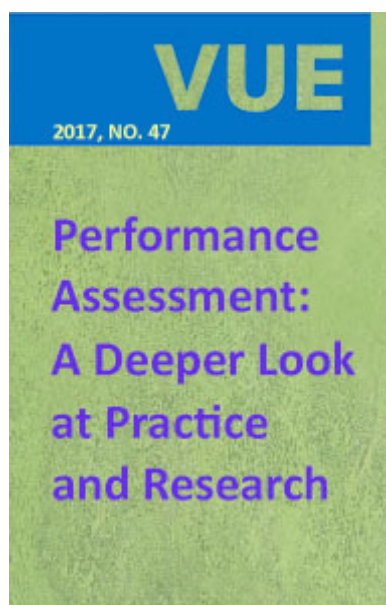


Performance Assessment: A Deeper Look at Practice and Research

2017, NO.47



<http://vue.annenberginstitute.org/issues/47>

This issue is an online supplement to VUE 46, which addressed the topic of performance assessment – a personalized and rigorous alternative to standardized testing that allows teachers to build on individual students’ strengths and foster more equitable learning outcomes.

VUE 47 adds additional current materials, offers opportunities for additional voices, and provides more examples of performance assessment. Because performance assessment is an active national conversation, the work continues; following VUE 46’s publication, important national conferences and other milestones occurred that we’re able to share here. This issue also provides perspectives from students, educators, researchers, and policymakers.

About this issue

As with VUE 46, this issue was planned and produced in partnership with the Center for Collaborative Education (CCE), a nonprofit organization established to promote innovative models of schools and increase justice and opportunity for all learners, and was inspired by CCE’s work on a new approach to student performance assessment as an alternative to standardized testing. For more information on CCE and resources related to performance assessment, visit <http://cce.org>.

Related resources

Please see VUE 46, [*Performance Assessment: Fostering the Learning of Teachers and Students*](#), for comprehensive references and resources on the subject of performance assessment. In addition, for more, we recommend the following:

Readings:

Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L. (1994). [Performance-based assessment and educational equity](#). *Harvard Educational Review* 64(1):5-30.

Darling-Hammond, L., & Falk, B. (2013). *Teacher learning through assessment: How student-performance assessments can support teacher learning*. Washington, DC: Center for American Progress.

Resources:

[Performance Assessment Resource Bank](#)

Built by the Stanford Center for Assessment, Learning and Equity and the Stanford Center for Opportunity Policy in Education, in collaboration with members of the Council of Chief State School Officers' Innovation Lab Network, the Performance Assessment Research Bank is a curated selection of performance assessment tasks, supplemented by related design and implementation resources.

[Quality Performance Assessment](#)

Created by the Center for Collaborative Education, Quality Performance Assessment is a new generation of processes and tools that leverage assessment as an essential tool of equitable learning and teaching.

Articles

Why Not Push Us to Think?

By Ai-Zhen Tung

A high school sophomore reflects on the limits of and alternatives to high-stakes standardized tests. Ai-Zhen Tung wrote this reflection during her sophomore year at Boston's Fenway High School.

Performance Assessment Examples from the Quality Performance Assessment Network

By Christina Kuriacose

Four examples spanning grade levels and disciplines demonstrate the range and possibilities of performance assessment. These examples were compiled by Christina Kuriacose, Program Associate, Quality Performance Assessment, Center for Collaborative Education.

Seizing the Opportunity for Performance Assessment: Resources and State Perspectives

By Laura Gutmann, Christina Jean, and Joey Hunziker

This article reports from Stanford University's Innovative Assessments Institute on the development of performance assessment at scale, along with implementation recommendations.

GreenTalks at Boston Green Academy: Student Reflections on Performance Assessment

By Christina Kuriacose

High school students reflect on their participation in GreenTalks, a public exhibition of their research on food justice issues in Boston. Interviews conducted by Christina Kuriacose, Program Associate, Quality Performance Assessment, Center for Collaborative Education

Evaluating Comparability in the Scoring of Performance Assessments for Accountability Purposes

By Susan Lyons and Carla Evans

Researchers report on their evaluation of comparability claims in local scoring of performance assessments across districts participating in New Hampshire's Performance Assessment of Competency Education pilot project.

