



Teacher Learning Through Assessment

How Student-Performance Assessments
Can Support Teacher Learning

Linda Darling-Hammond and Beverly Falk September 2013

Center for American Progress



Teacher Learning Through Assessment

How Student-Performance Assessments
Can Support Teacher Learning

Linda Darling-Hammond and Beverly Falk September 2013

Contents

- 1 Introduction and summary**
- 3 Teacher engagement in assessment in high-performing countries**
- 7 'Sitting down' to score**
- 11 Current performance-assessment initiatives**
- 21 Learning from scoring**
- 29 Recommendations**
- 31 Conclusion**
- 35 About the authors**
- 37 References**
- 43 Endnotes**

Introduction and summary

As the internationally benchmarked Common Core State Standards, or CCSS, are adopted in states across the country, educators are seeking ways to support an increasingly diverse student population to meet these more demanding expectations. The likelihood that students will achieve the aims of the standards will be substantially shaped by how well teachers teach these challenging academic skills in ways that support a wide range of learners. Teachers' understanding of the standards and their grasp of how to teach them will also influence whether the new assessments provide useful insights, rather than harmful side effects, particularly for those students who have historically been least well served by their schools.

Designing professional-learning opportunities that can develop and improve the capacity of teachers to support more ambitious teaching and thus enhance learning for all students is challenging—and now even more so given the demands of teaching the Common Core State Standards. Research and experience have demonstrated the limitations of short-term training models such as the all-too-common one-shot workshop designed to transmit information to passive recipients.¹ In fact, research shows that what teachers have dubbed “drive-by” professional development has little effect on teacher practice and virtually no effect on student achievement, despite the fact that this form of professional development is ubiquitous in American schools.² By contrast, significant gains in student achievement can result from strategies that engage teachers in content-specific activities that are linked to collegial analyses of student work and learning over a more sustained period of time.³

Indeed, the performance assessments themselves may offer such an opportunity. Past experience suggests that teachers' involvement in developing, scoring, and analyzing the results of performance-based student assessments in the CCSS could be a powerful opportunity for teachers to learn about the standards, their students, and their teaching practice. This kind of professional learning can help teachers acquire the tools to teach the more complex skills and knowledge represented in the CCSS, which are crucial to preparing our citizens for the global workforce of the 21st century.

This paper describes how teacher learning through involvement with student-performance assessments has been accomplished in the United States and around the world, particularly in countries that have been recognized for their high-performing educational systems. We discuss how teachers' engagement with performance assessments influences their understanding of the standards and their students' abilities. This discussion includes comments from teachers about their experiences with performance assessments as provided through interviews conducted for this report as well as in previously published research.

Finally, we recommend how these kinds of performance-assessment opportunities can be planted and scaled up as states and districts implement CCSS and deepen their efforts to teach 21st-century skills. Specifically, we encourage states, districts, and schools to do the following:

- Ensure that performance assessment is integral to the learning system.
- Include performance tasks as part of assessments.
- Ensure that rubrics for scoring assessments are clear and explicit.
- Involve teachers in collaborative scoring of assessments.
- Expand opportunities for teachers to engage in assessment.
- Provide teachers with coaching and professional development around assessment.
- Build communities of practice to inform performance-assessment work.

Teacher involvement in the design, use, and scoring of performance assessments has the potential to powerfully link instruction, assessment, student learning, and teacher professional development. The use of high-quality standards and performance assessments over time has been shown to improve both teaching and learning. As teachers become more expert in their practice through involvement and engagement with performance assessments, the outcomes for students can be expected to improve. If used wisely, this approach has the potential to address multiple important education goals through one concentrated investment. Not only will overall pedagogical capacity be enhanced, but also teaching and assessing will remain focused on its central purpose: the support of learning for all involved.

Teacher engagement in assessment in high-performing countries

Contemporary efforts to raise education standards in the United States have developed in response to the more competitive expectations and outcomes for student learning around the globe. The Common Core State Standards, currently adopted by 45 states, the District of Columbia, and four U.S. territories, are benchmarked to learning objectives found in high-performing nations.⁴ These high-achieving jurisdictions, including Finland, Japan, Singapore, New Zealand, Hong Kong, and Australia, teach fewer but more in-depth topics, focus more on reasoning skills and applications of knowledge, and have a well-worked-out sequence of expectations based on how students typically progress in mastering specific skills over time.⁵ (In the CCSS, as in some other countries' standards, these expectations are called learning progressions.)

High-achieving jurisdictions implement their standards by developing systems that integrate curriculum, instruction, and assessment to improve both teaching and students' learning. Teachers are engaged throughout the assessment process in developing, reviewing, scoring, and analyzing the results of student assessments, which enables them to understand the standards and develop stronger instruction.⁶

Increasingly, high-performing nations use open-ended performance tasks to give students opportunities to develop and demonstrate 21st-century skills, such as the ability to find and organize information to solve problems, frame and conduct investigations, analyze and synthesize data, and apply learning to new situations. Students solve extended problems in mathematics and the sciences, showing and explaining how they are approaching the task; compare and synthesize evidence from different kinds of data and texts; and then compose essays that explain and defend their thinking.

The growing emphasis on project-based, inquiry-oriented learning by high-performing nations has also caused many of these countries to introduce school-based tasks into their assessment systems, such as research projects, science

investigations, and development of products ranging from software solutions to engineering designs. These tasks, incorporated into examination scores in contexts as far ranging as Britain, Canada, Singapore, Australia, New Zealand, and the International Baccalaureate program, focus teaching and learning on the development of higher-order skills and the use of knowledge to solve problems.⁷

The use of these types of assessments provides teachers with models of good curriculum and assessment practice, enhances curriculum equity within and across schools, and allows teachers to see and evaluate student learning in ways that can inform instructional and curriculum decisions. Such curriculum-embedded assessments can also build students' capacity to assess and guide their own learning.

Teachers score these open-ended tasks through a process called moderation, in which they receive training and then score and discuss model answers until their judgments are reliable; that is, their judgments accurately represent the standards and are consistent with one another. Sometimes these moderation processes occur within schools; at other times, teachers are assembled from across a region. Teachers use benchmark examples of student work at different levels along with a rubric or set of scoring criteria to calibrate their own judgments. As teachers learn to look for the key features of the work expressed in the criteria, they become more aware of the elements of strong student performance. As they continue to score and discuss the work, they fine tune their capacity to evaluate so that high rates of reliability are achieved.

Equally important, the scoring process and the discussions around student work help teachers to reflect on their curriculum, teaching, and assessment strategies, thus becoming more effective at teaching the standards.⁸ Such involvement heightens the probability that teachers—the critical players for enacting educational change—will come to understand and embrace the standards and be able to use the data from the new assessments. Lauren Resnick, professor and co-director of the Institute for Learning at the University of Pittsburgh, emphasizes this fact in her writings about teachers' work with the standards:

Standards documents, even elegant ones with benchmarks and commentary, can affect achievement only if the standards come to be held as personal goals by teachers and students. ... That will happen only if a concerted effort is made to engage teachers and students in a massive and continuing conversation about what students should learn, what kinds of work they should do, and how well they should be expected to do it.⁹

Involving teachers in scoring assessments is powerful professional development because it connects teacher learning directly to their examination of student learning, and gives them the opportunity to think together about how to improve that learning. It also sends an important message by signaling that teachers can be active participants in shaping the direction of school change. As this kind of professional development acknowledges the critical role of teachers in supporting student learning, it put teachers in their rightful place: at the center of the school-improvement process.

Teachers' involvement in performance assessment in the United States

Like behind-the-wheel driving tests, performance assessments require people to show what they know by demonstrating their skills in action. In education, performance assessments can include everything from extended written responses to questions to oral presentations of research and communication skills to graphical and written responses that explain and show the results of problem solving to the conduct and reporting of science experiments or even musical or artistic presentations. The original version of the National Assessment of Educational Progress included all of these kinds of performances. In some cases, performance assessments are collected into portfolios of student work that can show how students meet a comprehensive set of standards and how their learning has progressed over time.

The new multistate assessment consortia created to assess the Common Core State Standards—the Smarter Balanced Assessment Consortium, or SBAC, and the Partnership for Assessment of Readiness for College and Careers, or PARCC—plan to include some performance tasks. Initially, these will be limited to open-ended written responses to research, writing, and mathematical tasks. Some states, including those that are part of the Council of Chief State School Officers, or CCSSO, Innovation Lab Network, plan to introduce even more extensive performance assessments to complement the SBAC tests. These may include longer-term tasks that require students to undertake investigations over multiple weeks, and could result in a range of products—such as engineering designs, built objects, spreadsheets, and research reports—presented in a variety of forms, including oral, graphic, and multimedia presentations.

Many states used performance assessments in the 1990s, prior to the passage of the No Child Left Behind Act, or NCLB. These assessments were discontinued in most places in large part because of the costs associated with the NCLB require-

ment of testing every child, every year, and because of the rules imposed by the Department of Education for approving state testing plans.¹⁰ Most jurisdictions reverted to inexpensive, machine-scored, multiple-choice tests. Some states and localities, however, continued to use performance assessments because of their commitment to ensuring that students learn and demonstrate higher-order skills of analysis, deeper problem solving, and communication.¹¹ New initiatives are now underway as a consequence of the CCSS and the movement to expand students' opportunities to develop 21st-century skills. We describe some of these past and current initiatives along with their outcomes below.

