of all items, standards, or content, although that is neither the only plausible approach, nor even the most common approach in K-12 test construction, where items, standards, or content within a test are often differentially weighted and therefore contribute more or less proportionally to a student's total score. A final point related to this concern is the effect that adding items, standards, or content is likely to have on between-state comparisons of test performance. Such comparisons were a motivating factor for pursuing CCSS in the first place. An obvious policy stance would have seemed to be a requirement that only the *common* portion of CCA would come into play in the comparisons. However, it is not at all clear that this is the case. Consider the following advice and implied interpretation developed by Achieve:

> "States should be judicious about adding content and keep in mind the possible implications of doing so. Remember, a central driver in the creation of the CCSS was to develop standards that were common across states lines..." and that "the standards added... [should] have minimal implications on a common assessment effort." (Achieve, 2010, pp. 1, 2)

A final question related to assessment innovation and the option for states to add whatever percentage of something in creating their assessments is: Why would any state want to do it?" The incentives here would all appear to press toward even greater standardization. For example, consider a state that was interested in adding 15% of something to the particular CCA it adopted. It is not clear yet, of course, how the cost of participation in a CCA will be addressed, but the two clearest options that capture the extremes would include: a) each state pays some amount per student to participate in the CCA; or b) the CCAs are provided free of charge to

participating states. (For purposes here, "participation" is intended to cover all traditional aspects of testing that would be of-cost to a state; that is, test development, test form printing (if any), scoring, and reporting.)

Now, if states are required to pay per-student costs to participate in an operational CCA program, there would seem to be a strong disincentive for any state to also pay for the parallel development, administration, scoring, and reporting of the optional "15% test." The disincentive is strong because the costs for developing a state-specific 15% would not be that much less than what the state is currently spending to develop its own full-blown test currently. Because all of the same procedures must be followed (item development and review, pilot testing, administration, scoring, documentation, etc.), there is little savings realized in producing a "15% test" compared to a "100% test." It seems inconceivable that a state would *increase* its allocation to educational testing by paying for both CCAs and custom state assessments. In difficult economic circumstances, the motivation to eschew any supplemental "15% test" would be very strong indeed. By the same token, even if the CCAs were provided to states at no cost, that would seem to create a strong impulse to then divert the state's current test development allocation to fund other pressing budget items.

AC5 - Validation

Although they must be viewed as intended–as rough road maps–and cannot be critiqued on the grounds that fine levels of detail were not included, the CCA proposals (SBAC, 2010; PARCC, 2010) are particularly skimpy on how validation of the CCAs is likely to occur. Outlines of intended alignment and other content validity evidence-gathering studies are presented in sufficient detail to have some confidence that the kinds of information typically

supporting a traditional standards-referenced testing program will be collected and analyzed. A significant concern arises, however, because of one game-changing fact: *the CCAs are not primarily intended to be standards-referenced assessments.*

Such a claim might seem odd given the crisp focus to date on crafting the grade- and subject-specific content standards that resulted in the CCSS in, say, English Language Arts and Mathematics. Despite this context, the new assessments depart significantly and intentionally from limited standards-referenced score interpretations, instead intending to support inferences about future success. The italicized portions of the following excerpts clearly illustrate the shift in intended inferences:

* "The Common Core State Standards define the rigorous skills and knowledge in English Language Arts and Mathematics that need to be effectively taught and learned for students to be *ready to succeed* academically in credit-bearing, college-entry courses and in workforce training programs." (CCSSO & NGA, nd, p. 1)

* "These standards define the knowledge and skills students should have within their K-12 education careers so that they will graduate high school *able to succeed* in entry-level, credit-bearing academic college courses and in workforce training programs." (Common Core Standards Initiative, nd, p. 1)

* "The Common Core State Standards represent what American students need to know and do *to be successful* in college and careers" (CCSSO & NGA, 2010c, p. 1)

* "Through this collaboration, 26 states will share a common assessment system that annually reports on students' progress toward the Partnership's primary goal--*college and career readiness...*" (PARCC, 2010, p. 18)

* "The overarching goal of the SBAC is to ensure that all students leave high school

prepared for postsecondary success in college or career..." (SBAC, 2010, p. 5)