Voices from the Field: Performance Assessments in State Accountability as discussed at the National Conference on Student Assessment

By Susan Lyons and Yuxi Qiu

This field report from 2017's National Conference on Student Assessment shares possibilities for flexibility and innovation in assessment and accountability made possible by the Every Student Succeeds Act.

Exploring the Role of Performance Assessment in Competency-Based Education: Recommendations from the National Summit on K-12 Competency-based Education

By Laurie Gagnon

Based on reports created for the National Summit on K-12 Competency-based Education, this article explores how educators can take action to build momentum for and implement competency education and performance assessment.

A New Age of Implementation: Guiding Principles for Implementing Performance Assessment Systems

By Gary Chapin, Laurie Gagnon and Virgel Hammonds

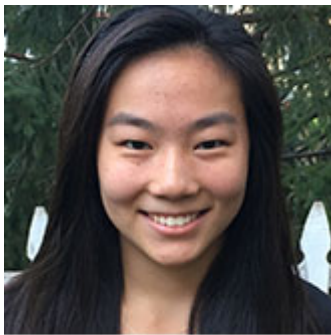
In an examination of the conditions required for the successful implementation of performance assessment, the authors draw on a range of personal experience and other insights to guide practitioners and policymakers.

Why Not Push Us to Think?

by Ai-Zhen Tung

<http://vue.annenberginstitute.org/issues/47/why-not-push-us-think>

A high school sophomore reflects on the limits of and alternatives to high-stakes standardized tests. Ai-Zhen Tung wrote this reflection during her sophomore year at Boston's Fenway High School.



Ai-Zhen Tung

After taking the Massachusetts Comprehensive Assessment System (MCAS) test this past spring, I'm sure all Boston Public School students would agree that they are a waste of time and an inaccurate representation of our intellect. For three days straight, we sit in solemn silence while taking a boring test; our brains definitely aren't at their best. In fact, the MCAS actually measures our lowest possible potential rather than our highest. What I wanted to do was fill an open response section ranting about how ineffective MCAS is. By publishing my thoughts here, I hope to gain the attention of people who have power to make change.

Two summers ago, I participated in a science and math camp where a group of students spent six hours solving a multistep problem that would, at the end, open a locked box. The activity tested us in multiple subjects and taught us life lessons like leadership, cooperation, group work, and persistence. As a student who learns best kinesthetically and visually, the MCAS is the worst type of test you can give me, because while I'm trying to focus on answering meaningless questions, I'm also doing anything I can to keep the boredom away. The questions on the MCAS are irrelevant. Why not push us to think about and address world issues like immigration, racism, and politics?

Performance Assessment Examples from the Quality Performance Assessment Network

by [Christina Kuriacose](#)

<http://vue.annenberginstitute.org/issues/47/performance-assessment-examples-quality-performance-assessment-network>

Four examples spanning grade levels and disciplines demonstrate the range and possibilities of performance assessment. These examples were compiled by Christina Kuriacose, Program Associate, Quality Performance Assessment, Center for Collaborative Education.



Christina Kuriacose

The following performance assessments are strong examples of teacher-developed performance assessments from schools within the Center for Collaborative Education's Quality Performance Assessment network. These performance tasks demonstrate the pedagogical decisions teachers made, as well as the ways the experience allowed for deeper learning. While these summaries do not address the full curriculum, expectations, or teaching context in which the tasks are embedded, they do offer a consistent format to explore task details and learn more about work resulting from teacher-created, student-centered design.

DEPTH OF KNOWLEDGE LEVELS

These performance assessment examples refer to Depth of Knowledge Levels from Quality Performance Assessment, which are:

Depth of Knowledge 1: Recall; memorization; simple understanding of a word or phrase.

Depth of Knowledge 2: Covers level 1 plus: paraphrase; summarize; interpret; infer; classify; organize; compare; and determine fact from fiction. There is a correct answer, but may involve multiple concepts.

Depth of Knowledge 3: Students must support their thinking by citing references from text or other sources. Students are asked to go beyond the text to analyze, generalize, or connect ideas.

Requires deeper knowledge. Items may require abstract reasoning, inferences between and across readings, application of prior knowledge, or text support for an analytical judgment about a text.