Studies of the implementation of performance assessments in California, Connecticut, Kentucky, Maine, Maryland, Missouri, New Hampshire, New York, Ohio, Rhode Island, Vermont, and the state of Washington found that the portfolios and performance tasks could ultimately be scored quite reliably by teachers. Furthermore, the assessments supported improvements in instruction and student learning.¹² Teachers assigned more writing and mathematical problem solving of the kinds demanded on these new assessments, and as a result, student achievement climbed on both traditional measures and these more challenging measures. The results were most positive when states or districts developed teachers' expertise for designing, scoring, and evaluating the results of the assessments.¹³

Researchers found that teachers scoring the assessments led to teachers working on instruction, which makes it excellent professional development. Examining and assessing students' work helps teachers learn more about what their students know and can do, as well as what they think. Doing this in the context of standards and well-designed performance tasks stimulates teachers to consider their own curriculum and teaching. Together, teachers can then share specific instructional approaches that can be used to support the strengths and needs of their students.

‘Sitting down’ to score

Scoring sessions typically begin with an orientation process that helps teachers learn to use standards as a reference for evaluating students’ responses. A scoring orientation typically goes something like this: Aided by a facilitator, teachers, working in small groups, look over each assessment task and discuss the specific standards that the tasks assess. Teachers then review each task’s criteria and discuss what students need to do to accomplish the task. Together, teachers examine sample student responses, referencing the scoring guide—also known as a rubric—for descriptions of what the completed work looks like at different levels of proficiency. Discussing the student work, detail by detail, they then compare the evidence in each response to the rubric’s indicators until they arrive at a consensus for a score.

Differences in opinions and perspectives among teachers can be mediated by using the rubric’s clearly articulated criteria for performance and the requirement that teachers always justify their evaluation using evidence from student work. Although viewpoints may initially vary, after going through several sample responses, teachers begin to agree on the scores that they assign. A recurring consensus signals completion of the orientation. Only then do teachers move on to begin the independent scoring of tasks used to assign “official” scores.

Using standards to guide evaluation

Learning how to use a rubric helps teachers evaluate students’ work based on evidence, rather than feelings or assumptions. In the course of scoring, teachers learn to apply common criteria and standards to the work of all their students, rather than just comparing one student’s work to another student’s work. Learning to use evidence as a result of participating in standards-based scoring often transforms the way teachers evaluate student work. As one elementary school teacher who participated in a statewide performance assessment of student work project put it:

I moved away from thinking about work in an A, B, C, or D way, to thinking about the criteria for performance and the evidence that would justify my evaluation.¹⁴

As teachers work with content standards, scoring rubrics, and student-benchmark papers during scoring, they come to think more deeply about their teaching. Content standards provide an overarching conception of the discipline itself, as they specify what are the important aspects of a subject area. Scoring rubrics identify the features of student work that are important, and guide teachers in looking for these features. Actual examples of student work at differing levels of performance offer reference points for understanding to what degree standards are met. In these ways, participating in scoring helps teachers clarify goals and expectations for their teaching and for their students' learning. In addition, it deepens their knowledge of their discipline, reveals important information about what their students know and can do, and offers them insights to improve their teaching.

Learning to evaluate student work with a clear, objective eye also helps safeguard against biases that teachers may have about students' capabilities. When teachers carry this approach back into the classroom, they can better recognize the varying strengths of diverse learners. This in turn makes it harder to attach labels and judgments to students, which often have the unintended effect of becoming self-fulfilling prophecies. In addition to safeguarding against bias, keener observation of students and their work helps teachers to make more informed decisions and provide better supports for learning.¹⁵

In addition, engagement in scoring sessions strengthens teachers' sense of professionalism, heightens their understanding about the workings of the system of which they are a part, and reaffirms the central importance of teachers to the evaluation process. Instead of having to rely on testing companies to judge the outcomes of students' work, teacher involvement in scoring places assessment back into the domain of teaching, where it can be readily accessed to inform and support learning. Several important aspects of this process are discussed next.

Conversing about the standards

One of the most valuable aspects of this work is the opportunity that scoring sessions provide for collegial conversations. These discussions—which can take place before, during, or after scoring—enable teachers to learn about state or district expectations for their students, to hear about how other teachers interpret

the standards, and to see how the big ideas embodied in standards play out in real student work. Working with the standards helps teachers gain a perspective about what is valued and valuable in their broader community.¹⁶ In addition, the scoring experience helps them develop a shared understanding and a common language about the essentials of their disciplines, which develops a sense of professional community and can facilitate more coherent instruction across classrooms.

Looking at performance on high-quality tasks

Scoring-worthy tasks gives teachers a window into what their students can do, as well as how their students think. Well-designed tasks are contextualized in real-world situations; they ask students to show and explain their work and they allow for multiple ways for students to demonstrate their abilities.¹⁷ From this evidence, teachers learn more about the variety of ways students approach and solve problems. Furthermore, because the expectations are publicly articulated, students have a better chance of achieving them. This makes the assessments fairer and more accessible for different kinds of learners.¹⁸

These features of assessments can help expand teachers' visions of what good work may look like when it comes from a wide range of students across many locales and backgrounds. According to an elementary school teacher who administered and scored a math/science/technology task:

Looking at student responses to the assessment tasks reinforces the idea that good work can look very different and can take on many forms.¹⁹

Exploring teaching

Many teachers say that participation in scoring motivates them to strengthen their practice, not only to better prepare their students for tests, but also to improve the teaching and learning that goes on in their classrooms. The following comments from an anonymous survey of New York public school teachers about their participation in a recent performance-assessment project illustrate the types of changes teachers planned to make to their practice as a result:

"I plan to give kids rubrics detailing what makes 'quality' work," said one elementary teacher

“I will provide more opportunities for revision, self-analysis, and evaluation,” said another.

A middle school teacher vowed to “make the open-ended questions I ask clear enough to get the information I want to get from the students. I will also make grading criteria very clear, very related to the question, and available to students ahead of time.”

A high school mathematics teacher who recognized the value of student explanations said, “I will do more testing requiring justifications—help students to become more comfortable explaining their understandings.”²⁰

Teachers’ discussions about students’ work in scoring conferences offer opportunities to learn from each other about new practices and educational processes and to validate their knowledge as competent professionals.²¹ High school teachers involved with scoring state performance assessments note how much they appreciate this opportunity:

“Meeting with dedicated, concerned teachers was most valuable to me. ... I learned from their positive attitudes, from discussing concerns about my students and about future directions for my discipline.”

“I don’t think you can underestimate the need that folks have for getting together and having quality time to reflect on all of the changes that are happening in our schools.”

“Scoring sessions provide valuable professional dialogue. It is a great way to do teacher in-service.”²²

Current performance-assessment initiatives

Performance-assessment initiatives are developing across the country as states have adopted the Common Core State Standards, which call for higher-order thinking skills not measured on traditional multiple-choice tests. At least a dozen states have some form of performance-assessment activity occurring at the state or local level,²³ and a number of others have plans in the works. In this section, we discuss initiatives in California, Ohio, New York, and New England that take distinctive forms, offering a range of lessons about how teacher engagement with performance-based assessments improves teaching and learning.

The Silicon Valley Mathematics Assessment Collaborative

Since 1998, a group of school districts in California's Silicon Valley have supplemented the state testing system with a common set of performance assessments.²⁴ The Mathematics Assessment Collaborative, or MAC, uses tasks designed by the Mathematics Assessment Resource Service, or MARS, in second grade through high school. The MARS assessments, which are developed by mathematics teachers working with researchers, feature tasks that test key mathematical concepts along with the mathematical practices now codified in the Common Core State Standards: problem solving, modeling, reasoning, and communication.²⁵ (see example below)²⁶

Teachers score the tasks using a rubric that takes into account how students approach the problem, their solutions, and their ability to justify or generalize their solutions. MAC also offers professional development for teachers and data supports for districts. Although it is still centered in California, MAC has grown to more than 90 school districts and charter-school networks located in several states. Between 40,000 and 80,000 student-assessment papers are hand-scored annually.²⁷

The ongoing professional development and coaching offered by the project begins with teachers coming together within their districts to score the MARS exams on professional-development days. They receive training and calibration support

to score reliably to the standards.²⁸ At the end of the scoring day, teachers spend time reflecting on students' successes, challenges, and implications for instruction. Teachers, supervisors, and coaches all see this aspect of the project as valuable for teacher learning:

*“Scoring the MARS test is the single most valuable professional development we have done with our teachers in mathematics. The full day of scoring the tests leads to rich conversations about what we expect from students and how our students think mathematically. We see real buy-in from teachers.” – Assistant superintendent of instruction from a suburban district**

“We joined the Silicon Valley Mathematics Initiative and decided to give the MARS test. We didn’t know what we signed on for or how much work it would be. At one point, I thought we were over our heads. But we continued to forge ahead with the scoring session. I have to say it is one of the most rewarding days I have had in education. We got them all scored and the teachers were great. They really felt they had had an opportunity to explore what was in the students’ heads. They came away convinced this way of scoring student work had changed forever the way they will teach.” – Math coach from a low-income, urban school district

