

the NASBE Interview

Linda Darling-Hammond is Charles E. Ducommun Professor of Education emeritus at Stanford University and faculty director of the Stanford Center for Opportunity Policy in Education. She is also the president of the newly founded Learning Policy Institute. She recently has written and lectured extensively on performance assessments—learning that requires students to use high-level thinking to perform, create, or produce something with transferable real-world application—and the role of such assessments in states' systems of assessment. Ace Parsi, NASBE's director of deeper learning, interviewed her on this topic on October 7, 2015.



Linda Darling-Hammond
Professor at Stanford University
and Director of the Stanford Center
for Opportunity Policy in Education

How does the conversation around measures of success account for the needs of disadvantaged students? Is subgroup accountability a realistic goal when it comes to the deeper measures of success that you, Pedro Noguera, and Diane Friedlaender have written about?

The deeper measures of success, which often include performance assessments and various kinds of performance tasks, are explicitly focused on equity in several ways. Over the 20th century, a “thinking” curriculum was reserved for a limited number of students who were thought to be going into more knowledge-based occupations while others were just getting basic skills to prepare them for factory line or menial work. Disproportionately, low-income students and students of color were not getting access to problem-solving, critical-thinking curricula. An approach that brings this curriculum to all students and engages them all in research, analysis, writing, scientific inquiry, mathematical modeling, and problem solving creates more access to the 21st century curriculum that virtually all kids need today.

Over the years, we have learned how to score these kinds of measures reliably—both in the US and in other countries where they are more commonly part of the assessment systems. We can score performance tests reliably for large-scale purposes, and we can score them reliably for small-scale and local purposes. And they can be reported with individual and collective scores overall and by subgroups, just like any other assessment can be. So you can also monitor progress for all groups on their performance within these

types of measures, just like you could with a multiple choice test.

Will performance assessment results be as easily communicated to parents in those populations so that it is actionable for them?

It certainly can be. Advanced Placement courses and tests in world languages, English language arts, and the arts all [have] open-ended tasks, essays, portfolios, and other types of tasks that are scored reliably and are reported in ways that people understand. Performance tasks are being introduced in the AP science courses now, and they have long been present in the International Baccalaureate program. You can analyze these data by different groups of students, by different schools, etc. There is nothing about performance assessment that makes it “un-scoreable” or “unanalyzable.”

Think about Singapore’s high school examinations: Those students design and conduct science investigations over a couple of months. They get numerical scores for those investigations, the same way they would for sit-down essay tests. These data can be analyzed by the kinds of students or kinds of schools, by the kind of instruction they have. And that is what we need. It does cost a bit more to score these kinds of tests, because most of the time, real people have to do it, but the cost is also producing benefits for student learning and teacher learning.

Part of the benefit is that, as teachers are involved in scoring that kind of task or test, they become more knowledgeable about the standards and what they mean in operation, how to teach to the standards, what kind of work is proficient, and how to develop that

work. So they become better teachers as they become involved in scoring. A lot of studies show that teachers find scoring of performance assessments to be excellent professional development. It helps them design their instruction and their curriculum in the classroom in ways that will enhance student's success.

What steps should states be taking, or thinking about taking? What's the low-hanging fruit?

If state boards are interested in developing performance assessments in their states, they are probably going to want to engage in some state-supported work and some local work. The Innovation Lab Network created the Performance Assessment Resource Bank that a number of states are using—New Hampshire, Colorado, California, Virginia, and others. Some states have teachers developing tasks that come into the bank; they and others can also download tasks. They can use those for state assessments or local assessments as part of a new system, and some teachers simply use them in their classrooms. These are matched to standards, topics, and concepts [and] have been through a quality control process. So there are tools to get started.

Policymakers need to think about where this fits in their assessment systems. In New Hampshire, which has a waiver from the federal government under ESEA flexibility, pilot districts use a traditional state sit-down test in some years (the Smarter Balanced test) and then they also have what they call common tasks. These are performance-based assessments used across all the districts that are included in the waiver, which are eventually going to be used statewide. They have kids engaged in these rich tasks; they then organize teachers to score the tasks in a way that will produce consistent scores. Beyond that, the state expects local districts to choose existing tasks or create their own tasks for other subject areas—for example, in history and social science, or science in years when there is not a state assessment. So the kids are getting a steady diet of performance-based, project-based learning. They are engaged in investigations and inquiry in multiple disciplines, writing, speaking, communicating, and using technology.

