New Student Assessments and Advancing Teaching as a Results-Oriented Profession

A White Paper Developed for the Race to the Top Assessment Consortia

BARNETT BERRY & ALESHA DAUGHTREY
WITH RENEE MOORE, DAVID ORPHAL AND MARSHA RATZEL

Executive Summary

The Common Core State Standards Initiative (CCSSI) and the emerging efforts of the two Race to the Top (R2T) assessment consortia offer an enormous opportunity to leverage higher academic expectations for our nation's public school students as well as transform teacher evaluation. Counter to prevailing media accounts, researchers have documented—for decades—the myriad of organizational, technical, and political problems associated with how teachers are assessed. The call to turn teacher evaluation systems on its head and focus on student results is needed, but not new.

Although value-added models (VAMs) are fast becoming the preferred method to estimate the effects of individual teachers on gains in student achievement, even highly accomplished teachers like our classroom-based co-authors Renee Moore, David Orphal and Marsha Ratzel who embrace accountability are skeptical of using VAMs as a central measure of their effectiveness. They express concerns about the narrowness of what they measure and reports from research of significant error rates.

However, teachers call for the *strategic* use of value-added data, with the models' limitations in mind—and request the engagement of expert teachers in efforts to sharpen those models and their underlying student assessments to improve accountability systems in ways that support more effective teaching. The consortia must offer opportunities for teacher leaders to build and use new student tests for assessing teaching effectiveness, drawing on the classroom context in which they teach. Without such teacher leadership opportunities, the

Dr. Barnett Berry is President and founder of the Center for Teaching Quality, Inc. (CTQ), a nonprofit organization working to advance the teaching profession by cultivating teacher leadership, conducting research, and crafting policy. Alesha Daughtrey is Director of Research and Policy at CTQ.

This paper was produced in partnership with Arabella Advisors. Arabella Advisors is a philanthropy consulting firm supporting the efforts of individual, family, institutional, and corporate donors worldwide. We are committed to unbiased analysis that helps donors support issues and nonprofits with confidence. Our expertise and insights transform philanthropic goals into results. For more information on our firm, please visit our website at www.arabellaadvisors.com.

United States will not create and sustain the internationally benchmarked curriculum and assessment systems that can drive 21st century teaching and learning

This paper, developed with the on-the-ground insights from three highly accomplished teacher co-authors, raises concerns about the inadequate contextual knowledge possessed by those who create and administer evaluation instruments, and about the documented effects on teacher effectiveness and student achievement of certain school working conditions. But if policymakers can begin to develop ways for teachers to own the standards and assessments and offer time and training for implementation with fidelity and rigor, then classroom practitioners will not only embrace 21st century school reforms, they will help lead them. Investing in formative assessments, and valuing (and compensating) teacher leaders are key levers for change. So are using assessment data as tools, not carrots or sticks. Finally, states and districts must invest in serious professional preparation (in meaningful pre- and in-service programs) that prepares teachers specifically as expert leaders in assessment and evaluation, and give them the time and tools they need to execute that leadership.

We conclude that the vast majority of teaching professionals are not questioning the idea that individual teachers need to be evaluated rigorously. If anything, teachers want more, and more accurate, feedback than they currently receive in order to inform their instruction. But if policymakers do not acknowledge the importance of the issues raised here, they will not address a central concern of teachers who are seeking ways to improve teaching and learning through authentic classroom assessment, regular collaboration, and the sharing of expertise—in order to create a results-oriented profession. The first step must be to tap into the assessment expertise of many current teachers today and cultivate the knowledge and skills of others, now and in the future.

RECOMMENDATIONS

1. Leaders of the Common Core State Standards initiative must work with teachers to clearly define the assessment expertise needed by classroom practitioners, drawing on experts who already have the skill to lead the reforms.

2. Key universities and school districts in targeted states must work with teachers to develop the pre- and in-service professional preparation curriculum needed to cultivate assessment leaders who can implement the Common Core Standards.

3. States and the assessment consortia must work with teachers to develop templates for districts to use in organizing time and tools for classroom practitioners to engage in assessment and teacher evaluation reforms.

4. All parties connected with implementation—from leaders of the national consortia to school administrators—must create opportunities for large numbers of accomplished teachers to lead the development and implementation of Common Core-aligned tools, leveraging their expertise and engaging them as arbiters of results-oriented teaching evaluation systems. **We call for 30,000 classroom-based, assessment experts by 2014—one-tenth of those needed to lead assessment reforms well into the 21st century.**

Introduction

In 2010, 48 states[1] signed on to the Common Core State Standards initiative, which calls for fewer, clearer, and higher academic expectations so that all students are prepared to succeed in our global economy and society. At the same time, the federal government's $4 billion Race to the Top initiative pushes states to more precisely assess teacher effectiveness, using standardized tests and statistical tools to determine who is successful in raising student achievement. As part of the RTTT efforts, the federal government is supporting two multistate consortia, the Partnership for Assessment of Readiness for Colleges and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC), to transform how students are tested for academic progress as well as how educators can use the data to improve teaching and learning.

This aggressive policy push is understandable. Students from the United States have fared poorly on the Programme for International Student Assessment, ranking 35th among the top 40 countries in mathematics and 29th in science in 2006. The achievement gap is much wider in the United States than in the most highly ranked nations, and in each disciplinary area tested, U.S. students, compared to their international counterparts, scored lowest on the problem-solving items.[2]

Today, new reports indicate that U.S. teacher evaluation systems are in woeful condition, with 99 percent of teachers in a given district often rated "satisfactory" and excellence seldom identified. In this context, enthusiasm for using test score gains to determine who is effective is no surprise.[3] However, unlike what is commonly portrayed in the media, the organizational, technical, and political problems of teaching have been long-standing. For example, in 1984, Arthur Wise and Linda Darling-Hammond, then social scientists at the RAND Corporation, carefully documented how administrators "lacked sufficient resolve and competence to evaluate [teachers] accurately."[4] A decade earlier, other researchers pointed to how teachers—many of whom were seeking guidance on how to improve their teaching—saw "nothing to be gained from evaluation."[5]

The achievement gap is much wider in the United States than in the most highly ranked nations, and in each disciplinary area tested, U.S. students, compared to their international counterparts, scored lowest on the problem-solving items.