“At first when we were training to learn to score the MARS task, I was very skeptical about the process. There were a lot of concerns among the teachers. Some of us really pushed back on the facilitator. But after going through the standardizing papers and especially after spending the full day scoring tests, it became very obvious we were focusing on what the students really knew and could explain. We all seemed to discover the same problems the students were having doing real math.” – Sixth-grade teacher

Researchers who have evaluated the MARS process explain how this learning occurs:

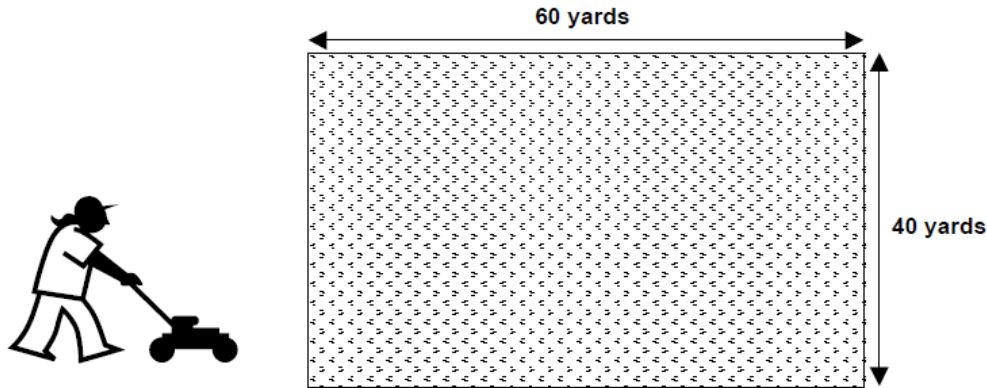
To be able to score a MARS exam task accurately, teachers must fully explore the mathematics of the task. Analyzing different approaches that students might take to the content within each task helps the scorers assess and improve their own conceptual knowledge. The scoring process sheds light on students’ thinking, as well as on common student errors and misconceptions. As one teacher said, “I have learned how to look at student work in a whole different way, to really say, ‘What do these marks on this page tell me about [the student’s] understand-

*Author’s note: Quotes from educators involved in the MARS project were collected anonymously by researchers studying the assessment and were published without attribution.

Lawn Mowing

This problem gives you the chance to:

- solve a practical problem involving ratios
 - use proportional reasoning
-



Dan and Alan take turns cutting the grass.
Their lawn is 60 yards long and 40 yards wide.

1. What is the area of the yard? _____ square yards

Dan takes an hour to cut the lawn using an old mower.

2. How many square yards does Dan cut in a minute? _____
Show your work.

Alan only takes 40 minutes using a new mower.

3. How many square yards does Alan cut in a minute? _____
Show your calculation.

4. One day they both cut the grass together.
How long do they take? _____
Show how you figured it out.

Source: Inside Mathematics, "Performance Assessment Task: Lawn Mowing, Grade 7," available at <http://insidemathematics.org/common-core-math-tasks/7th-grade/7-2005%20Lawn%20Mowing.pdf> (last accessed September 2013).

ing?” Recognizing misconceptions is crucial if a teacher is to target instruction so that students can clarify their thinking and gain understanding. The emphasis on understanding core ideas helps teachers build a sound sequence of lessons, no matter what curriculum they are using.²⁹

The learning does not stop there. Once the papers are scored, they are returned to the schools, along with a copy of the master scoring sheets, for teachers to review and use as a guide for further instruction. It is important to note that teachers receive the tasks and rubrics, along with real student work, not just abstract scores. Districts also receive data for both the MARS tests and the state exams, along with a set of reports that can inform professional development, district policy, and instruction.

These reports include an annually produced document titled “Tools for Teachers,” which includes an analysis of student thinking, understanding, errors, and misconceptions derived from an in-depth study of the tasks. The study’s findings are combined with sample student work and a set of suggestions, strategies, and questions for teachers to use to inform and improve instruction.³⁰ MAC offers training and coaching for teachers, teacher leaders, and coaches around these tools. This combination of assessment materials, student work, and professional supports contributes to improvements in curriculum development and instruction, as participants explain:

“We believe the value of the MARS tasks is for formative assessment. We coaches meet with our teachers during early release days. They bring the student work from a MARS task they have given to their classes. We score it, analyze it, and discuss it with the teachers. Afterwards, we plan re-engagement lessons. They are designed to address students’ misconceptions and errors. This process has caught on and teachers are having success deepening students’ understanding.” – Math coach from a low-income, urban school district

“Our grade-level team uses the MARS tasks a lot. We use them to make sure the students can do real-life math problems. We also like reading the students’ explanations. At the start of the year the students do not know how to explain and they barely write anything. So we use them regularly, and by December they are writing more and it becomes insightful for our team to go through their solutions.” – Third-grade teacher

Research over a 13-year period shows that as teachers and schools participate in this process, students’ performance in mathematics improves significantly, both

on the MAC tasks and on the more traditional state tests.³¹ As teachers learn how to evaluate student needs and design their instruction to produce stronger mathematical understanding, their students' improvement is greater. Across 35 districts participating in one recent study, students in all grades—from third grade through 10th grade—had better outcomes the longer they had been taught by teachers who had been involved in scoring, coaching, and professional development through the Mathematics Assessment Collaborative. Furthermore, students of teachers who had received more intensive coaching around formative assessment uses of the tasks had stronger results.³²

These outcomes result from a combination of factors: performance assessment; scoring sessions; performance reports; and the use of MARS assessments and rubrics for formative assessment, instruction, coaching, and professional development. Researchers note:

*The scored tests themselves become valuable curriculum materials for teachers to use in their classes. MAC teachers are encouraged to review the tasks with their students. They share the scoring information with their students and build on the errors and approaches that students have demonstrated on the exams. ... The Mathematics Assessment Collaborative fights teachers' sense of isolation and helplessness by sharing everything it learns about students. It identifies common issues and potential solutions. It helps teachers understand how learning at their particular grade level is situated within a continuum of students' growing mathematical understanding. It promotes communication across classrooms, schools, and grade levels. It encourages teachers to take a longer, deeper view of what they are working to achieve with students.*³³

Together, these supports result in strong teaching and stronger learning within and across classrooms, schools, and districts.

The Ohio Performance Assessment Pilot Project

Extended, curriculum-embedded performance tasks are the basis for the Ohio Performance Assessment Pilot Project, or OPAPP. In 2008, Ohio undertook a statewide project to develop and pilot performance-based assessments developed in partnership with the Stanford Center for Assessment, Learning, and Equity, or SCALE.³⁴ The assessments in English language arts, mathematics, and science are mapped to the Common Core State Standards and the anticipated Next

Generation Science Standards, and are designed to promote and evaluate students' learning of content and skills that will prepare them to be successful in college and in their careers. Within the context of Ohio policy, the assessments may be used as project components for course examinations, as proficiency measures to grant credit based on competence instead of seat time, or as options for the senior project encouraged in Ohio schools. They are intended not only to contribute to the development of a new, multiple-measures assessment system, but also to support improvements in instructional practice.³⁵

The project has involved educators from 30 schools in constructing and piloting a set of curriculum-embedded, one- to four-week-long formative “learning tasks” and conceptually related, but shorter, one- to three-day “assessment tasks.”³⁶ The Ohio Department of Education is now expanding the project to include history and social science and career-technical education, and will build a task bank. In addition, it plans to expand the network of districts in the pilot project; build the capacity of local educators, administrators, coaches, and regional assistance centers to carry on this work into the future; and build a technology platform that will support the scaling up and implementation of a performance-based assessment system.³⁷ The curriculum-embedded performance tasks include in-class, collaborative components, which are supported through a set of lessons and interim products, followed by individual student-work products that are scored. In the English language arts tasks, for example, students read and take notes on required or self-selected texts, and have an opportunity to engage in discussions with classmates on those texts prior to using the texts for their final essays, which are developed individually. Some tasks have group collaboration built in, along with peer and teacher feedback on drafts of products. An example of one of such task, “*Got Relieve IT?*” is shown below.

The tasks are standardized with respect to guidelines for the kinds of supports, instructional scaffolds, and feedback teachers are allowed to provide to students without decreasing the rigor and challenge of the tasks. In the learning tasks, there is also some flexibility: Teachers can choose when to implement the tasks and students or teachers may have choices within the task such as choices of texts. The shorter assessment tasks, which are focused more explicitly on measurement, are more standardized.³⁸

Teachers initially receive two days of professional development to help them understand how to use the tasks, including an opportunity to complete key elements of the tasks themselves, and to reflect together about their approaches or solution strategies.³⁹ In addition, the Ohio Department of Education and their districts provide ongoing coaching during the year to support implementation.

“Got *Relieve IT?*”

A curriculum-embedded performance task

In the science task “Got *Relieve It?*” students are cast as employees for a company called Achoo-B-Gone. They are told their team has been working for the past year to create a new drug that will instantly relieve cold symptoms.

The new product, *Relieve IT*, is in the final testing stages before being sent to the Food and Drug Administration, or FDA, for human trials. Part of the FDA-approval requires the team to share their current knowledge about acids and bases and to provide all of their experimental data on the pH levels of *Relieve IT*. The FDA provides a preliminary report identifying concerns of potential negative human side effects due to the product’s pH levels. The FDA wants to know what the team will do to “fix” the product before initiating human trials.