Depth of Knowledge 4: Requires higher-order thinking, including complex reasoning, planning, and developing of concepts. Usually applies to an extended task or project. Examples: evaluates several works by the same author; critiques an issue across time periods or researches topic/issue from different perspectives; longer investigations or research projects.

See www.qualityperformanceassessment.org for additional information.

PERFORMANCE ASSESSMENT EXAMPLE 1

Background

Your Community Project: Understanding and Describing a Community

School: Plymouth Elementary School, Plymouth, NH

Grade level: 1

Content area: Social Studies, ELA

Teacher Authors: Sarah Carlson, Kristen Kilduff, and Karen McLoud

Plymouth Elementary School, part of SAU 48 in New Hampshire, entered performance assessment work by focusing on assessment literacy. Over three years, the school's professional development plan focused on building the full staff's capacity to design, validate, and implement performance assessments. As part of their collaborative practice, Plymouth teachers engage in monthly validation sessions that are facilitated by a volunteer group of teachers. Incorporating feedback is a key focus as the faculty members build their assessment literacy. Plymouth Elementary School is also a Tier 2 (building capacity) member of New Hampshire's Performance Assessment for Competency Education (PACE), a statewide initiative in which member districts and schools are exempt from most state standardized testing and use teacher-generated local and common performance assessments in its place to assess student learning. Beginning in the 2017-2018 school year, the school will become a Tier 1 PACE school; the PACE performance assessment system becomes their state accountability model.

The Task

The Your Community task is designed to help students see the diversity of people, places, and things that are included in communities. After completing a unit on the community, students draw a picture of a community, real or imagined. They are asked to include and label people, buildings, animals, and nature. Once the maps are completed, students present their maps to the class, explaining how people use the different spaces in their communities, what makes their community inviting or uninviting to live in, and what people do in their community.

- Topic: What is a community?
- Genre: Social Studies and ELA

- Depth of Knowledge: 3 – requires students to synthesize information from multiple texts and develop a complex model
- Habits/Skills/Dispositions
 - Communication
 - Creativity
 - Self-Direction
- Voice and Choice: Students create their own communities
- Audience: Classmates and teacher
- Time Frame: Three weeks, including current reading unit, the drawing process, and the presentation process

Teacher observation on the Your Community performance task

“The prior unit fell around Thanksgiving and Christmas time, so we had been focusing on community in general. Starting in October, we learned about community helpers. For Thanksgiving, we participated in creating centerpieces for lunches. We were active in our community, not just learning about community. I had told the students we would be creating our own communities later in the unit, but halfway through the community unit, they were already asking to create their own. When I introduced the task, they cheered!”

PERFORMANCE ASSESSMENT EXAMPLE 2

Background

Building Bold Bills: Integrating Civics into Persuasive Writing

Daniel J. Bakie Elementary School, Kingston, NH

Grade level: 4

Content area: ELA and Social Studies

Nels Tooker, Kathleen McLaughlin, and Jillian Zeeben

Bakie Elementary School is part of the Sanborn School District in New Hampshire. Sanborn is a vanguard district in New Hampshire in terms of competency-based education (competency-based education sets broad learning targets of what students should know and be able to do; students progress as they demonstrate proficiency over them), and was one of the initial districts in the founding cohort of NH PACE schools. Since 2012, Sanborn has been hosting a summer symposium on competency-based learning, inviting schools from all over the state and beyond to attend. Because their approach to competency-based learning is vertically aligned from kindergarten through 12th grade, the Sanborn School District frequently hosts site visits for schools exploring the shift towards competency-based learning.

The Task

With this task, three teachers aim to integrate civics standards into their fourth grade lesson about government through a persuasive writing task. Pairs of students are asked to research an issue that is important to them. Based on the research, they individually write persuasive essays proposing a state bill. In their essays, they must consider arguments and counter-arguments. Finally, the students go back to their pairs to draft a final proposal, combining their ideas. These proposals are brought to a mock New Hampshire Senate session where the students proposing the bill act as co-sponsors and their peers as Senators. In this formal mock Senate session, the co-sponsors stand up and present their bill. If the bill is passed after a Senate vote, it goes on to the governor for veto or signing into law. This task introduces students to the law-making process, the balance of power in government, and the process of decision-making about civic issues that are important to them.