Today, with a strong push from the Obama administration's Race to the Top initiative, states and school districts are expected to develop and use results-oriented teaching evaluation systems that differentiate more and less effective teachers. Over the past decade, there has been growing support to draw on statistical data, generated by "value-added" estimates of individual teachers based on annual standardized achievement tests of their students. Different researchers estimate teacher effects in a variety of ways, but their formula usually includes some variant of a "change score" from one year to the next and making statistical adjustments to account for differences in student learning that might result from home background.

Currently, value-added models have surfaced as many policy makers' preferred method to estimate the effects of individual teachers on gains in student achievement, since they are (at least in theory) based on growth rather than status measures and can include controls for various school or student characteristics. However, even highly accomplished teachers who embrace accountability are skeptical of using value-added methods to determine their effectiveness, no matter how sophisticated those models may be. This perspective is understandable, given that there has been mixed empirical evidence about the stability of these measures.[6]

Based on our experience as researchers and classroom teachers, we believe that forcing teachers to embrace the use of value-added data in either/or terms is counterproductive. We believe that the vast majority of teachers will embrace a results-oriented evaluation system—and the strategic use of value-added data—under certain conditions. This means engaging expert teachers in efforts to sharpen those models and their underlying assessments so that they are more valid and reliable tools for evaluating student growth and teacher effectiveness. It does not mean simply having some high-profile or award-winning teachers to sign off or act as advocates of the tests, without making any substantive change or involvement in their design, implementation, or analysis. This more measured and inclusive approach will not only encourage teachers to embrace the much-needed student assessment and teaching evaluation reforms, but ensure that they are implemented with veracity. (See Appendix A, capturing a virtual conversation with our three teacher colleagues.)

We hope to see student results used effectively in the evaluation of teaching and teachers, but we must do so without identifying too many false positives and false negatives and pushing classroom practitioners to "game the system." As the eminent social scientist Donald Campbell said many decades ago: "The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." Evidence abounds in other professions, such as in medicine where "cardiac surgeons, held accountable for raising surgical survival rates, refuse to operate on the sickest patients most in need of intervention."[7] But this does not mean clear indicators of success need to be generated and used in assessing teaching effectiveness. We believe that if more opportunities are made available for teachers to understand and use the results, in light of both the assessments upon which they are built and the classroom context in which they are generated, many more of them will be responsive to the evidence.

Understanding Concerns about Value-Added Measures

Weak Data Systems to Support Value-Added Models

The reasons that teachers (and other education experts) hesitate to immediately implement value-added measures that rely on *current* student assessment tools are as complicated as the models themselves. The first challenge is the lack of comparable and linked student and teacher data for every grade level and subject, making it impossible to apply value-added methodologies uniformly to all teachers. In addition, a number of value-addded models rest on the assumption that teachers and students are randomly assigned. In fact, they are not, and the statistical models typically do not adequately control for this fact. Furthermore, to judge a teacher's value-added effects, test data must be available for sufficient numbers of students for full school years, but in low-income urban schools with highly mobile student populations, this is not often possible.

As a consequence of these shortcomings, some researchers have found that virtually all value-added models are unstable in distinguishing among teachers in the middle ranges of performance. Others have discovered that the same teacher's effectiveness rating can change depending on the school where the teacher is working. In fact, a recent U.S. Department of Education report revealed that in using value-added models to judge teachers, about 25 percent of the teachers would be classified incorrectly as either effective or ineffective.[8] Even as many school districts are now rushing to use value-added data to publicly identify effective and ineffective teachers, the *New York Times*, which had been promoting use of the data on its editorial pages, pointed to the serious technical "hurdles" that need to be cleared if classroom practitioners can be accurately rated using the methodology.[9]

The current tests used to assess student learning have not been designed so that teachers can be assessed fairly on how much they help students learn content in the same subject area over time.

Problems with Underlying Assessments

The current tests used to assess student learning have not been designed so that teachers can be assessed fairly on how much they help students learn content in the same subject area over time. For example, most teachers teach students whose academic proficiencies span many grade

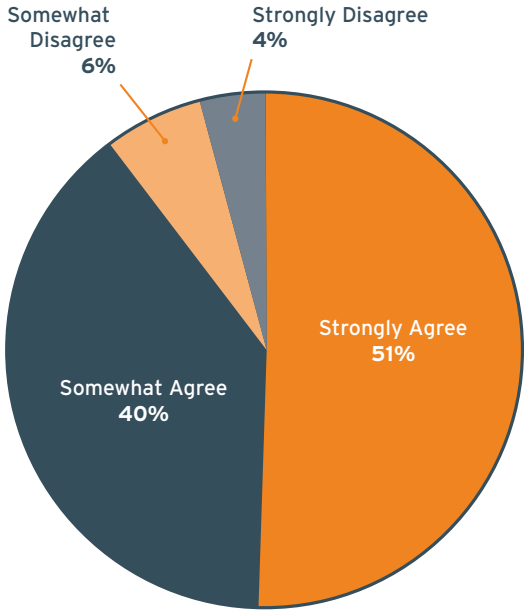
levels, yet the tests upon which they are judged assess only a narrow band of content. As a result, even value-added tools cannot adequately adjust for the seventh-grade English/language arts teacher whose students' reading skills range from grades 3 to 12, when the standardized test they take has mostly seventh-grade items. The underlying assessments on which the value-added models are based simply lack sufficient "stretch." With only a few test items benchmarked for sixth- and eighth-grade proficiency, the test does not come close to reaching the level of sophistication needed to accurately measure the students' academic growth. As one school principal interviewed as part of the *New York Times* article on using value-added models noted: "If I thought they gave accurate information, I would take them more seriously. But some of my best teachers have the absolute worst scores."[10]

Balancing Individual and Collaborative Teacher Effects

What's more, many students are taught the same concept or skill by multiple teachers. For example, it is common in many schools for teachers of history and English to teach academic writing skills, and for both science and math teachers to teach quantitative reasoning. Many teachers now share the teaching of students in the upper elementary grades, both across subjects and in co-teaching models—and thus are not sure who is statistically responsible for raising student achievement in a given class. Indeed, the recent *New York Times* article pointed to the fact that two teachers successfully team-teach a fourth-grade class in Park Slope, Brooklyn, but the statisticians who ran the teacher effects estimates credited only one of them. The other was not even identified as a teacher in the school.