The bottom line is that the stated purpose of supporting inferences related to readiness, future performance, likelihood of success in subsequent endeavors, and so on necessitates a major shift in the design and conduct of validation efforts from one of a (nearly) exclusively content-based evidence process to one of gathering predictive validity evidence. Whereas predictive validation procedures are used routinely to support tests in various employment, licensure and certification, college admissions and other testing contexts, predictive validation efforts have not routinely been applied in the K-12 achievement testing context. The CCA proposals contain modest suggestions for the beginnings of predictive validation evidence gathering, such as the regression-based standard setting method in one proposal (PARCC, 2010) and the intention to involve institutions of higher education and representatives from the workplace in the other (SBAC, 2010). Nonetheless, although the CCAs still retain a standards-referenced flavor, the new and clear predictive-inferential aspect will require new thinking and new approaches to validation in the K-12 assessment context before CCA scores can be confidently interpreted to mean what they are purported to represent.

AC6 - Diagnostic/Instructional Value

In contrast to the fairly consistent expenditures and support for implementing assessment programs tied to accountability reforms, there is a much less consistent or successful track record of investment in designing and delivering *planned assessment systems* (Cizek, 1995; Cizek & Rachor, 1994). As I have observed previously:

"The most striking deficiency of the standards-referenced testing movement is that

> resource allocation and emphasis on alignment and educationally useful information has been focused almost exclusively on large-scale assessments. That emphasis has yielded high-quality measures for gauging overall achievement, but little information that educators can use to meet the classroom instructional needs of individual students" (Cizek, 2007, p. 105)

In planned assessment systems, the distinct purposes of assessments are clearly articulated and integrated with other sources of information about student achievement, test scores are translated into actionable information, and assessment results are linked to specific instructional interventions that are relevant and effective vis a vis the specific strengths and weaknesses of individual students. Although many details remain to be worked out, there would appear to be only modest indications that the deficiency noted above will be adequately addressed. In my opinion, one such source of encouragement can be found in the SBAC proposal where it appears to be recognized that a key requirement for an effectively integrated instruction and assessment system will be the development and dissemination of "exemplary instructional materials linked to the CCSS" (SBAC, 2010, p. 6).

That glimmer notwithstanding, it appears that most of the work supported by the CCA grants will be toward producing better and more frequent assessments. That, of course is a worthwhile goal, and some realignment of resources to include greater emphasis on formative and benchmark assessments may well prove to be an effective means of improving student achievement. As I have also suggested previously, the strategy of using large-scale summative testing to leverage large leaps in learning has likely reached a point of diminishing returns, and alternatives, such as meaningful formative assessments, likely represent "the next best hope for stimulating gains in student achievement" (Cizek, 2010b, p. 3).

A prominent feature of both CCA proposals is the promise to deliver more such actionable, individualized, and diagnostic information than is yielded by current assessment systems–to go "beyond the bubble" to use the relevant rhetoric. Unfortunately, in contrast to what most educators have in mind when they ask for diagnostic assessments, the CCA proposals offer little hope that the next generation of assessments will be any more diagnostic than those currently administered by the states.

Truly diagnostic assessments would be longer, narrower in focus, more specific in content coverage, more reliable and more frequent. In some ways, the CCA proposals move toward these requirements only very modestly (e.g., with more discrete, roughly quarterly testing); in other ways, the proposals actually would move away from these requirements. To illustrate the very real potential for a *reduction* in useful information about student achievement, it is instructive to consider the proposals for test design. (For simplicity, all of the following considers only the Grade 3 English language arts components of each proposal; although design specifics differ across grades and subjects, the major design features are essentially the same.)

Both proposals include formative and summative assessment components. For example, a formative component in the PARCC (2010) proposal includes what are called "through-course" assessments, administered on roughly a quarterly basis throughout the school year. In Grade 3 English language arts, a typical through course assessment consists of only 1 or 2 on-demand, constructed response items. It is asserted that a student's response(s) to the 1 or 2 items will allow monitoring of student progress, support inferences regarding whether a student is "on track," inform instructional decisions, signal interventions for individual students, permit aggregation to assess teacher effectiveness (p. 78). On reading what was promised based on 3rd graders answers to one or two questions, I couldn't help but wonder how NIH peer reviewers

would evaluate a proposal that indicated 1 or 2 minutes of light exercise per quarter would lower LDL, elevate mood, prevent aging, increase social desirability, boost IQ, reverse hair loss and cure cancer.