Students must synthesize their current knowledge about acids, bases, and neutralization reactions. Then as a group they design and conduct an experiment to determine the pH of the *Relieve IT* product and to determine which unknown solution (1, 2, 3, or 4) or combination of solutions can be used to neutralize any excess acid or base. Finally they prepare an individual formal lab report, which includes recommendations for “fixing” the pH levels of *Relieve It*.

Source: Ruth C. Wei, Susan E. Schultz, and Raymond Pecheone, “Performance Assessments For Learning: The Next Generation of State Assessments” (Stanford, CA: Stanford Center for Assessment, Learning, and Equity, 2012).

Teachers are trained to score, and their scoring is calibrated until it becomes consistent. They are allowed time after scoring to analyze student work, as well as to reflect on their experiences and implications for future instruction.

The Ohio Performance Assessment Pilot Project performance assessments are scored using genre-specific, descriptive four-level analytic rubrics. These are introduced to students before they undertake the task and are used to provide detailed feedback to students about their performance after they finish. Furthermore, as the developers of the OPAPP assessments explain:

Since the rubrics are genre-specific rather than task-specific, and the same dimensions are scored across tasks, it becomes possible to track a student’s progress along the same dimensions of performance over time across years and courses in the same discipline (e.g., science inquiry, math problem-solving).⁴⁰

Many teachers were involved in developing and fine tuning the OPAPP assessments’ tasks and rubrics, as well as piloting and scoring the tasks. Surveys revealed that virtually all of the teachers felt very positively about the professional development they received, including participating in doing the actual tasks themselves

and planning together for conducting the tasks.⁴¹ Teachers also reported learning from conducting the tasks, as well as scoring them. In science, for example, many teachers had not previously been involved with inquiry-based teaching and the project provided insights into students' capacities and about how to provide more engaging instruction. Here is what some teachers had to say:

"I learned that I need to have higher expectations for my students and to do more inquiry based labs. My students exceeded my expectations. They were able to complete the task with very little direction from me."

"I saw students who were very engaged and invested in the designs of their cars. They were having great evidence-based discussions to determine which changes to make on their cars. I asked the kids if they liked what they were doing and they answered a resounding 'yes.'"

"My students came in voluntarily during lunch and after school to research the normal values for each of the diagnostic tests and what the results meant related to the 'Medical Mystery task' ... that is the first time I have seen that in 20 years of teaching."⁴²

Some students also expressed their enthusiasm for tasks that allowed them to think critically, make decisions, and learn for themselves. According to one student, "This task was better than what my teacher usually does because we were able to make decisions and figure out how to solve the problem, instead of just mindlessly following his directions."⁴³

As in other performance-assessment contexts, the scoring sessions were often viewed as "the most useful for improving my teaching and understanding of where my students are in comparison to where they should be," as one teacher says. More than 97 percent found the opportunity to discuss their experiences with other teachers highly useful. According to teachers, many kinds of learning resulted from the scoring sessions:

"I learned how better to apply rubric expectations in my instruction by clarifying what those things 'look like' within student examples."

"I learned just how difficult it is for students to do the reflection piece of the project. I will try to construct some models and activities to help students to better

accomplish this task. Students did a lot of good work, but had trouble labeling and explaining. [I] need to emphasize that more in the classroom.”

“I will be incorporating communication skills in ‘bite-sized’ chunks into day-to-day teaching; and incorporating smaller ‘rich’ problems into homework.”⁴⁴

The researchers noted that in order to support professional learning and improvements in practice, it is important that teachers have the following:

- Opportunities to learn and practice the content and skills necessary to implement the performance tasks, as well as administrative support
- Access to communities of practice engaged in performance-assessment work, with opportunities to collaborate in planning for and reflecting on implementation
- Opportunities to analyze student-work samples and scores, and to learn to score using a common set of evaluation criteria, engage in calibrating conversations with colleagues, and score at least one class set of their own students’ work in order to analyze patterns of performance across students.

The Quality Performance Assessment initiative

The Quality Performance Assessment, or QPA, program was launched by the Center for Collaborative Education, or CCE—a Boston-based nonprofit organization that partners with public schools and districts in several New England states to create and sustain effective and equitable schooling.⁴⁵ CCE aims to strengthen and document local assessment systems by introducing common tasks and moderated scoring processes. The QPA work has led to the creation of a set of Common Core State Standards-aligned performance tasks with teacher guides and student-work samples that have been field tested in schools and are documented in the book, *Quality Performance Assessment: A Guide for Schools and Districts*. In addition to the disciplinary focus of the CCSS, the initiative has placed an explicit emphasis on “habits of mind”—a term that refers to the complex, critical-thinking, problem-solving, and communication skills students will need in college and throughout their careers and civic life.

The original initiative started with schools that practiced performance assessment at the individual classroom and school level. These schools developed the QPA model and field tested the process with common performance assessments used across schools. The work of QPA has since expanded beyond the original network to schools, districts, and states where QPA provides professional development and resources for implementation. These partners develop and use their own unique assessments while also working to develop and score common tasks that can contribute to high-quality, reliable, and large-scale performance assessment systems. QPA brings greater rigor to the process of task development and scoring, as well as analysis of student work and instruction. Source: Personal communication from Anne Clark, headmaster, Boston Arts Academy, October 2012.

Performance assessment at Boston Arts Academy

The Boston Arts Academy has long used performance assessments as a central element of the teaching and learning process. Portfolios and exhibitions are a primary means of assessment for all content areas. According to Anne Clark, a veteran teacher who is now the academy's headmaster, because performance assessment "asks students to demonstrate depth of understanding over time rather than just choosing an answer from a given set, it allows us to really understand what kids understand. They have to share, show, and even teach what they know."⁴⁶

The school's senior capstone experience—a graduation requirement—is an example of how a project and accompanying presentation are evaluated by a group of professionals, much as would be done in the world outside of school.

Created by the school staff after consulting with artists and community groups to determine what practicing artists need to know and do, the assessment task asks students to write a grant proposal for a community service arts project. The project must be interdisciplinary, conveying the artist's vision, but also must employ real-world skills, including creating a budget, timeline, and rationale. Each student presents his or her proposal to a team of outside judges made up of artists, corporate and community partners, and students' family members. These judges, along with community members who come in to hear the presentations (sometimes the group is as large as 100 people), score the proposals. Students who get the highest scores are invited to go to a finalist round. The winner's project is actually funded. All students, however, have to keep working on their respective projects until they score at least a 3 on the rubric. According to Clark:

This is one of the most important experiences our students have ever had, not just in school, but in life. It is also a practice that impacts my teaching and is an internal accountability mechanism with my colleagues. In our building, we have always co-taught, swapped kids—I teach them for one term, you teach them for another.

But for this [common] assessment, we co-plan, look at formative assessments while planning, and cross-score together. This has helped us move forward in developing a common vision of who we are and what we are trying to teach. It has strengthened our professional identity.

If you are just implementing someone else's curriculum, you don't have the same ownership and accountability as if you create it yourselves. We created our assessments, guides, and shared that standard with our university/community partners. We have to look at each other every day and say, 'why can't so and so do that? What do we need to do differently?' Working in this way makes sure that our assessment informs our instruction and that our students really learn.

In addition to their unique school-based assessments, Clark and her Boston Arts Academy colleagues share common assessment frameworks and rubrics with other schools that are part of the QPA network. Together with colleagues from these schools they score their assessment with common rubrics. Clark has found this cross-school scoring to be powerful.

Scoring other papers in other schools and having other people in other schools score ours has led us to the conversation of 'What should a tenth grader be able to do?' This has informed our work even more deeply, [considering] what kids in other schools are doing, how we are teaching to the standards, and how the teachers are getting results.

Source: Personal communication from Anne Clark, headmaster, Boston Arts Academy, October 2012.

Learning from scoring

Many teachers who have attended QPA scoring sessions note that their experience deepened their understanding about how to teach 21st-century skills. This quote from one of them captures the sentiment of the group:

[This work] has got me thinking about how to use 21st century skills in my assessments and [how to] grade the work not the kid.⁴⁷

William Hart, assistant superintendent of the Pentucket Regional School District in Massachusetts explains:

Teachers are now using the common rubrics to guide the type of project or task they can develop to marry concept/content acquisition and 21st-century skills. They are asking questions like: “How do I shift the instructional environment to do both? How am I helping kids develop as collaborators?”