As a result, it is no wonder that many teachers raise serious questions about using value-added methods to identify individual teaching effectiveness. In fact, more than 90 percent of the nation's teachers report that their colleagues contribute to their teaching effectiveness[11] (figure 1). New teachers, in particular, were more likely to strongly agree that their success in the classroom hinged on the effectiveness of others.

FIGURE 1: TEACHERS WHO AGREE THAT "OTHER TEACHERS CONTRIBUTE TO MY SUCCESS IN THE CLASSROOM"



These teacher sentiments have been validated by scholarly evidence. In a recent study, using 11 years of matched teacher and student achievement data, researchers were able to isolate and quantify the added value generated by such collective expertise. They found that most value-added gains are attributable to teachers who are more experienced and better qualified, and who stay together as teams within their schools. Drawing on sophisticated analyses, the researchers found that peer learning among small groups of teachers seems to be the most powerful predictor of improved student achievement over time[12] (table 1). *Education Week*, in reporting on this groundbreaking research, concluded: "[T]eachers raise their games when the quality of their colleagues improves."[13]

TABLE 1:
DEVELOPMENT OF SHARED EXPERTISE LEADS TO SIGNIFICANT STUDENT GAINS

	Reading Score Impacts (in standard deviations)	Math Score Impacts (in standard deviations)
Teachers' collaboration with effective peers correlates with increase in students' achievement.	+2.6%	+4.0%
Teachers' value-added scores remain higher even years after active collaboration with effective teachers ends.	+7.2%	+7.8%

SOURCE: C. Kirabo Jackson and Elias Bruegmann, "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal*: Applied Economics 1(4) (2009).

Our co-author, Renee Moore, a Carnegie Scholar and 20-year English teacher in the Mississippi Delta, points out the educational conundrum emerging from these findings: balancing the tension between systems focused on individual teachers' value-added scores and trends toward greater collaboration among teachers. Teaching professionals are not questioning the idea that individual teachers need to be assessed. But, Moore notes, if policy makers do not acknowledge the importance of this issue, they will not address a central concern of teachers who are seeking ways to improve teaching and learning, including collaboration, in order to create a results-oriented profession.

Lack of Context to Inform Interpretations of Value-Added Data

Many teachers do not trust those who are making the decisions about their effectiveness. Researchers who generate value-added effects for schools and teachers may be better suited to make summative judgments about the former, but not the latter. All too often, an individual teacher's effectiveness cannot be interpreted unless more information is made available about the context of a specific class (including the circumstances of individuals and groups of students in a given class, known only to those working in a school). And while many school administrators do have access to knowledge of context within a classroom, they often have insufficient time or skill in a teacher's specific subject area and how to teach that content to make fine-tuned decisions.

School Conditions that Undermine the Results We Hope to Measure

Finally, while many teachers embrace accountability for results, they recognize that their effectiveness is in large part due to the conditions under which they work. With the support of the Ford Foundation, the Center for Teaching Quality convened a TeacherSolutions team of 14 accomplished teachers to embark on a one-year study of school conditions and their impacts on teaching effectiveness in high-needs contexts. Their recent report pointed to a number of working conditions that help determine teaching effectiveness. For example, teacher mis-assignments are a common problem in high-needs schools. In times of shortages, elementary-trained teachers, with weak math backgrounds, are forced to teach middle school algebra. And then there are effective first-grade

teachers who, by administrative fiat, are assigned to the fifth grade and given little time or support to prepare to teach a new curriculum to much older students.[14]

In addition, teachers need more time to work with each other in using student data, designing instruction, and refining their lessons. They need more opportunities to integrate what they teach between 7:30 a.m. and 3 p.m. with the growing opportunities their students have in extended day and year programs. They need tools to cope with growing numbers of highly mobile students, to ensure they can teach effectively even when those whom they teach enter and exit their classrooms throughout the year. They need new skills and strategies to work with parents—their students' first teachers—even when there is less time to do so. Along with more rigorous standards, teachers must have access to instructional materials that align to the standards and must be offered resources, tools, and time to adjust their classroom teaching. Some analysts and reformers are beginning to recognize that while more rigorous core content standards are essential to college and career preparation, they are inadequate by themselves. Advancing teaching as a results-oriented profession will also require new student assessments and a much greater commitment to continuous professional growth.

Teachers can embrace the use of the Common Core Standards and new student assessments as part of a more robust and rigorous evaluation if they can "own" the results in ways that improve their instruction. Doing so will require new investments in teacher education as well as leadership opportunities for classroom practitioners, in and out of cyberspace, to transform teaching evaluation as part of a results-oriented profession.

Educators Owning the Standards and Assessments

Most teachers, if engaged, readily would understand the need for the Common Core Standards, given the globalized economy and the demands on the 21st-century workforce, coupled with disparate curriculum standards across states and increasing student mobility. We suspect that most teachers, at first glance, would find little to criticize in the Anchor Standards of the Common Core—for instance, the expectation in English/language arts standards that students must be able to "determine central ideas or themes of a text and analyze their development" and "assess how point of view or purpose shapes the content and style of a text." Indeed, most accomplished teachers would tell you that this is what they already do in their classrooms, and they would expect their colleagues to do so also.

Most also would be sanguine about the PARCC and SBAC approach to assessment that will combine on-demand (sit-down test) measures with curriculum-embedded performance tasks—a system used widely in high-achieving countries. Linda Darling-Hammond, one of the intellectual architects of the SBAC approach, has called for curriculum-embedded tasks to count for 20–50 percent of the total student score. Teachers believe that performance tasks, such as the English/language arts sample from Ohio (next page), can tell them a lot more about what students know and why. We also are cautiously optimistic about the SBAC proposal to create state-of-the-art adaptive online exams, using open source technology that offers teachers a "variety of tools, processes and practices to use in planning and implementing informal, ongoing assessment."[15] However, teachers have noted several issues to which policy makers, and the consortia, should attend in order to increase educators' role in implementing the Common Core Standards.