Table 1 provides a summary of the designs for Grade 3 English language arts summative assessments from each of the CCA proposals. The PARCC (2010) proposal indicates that the summative assessment will contain 45-55 items. The specifications for that test roughly mirror those of current state assessments and would result in scores of approximately the same reliability and diagnostic information as current tests. The PARCC design also indicates however that the test will draw on higher order skills, assess hard-to-measure skills, use digital technologies to include innovative item types... [and] ...will be computer-based and computer-scored" (p. 82). Two concerns here are: 1) that the roughly equal item count but intended greater complexity of this design will prove both challenging and taxing for typical 3rd grade students, requiring even greater testing time for approximately the same diagnostic benefit; and 2) large-scale, computer-scored, every-pupil testing with one-week turnaround time for student scores as indicated in the proposal has not been accomplished to date in any statewide K-12 assessment context, much less on the scale represented by the several state members of the PARCC consortium. It is likely to require substantial investment beyond the $170 million already allocated to ramp up the technological and psychometric infrastructures to accomplish those aspirations.

The SBAC design proposal actually substantially *reduces* the achievement information provided by the summative assessment component. With a total of only 19 items/tasks, the SBAC assessment is substantially shorter than current, typical state assessments. Assuming that current tests have a length of approximately 50 items and a reliability of approximately .90, applying a common psychometric formula for estimating the reliability of a 19-item test yields a

reliability estimate of only .77–a level that would ordinarily be considered low for making any important decisions about students. The SBAC proposal also indicates that human scoring will be used with two-week turnaround for scores–another example of a worthwhile goal that has not as yet been shown to be routinely feasible in large-scale K-12 achievement testing contexts.

Finally, whereas it is true that each proposal indicates, in one way or another, that information from the summative assessment will be combined (in as yet unspecified ways) with other information sources to provide more diagnostic information, the state of the art in educational measurement lags far behind this aspiration and, in many circumstances, would caution against combining information sources that are as highly disparate as those contemplated for synthesis in the CCA proposals. A clear focus going forward should be on ensuring that the CCAs will not only be larger-scale in terms of sample sizes compared to current single-state assessment programs, but also truly larger-scale in informational value.

AC7 - Accountability and Consequences

Perhaps the biggest policy void at present is the relative absence of detail regarding how the planned CCAs will be incorporated into current accountability systems, or proposals for new accountability mechanisms. To be sure, each of the CCA proposals include mentions of how the results might be used to help monitor student achievement growth; equally certain is that accountability systems that rely more on individual and aggregated growth targets represent an improvement over status models and percentage targets such as those that were a feature of the NCLB system.

The lack of policy guidance for a next generation of accountability systems is a conspicuous void in the CCSS and CCA initiatives that must be addressed immediately. For

example, long-recognized deficiencies in current accountability systems, such as the asymmetry in consequences described by Porter and Chester (2002), are not discussed. It is also well-known that student motivation on assessments must be assured in order for the assessment results to be taken as valid measures of their real levels of knowledge and skill; at minimum, it would seem that a menu of policy options would need to be considered with respect to ensuring students engage in CCAs at least to the same extent that they engage in current accountability tests–and preferably to a much greater extent. Innovative test item formats, challenging tasks, and computer administration alone are not likely to provide the impulse necessary–particularly for older students–to elicit high levels of motivation.

To be fair, the "A" in CCA refers to assessment only. However, it has also been long recognized that to function most effectively and to have the most positive influence on educational practice and outcomes, standards, assessments, and accountability mechanisms should not developed in isolation, but as a planned, integrated system. There are essentially no specifics on linkages to accountability mechanisms in the CCA proposals, with the proposals only hinting in guarded terms that some voluntary accountability uses might be possible. For example, the PARCC proposal indicates: "States that consider results from the common assessment for school accountability and for teacher and principal evaluations will signal the kinds of good instructional practices that will help students meet the CCR standards" (2010, p. 37).

Conclusions

The development of CCAs aligned to clear, narrow, and teachable content standards represents a desirable next step in sustaining a the national policy focus on improving instruction

and achievement of American students. However, like its NCLB predecessor, the current CCSS and CCA reform initiatives have missing elements that fail to address key aspects necessary for the reforms to be as effective as they can be. The current reforms are at a critical juncture. At this time, the opportunities are greatest for integrating in a systematic and comprehensive manner the interrelated aspects of effective educational practice: instruction, assessment, and accountability. The timing and extent to which these aspects are recognized and addressed via both sound psychometric approaches and appropriate policy will significantly affect not only the efficacy of the common core strategy but the ultimate goal of improving the achievement of all students.