Measuring worthy skills and knowledge in this way has driven the context of the classroom.⁴⁸

As teachers define these worthy skills, they also learn to measure them based on evidence, rather than subjective hunches. Laurie Gagnon, QPA director, notes:

There is great power in grounding the conversation in evidence ... in discussing, “What do I mean by a well-chosen and supported quote? What does it mean to write a good thesis statement? What does this really look like?” Conversations around such questions have yielded big learning for teachers.⁴⁹

Christina Brown, director of QPA’s Principal Residency Network, compares the process of learning to score student work with what happens in umpire school:

[Just as] the prospective umpire learns to distinguish between a ball and a strike, and to know the criteria for what each means, [scorers] learn the details of what proficiency actually looks like.⁵⁰

Using common assessments and rubrics and engaging in collective scoring of student work helps create coherence for the teaching that takes place across grades in a school. Jeanne Sturges, a staff developer at Souhegan High School in Amherst, New Hampshire, explains:

Before [working with the QPA initiative] we did not always have consistent alignment of learning outcomes across teams and classes. Teachers might have had rubrics aligned with standards, but the work was not necessarily comparable across classes. The work we have done with QPA over the past two years has focused on trying to ensure that if ninth-grade science teachers are all doing the same project with the same rubrics, they will all make a similar judgment about the students' work. Using common rubrics and performance tasks has created great opportunities for teachers to push their thinking about the level of rigor we ask of students and the level of equity we provide. Although this work presents a huge challenge, the commonness of our work—the shared accountability—offers the best of what standards can bring.⁵¹

Addressing the opportunity gap

In addition to developing coherence across schools and districts, a consequence of using common tasks and rubrics is that teachers begin to understand how their own students match up to commonly held expectations about necessary knowledge and skills. This can be an eye-opening experience for teachers who work in varying contexts. They sometimes find discrepancies between what has been considered proficient in one type of community and what has been considered proficient in a community with different demographics. Again, according to Gagnon:

Surfacing these differences can lead to tough conversations among educators about issues of equity—conversations that may be hard but important to think about. Naming the challenge and having the evidence to substantiate it can lead to positive conversation and change.⁵²

Brown underscores the importance of having such conversations as a means of providing access to excellence for those who have not had equitable access to educational opportunities:

To determine proficiency in learning, it doesn't matter in the long run, if a kid comes from a hard life. We do kids a disservice if they are not held to a high stan-

dard. This is an equity issue. We need to hold everyone to the same high standard. We need to give accurate feedback—tell them when they don’t meet the standard and help them understand what we are looking for in actual work and in defense of that work. We need to get away from the soft and mushy. When teachers do this together they are developing knowledge of an agreed-upon standard for the community. They are holding students accountable and themselves accountable.⁵³

Building communities of practice

Hard but necessary conversations about equity are most productive when they are supported by a community of practitioners who have common goals. The work of sharing common resources and assessments, as well as engaging in cross-school scoring, nurtures this kind of professional community.

Todd Wallingford, curriculum director for secondary English language arts and social studies at Hudson Public Schools in Massachusetts, which is part of the QPA network, underscores the point:

For the past three years, Hudson’s collaboration with QPA has steered us toward building a stronger professional culture founded upon the development and scoring of common performance assessments aligned to Common Core standards.⁵⁴

Likewise, Priti Johari, redesign administrator at Chelsea High School in Chelsea, Massachusetts, notes that using common performance assessments and rubrics in her school has nurtured not only collaboration but also a culture of inquiry among teachers:

Our work of creating common assessments and rubrics and scoring them across classrooms has created a culture of inquiry and a collaborative atmosphere. Four years ago, classroom doors were closed and there was no collaboration. Twenty-five percent of the teachers in the school were a professional learning community. Now I believe 100 percent of the teachers experience themselves that way. This is a result of our process of learning about the Common Core, unpacking standards, writing lesson plans and tasks, sharing those plans, giving each other feedback, creating common rubrics, and collectively examining student work.⁵⁵

Moving from a culture of testing to a culture of teaching

Engagement with performance assessment has the potential to change discussions about what teaching and learning should be among educators. For example, in the Pentucket Regional School District in Massachusetts, every student in every school in addition to completing assignments and assessments based on state standards also has a portfolio of work that demonstrates standards for the district agreed-upon habits of learning. This portfolio is presented in a public forum in grades 4, 6, 8, and 11, and is evaluated using rubrics that are common across the district. William Hart, currently in the sixth year of this work, notes the power of this work:

Like a lot of districts, for many years, standardized testing had dominated the thinking of teachers in our district and defined the practices about what kids should know and be able to do. Teachers and administrators had focused their energies on finding the most expedient way to prepare kids for those exams. The unintended impact of this focus was more didactic teaching. As we have engaged in performance assessment development and scoring, a new balance has been brought to people's efforts. Now we assess our students' thinking, collaboration, independence, creative exploration as well as state standards.⁵⁶

The public nature of this work helps parents and family members understand and support their children's learning. Hart explains:

We have parents attend their children's presentations so that they can see what we do. It is not infrequent that parents leave these events in tears because they are just blown away by the deeper kind of learning and the broader set of skills, attributes, and habits they see that they never before saw in work for the old tests.⁵⁷

Improving teaching and student learning

The effort and time invested in having teachers design common assessment tasks and then score them through common rubrics yields dividends in regard to both teacher learning and the quality of student outcomes. Amy Woods, an eighth-grade English teacher at Cape Cod Lighthouse Charter School in East Harwich, Massachusetts, speaks to this point:

We have been doing performance assessment in our school since its inception. At the beginning, each teacher developed his/her own assessments and rubrics and

was scaffolding teaching and student assignments differently so that each class was being prepared differently. Now, with our common assessments, we have developed continuity in our rubrics across the grades. Our collective scoring of that work has given us a common language and more coherence in the school in terms of preparing kids across the continuum of development.⁵⁸

The Pentucket Regional School District in Massachusetts had a similar experience. As Hart notes:

Using common performance assessments with common rubrics has had a positive impact on our teachers, as well as our students. Teachers are now using the common rubrics to guide the type of project or task they can develop to marry concept/content acquisition and 21st-century skills. They are asking questions like: How do I shift the instructional environment to do both? How am I helping kids develop as collaborators? Measuring worthy skills and knowledge in this way has driven the context of the classroom. The work takes time and collaboration, but huge dividends are evident in the student work. Our high school state test performance is extremely positive. And what's interesting is that where performance assessments are being implemented with the greatest fidelity, we are getting the best test performance: One of our high schools was identified as an exemplar by our state commissioner. Our elementary school is the top in the state. The message I take from this is: If you do performance assessment well, then it is just good teaching and learning and kids are going to achieve.⁵⁹

Collaborative work in performance assessment that is managed skillfully can impact all layers of the school system. Laurie Gagnon, QPA director, explains:

The work takes a systems approach. It gives kids opportunities to learn and demonstrate their learning, embedded in a cycle of teacher learning and in what teachers need to be supported by their district and community. It touches upon all the different pieces of the system. It has led to deep thinking amongst teachers and administrators about rubrics and how to communicate with kids about what their next steps are.

Before [performance assessment work such as ours] teachers used to teach something and then move on. Now they are thinking about re-teaching and about making connections within a single source and across disciplines. They are identifying skills that are common across disciplines and helping kids understand how to read and use rubrics. They are learning from what kids are doing and

*helping them get to the next level. This work is an ongoing cycle of inquiry, both within and across disciplines. It takes a long time to get this right, but the impact is clear on the kids.*⁶⁰

A belief that performance assessment with extensive teacher and leader involvement can systemically strengthen learning for students and teachers is central to the policy initiatives of New Hampshire's State Department of Education. In New Hampshire, there is an effort underway to create an accountability system that includes performance assessment in addition to other paper-and-pencil tests. The goal is for all students to be involved in complex performance tasks that measure deeper learning over time, to expect teachers to be involved in the development and scoring of these tasks, and to create a pool of tasks for teacher use. Paul Leather, deputy commissioner of education in New Hampshire, explains:

*We have been involved in performance assessment for years and looking at student learning for years. But what we found was that it was very spotty. Some folks did a great job with it and others didn't know what we were talking about. Realizing this has led us to ask what we need to do better. We came to understand that preparation is not set up to help [teachers] support the degree of expected learning for students. And schools are not set up to encourage personalized learning going deeper. So we are trying to deconstruct and reconstruct our whole system so that teachers feel more supported. We want to move forward on a continuum toward deeper assessment that is more challenging for students and teachers too. We are aiming eventually to have a system where the students create their own tasks and teachers score them with common rubrics. Right now, though, teachers are creating the tasks and developing common rubrics that they use to assess against established competencies.*⁶¹

Leather believes that this performance-assessment development and collaborative scoring work has had a positive effect not only on teacher learning but on student outcomes as well. As evidence for this claim, he points to a decline in New Hampshire's high school dropout rate, an increase in the state's high school graduation rates, and an increase in the number of the state's graduates who go on to college. Leather attributes this to the more personalized teaching that results from the process of teachers' involvement with collaboratively developing and scoring complex performance assessments. According to Leather:

They are placing a lot more attention on depth of knowledge of the learning process. They are looking at assessment questions—are we asking students to do things that are going to be asked of them in the realities of their lives? We are encouraging teachers and students to take on deeper learning. We want to make sure that the assessments we use will incent the kind of teaching we want. This [whole process] has been a breath of fresh air for our teachers as well as our students.⁶²

Recommendations

In order to ensure that the kinds of performance-assessment opportunities outlined above can be successfully planted and scaled up as states and districts implement the Common Core State Standards and deepen their efforts to teach 21st-century skills, we suggest states, districts, and schools do the following:

- **Ensure that assessment is embedded in a learning system.** At the state and local levels, assessment should be considered part of a learning system for both students and adults, connected to curriculum, instruction, and professional development.
- **Include performance tasks as a part of assessment.** States and multistate consortia should include performance tasks as part of their systems of assessment and involve teachers in the design, use, and scoring of these tasks.
- **Make sure that criteria and rubrics for scoring tasks are clear and explicit for both students and teachers.** Sample tasks evaluating key standards should be publicly available for formative use in classrooms.
- **Involve teachers in collaborative scoring sessions.** State and districts should bring teachers together for training and moderation to learn to score reliably, and should include opportunities for teachers to discuss the implications of the standards and assessment results for their teaching.
- **Expand opportunities for teachers to engage in analysis of student work.** Although teachers may score the work of students other than their own for the purpose of summative evaluation, they should have timely opportunities to see the work of their own students—completed tasks and rubrics—to inform and guide their practice.
- **Provide teachers with coaching and professional development around assessment.** Teachers' engagement in examining and scoring student work should be

supplemented with coaching and professional development focused on teaching strategies to implement the new standards.