Ohio Performance Assessment Project English Language Arts Performance Task

Imagine that you are editing an online digital anthology for 12–12th graders entitled "Perspectives on the American Dream." Your job is to prepare the introduction to this anthology. In your introduction, please do the following things:

a) Decide which texts you want to include and in which order (you must include at least six texts). Texts can include books, poems, songs, short stories, essays, photographs, articles, films, television shows, or Internet media. The six texts must represent at least two different perspectives and must include at least two different types of text (e.g., print text, visual media, multimedia, digital media). At least two of the six texts must be print (written) texts.

b) Identify and discuss perspectives on the American dream represented in the six texts you selected.

c) Write a short paragraph about each text, in which you make clear why you have included it and how it relates to the other texts in your anthology.

d) Propose a set of questions to focus readers as they consider the perspectives represented in these texts.

Teachers' Input into Development and Adoption of New Standards

To date, teachers have not been party to creating the standards that will be implemented. Other professionals, such as medical and legal practitioners, typically play leading roles in developing standards for their professions, and teachers want to act as similarly capable professionals in their own field. Renee Moore notes that teachers are, however, not of one mind, and have not been informed, much less consulted, to date:

> *Teachers' reactions to the Common Core are spread across a wide spectrum—and understandably so. After all, we are a very large and diverse group of professionals. Policy makers must recognize that some teachers don't even know there is a Common Core, even though their states have already signed on to use them.*

Leaders of the Common Core Standards did seek input from a variety of "teacher groups," and classroom practitioners had opportunities to review the standards. But as Renee noted:

> *Policy makers need to recognize that many teachers are not satisfied with the level of participation in the review process and how their input was handled or used, if at all.*

Time and Training to Implement with Fidelity and Rigor

Effective teachers have deep and unique knowledge of the content they teach, the pedagogical tools needed to help students learn that content, and the needs of their specific students and communities. They are thus best placed to lead educational innovation and implementation, including of the Common Core and its related curricular and assessment components. To do this well, they need the time and tools to revise instructional materials, unit and lesson plans, and access to the assessment instruments accordingly.

David Orphal, a social studies teacher of 15 years from inner city Oakland, California, made it very clear that more time and training are needed if teachers are to design, implement, and teach curricula that support the new Common Core Standards. He explains:

> *Teachers need time to analyze the standards in collaborative teams and determine how they can be broken down into several small steps that lead logically and sequentially to the main goal. While state or local district education officials will perhaps design the summative assessments, teachers will need to learn how to design formative assessments, as well as daily activities and educational tools that will allow students the opportunity to gain or practice the knowledge and skills.*

Marsha Ratzel, a veteran middle school science and math teacher from Kansas, recognized the serious need for formative assessment training among teachers today and the potential of more open-ended performance assessments to facilitate embedded professional development and leadership opportunities for teachers:

> *A real opportunity presents itself in the field-testing of these assessments because teachers can play a vital role in identifying the strengths and weaknesses of the tasks. Teacher collaboration could be used to not only strengthen and improve the tasks but to serve as a basis for the kind of professional development, delivered in the context of actual classroom practice, that is so needed for all of us around formative assessments.*

Serious Recognition of Teacher Knowledge in Implementing Assessment Reforms

Renee Moore pointed out that it is not just the assessment systems that need to be developed:

> *Up until this point, I haven't seen any discussion of how and when curricula will be created. We've gone from the Common Core to the assessments without any discussion of the time needed to create curricula. That's important and the part in which teachers must really be involved. Are they saying we can just do this as we go along?*

Moore's concern is well-placed and was echoed by prominent researchers and assessment experts with whom we spoke in the development of this paper. States must be very clear about not simply layering Common Core goals and tools on top of what is already in place, but rather carefully overhauling *all* curricula, instructional programs, and assessment and evaluation systems to ensure that teachers are not working under conflicting or confusing standards. This process requires considerable time and a disciplined focus from state- and district-level administrators and policy makers; without it, the layering will certainly occur. Researchers have documented the failures of past school reforms, in large part due to policy makers and administrators not providing time for teachers to learn to use new practices while also adding to teachers' workloads without taking anything away from their assigned tasks.[16]

It's challenging, but our schools can meet the best practice demands that our teaching colleagues outline. For example, Nebraska's unique School-based Teacher-led Assessment and Reporting System (STARS) created a statewide policy framework that honored local teaching and assessments, which accurately measured and reported student performance on state

content standards. From 2000 to 2008, each school district developed its own system, with rigorous oversight from educators themselves. Standardized test scores rose during this period. Most important, teacher knowledge and use of more fine-grained student assessments also rose, as did the quality of the formative tests designed locally.[17] In part, Nebraska developed expertise within the teaching profession by offering graduate-level college coursework for practicing teachers who conducted action research projects focused on the standards and assessments. In 2007, *Time* magazine praised Nebraska for "leaving no child behind" with teacher-driven student assessments.[18] Unfortunately, the Nebraska legislature ended the program in light of pressures for a more standardized system of assessment that did not have to rely on knowledgeable and empowered teachers.

For an international comparison, consider Singapore, one of the top-performing nations on internationally benchmarked assessments. Expert teachers there are given time to serve as mentors to help new recruits learn to teach, including how to examine student work and review test data. The government pays for 100 hours of professional development each year for all teachers. Critically, most teachers teach students only during half of a typical working day; as a result, they have time to visit one another's classrooms as well as develop assessment skills. In addition, the country requires its teachers to "Master the Art of Authentic Assessments" by systematically learning how to analyze the quality of assessment tasks and student work using standards-based criteria.[19] Singapore is well known now for its investments in the teaching profession (including pre-service preparation) and respect for teacher leaders.[20]

Policy makers at all levels must allocate enough resources to develop appropriate systems for scoring new assessments and for actually completing that scoring each year. Even after all curricula are aligned and new assessments are created and validated, they still must be scored. In the 1990s, Kentucky and Vermont attempted to implement new assessment systems that moved toward assessment of higher-order skills through the use of less standardized student responses via essay prompts or performance tasks. But, in part because policy makers failed to reckon with the cost of taking such work to scale, the programs ended before their promise could be realized. These stories offer a critical cautionary tale for the consortia seeking assessment reforms today.