References

Achieve. (2010, August). *Adding to the common core: Addressing the "15%" guideline.* Washington, DC: Author. Retrieved from http://www.achieve.org/files/15Percent Guideline.pdf

Cizek, G. J. (1995). The big picture in assessment and who ought to have it. *Phi Delta Kappan, 77*(3), 246-249.

Cizek, G. J. (2001, March). *High-stakes testing and accountability systems: Unintended consequences, unrecognized benefits.* Invited presentation to the New York State Education Department Conference on Validity in Testing, Albany, NY.

Cizek, G. J. (2007). Formative classroom assessment and large-scale assessment: Implications for future research and development. In J. A. McMillan (Ed.), *Formative classroom assessment* (pp. 99-115). New York: Teachers College Press.

Cizek, G. J. (2010a, 30 March). Eight questions for Gregory Cizek: On the role of testing in America's education system. *The Economist*. Retrieved from: http://www.economist.com/blogs/democracyinamerica/2010/03/testing_and_assessment

Cizek, G. J. (2010b). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. Cizek, (eds.) *Handbook of formative assessment* (pp. 3-17). New York: Taylor and Francis.

Cizek, G. J. & Rachor, R. E. (1994). The real testing bias: The role of values in educational assessment. *NASSP Bulletin, 78*(560), 83-93.

Common Core State Standards Initiative. (nd). *About the standards.* Retrieved from http://www.corestandards.org/about-the-standards

Council of Chief State School Officers [CCSSO] & National Governors Association [NGA] Center for Best Practices. (nd). *Common core state standards initiative standards-setting criteria.* Retrieved from http://www.corestandards.org/assets/Criteria.pdf

Council of Chief State School Officers [CCSSO] & National Governors Association [NGA] Center for Best Practices. (2010a). *The standards: English language arts standards.* Retrieved from: http://www.corestandards.org/the-standards/english-language-arts-standards

Council of Chief State School Officers [CCSSO] & National Governors Association [NGA] Center for Best Practices. (2010b). *The standards: Mathematics.* Retrieved from: http://www.corestandards.org/the-standards/mathematics

Council of Chief State School Officers [CCSSO] & National Governors Association [NGA] Center for Best Practices. (2010c, June). *Reaching higher: The common core state standards validation committee; A report from the National Governors Association Center for Best Practices & the Council of Chief State School Officers.* Washington, DC: National Governors Association.

Iowa Department of Education. (nd). *Common core state standards & Iowa fact sheet.* Des Moines, IA: Author. Retrieved from http://www.iowa.gov/educate/index.php?option=com_ docman&task=doc_download&gid=9935&Itemid=1507

Improving America's Schools Act. (1994). P. L. 103-382, 20 USC 8001.

No Child Left Behind Act. (2001). P. L. 107-110, 20 U.S.C. 6301.

Partnership for Assessment of Readiness of College and Careers. (2010, May 28). *Race to the top assessment program application for new grants.* Tallahassee, FL: Author.

Porter, A. C., & Chester, M. (2002). Building a high-quality assessment and accountability program: The Philadelphia example. In D. Ravitch (Ed.), *Brookings papers on education policy 2002* (pp. 285-337). Baltimore, MD: Brookings Institution Press.

Shepard, L., Glaser, R., Linn, R., & Bohrnstedt, G. (1993). Setting performance standards for student achievement: A report of the National Academy of Education Panel on the evaluation of the NAEP trial state assessment. Stanford, CA: National Academy of Education.

Smarter Balanced Assessment Consortium. (2010, May). *Race to the top assessment program application for new grants.* Olympia, WA: Author.

Table 1

Grade 3 English Language Arts Summative Assessment Designs [1]

CONSORTIUM	ASSESSMENT DESIGN SPECIFICATIONS
SBAC	6 selected-response items
	6 technology-enhanced items
	2 on-demand writing prompts
	1 performance event
Total Items/Tasks	15
PARCC	45-55 items [2]
Total Items/Tasks	45-55

Notes:

1. Includes reading and writing ELA components only; other skills such as speaking, listening, viewing, etc., are not represented. For SBAC design, inclusion of speaking and listening adds 4 items.

2. The PARCC proposal indicates that the 45-55 items "will draw on higher order skills, assess hard-to-measure skills, use digital technologies to include innovative item types..." The assessment will be computer-based and computer-scored" with "one week turnaround time" for scores (p. 82).