- **Build communities of practice.** States, districts, and schools should build communities of practice engaged in performance-assessment work with iterative opportunities to build tasks, collaboratively plan instruction, analyze student work and scores, and continually fine-tune practice.

Conclusion

Anthony Alvarado, the renowned educator and administrator—whose successful school reforms in New York City and San Diego, California, have been widely researched—several years ago called attention to the fact that in order to support improved student learning, educators need “to find ways of getting deeply into the specifics of how to help students master subject matter ... [and to] create contexts that support changes in thinking and pedagogy on the part of teachers.”⁶³

Alvarado wisely understood that a focus on how to better support teacher learning is critical to efforts aimed at improved student learning. What’s more, he knew that the ability of schools to develop students to meet the challenging demands of the future depends on the existence of teachers who are knowledgeable about the critical elements of learning and who can employ the strategies that are needed to connect these elements with the understandings of diverse learners. Involving teachers in the design, use, and scoring of standards-based, performance-based assessments is a powerful way to help teachers develop this knowledge and these skills.

Thus, the use of common standards-based performance assessments that are designed and collaboratively evaluated by teachers can have many benefits, including:

- Providing teachers with more direct and valid information about student progress than is offered by traditional assessments, especially on the deeper learning skills that characterize the Common Core State Standards
- Enabling teachers to engage in evidence-based work: reflecting more clearly and analytically on student work to inform their instructional decisions
- Yielding information that enhances teachers’ knowledge of students, standards, curriculum, and teaching, especially when scoring is combined with debriefing and discussing next steps with other teachers

By examining the work of their students, teachers increase their knowledge of individual students, become better informed about their students' capacities, and receive guidance about what they need to do next to support students' forward development. Teacher involvement in assessment helps teachers clarify their goals for teaching, make expectations of students explicit, create learning experiences that apply knowledge to real-life contexts, and provide many different ways for students to demonstrate their abilities and skills. It supports teachers' learning about state standards, their discipline, their students, and their teaching practices. The approach also offers teachers a forum for collaboration and an opportunity to learn. In other words, teacher involvement in standards-based and performance-based assessments lays the groundwork for better teaching and learning. As leaders in the Mathematics Assessment Collaborative note:

Teachers benefit from this approach as much as students do. ... Teachers laboring to improve student performance on a high-stakes exam can come to feel isolated, beaten down, and mystified about how to improve. ... The exigencies of test security mean that teachers often receive little specific information about where their students' performance excelled or fell short. When high-stakes test results come back, often months after the exam, teachers can do little with the results but regard them as a final grade that marks them as a success or failure.

Assessment that requires students to display their work ... is a tool for building the capacity of the teaching community to improve its work over time. The discipline of exploring together [what] we want students to know and the evidence of what they have learned is simultaneously humbling and energizing. Knowing that they are always learning and improving creates among educators a healthy, rich environment for change. To improve instruction requires that teachers become wiser about the subject they teach and the ways that students learn it. Performance assessment of students, with detailed formative feedback to teachers accompanied by targeted professional development, helps to build the teacher wisdom we need.⁶⁴

While standards and assessments provide teachers and students with explanations of and access to images of excellence, this awareness is not sufficient. An open and public guide to expectations for teaching and learning offers one way to help level the playing field between students who have experienced vastly unequal opportunities, resources, and supports for their learning. But adequate supports also need to be provided to build the capacities of all teachers and students to achieve these new and more challenging standards. Special attention and resources need to be

allocated to provide teachers and students, especially those in historically underserved communities, with the appropriate opportunities to learn so they can be sufficiently prepared to reach higher levels of success.

In addition, if tests are to support system learning, they must be used for information rather than for punishment. If they are to be educative, assessments cannot be used to allocate sanctions for teachers or schools that create competition where collaboration is needed, that use fear to shut down teacher learning, or that create incentives for keeping or pushing struggling students out of schools in order to boost scores.

Teacher involvement in the design, use, and scoring of performance assessments has the potential to powerfully link instruction, assessment, student learning, and teachers' professional development. If used wisely, it has the potential to address multiple important goals through one concentrated investment. It also offers a powerful way to evaluate what students know and can do while at the same time affirming teachers' knowledge and supporting their learning.

Continued use of high-quality standards and performance assessments over time has been shown to improve teaching and learning. As teachers become more expert about teaching, continued improvement and progress on the part of students can be expected. Not only will overall pedagogical capacity be enhanced, but also teaching and assessing will stay focused on its central purpose: the support of learning for all involved.

About the authors

Linda Darling-Hammond is the Charles E. Ducommun Professor at Stanford University, where she founded and co-directs the Stanford Center for Opportunity Policy in Education. She was the founding executive director of the National Commission on Teaching and America's Future and headed President Barack Obama's education policy transition team. Her award-winning book, *The Flat World and Education*, examines the U.S. education system in comparison to those of successful countries around the world.

Beverly Falk is a professor at the City College of New York's School of Education and a senior scholar at the Stanford Center for Assessment, Learning, and Equity, or SCALE. Falk has served in a variety of educational roles, such as classroom teacher; public school founder and director; district administrator; researcher; and consultant at the school, district, state, and national level. The founding and current editor of *The New Educator*, a quarterly journal about educator preparation, Falk is the author, co-author, and editor of many publications. Her most recent books are: *Defending Childhood: Keeping the Promise of Early Education*, *Teaching Matters: Stories from Inside City Schools*, co-authored with Megan Blumenreich, and *Teaching the Way Children Learn*.

References

- Abedi, Jamal. 2010. "Performance Assessments for English Language Learners." Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Abedi, Jamal, and Joan L. Herman. 2010. "Assessing English Language Learners' Opportunity to Learn Mathematics: Issues and Limitations." *Teacher's College Record* 112 (3): 723–746.
- Allen, David. 1998. *Assessing Student Learning: From Grading To Understanding*. New York: Teachers College Press.
- Alvarado, Anthony. 1998. "Professional Development Is The Job." *American Educator* 22 (4): 18–23.
- Barrs, Myra, and others. 1989. *The Primary Language Record: Handbook For Teachers*. London: Centre for Language in Primary Education.
- Beaver, Joetta, and Mark A. Carter. 2001. *Developmental reading assessment*. Parsippany, NJ: Celebration Press.
- Black, Paul, and Dylan Wiliam. 2010. "Kappan Classic: Inside the Black Box: Raising Standards Through Classroom Assessment Formative assessment is an Essential Component of Classroom Work and Can Raise Student Achievement." *Phi Delta Kappan* 92 (1): 81–90.
- Borko, Hilda, Rebekah Elliott, and Kay Uchiyama. 2002. "Professional Development: A Key to Kentucky's Educational Reform Effort." *Teaching and Teacher Education* 18 (8): 969–987.
- Borko, Hilda, and Brian M Stecher. 2001. "Looking at reform through different methodological lenses: Survey and case studies of the Washington state education reform." Paper presented as part of the symposium, *Testing policy and teaching practice: A multi-method examination of two states* at the annual meeting of the American Educational Research Association. Seattle, WA: April.
- Chapman, C. 1991. "What have we learned from writing assessment that can be applied to performance assessment?" Presentation at ECS/CDE Alternative Assessment Conference. Breckenbridge, CO.
- Clay, Marie. 2000. *Running Records For Classroom Teachers*. Portsmouth, NH: Heinemann.
- Cohen, Dorothy, Vivian Stern, and Nancy Balaban. 2008. *Observing And Recording The Behavior Of Young Children*. New York: Teachers College Press.
- Common Core State Standards Initiative. 2012. "In the States." (<http://www.corestandards.org/in-the-states> [July 11, 2013]).
- Darling-Hammond, Linda. 2010. "Performance Counts: Assessment Systems That Support High-Quality Learning." Washington: Council Of Chief State School Officers.
- Darling-Hammond, Linda. 2004. "Standards, Accountability, and School Reform." *Teachers College Record* 106 (6): 1047–1085.
- Darling-Hammond, Linda, and Frank Adamson. 2010. *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