Formative Assessments Are the Key

The key to teachers "owning" the Common Core Standards and the new assessments will be to focus not just on the summative, on-demand tests, but also on the formative ones and the curricula that underlie them, says Marsha Ratzel:

> *The first part of owning something is to tear it apart into its component pieces. Teachers must have opportunities to dissect the standards until they understand all the underlying and foundational skills, knowledge and processes. The second part is to develop curricula that map to those assessment tools. All this must then be customized for individual classrooms and students. Finally, teachers debrief with colleagues, using data to determine the next steps. If this kind of careful planning is utilized and the formative feedback is used to drive the instructional pacing, then teachers will embrace the evidence and the assessments and standards.*

In other nations, a wide range of classroom- and school-based evidence—from exams, projects, "learning records," and portfolios—is used

formatively, but the evidence is also audited and moderated by practitioners who use standardized criteria for assembling and judging work products.

Renee Moore goes further, explaining why formative assessments are so important to teachers:

> *Suppose for the sake of argument, student learning is relatively linear—although I have serious doubts about that assertion. There is still no evidence that the learning occurs at the same rate for all students. Regular assessments of individual students over extended periods should reveal each student's personal trajectory toward a specific set of standards and benchmarks.*
>
> *In my classroom I don't generally write my assessments first. I study the curriculum and standards and decide what I want the students to know or do. Then I plan the teaching and learning activities that I believe will get the students there. I stop to consider all the variations among students, who will need different approaches, accommodations, supports, etc. We plan backwards from the standards, and not from the tests.*

Effective, Compensated Teachers Leading Common Core Implementation

Renee Moore, in reviewing the approved SBAC framework submitted to the U.S. Department of Education, asked:

> *Their proposal calls for 'teachers' roles (to) include the construction and review of items/tasks, the definition of scoring guides, selection of student work exemplars, and scoring. This is admirable, but how will other teachers know that teachers were included in these roles and on the basis of what criteria? How many such roles will be available, and will the teachers be properly compensated for this work? These issues are keys to getting teachers to accept the standards and the assessments.*

Again, compensation may mean paying teachers more for their professional leadership in developing assessments, training teachers to use and score them, and having teachers serve as external raters to ensure reliability in the scoring. Additional time to perform this work—whether through part-time or temporary full-time release from classroom work—could be another means of "compensation."

Assessment Data as Tools, Not Carrots or Sticks

Marsha Ratzel argues that formative assessments are lower-profile tools, but tools of the highest importance and usefulness for teachers:

> *It is not the summative assessments that give teachers powerful information that fuels student learning… it is the immediate impact that formative assessments have on the daily instructional sequence and the power to deliver instruction that is based on what students already know and what they still need to learn.*

Renee Moore also pointed to how classroom practitioners might embrace new forms of student results:

> *I would suggest that one important component of new assessments for students is that the results not just be given to teachers, but that teachers in various configu-*

rations be involved in the interpretation of test data before it is released to either students or the general public. This type of data interpretation would be, in fact, a form of de facto peer evaluation.

We agree that access to such rich formative data would primarily be important to teachers as a tool for strengthening their professional practice. We also assert that supplying such tools to teachers could encourage them to see the importance of using value-added data and stronger assessments, thus improving their understanding and use of the Common Core Standards.

Investing in Teacher Education

Developing the requisite skills for teachers to engage in assessment and evaluation work also means going beyond narrow definitions of implementation needs, to include new investments in teacher education, both pre-service and in-service. About 10 years ago, scholars writing for the National Academy of Sciences released a hard-hitting report on student learning and assessment. Their conclusions were clear: The future of student assessments was, just as our colleagues have described, tools for students and teachers to improve teaching and learning.[21]

We know, as described in detail in *Teaching 2030*, a new book written by Barnett Berry and a TeacherSolutions team, that teachers must be prepared with the ability to find and adapt new technologies to engage iGeneration students. Teachers must also work across traditional subject area lines to broker 21st-century project-learning opportunities for students; they must be able to use data and evidence to inform their practice, and they must possess the skills to work in both virtual learning environments and brick-and-mortar schools.[22]

Policy makers and teacher preparation programs—both traditional higher education programs and alternate-route pathways—have done little so far in advancing the assessment knowledge of future and current teachers. In fact, the U.S. Department of Education, with its Race to the Top framework, may be undermining much-needed efforts in this regard by advancing alternative certification programs that encourage states to lower the bar for what teachers need to know and be able to do before they enter the classroom. Other countries known for educational excellence, such as Finland and Singapore, offer extensive, graduate-level study in assessment and other "education sciences" alongside preparation for instruction as well as the time (once they are in a classroom) to put these skills to work.

In reflecting on the kind of teacher education and professional development system needed for the Common Core Standards and the new assessments, David Orphal suggested:

Every teacher-preparation program should have a position on staff for several teaching professionals with assessment expertise, as a one- or two-year professorial role that would then rotate to another local teacher in order to spread the sharing of expertise. The teacher-fellow would help to design and implement the kinds of assessment experiences needed to get teachers ready to work with Common Core assessments.

These preparation and professional development innovations are important because teachers will need to possess a fair amount of technical knowledge regarding scoring and calibrating tests and tasks. They will need to know more about how to judge, with their colleagues, whether their students met standards and why (or why not). New testing technologies will help them make these judgments aca-

demically as well as emotionally and socially in the context of the specific students they teach. Moreover, offering such training and support to teachers, both pre-service and in-service, may increase not only their skills but also their confidence, and thus allow them to embrace and own the Common Core Standards assessments more readily.

