

Making Sense of Educational Assessment: Bridging Research and Practice —Webinar Q&A

This document includes an edited transcript of questions asked and answered during the webinar, held on Dec. 11, 2023.

Which assessments are best suited to continuous improvement cycles and what protocols did you pair with them?

Chris Brandt: In terms of which ones are best, it really depends on purpose and use. So, what's the purpose of the assessment? What's the intended use of the assessment? That should guide your decision on how to use it.

Just to give a concrete example: Interim assessment, for example, like Measures of Academic Progress (MAP), or an assessment like that might be better used at a district level to monitor changes, whereas a more practical measure of protocol—for example, a Google spreadsheet that multiple teachers can get into and that have common questions that they can fill out that can then be aggregated on a biweekly basis for them to use as a way to discuss commonalities, differences, and ways to improve—would be an example of one that would be more formative at a classroom level.

Are common formative assessments across the district a good idea? What are the pros and cons?

Bill Penuel: I'll start by speaking about one of the impacts of our particular course that we found, which is that for some teachers who do not have highly aligned curriculum materials, they realized, "Wow, in order to actually help our students meet this five-dimensional vision for science learning, we need high quality instructional materials." I start there because I think a lot of districts try to implement common assessments without having common high-quality instructional materials. And I think that's a mistake. You can implement common assessments when you have an adopted set of materials, when you have a common set of materials that help teachers know to do what's next with their assessment information.

High-quality instructional materials, by definition, have good assessments and they have suggestions for rubrics, for feedback, for strategies for what to do next. That's what we built in to OpenSciEd for high school and middle school and I'm sure are being built in right now into elementary school.

The evidence is that when people use assessment data that is collected at the state level or even interim data, the use of those things is too hard to actually have a positive impact on student learning. The rigorously designed studies don't show a positive impact of that kind of data use.

So I would emphasize the value of common assessments. When there's an adopted common set of instructional materials, the teachers can then use to know what to do next. Without that, you don't see the hoped for benefits of these other forms of assessment and the value of doing common assessments.

How did you select your control group in your study?

Bill Penuel: Everyone volunteered because they wanted to participate in our course. Once they had signed up for our course, we did a block random assignment basically by state to either the treatment or control. So, teachers signed up individually for this rather than schools.

How does project-based learning and performance assessment fit into the assessment context?

Bill Penuel: In our particular context, like a glove in the sense that one of the things that we're trying to do is have people engage in a complex performance, we call it a transfer task. By transfer, what we mean is that students are having to explain a related phenomenon to the one that they would have learned about in their class, usually as part of a problem-based or project-based unit.

Problem-based learning is a particularly important or central kind of instructional model for the Next Generation Science Standards and framework-based standards, and so the kinds of tasks and performance tasks that we're focusing on preparing teachers to do are really intended to fit because they serve the same goals of having students synthesize and make use of knowledge that they have developed through an extended unit to explore related phenomenon or a problem in a new context.

Chris Brandt: I'll just concur with what Bill said and say performance tasks are essential for understanding deeper learning. So, assessing skills that require students to demonstrate deep understandings that can't be elicited through multiple choice questions would be essential if you really want to understand whether the types of instructional strategies that you implement elicit evidence of deeper learning from students.

The performance expectations often contain more than one skill or learning target. Does the course provide them with the know-how to narrow the focus and align the instructions to that focus?

Bill Penuel: The course really focuses on helping them to think about what they should be assessing and to narrow the focus on what we call the different elements of each of the dimensions and the targeted performance expectation. We focus on a single performance expectation to kind of constrain the space. But most phenomenon actually would probably invite you more naturally to mix some of those elements, but the key is to make sure that those are 3D.

The focus is not on helping them to narrow and focus and align instruction because one of the things that we're assuming or hoping in the future is that people have high-quality instructional materials that they're working with and adapting for use that actually address a range of the performance expectations through a compelling phenomenon or design problem.

Are your instructional materials tied to assume curriculum for any particular grade level? Are the assessments related to actual teaching of those curricular concepts?

Bill Penuel: For the current version of the course, no. Teachers used a wide range of materials, and the assessments were related to their own teaching of curricular concepts. So, we asked teachers in the course

to choose a performance expectation that they would be teaching towards as part of their instruction. So, they picked a wide range of performance expectations to work with. And this was secondary, so it's middle and high school.

Is there a rubric used to measure the performance-based outcomes? Are these rubrics standardized, teacher created, or both?