- Darling-Hammond, Linda, and Jacqueline Ancess. 1994. "Authentic Assessment and School Development." New York: Columbia University National Center for Restructuring Education, Schools, & Teaching.
- Darling-Hammond, Linda, Jacqueline Ancess, and Beverly Falk. 1995. "Authentic Assessment In Action." New York: Teachers College Press.
- Darling-Hammond, Linda, and Beverly Falk. 1997. "Using Standards and Assessments to Support Student Learning." *Phi Delta Kappan* 79 (3): 190–199.
- Darling-Hammond, Linda, and others. 2009. "Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad." National Staff Development Council.
- Darling-Hammond, Linda, and Elle Rustique-Forrester. 2005. "The consequences of student testing for teaching and teacher quality." In Joan Herman and Edward Haertel, eds., *The Uses And Misuses Of Data In Accountability Testing*, pp. 289–319. Malden, MA: Blackwell Publishing.
- Darling-Hammond, Linda, and Laura Wentworth. 2010. *Benchmarking Learning Systems: Student Performance Assessment In International Context*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, Linda, and George Wood. 2008. *Assessment for The 21st Century: Using Performance Assessments to Measure Student Learning More Effectively*. Washington: The Forum for Education and Democracy.
- Dorfman, Aviva. 1997. "Teachers' Understanding of Performance Assessment." Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.
- Duncan, Teresa, and others. 2007. "Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement." National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- European Commission. 2007/2008. "The Education System In Finland." *Eurybase*.
- Falk, Beverly. 2001. "Professional Learning Through Assessment." In Ann Lieberman and Lynne Miller, eds., *Teachers Caught in the Action: The Work of Professional Development*. New York: Teachers College Press.
- Falk, Beverly, and others. 1999. "The Early Literacy Profile: An Assessment Instrument." New York: New York State Education Department.
- Falk, Beverly, and Linda Darling-Hammond. 1993. *The Primary Language Record at P.S. 261: How Assessment Transforms Teaching and Learning*. New York: National Center for Restructuring Education, Schools, and Teaching.
- Falk, Beverly, and Linda Darling-Hammond. 2010. "Documentation and Democratic Education." *Theory Into Practice* 49 (1): 72–81.
- Falk, Beverly, Susan MacMurdy, and Linda Darling-Hammond. 1995. *Taking A Different Look: How The Primary Language Record Supports Teaching for Diverse Learners*. New York: National Center for Restructuring Education, Schools, and Teaching.

- Falk, Beverly, and Suzanna Ort. 1998. "Sitting Down To Score: Teacher Learning Through Assessment." *Phi Delta Kappan* 80 (1): 59–64.
- Fiore, Lisa, and Stephanie Cox Suárez, eds., 2010. Observation, Documentation, and Reflection to Create a Culture of Inquiry." *Theory Into Practice* 49 (1).
- Finnish National Board of Education. "Background for Finnish PISA Success." (<http://www.oph.fi/english/SubPage.asp?path=447,65535,77331> [September 8, 2008]).
- Finnish National Board of Education. "Teachers." (<http://www.oph.fi/english/page.asp?path=447,4699,84383> [April 30, 2008]).
- Finnish National Board of Education. "Basic Education." <http://www.oph.fi/english/page.asp?path=447,4699,4847> [September 11, 2008].
- Firestone, William A., David Mayrowetz, and Janet Fairman. 1998. "Performance-Based Assessment and Instructional Change: The Effects of Testing in Maine and Maryland." *Educational Evaluation and Policy Analysis* 20 (2): 95–113.
- Foster, David, Pendred Noyce, and Sara Spiegel. 2007. "When Assessment Guides Instruction: Silicon Valley's Mathematics Assessment Collaborative." In A. H. Schoenfeld, ed., *Assessing Mathematical Proficiency*, pp. 137–154. Cambridge, MA: Cambridge University Press.
- Goldberg, Gail Lynn, and Barbara Sherr Rosewell. 2000. "From Perception to Practice: The Impact of Teachers' Scoring Experience on the Performance Based Instruction and Classroom Practice." *Educational Assessment* 6 (4): 257–290.
- Hatch, Thomas, and others. 2013. *Lessons from New York City's Local Measures Project*. New York: National Center for Restructuring Education, Schools, and Teaching.
- Herman, Joan L., and others. 1991. "A first look: Are claims for alternative assessment holding up." *CSE Technical Report*. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.
- Himley, Margaret, and Patricia Carini. 2000. *From Another Angle: The Prospect Center's Descriptive Review of the Child*. New York: Teachers College Press.
- Kates, Laura. 2011. "Toward Meaningful Assessment: Lessons From Five First-Grade Classroom." Working Paper 26. New York: Bank Street College.
- Koretz, Daniel, and others. 1996. "Perceived Effects of the Maryland School Performance Assessment Program." *CSE Technical Report*. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.
- Koretz, Daniel, Brian Stetcher, and E. Deibert. 1994. "The Vermont Portfolio Assessment Program: Findings and Implications." *Educational Measurement: Issues and Practice* 13 (3): 5–16.
- Lane, Suzanne, and others. 2000. "Consequential evidence for MSPAP from the Teacher, Principal and Student Perspective (Mathematics, Reading, Writing, Science, and Social Studies)." Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans, LA.

- Lieberman, Ann, and Lynne Miller, eds. 2001. *Teachers Caught in the Action: The Work of Professional Development*. New York: Teachers College Press.
- Little, Judith Warren. 1999. "Organizing schools for teacher learning." In Linda Darling-Hammond and Gary Sykes, eds., *Teaching as the Learning Profession: Handbook of Teaching and Policy*, pp. 233–262. San Francisco: Jossey-Bass.
- Little, Judith Warren. 1993. *Teachers' Professional Development in a Climate of Educational Reform*. New York: National Center for Restructuring Education, Schools, and Teaching.
- Little, Judith Warren, and others. 2003. "Looking at Student Work for Teacher Learning, Teacher Community, and School Reform." *Phi Delta Kappan* 85 (5): 184–192.
- Mathematics Assessment Resource Service. 2000. *Balanced Assessment for the Mathematics Curriculum*. Upper Saddle River, NJ: Dale Seymour Publications.
- McDonald, Joseph P. 2001. "Students' Work and Teachers' Learning." In Ann Lieberman and Lynne Miller, eds., *Teachers Caught in the Action: Professional Development That Matters*, pp. 209–235. New York: Teachers College Press.
- McLaughlin, Milbrey. 2005. "Listening and Learning From the Field: Tales of Policy Implementation and Situated Practice." In Ann Lieberman, ed., *The Roots Of Educational Change*, pp. 58–72. New York: Teachers College Press.
- Meisels, Samuel J., and others. 1996. *The Work Sampling System*. New York: Pearson.
- Murnane, Richard, and Frank Levy. 1996. *Teaching the New Basic Skills*. New York: The Free Press.
- Odden, Allen, and others. 2008. "The Cost of Instructional Improvement: Resource Allocation in Schools Using Comprehensive Strategies to Change Classroom Practice." *Journal of Education Finance* 33 (4): 381–405.
- Organisation for Economic Co-operation and Development. 2007. "PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis." (<http://www.pisa.oecd.org/dataoecd/30/17/39703267.pdf> [July 11, 2013]).
- Paek, Pamela L., and David Foster. 2012. "Improved Mathematical Teaching Practices and Student Learning Using Complex Performance Assessment Tasks." Paper presented at the annual meeting of the National Council on Measurement in Education, or NCME. Vancouver, Canada.
- Pecheone, Ray, and Stuart Kahl. 2010. "Through a Looking Glass: Lessons Learned and Future Directions for Performance Assessment." Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Pianta, Robert C., Karen La Paro, and Bridget K. Hamre. 2007. *The Classroom Assessment Scoring System (CLASS)*. Baltimore: Brookes Publishing.
- Resnick, Lauren. 1995. "Standards for Education." In Diane Ravitch, ed., *Debating the Future of American Standards*. Washington: Brookings Institution.
- Schmidt, William H., Hsing Chi Wang, and Curtis C. McKnight. 2005. "Curriculum Coherence: An Examination of US Mathematics and Science Content Standards from an International Perspective." *Journal of Curriculum Studies* 37 (5): 525–559.

- Seidel, Steve. 1998. "Wondering to be Done: The Collaborative Assessment Conference." In David Allen, ed., *Assessing Student Learning: From Grading to Understanding*, pp. 21-39. New York: Teachers College Press.
- Sheingold, Karen, K., J.I. Heller, and S.T. Paulukonis. 1995. "Actively Seeking Evidence: Teacher Change through Assessment Development." Princeton, NJ: Educational Testing Service.
- Sheingold, Karen, J.I. Heller, J.I., and B.A. Storms. 1997. "On the Mutual Influence of Teachers' Professional Development and Assessment Quality in Curricular Reform." Paper presented at the annual meeting of the American Educational Research Association. Chicago: April.
- Stecher, Brian, and others. 2000. "The Effects of the Washington State Education Reform on Schools and Classrooms." *CSE Technical Report*. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.
- Stecher, Brian, and others. 1998. "The Effects of Standards-Based Assessment on Classroom Practices: Results of the 1996-97 RAND Survey of Kentucky Teachers of Mathematics and Writing." *CSE Technical Report*. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.
- Thomas, William H., and others. 1995. "California Learning Assessment System Portfolio Assessment Research and Development Project: Final report." Princeton, NJ: Center for Performance Assessment, Educational Testing Service.
- Wei, Ruth C., Linda Darling-Hammond, and Frank Adamson. 2010. "Professional Development in the United States: Trends and Challenges." Dallas: National Staff Development Council and Stanford: Stanford Center for Opportunity Policy in Education.
- Wei, Ruth C., Susan E. Schultz, and Raymond Pecheone. 2012. "Performance Assessments For Learning: The Next Generation of State Assessments." Stanford: Stanford Center for Assessment, Learning, and Equity.
- Wolf, Shelby, and others. 1999. "No Excuses': School Reform Efforts in Exemplary Schools of Kentucky." Los Angeles: Center for the Study of Evaluation.
- Yoon, Kwang Suk, and others. 2007. "Reviewing the evidence on how teacher professional development affects student achievement." Washington: U.S. Department of Education.