Teacher Leadership for New Evaluations of Teaching

For teachers to embrace the Common Core Standards and a results-oriented teaching evaluation system, they must be afforded plentiful opportunities to lead innovations in their profession. While teacher leadership is sometimes discussed in school reform circles, and occasionally touted, it is rarely enacted in any serious way. A literature search can surface innumerable cases of how teachers develop their *leadership legs* and work toward personal self-renewal and school change by collaborating with their colleagues, yet most often these collaborations emerge in spite of the dysfunctional system in which the teachers work.

A number of recent reform efforts have focused on a new brand of school leadership in which authority and responsibility are distributed among administrators and teachers alike. But these systems often rectify long-standing organizational hierarchies where teacher leadership continues to be viewed through the prism of in-classroom support, professional development activities, and indirect support to instruction. Most leadership activities are still best classified as *tasks du jour*, and as such, teachers are expected to represent administrative purview, not to transform schools.[23] There have been few efforts, as we call for in *Teaching 2030*, to "blur the lines of distinction between those who lead schools and those who teach in them."[24]

We do not believe teachers will embrace the kinds of results-oriented teaching evaluation systems imagined by advocates of reform until those distinctions do indeed begin to blur. And classroom professionals are ready to lead. The U.S. teaching profession is large—well over 3.5 million strong—and highly diverse. A number of years ago, one researcher discovered that teachers fell into four categories: the "cynics," the "sleepy people," the "yes-but" people, and the "leaders"—with rarely more than 25 percent falling into the leadership group.[25] But even if that estimate were off by twofold, there is still likely a corps of more than 300,000 teachers ready to lead curriculum, assessment, and evaluation reforms—including a smart and sustainable system of implementation for Common Core-aligned curricular and assessment tools.

The most powerful teacher evaluation systems—ones that will push and challenge all of our nation's classroom practitioners—will be those that require teachers to explain who learned and who did not, and why, and what they will do next to ensure that all students are growing academically.

Elements of what we envision can be found in teacher-led schools like the Math and Science Leadership Academy (MSLA) in the Denver Public Schools. At MSLA, multiple measures are used to assess student progress, and peer assistance review is used to judge teachers and their effectiveness. Indeed, teachers make all the decisions at MSLA, including which applicants for teaching positions are effective and collegial enough to be invited to join the staff; how they will prioritize resource needs to serve their students; and how they will organize time during the school day for instruction, planning, collaboration, data analysis, and construction of assessments. MSLA, in all these respects, functions as a best practice model for many of the innovations and changes that we recommend—all within the context of a regular public school.

In the 21st century, the effective teacher is not someone who can be credited just for generating a value-added gain for the majority of that teacher's students. An effective teacher is someone who can interpret a wide range of evidence of student learning and act on it. Therefore, the most powerful teacher evaluation systems—ones that will push and challenge all of our nation's classroom practitioners—will be those that require teachers to explain who learned and who did not, and why, and what they will do next to ensure that all students are growing academically. Teachers who can explain their results and know what to do next are effective. Those who can spread their expertise to others and help them get better results for students are highly effective. In such a system, some teachers will not pass muster. When they do not, we will have much more evidence to either help them improve or remove them from teaching. We will also have a system that will elevate teacher leadership, ensuring that all practitioners embrace the kind of results-oriented teaching evaluation system that students of the 21st century demand and deserve.

Recommendations

The new assessments under development are surely a critical component in creating an educational system that can offer greater educational equity to students nationwide, as well as helping policy makers assess the growth of students more meaningfully and fully across subject areas and grade levels. The process of developing these assessments, including serious conversations at the federal, state, and local levels, may also help to clarify and focus national goals for public schools and the teaching profession.

However, we consider it critically important that the consortia leading these efforts, as well as the states and districts they represent, not forget about the steps that necessarily come *before* the creation of these new assessments. Specifically, we recommend that:

1. Leaders of the Common Core State Standards initiative must work with teachers to clearly define the assessment expertise needed by classroom practitioners, drawing on experts who already have the skill to lead the reforms. At present, such leadership opportunities are not clearly designed and described, and teachers selected to fill them may or may not have the competencies and dispositions needed to fill them, in terms of understanding assessment well themselves or being able to spread that expertise among their peers in effective and collegial ways. This definition is important for buy-in among teachers as well, since most people tend to mistrust what they cannot easily understand. More information and transparency will facilitate the building of trust and of strong assessments and standards over the long run.

2. Key universities and school districts in targeted states must work with teachers to develop the pre- and in-service professional preparation curriculum needed to cultivate assessment leaders who can implement the Common Core Standards. Once teachers fully understand what is being asked of them, they need opportunities,

both pre-service and in-service, to develop or enhance their assessment-related skills and knowledge. Just as computer programmers wouldn't be asked to tackle a new project without being offered opportunities to learn the right programming languages, teachers should not be asked to support or execute new assessments without learning the "language" of the Common Core Standards.

3. States and the assessment consortia must work with teachers to develop templates for districts to use in organizing time and tools for classroom practitioners to engage in assessment and teacher evaluation reforms. As our teaching colleagues remind us, higher standards and more rigorous assessments are easy to support. But they are meaningless without considering what happens in the middle: curriculum and lesson development that drive teaching and learning processes. No matter how comprehensive the tools teachers are given for Common Core implementation, they need time to make the transition in their classrooms.

4. ***All parties connected with implementation, from national consortia to school assistant principals, must create opportunities for large numbers of accomplished teachers to lead the development and implementation of Common Core-aligned tools, leveraging their expertise and engaging them as arbiters of results-oriented teaching evaluation systems.*** Again, a lack of transparency tends to be the first concern that teachers cite when asked why educators' buy-in for the new standards and assessments is in question. We believe, however, that it is important to use the opportunity to go beyond keeping teachers informed. We need to get them engaged. Public education faces increasing challenges and shrinking budgets, and society cannot afford any longer to sideline the expertise of the professionals who work with students, assessments, and curricula on a daily basis. ***We call for 30,000 classroom-based, assessment experts by 2014—one-tenth of those needed to lead assessment reforms well into the 21st century.***

This paper, developed with the on-the-ground insights from highly accomplished teachers, raises concerns about the inadequate contextual knowledge possessed by those who create and administer evaluation instruments, and about the documented effects on teacher effectiveness and student achievement of certain school working conditions.