Chris Brandt: Just broadly speaking, yes, rubrics are often used to measure performance-based tasks and outcomes. What's really essential with these rubrics is they need to be connected to an underlying growth target or growth trajectory, so demonstrating what adequate performance looks like, but also what evolving performance looks like. So, what would I expect to see in a student who maybe is meeting these learning targets? What would I expect to see of a student who is going beyond those performance targets? What would I expect from a student who maybe is approaching those targets but hasn't quite gotten there yet? And so, those are the kinds of things to think about when you're designing rubrics.

The rubrics can be either standardized or teacher created. Again, it kind of comes back to purpose and use. So if the intention of the rubric is to elicit evidence of student knowledge and understanding across a group of students, then you would want to make sure that was standardized. But it doesn't necessarily have to be.

Bill Penuel: [In our project] there are two kinds of rubrics, one that we used to evaluate the quality of teachers' assessment. That was a common one that we used in our research team and spent more than a couple of years actually working towards a final rubric. But then teachers also generated scoring guides for their particular assessments that they developed.

What we're really looking for in those kinds of scoring guides especially is addressing each of the three dimensions of standards. I would also say one ideal is to really think about, related to Chris's comment about learning targets, is what kind of feedback would you give to students that would actually help them improve or deepen their thinking? So, again, I mention a lot of high-quality instructional materials there. Embedded assessments will do that and provide some guidance that can be useful to teachers.

How do you measure or assess the diverse ways of learning that students have?

Chris Brandt: Just generally, one way to represent the diverse ways of knowing and doing is through performance tasks, because performance tasks allow for voice and choice. So, for example, you can embed voice and choice into how students demonstrate what they know and can do. It could be done by allowing a student to orally present their learning, to write a paper, or to provide evidence through other modes.

Katie West: Two things come to mind. One was a story that I just heard from a kindergarten teacher that said she had a student who was having to go to the principal's office repeatedly, was having a number of behavior difficulties that were getting reported. And then once they started the unit of going outside and developing the relationship with the maple tree, the student who had been throwing chairs every day of school completely stopped. So, they saw a behavioral intervention decrease both in their reporting and then also observational and this child's demeanor completely changed as long as they could go visit the tree every day.

Secondly, we've also seen that students, once we give them these field experiences, then want to bring—and do bring—their families to go see the river, so a place that families may not have had access or experience before then become a part of the family and then the community experience.

Katie Coppens: For our Rethinking Responsive Education Ventures (RREV) grant, students ended up creating a portfolio that was really reflective of their own individual journey throughout their learning. For each of our domains that we were looking at, they had three different paths they could take to choose, and we had four different domains in science that they were examining. So, if you look at those combinations, they had a lot of opportunities for that voice and choice to decide how they wanted to individualize their learning path, of what area was of interest for them. I think that was really important for that performance-based assessment, that they were doing things but also that the choice was built into it. I think students really responded to that.

Another layer of it was the fact that they were a pilot group for us, of our first time doing this unit. I think students really felt honored to give us feedback on those choices and how they felt to them, and to give them that added layer of helping us reflect on our teaching to make it better for this current year that we're going into.

Bill Penuel: One of the things that is really important is choosing a context or phenomenon for a scenario that drives learning. So we really encourage folks to take an inventory of students' interests and also community priorities and concerns.

In our instructional work, we also conduct work to learn about relative interests of different kinds of phenomenon related to a set of standards.

There are other ways we can do this formatively, that often we don't think about. For example, inviting students to make connections between the day's lessons and something that's important to them in their lives, something that they're interested in, and then bringing that the next day and actually present and have some kind of discussion about that to help build the connection, because interest is something that we can co-construct with students. So just because a student doesn't start out interested in something doesn't mean they can't become interested. But they have an active role in that process, and we can use assessment actually to generate and have students generate those kinds of connections and bring them into the classroom as part of the class discussion.

What types of assessments have you developed or used to measure student content knowledge or practices?

Katie Coppens: One of our domains involved river testing, where students were looking at abiotic and biotic indicators in the Presumpscot River. And they had three different options of what to do. This was the area that was the heaviest in terms of data analysis for them.

So, some of them were looking at river samples that we had analyzed as a class and looking at patterns over time from different locations. We had some that were in Casco Bay, so higher levels of salinity compared to upriver, obviously. And seeing those changes.

So, when we designed our different areas of the portfolio, we really thought about different opportunities for students to show their knowledge. There was another option that they could choose between microorganisms, where we had them set up in our classrooms with microscopes and they could look at them and do some research to understand them more. Many students ultimately chose to do more than one because they were really engaged by the different options. But I think having that data analysis as a layer of what they were working on really connected well to the work that they were doing at that time in math, which was exciting for everybody, especially the math teachers, to see that connection.