Endnotes

- 1 Ann Lieberman and Lynne Miller, *Teachers Caught in the Action: The Work of Professional Development* (New York: Teachers College Press, 2001); Milbrey McLaughlin, "Listening and Learning From the Field: Tales of Policy Implementation and Situated Practice." In Ann Lieberman, ed., *The Roots Of Educational Change*, pp. 58–72 (New York: Teachers College Press, 2005); Linda Darling-Hammond and others, "Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad" (Dallas: National Staff Development Council and Stanford, CA: Stanford Center for Opportunity Policy in Education, 2009).
- 2 Kwang Suk Yoon and others, "Reviewing the evidence on how teacher professional development affects student achievement" (Washington: U.S. Department of Education, 2007), available at http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf.
- 3 For a summary of research, see Ruth C. Wei, Linda Darling-Hammond, and Frank Adamson, "Professional Development in the United States: Trends and Challenges" (Dallas: National Staff Development Council and Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010).
- 4 Common Core State Standards Initiative, "In the States," available at <http://www.corestandards.org/in-the-states> (last accessed September 2013).
- 5 William H. Schmidt, Hsing Chi Wang, and Curtis C. McKnight, "Curriculum Coherence: An Examination of US Mathematics and Science Content Standards from an International Perspective," *Journal of Curriculum Studies* 37 (5) (2005): 525–559.
- 6 Linda Darling-Hammond and Laura Wentworth, "Benchmarking Learning Systems: Student Performance Assessment in International Context" (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010).
- 7 Wei, Darling-Hammond, and Adamson, "Professional Development in the United States."
- 8 Linda Darling-Hammond, "Performance Counts: Assessment Systems That Support High-Quality Learning" (Washington: Council Of Chief State School Officers, 2010).
- 9 Lauren Resnick, "Standards for Education." In D. Ravitch, ed., *Debating the Future of American Standards* (Washington: Brookings Institution, 1995).
- 10 Linda Darling-Hammond and Elle Rustique-Forrester, "The consequences of student testing for teaching and teacher quality." In Joan Herman and Edward Haertel, eds., *The Uses And Misuses Of Data In Accountability Testing*, pp. 289–319 (Malden, MA: Blackwell Publishing, 2005).
- 11 Ray Pecheone and Stuart Kahl, "Through a Looking Glass: Lessons Learned and Future Directions for Performance Assessment" (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010).
- 12 For a summary, see Darling-Hammond and Rustique-Forrester, "The consequences of student testing for teaching and teacher quality."
- 13 Hilda Borko, Rebekah Elliott, and Kay Uchiyama, "Professional Development: A Key to Kentucky's Educational Reform Effort," *Teaching and Teacher Education* 18 (8) (2002): 969–987; Hilda Borko and Brian M. Stecher, "Looking at reform through different methodological lenses: Survey and case studies of the Washington state education reform," Paper presented as part of the symposium, "Testing policy and teaching practice: A multi-method examination of two states," at the annual meeting of the American Educational Research Association, Seattle, WA, April 2001; Beverly Falk and Suzanna Ort, "Sitting Down To Score: Teacher Learning Through Assessment," *Phi Delta Kappan*, 80 (1) (1998): 59–64; Linda Darling-Hammond, "Standards, Accountability, and School Reform," *Teachers College Record* 106 (6) (2004): 1047–1085; Karen Sheingold, J.I. Heller, and B.A. Storms, "On the Mutual Influence of Teachers' Professional Development and Assessment Quality in Curricular Reform," Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1997); Karen Sheingold, J.I. Heller, and S.T. Paulukonis, "Actively Seeking Evidence: Teacher Change through Assessment Development" (Princeton, NJ: Educational Testing Service, 1995); Shelby Wolf and others, "No Excuses: School Reform Efforts in Exemplary Schools of Kentucky" (Los Angeles: Center for the Study of Evaluation, 1999).
- 14 Falk and Ort, "Sitting Down To Score."
- 15 Beverly Falk, "Professional Learning Through Assessment." In Ann Lieberman and Lynne Miller, eds., *Teachers Caught in the Action: The Work of Professional Development* (New York: Teachers College Press, 2001).
- 16 Falk and Ort, "Sitting Down To Score."
- 17 Linda Darling-Hammond and George Wood, "Assessment for The 21st Century: Using Performance Assessments to Measure Student Learning More Effectively" (Washington: The Forum for Education and Democracy, 2008).
- 18 Jamal Abedi and Joan L. Herman, "Assessing English Language Learners' Opportunity to Learn Mathematics: Issues and Limitations," *Teacher's College Record* 112 (3) (2010): 723-746.
- 19 Falk and Ort, "Sitting Down To Score," p. 62.
- 20 Ibid.
- 21 Judith Warren Little and others, "Looking at Student Work for Teacher Learning, Teacher Community, and School Reform," *Phi Delta Kappan* 85 (5) (2003): 184–192.
- 22 Falk and Ort, "Sitting Down To Score."
- 23 Performance-assessment projects are underway at the state or local level in at least the following states: California, Colorado, Connecticut, Kentucky, Maine, Massachusetts, New Hampshire, New Mexico, New York, Ohio, Rhode Island, Vermont, and Washington.
- 24 David Foster, Pendred Noyce, and Sara Spiegel, "When Assessment Guides Instruction: Silicon Valley's Mathematics Assessment Collaborative," *Assessing Mathematical Proficiency* 53 (2007): 137–154.
- 25 Common Core State Standards Initiative, "Standards for Mathematical Practice" available at <http://www.corestandards.org/Math/Practice> (last accessed September 2013).

- 26 Inside Mathematics, "7th Grade," available at <http://insidemathematics.org/index.php/7th-grade> (last accessed September 2013); Inside Mathematics, "Performance Assessment Task: Lawn Mowing, Grade 7," available at <http://insidemathematics.org/common-core-math-tasks/7th-grade/7-2005%20Lawn%20Mowing.pdf> (last accessed September 2013); Sonoma County Office of Education, "MARS Tasks Grade 7," available at <http://www.scoe.org/files/mars-grade7.pdf> (last accessed September 2013).
- 27 Pamela L. Paek and David Foster, "Improved Mathematical Teaching Practices and Student Learning Using Complex Performance Assessment Tasks," Paper presented at the annual meeting of the National Council on Measurement in Education, or NCME, Vancouver, Canada, 2012.
- 28 Annual audits of a randomly selected 5 percent of the student papers find very high reliability; a recent analysis found that the mean difference between the original score and the audit score was only 0.01 point. Foster, Noyce, and Spiegel, "When Assessment Guides Instruction," pp. 137–154.
- 29 Foster, Noyce, and Spiegel, "When Assessment Guides Instruction," p. 141.
- 30 Ibid.
- 31 Paek and Foster, "Improved Mathematical Teaching Practices and Student Learning Using Complex Performance Assessment Tasks."
- 32 Ibid.
- 33 Foster, Noyce, and Spiegel, "When Assessment Guides Instruction," p. 141.
- 34 Ruth C. Wei, Susan E. Schultz, and Raymond Pecheone, "Performance Assessments For Learning: The Next Generation of State Assessments" (Stanford, CA: Stanford Center for Assessment, Learning, and Equity, 2012).
- 35 Ibid.
- 36 Ibid.
- 37 Ibid.
- 38 Ibid.
- 39 Ibid.
- 40 Ibid., p. 11.
- 41 Ibid.
- 42 Ibid., p. 42.
- 43 Ibid., p. 41.
- 44 Ibid., p. 45.
- 45 "Quality Performance Assessment," available at <http://www.qualityperformanceassessment.org> (last accessed September 2013).
- 46 Personal communication from Anne Clark, headmaster, Boston Arts Academy, October 2012.
- 47 Personal communication, October 2012.
- 48 Personal communication from William Hart, assistant superintendent, Pentucket Regional School District, October 2012..
- 49 Personal communication from Laurie Gagnon, director, Quality Performance Assessment, December 17, 2012.
- 50 Personal communication from Christina Brown, director, Quality Performance Assessment's Principal Residency Network, October 23, 2012.
- 51 Personal communication from Jeanne Sturges, staff developer, Souhegan High School, October 2012.
- 52 Personal communication from Gagnon.
- 53 Personal communication from Brown.
- 54 Quality Performance Assessment, "Clients."
- 55 Personal communication from Priti Johari, redesign administrator, Chelsea High School, November 14, 2012.
- 56 Personal communication from Hart, November 6, 2012.
- 57 Ibid.
- 58 Personal communication from Amy Woods, eighth-grade English teacher, Cape Cod Lighthouse Charter School, October 2012.
- 59 Personal communication from Hart.
- 60 Personal communication from Gagnon.
- 61 Personal communication from Paul Leather, deputy commissioner, New Hampshire Department of Education, November 20, 2012.
- 62 Ibid.
- 63 Anthony Alvarado, "Professional Development Is The Job," *American Educator* 22 (4) (1998): 18–23.
- 64 Foster, Noyce, and Spiegel, "When Assessment Guides Instruction," pp.

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just, and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is “of the people, by the people, and for the people.”

Center for American Progress