- Topic: The United States' law-making process
- Genre: Persuasive writing – the goal of students' essays is to convince their peers to accept their proposal
- Depth of Knowledge: 3 – requires synthesis of multiple sources and evidence to support arguments
- Habits/Skills/Dispositions
 - Communication
 - Collaboration
 - Creativity
 - Technology use
- Voice and Choice: Students select their topic
- Audience: Classmates in a mock Senate
- Time Frame: 7 days (45- to 60-minute sessions each day)

Teachers' observations on the Building Bold Bills performance task

"Collaboration, communication, and creativity are important skills that students employ in this task.... Students come up with amazing things. One of the groups that we had this year wanted to limit tattoos. They researched the topic and called the bill "Think Before You Ink," which I think is really catchy. The persuasive writing that they did was top-notch for what they were capable of, because they cared about the issues they chose and this choice is really important for engagement with the task."

"This task elicited our students' best work because we worked on creating a task that is both a lot of fun and grounded in a real situation. Hopefully the leaders in our classrooms become the leaders in our country and in our states, and will use the skills they learned in my fourth-grade classroom to make the world a better place!"

PERFORMANCE ASSESSMENT EXAMPLE 3

Background

Designing Lincoln's Playground: Combining Mathematical, Writing, and Social Skills

Abraham Lincoln School, Revere, MA

Grade level: 4

Content area: Math/ELA/Art

Lani Pini-Cabral, Rebecca Cohen, Marisa LeManquis

Abraham Lincoln School is one of three Revere schools in the first cohort of Massachusetts Consortium for Innovative Education Assessment (MCIEA) schools, a partnership of Massachusetts public school districts and their local teacher unions to create fair and effective accountability systems through holistic measures for school quality and performance assessment development. During the second half of the 2016-2017 school year, a team of teachers attended four MCIEA Quality Performance Assessment Institute days in order to develop assessment literacy and experience the design-align-analyze cycle through creating an original task. The Lincoln team created and piloted four performance assessments in fourth grade math, second grade ELA, third grade science, and fifth grade social studies as a means of introducing the school to performance assessment. Performance assessment design is a key focus for the school's Professional Learning Groups as the school builds out their assessment system. Also, assessment of and for student learning is one of three core foci for the Revere Public Schools' 2016-2021 Improvement Plan.

The Task

With the Designing Lincoln's Playground task, three teachers integrate mathematical skills into a performance assessment that involves social skills and persuasive writing. Students create a blueprint of the playground for the Lincoln School, including various structures they would like to see in their park. They measure the recess yard in meters and create a 2D version with a partner in math class, measuring the area and perimeter in square meters and meters. Using their math skills, they draw a detailed square/rectangle of each playground structure on a poster. Each of these squares/rectangles symbolizes the play structure's scale in real life. Then the students write a persuasive essay in ELA to convince an audience that their playground is the best design for the school. Students present their designs to the Lincoln School families and community. Selected designs are built as a 3D model collaboratively, to scale, in art classes.

- Topic: Application of area and perimeter
- Genre: Mathematical Skills/Persuasive writing
- Depth of Knowledge: 3 – requires synthesis of multiple sources and evidence to support arguments
- Habits/Skills/Dispositions
 - Communication

- Collaboration
- Creativity
- Voice and Choice: Students design their own playgrounds
- Audience: Classmates, Family, Community
- Time Frame: 9+ days (additional time for building the winning design)

Teacher observation on the Designing Lincoln’s Playground performance task:

“One visiting teacher from Boston noted that I could also take this assessment and connect it to our science curriculum. She suggested that I have students think about the playground structure and materials if the playground were moved to a desert area or an area where it rained often. She suggested pushing students to think of the effects of erosion. I loved this idea! This connection had never crossed my mind. Within minutes, I had a valuable suggestion on how to strengthen this assessment in the future.”