In acting on these recommendations, the consortia will strengthen the assessment systems by cultivating the teacher leaders needed to institutionalize and sustain the reform in curriculum and assessment as well as in teacher evaluation. In doing so, the consortia will make it possible for the new assessments to be implemented with fidelity and rigor—and teachers, millions of them, will be ready to embrace the results-oriented profession we all seek.

If policy makers can begin to develop ways for teachers to own the standards and assessments, and offer time and training for implementation with fidelity and rigor, classroom practitioners will not only embrace 21st-century school reforms, they will help lead them. Investing in formative assessments and valuing (and compensating) teacher leaders are key levers for change. So are using assessment data as tools, not carrots or sticks, as well as investing in serious professional preparation (in meaningful pre- and in-service programs) and teacher leadership.

The vast majority of teaching professionals are not questioning the idea that individual teachers need to be assessed. But if policy makers do not acknowledge the importance of the issues raised in this paper, they will not address a central concern of teachers who are seeking

ways to improve teaching and learning through authentic classroom assessment, regular collaboration, and the sharing of expertise in order to create a results-oriented profession. The first step must be to tap into the assessment expertise of many current teachers today and cultivate the knowledge and skills of others, now and in the future.

APPENDIX A

Teacher Leaders Talk About Teaching Evaluation Reforms

DAVID ORPHAL: Teachers seem to be at war with the so-called reform community, which is led by many individuals with little, if any, teaching experience. I find this situation unsettling because both sides desire the same goal: children engaged in meaningful learning in great schools with excellent teachers.

Teachers may come to accept the Common Core Standards as they learn more about them, but many—including me—will continue to be leery about the new student assessments that will be created around those standards and the ways in which the results of those assessments will be used to evaluate teachers. We do not see serious classroom teachers involved in developing them. To us, it feels like the politicians, foundations, and administrators are going to tell us how to do our job—again.

RENEE MOORE: Part of the evaluation problem traces back to how we have traditionally prepared teachers and the conditions under which we have been expected to work. Both have bred isolationism and the illusion of egalitarianism.

I think if we really want to improve teaching across the board, teaching practice must be made public, and we should be constantly sharing what we do inside our classrooms with each other and the general public. But we have to be strategic about how we get there. In the book I co-author with Barnett Berry and 11 other accomplished teachers, *Teaching 2030*, we make the claim that it is time to blur the boundaries between classrooms, schools, and the larger community. And to do so means we need to begin "blurring the lines of distinction between those who lead schools and those who teach in them."

Just because teacher evaluation has always been handled by administrators doesn't necessarily mean that's the best way or that it should remain primarily their responsibility. We can and must give teachers more professional opportunities to make judgments about their colleagues' practices and outcomes for students. Doing so will require corresponding shifts in how schools are staffed and how teachers are paid.

MARSHA RATZEL: It's not that I'm not for teachers reviewing teachers. But I think that many teachers are so concerned about who evaluates them because it has been done so poorly in the past. I am—like many of my colleagues—against anyone reviewing and assessing us without [the evaluators having] adequate time and training. The simple fact that they are teachers and not administrators won't allay those concerns.

RENEE MOORE: When I was lead teacher at my high school, I spent half of each day conducting peer coaching, leading one-on-one and small group professional development, and analyzing lesson plans. Opening up our work for critique by our colleagues is one of the major hallmarks of a profession—and the failure to do this consistently is one reason why teaching has had to fight to be seen as a profession. Regular feedback and review of both curriculum content and each other's teaching practice should be as much a part of our normal work day as taking attendance. As we used to say in

my time with the Carnegie Foundation, teaching is scholarly work; and scholars have to be willing to get and give substantive peer review.

MARSHA RATZEL: If teachers in evaluator roles were fully trained in how to do evaluations—not just give surface feedback—and then were prepared to coach you forward in your teaching practice, then I'd be for it. But I have serious questions about anyone in the building where I teach coming in, making some observations, and evaluating my performance. In our current circumstances, they would have far too little training or understanding of the context in which I teach.

RENEE MOORE: Teacher evaluation should be a shared responsibility. I don't think every teacher in every building will or should be involved in peer evaluation—at least not yet, until we have completely changed the way we prepare teachers and administrators for this vital work. What I *can* envision are leadership teams that include both teachers and administrators for whom teacher evaluation is their focus. They must, of course, be teams that have the necessary tools for such evaluation. We have to work to bring our profession to that level of maturity if we want the respect and compensation of true professionals.

MARSHA RATZEL: Whether it's administrators, teachers, or teams made up of both, evaluators have to have a deep knowledge of teaching and how to help those they evaluate improve their practice. There's a trust barrier there that is not going to be easy to overcome.

DAVID ORPHAL: Teachers are cautious that the Common Core Standards movement will be NCLB dressed up in new clothing. We're nervous that the next generation of assessments will once again narrow the curriculum in perilous ways. Many teachers are concerned that the new assessments being developed by PARCC and SBAC could become nothing more than another "gotcha" tool to document the failure of teachers and public education.

RENEE MOORE: And just because statisticians develop value-added estimates of teacher effects, using this or any other data does not mean they should be the arbiters of who is effective or not.

I would suggest that one important component of new student assessments is that the results be given not just to the individual teachers of those students, but that teachers be involved in the interpretation and discussion of test data together, in various configurations, *before* it is released to anyone else. The ability to examine data deeply in collaborative settings will take learning and skill, for which teachers should be prepared and compensated. This type of data interpretation would be, in fact, a form of de facto peer evaluation and true job-embedded professional development.