So a short-term step would be tapping into an existing bank, a medium-term step is thinking

about how this fits within a state's comprehensive assessment and accountability system, and a long-term step is making sure educators are using this information to improve instruction?

Right. A lot of states have started to create graduation competencies, and a lot of districts have begun to create a list of competencies that reflect the 21st century skills they want to be sure students have. A number of schools and networks of schools and districts have begun to create assessments to measure those competencies and portfolios of student work that are evaluated against those competencies. When they reach a high enough standard to be part of the portfolio, they use that as evidence that they have accomplished certain competencies. Quite often, states will say to local districts, "Look, here are the competencies we will agree to. You need to develop a system locally to assess these competencies," and often that takes the form of a portfolio. And so another thing that [can be] shared are portfolio systems. That is another way states can begin to encourage higher order thinking and performance.

Having teachers score these kinds of performance tasks together, as many schools do, is helpful. For example, where they are involved in a portfolio assessment [group] like the New York Performance Standards Consortium, teachers are creating a shared practice around the work that students are engaged in and the standards that are used to judge that work. They calibrate that with each other and share curriculum ideas to improve the work. You can encourage better practice through that lens. You can further improve collaboration and collegiality because teachers are designing and scoring the work together. A good example of this is the Quality Performance Assessment group in New England, where many educators are doing this kind of work schoolwide.

How do you think this focus on performance assessment affects how we evaluate teachers and leaders and how we support them?

Districts should use some standards-based observation tool to look at teachers' practices, in order to have good evidence of what they are doing. We should also want to understand something about what kids are learning. And we should want to understand how [teachers and leaders] are working collegially to improve

the school. This kind of assessment addresses all three of those areas.

You can also use the data from performance assessments to inform the evaluation of teachers and leaders, where they are setting goals and collecting evidence of student learning, which illustrates that they are making progress toward those goals. These kinds of tests can provide some of that evidence.

When we are talking about teacher and leader effectiveness, we not only want to measure it, we want to develop it. So we want the tools used in evaluation to cause teachers to deepen their instruction, improve their curriculum, and support students more completely.

Is this the instrument to deal with questions of hiring and firing at the district level?

Districts have different data that they use to hire teachers initially. But subsequent personnel decisions have to be made based on a holistic appraisal of a teacher's or leader's accomplishments. We got off on a problematic foot when districts were asked to rate teachers based on a value-added test score. Researchers have found those scores to be unreliable, unstable, and inaccurate, particularly for teachers whose kids are in the groups that traditionally score above or below the grade level. In addition to the fact that there are many things that affect test score gains beyond teachers—for example, administrative and instructional supports, curriculum materials, class sizes, parent supports, and home contexts—the learning gains of some groups of students cannot be measured fully on current state tests because they lack items above and below the grade level.

Instead, we are rating teachers against each other with a tool that is not accurate. We have so many unexpected and unintended surprises. For example, in some states and districts virtually all the teachers of “gifted and talented” students were rated low on value added because the kids had already maxed out on the tests and couldn't go any higher than 100 percent. The decision of retaining or firing a teacher should be based on a holistic appraisal of what they do in the classroom, in the school, how their students learn on multiple indicators, and how they interact with students, parents, and others. Those processes [should be] well informed by rich data. If we just focus on getting one thermometer score, for example, we can miss some very important

things and fail to diagnose the problem, if there is one.

In what ways do current state assessment and accountability systems promote the type of learning you envision, and how do they inhibit it? Is there an exemplary state in this regard?

During the pre-NCLB era, Massachusetts continued to evaluate student thinking, writing, and communication on its assessments. Connecticut and Kentucky held out for a long time with their curriculum-embedded performance assessments, even though the law and the then Department of Education's way of implementing [ESEA] were making it very difficult. Right now, we are rebooting. The assessments that Smarter Balanced and PARCC have created are richer than the tests most states were using; they include more performance-oriented tasks.