PERFORMANCE ASSESSMENT EXAMPLE 4

Background

Senior Capstone

Regional School Unit 2, Maine

Grade level: 12

Content area: All

All teachers

Regional School Unit 2 (RSU 2) serves the towns of Dresden, Farmingdale, Hallowell, Monmouth, and Richmond, and the Senior Capstone is the culmination of its proficiency-based, learner-led system. The Senior Capstone for Maine RSU 2 was developed at Hall-Dale High School (one of three high schools in the district) as part of the school’s early exploration of personalized and agency-building practices. The entire faculty participates as mentors. “Senior Capstone” is a course that seniors take, so that the time and coaching they need is available to them. The creation of the course – and the first two years of the program – was the work of teacher Linda Aronson, who later went on to write about the experience and philosophy of Senior Capstone in her book, Unleashed to Learn.

The Task

The Senior Capstone at RSU2 is designed to support the senior student in pursuing a fascination and demonstrating it to faculty, students, and community. Students choose a topic (e.g., tiki carving, boat building, business creation). They research that topic in depth, write a research paper, deliver a 20-minute public talk/exhibition that is open to the public, and “enact” their topic in an authentic way (e.g., they carve a tiki, build a boat, start a business). Students begin in December and present throughout the end of their senior year. Every student has two mentors: a

faculty member in the school and a community member who can coach the student in their area of expertise.

As can be seen in [this video](#) featuring the Senior Capstone at Hall-Dale High School in Regional School Unit 2 in Maine, the energy and import attached to the annual celebration is powerful. Students learn in the eighth grade what will be expected of them in the 12th, and they prepare and are coached through their high school career.

- Topic: How do I learn?
- Genre: Multiple
- Depth of Knowledge: All are touched on in the process, but the structure of the capstone easily lends itself to Depth of Knowledge 4, advanced application and transfer
- Habits/Skills/Dispositions
 - Communication
 - Creativity
 - Self-Direction
 - Collaboration
- Voice and Choice: Students have choice at nearly every step of the way
- Audience: Faculty, students, community
- Time Frame: Three to five months: capstone starts in December and students present from March to May

District superintendent observation on the Senior Capstone performance task

“In the learner-centered model, we want kids to apply the content to a subject matter that is meaningful to them every single day, not just in their senior year. Initially, the kids struggle with it. ‘What do I want to do? What do I want to tackle? What goals do I have for myself?’ So they struggle through the planning process, much like we, as adults, struggle when we start something new. When it comes to the point in the year when they're presenting their culminating project, it's amazing to see the pride and the confidence that they exude when they're making their presentations.”

Student observation on the Senior Capstone performance task

“For my Capstone project, I did animal adoption. For my fieldwork, I organized a dog speed-dating event for the Kennebec Valley Humane Society. I realized how much I needed to be self-reliant. It was really up to me to make my own deadlines, to be assertive in finding my expert mentor, to get what I wanted done, and getting the results I wanted. And I learned I really enjoyed organizing events. So that was fun.”

Seizing the Opportunity for Performance Assessment: Resources and State Perspectives

by [Laura Gutmann, Christina Jean, and Joey Hunziker](#)

<http://vue.annenberginstitute.org/issues/47/seizing-opportunity-performance-assessment-resources-and-state-perspectives>

This article reports from Stanford University's Innovative Assessments Institute on the development of performance assessment at scale, along with implementation recommendations.



Laura Gutmann

Christina Jean

In recent years, as states adopted higher standards to define student success in college and careers, they discovered a significant challenge: how to assess higher-order thinking and complex skills. For the most part, standardized tests typically in use did not sufficiently answer that challenge.

Assessment systems such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced attempted to go beyond basic skills and look for deeper learning – that is, not just the right answer, but also evidence that the student understood and could apply knowledge. Open-ended and “constructed response” questions allow students to show how they thought critically and analytically, how they organized their thinking, and how they constructed a coherent argument.¹ But limitations remained to the depth and function of the information about student learning gleaned from these assessments.

This is not a new challenge. For years, while researchers, policy makers, superintendents, principals, and teachers have understood the limitations of standardized tests, they wracked their brains trying to develop ways to better assess students’ harder-to-measure higher-order skills. One potential solution that many state and local leaders have been exploring emerges from the significant advances over the last two decades in the fields of performance assessment and student portfolios. Across the country, state education agencies (SEAs) and local education agencies (LEAs, or districts) have begun to collaborate with teachers on the ground to develop *performance-based assessment systems* that increase our understanding of how students process complex texts, respond to challenging prompts, and employ important skills to solve real-world challenges.