APPENDIX B

Contributor Biographies

RENEE MOORE, a National Board-certified teacher and the 2001 Mississippi Teacher of the Year, teaches English to adolescents and young adults at Mississippi Delta Community College in Moorhead. She is a member of Mississippi's Teacher Licensure Commission, the Board of Directors of the National Board for Professional Teaching Standards, the National Council for Accreditation of Teacher Education's Blue Ribbon Panel on the Clinical Preparation of Teachers, and the Center for Teaching Quality's Teacher Leaders Network.

DAVID ORPHAL is a 15-year teaching veteran, currently teaching and designing curriculum for ninth-grade history at Skyline High School in Oakland, California. He has leaped at the chance to promote the Common Core Standards, in an effort to assure that all public school students can experience rigorous and enjoyable academic work as his students already do.

MARSHA RATZEL has taught middle school math and science for more than 15 years. She is National Board-certified in early adolescent science, a member of the Kansas Exceptional Educators Network for highly accomplished teachers, and was one of the founding members of the Center for Teaching Quality's Teacher Leaders Network.

Endnotes

1) In addition, the District of Columbia and two U.S. territories have also signed on to the Common Core State Standards initiative.

2) Organisation for Economic Co-operation and Development, *Programme for International Student Assessment 2006* (Paris: OECD, 2007).

3) Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, The *Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (New York, NY: New Teacher Project, 2009).

4) Arthur Wise, Linda Darling-Hammond, Milbrey Wallin McLaughlin, and Harriet Tyson-Bernstein, *Teacher Evaluation: A Study of Effective Practices* (Santa Monica, CA: RAND, 1984).

5) R. Wolf, "How Teachers Feel toward Evaluation," in *School Evaluation: The Politics and Process*, ed. Ernest House (Berkeley, CA: McCutchan, 1973).

6) Center for Teaching Quality, *Measuring What Matters: The Effects of National Board Certification on Advancing 21st Century Teaching and Learning* (A TeacherSolutions Report by 10 National Board Certified Teachers) (Hillsborough NC: Center for Teaching Quality, 2008).

7) Richard Rothstein, "Moment of Clarity," *The New Republic* (March 15, 2010). Retrieved on January 2, 2010 at http://www.tnr.com/article/politics/moment-clarity-0

8) Peter Schochet and Hanley Chiang, *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, IES National Center for Education Evaluation and Regional Assistance for US DOE (2010). Retrieved September 23, 2010 at http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf

9) Sharon Otterman, "Hurdles Emerge in Rising Effort to Rate Teachers," *New York Times* (December 26, 2010). Retrieved on December 27, 2010 at http://www.nytimes.com/2010/12/27/nyregion/27teachers.html?_r=1&nl=todaysheadlines&emc=a2

10) Ibid.

11) *The MetLife Survey of the American Teacher: Collaborating for Student Success* (New York, NY: MetLife Foundation, 2009).

12) C. Kirabo Jackson and Elias Bruegmann, "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics* 1(4) (2009).

13) Debra Viadero, "Effective Teachers Found to Improve Peers," *Education Week*. Retrieved September 1, 2009 at http://www.edweek.org/ew/articles/2009/09/01/03peer.html?tkn=VQ[F91pv4%2Fm1H05QrumV3xEwIqnZkr5Dl8iG

14) Barnett Berry, Alesha Daughtrey, and Alan Wieder, *Teaching Effectiveness and the Conditions that Matter Most in High-needs Schools*, Center for Teaching Quality (2009). Retrieved December 26, 2010 at http://www.teachingquality.org/sites/default/files/Tch_effective_twc_final.pdf

15) Hawaii Department of Education, "31-state Consortium Awarded RTTT Grant," news release (September 2, 2010). Retrieved September 26, 2010 at http://lilinote.k12.hi.us/STATE/COMM/DOEPRESS.NSF/a1d7af052e94dd120a2561f7000a037c/1d7e47a868a2f6330a257792006e5c1b?OpenDocument

16) David Cohen, "Teaching Practice: Plus Ça Change...," in *Contributing to Educational Change: Perspectives on Research and Practice* (National Society for the Study of Education Series on Contemporary Issues), ed. Philip Jackson (Berkeley, CA: McCutchan, 1988). Retrieved on December 26. 2010 at http://ncrtl.msu.edu/http/ipapers/html/pdf/ip883.pdf

17) Chris Gallagher and Suzanne Ratzlaff, "The Road Less Traveled," *Educational Leadership* 65(4) (2008): 48-53.

18) Sonja Steptoe, "How Nebraska Leaves No Child Behind." *Time* (May 30, 2007). Retrieved September 25, 2010 at http://www.time.com/time/nation/article/0,8599,1626423,00.html

19) Kim Koe and Raja LakshmiVelayutham, *Improving Teachers' Assessment Literacy in Singapore Schools: An Analysis of Teachers' Assessment Tasks and Student Work*, National Institute of Education, Singapore. Research Brief No. 09-

002. Retrieved September 25, 2010 at http://www.nie.edu.sg/files/oer/NIE_research_brief_09_002.pdf

20) Byron Auguste, Paul Kihn, and Matt Miller, *Closing the Talent Gap: Attracting and Retaining Top-third Graduates to Careers in Teaching: An International and Market Research-based Perspective* (New York, NY: McKinsey and Company, 2010). Retrieved September 25, 2010 at http://tinyurl.com/2wno7xv

21) James Pellegrino, Naomi Chudowsky, and Robert Glaser, eds., *Knowing What Students Know: The Science and Design of Educational Assessment*, National Research Council Committee on the Foundations of Assessment. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education (Washington, DC: National Academy Press, 2001). Retrieved November 20, 2010 at http://www.nap.edu/openbook.php?isbn=0309072727

22) Berry Barnett and the TeacherSolutions 2030 Team, *Teaching 2030: What We Must Do for Our Students and Our Public Schools … Now and in the Future* (New York, NY: Teachers College Press, 2011).

23) Brian Lord Barbara Miller, *Teacher Leadership: An Appealing and Inescapable Force in School Reform?* (Boston: Education Development Center, 2000).

24) Barnett and the TeacherSolutions Team (2011). Ibid.

25) Roland Barth, "Teacher Leader," *Phi Delta Kappan* 82(6) (2001): 443-449.