California, New Hampshire, and Iowa were involved in a project that had teachers score the Smarter Balanced performance tasks together and reflect on the implications of those tasks for their instruction and curriculum. Teachers gave that experience rave reviews. More than 95 percent said they learned more about the standards, they learned more about the assessments, that it helped them improve their instruction and their curricula, and they feel that teachers should be involved in scoring these tasks. States can consider this kind of practice when they issue contracts for administering their state tests. In California, for example, teachers will be asked to score the performance tasks, and this will be emphasized as a part of the new system.

So you envision this as the beginning of the system of assessment for states and not the whole system of measures?

Exactly. For many states, that will be the spine of the system, which will include other, deeper measures that supplement and augment what you can learn from a sit-down test. States like New Hampshire, Kentucky, and Colorado are augmenting those systems with performance assessments that may be the harbingers of the next era of deeper learning. In the real world, people do not give you a sit-down test to gauge and shape your activities. You need to know how to plan and implement some kind of investigation or problem-solving activity, including sometimes a collaborative activity. We need

these kinds of assessments in our systems as well because students need to learn what they are going to be doing in college and careers.

There is a very interesting new Advanced Placement senior project/seminar, in which kids design and conduct a collaborative project. There are different papers, projects, and activities that students need to do that are scored by teachers. The folks at the College Board are training teams of teachers to offer this new course, and they are overwhelmed by the number of requests to bring this new AP course into a range of schools. They can only incorporate a certain number of teachers each year because they have to train teachers to do it, so they have more demand than they can meet right now. And for the first time, the demographics of the students being served by this course match exactly the demographics of the students in the country. So this new course is getting much wider reach. That is the direction of the future. You see it already in International Baccalaureate, which is really the gold standard for project-based learning and performance assessment. These kinds of tests help teachers to be better instructors, allow kids to do more complex work, and help create systems to incentivize and evaluate national progress toward more productive learning.

How can states avoid making it seem like performance assessment is just a fad that people can ride out?

If you are going to initiate complex strategies like this, it is going to take time, and people need to understand that. There will be folks who will raise their hands right away to volunteer, saying, “Please, we want to be involved in the early piloting and testing.” That’s a particularly good strategy if people can watch and be informed by their colleagues across schools about what is working. It is important that these and other assessments not be used for sanctions and punishment and are [couched] clearly in terms of improvement.

What I have seen in so many of the schools that I have worked with are lots of kids saying, “This is so much better than a multiple choice test. We really like doing these projects, we really feel like we are learning.” As long as they have their learning supported and have the time to do it well, they are enjoying it, and their teachers feel enriched by the experience.

In your 51st State accountability paper, you discuss “reciprocal accountability.” Can you elaborate on it here?

We are coming out of an era where accountability was conceptualized as very top down: The state holds schools and teachers accountable for achieving test score gains or other outcomes. Reciprocal accountability says that everyone in the system has some responsibility. The state’s responsibility is to provide useful guidance and adequate, equitable distribution of resources for schools and districts to meet the standards that the state has articulated. The districts then have to organize those resources in intelligent ways to accomplish the goals and evaluate how they’re doing. Further, they have a responsibility to look at data around students and schools so that they can manage and continue to improve the system. Schools have to use the resources well and fairly on behalf of the kids, to invest in developing high-quality instruction, and to manage the process of engagement with parents. Teachers have the responsibility to teach productively toward the standards, to take advantage of the professional learning that the states and districts are responsible for providing, and to continue to help the students to reach further. Students and parents have to take advantage of educational opportunities that are available to them and to work collaboratively with educators. Reciprocal accountability is thinking about how everyone in the system is accountable and responsible for the parts of the system they can control, and how everyone is responsible for continuous improvement.

Are there examples of places moving in that direction?

Singapore is a really interesting place, and Finland is another. They have this idea of continuous improvement and ongoing evaluation, and they are always [collecting] data about what’s working, what’s not working, and how it is working. They are always giving that data and information back to the school-based educators and those who lead the system, including the ministry of education. They are always thinking about how they can support better, stronger, more equitable learning. Those are good examples of places where we can see tremendous gains in achievement due to the commitment to continual learning. ■