An accountability system built on the implementation of performance assessments has the potential to foster deeper and more authentic learning for students and more agency and assessment literacy for educators and school leaders. By investing in educators' capacity for performance assessment, states and districts support the development of better-prepared, more-empowered educators who use curriculum-embedded performance assessments as part of an instructional cycle.

REAL ASSESSMENT FOR REAL STUDENTS

During the autumn of 2016, 120 teacher leaders, coaches, district coordinators, professional development providers, and administrators from 16 states and a number of districts gathered at Stanford University to build their capacity to measure deeper learning. They had been selected to participate in Stanford's Innovative Assessments Institute in order to further leverage their roles in facilitating the effective implementation of K-12 performance assessments within their local contexts. Stanford Center for Assessment, Learning and Equity (SCALE) assessment experts from across the core content areas had organized the conference with support from the Hewlett Foundation and the Stanford Center for Opportunity Policy in Education (SCOPE), with the goal of demonstrating how newly developed resources to guide the design and development of performance tasks could be utilized to bolster a growing interest in performance assessment across the country.

Conference attendees listened to students enthusiastically describe how engaged they were with the meaningful instruction and assessment at their schools, which all had a history of developing assessments that ask students to perform, create, or produce something that would authentically demonstrate their learning.

Yet, familiar themes began to emerge as the conversation turned towards the practical. Participants wondered how to vet existing tasks for quality, better link such assessments to instructional units, and align their efforts to broader accountability standards. Teachers were providing students with opportunities to apply their knowledge, skills, and content understandings to novel problems and issues in the world, but lacked a consistent, coherent assessment system. States, districts, and teachers needed resources to build, scale, grade, and validate performance-based assessments on a large scale.

RESPONDING TO THE GROWING NEED FOR PERFORMANCE ASSESSMENT RESOURCES: PARB

One tool for implementing performance-based assessments at scale is the [Performance Assessment Resource Bank \(PARB\)](#), a selection of performance assessment tasks, supplemented by related design and implementation resources, launched in October 2016. PARB was created by SCALE and SCOPE, in collaboration with members of the Council of Chief State School Officers' Innovation Lab Network (ILN). PARB serves as a complement to emerging assessment policies and practices that foster deeper learning experiences for all students (Cook-Harvey & Stosich, 2016; Darling-Hammond, et al., 2016).

The resource bank was built over the course of three years, spurred by the need to fill a gap in readily accessible performance assessment tools that had been vetted for quality. The partnership between SCALE, SCOPE, and the ILN was formed as a coordinated response to address this issue, based on feedback from state leaders who recognized that their local capacity to establish

quality tools was lacking. Although there were pockets of progress where educators had developed their own ground-level assessments, they could not adequately respond to the increased demand for performance-based tasks that met uniform learning and design standards. Instead, a mixed bag of "home-cooking" had left consumers looking online for tasks with no way to distinguish between high and poor-quality tasks or to easily find assessments that aligned with their instructional objectives.

SCALE solicited contributions from dozens of like-minded organizations at the forefront of performance assessment, such as the Literacy Design Collaborative (LDC), Educational Policy Improvement Center (EPIC), and Center for Collaborative Education (CCE) in order to collect and curate curriculum-embedded tasks that allow K-12 students to more authentically demonstrate their learning in ELA, math, science, and history/social studies. PARB staff standardized a review process and calibrated reviewers according to set quality metrics in order to create a user-friendly bank of searchable rubrics, tasks, tools, and relevant policy research. Individual educators also played an important part in developing PARB materials, both as beta testers of the bank and as authors of performance assessments that were then reviewed according to the bank's criteria. PARB community members can also submit resources and tasks for potential inclusion in the bank and receive feedback according to quality criteria. States can take advantage of PARB's existing performance assessment resources while submitting their own tasks to the bank for validation.

LESSONS LEARNED ABOUT IMPLEMENTING PERFORMANCE ASSESSMENTS

Educators and assessment experts from states, districts, and education organizations worked hard to make high-quality performance assessments and exemplars widely available by contributing resources to PARB. However, pilot testing during the beta phase of bank development leading up to its launch revealed that building a collection of resources is only the first step towards promoting effective use of those materials. To facilitate effective use of performance assessments to drive instructional decisions, we must also develop deeper district and network-level partnerships within the field and offer customized support to educators trying out bank resources that considers their varied contexts for teaching and learning and makes use of their on-the-ground expertise. The work samples and benchmarks provided through the PARB are most useful when positioned as a starting point for a deeper process.

For example, one day of the Stanford event was dedicated to examining sample student work products generated by performance assessment tasks, so participants could see firsthand the connection between the implementation of PARB resources and the impact on student learning. Facilitators emphasized to the educators the importance of examining student work to check the assessment's instructional effectiveness and identify adjustments to the task that would increase its effectiveness in eliciting high quality student work. A participant noted that "having teachers review samples of student work, apply rubrics, and develop shared understandings of intended learning outcomes (as well as inter-rater reliability) is vitally important" and emphasized that these practices should be incorporated early in the process of developing performance-based assessments.

One enthusiastic state leader pointed out that discussing student work products relative to assessment expectations makes the most impact when multiple teachers are able to come together to share the results of a common task that was applied across a network of schools or

classrooms. When that process is established, “examining student work and calibrating with colleagues is paramount in helping identify misconceptions and can transform the use of assessment to inform instruction and to provide useful feedback.” Educators can further analyze performance assessment results to hone in on student needs that might otherwise be overlooked, such as providing scaffolds for English learners that address the language demands embedded within performance task assignments.

Thus, teachers are most effective when they go beyond the bank to consider the application of performance tasks to their particular classrooms, within a structure for assessment implementation that provides time and space for collaboration, conversation, and adaptation. Ideally, as the participants in the Stanford event hoped for, “analyzing the content of the student work” will reveal “patterns, trends, strengths, and needs” that teachers can then respond to, rather than placing “too much of an emphasis on the grade.”

RECOMMENDATIONS FOR LEADERS

As LEA and SEA leaders consider their next steps related to performance assessments and accountability, we offer the following considerations for shaping future work:

- **Balance maintaining high standards for both academic and non-academic learning with flexibility to allow for local innovation.** If we are to continue to drive gains in equity for all students, LEAs and SEAs must build systems of accountability and assessment that ensure readiness for college and career with regard to both academic and non-academic skills. This balance will require a reorientation away from the “traditional” role of the district and state focused on compliance and toward allowing local innovation to flourish in the development and implementation of assessment and accountability policies.
- **Support educators and school and district leaders in developing assessment and data literacy.** To engage networks of educators in developing quality assessments across a region or state, SEAs and LEAs must invest in supporting practitioners systemically and systematically in order to enable reliable, valid, comparable, and equitable performance assessment implementation and scoring, as has been the case in New Hampshire’s PACE pilot.
- **Seek out established resources to guide performance assessment goals.** The creation of digital libraries like PARB with a range of formative assessment instruments, curriculum resources, and instructional modules has the potential to provide educators with robust tools for assessment implementation, while moving away from a one-size-fits-all approach to measuring learning.
- **Make connections with educators and leaders engaged in similar efforts.** States and districts building systems of quality performance assessment can learn from each other’s work as they share best practices, contribute to open-ended educational resources, and receive coordinated support.

PERFORMANCE ASSESSMENT RESOURCE BANK (PARB)

Educators from across the globe can access the Performance Assessment Resource Bank (PARB) without charge at www.performanceassessmentresourcebank.org. Thousands of educators are already engaged in implementing PARB resources, and the bank's creators continue to collaborate to meet the needs of the field and create equitable access to tools for developing meaningful instruction and assessment practices, especially for English language learners and other students who have been traditionally underserved.

FOOTNOTES

1 PARCC and Smarter Balanced were developed through collaborations between groups of states and educators in response to new, more rigorous Common Core academic standards adopted by most states in 2010 and 2011. See <http://parcc-assessment.org/> and <http://www.smarterbalanced.org>. A “constructed response” question is one that requires students to supply their own response, rather than pick among multiple-choice answers.

REFERENCES

Cook-Harvey, C. M. & Stosich, E. L. (2016). *Redesigning school accountability and support: Progress in pioneering states*. Stanford, CA: Learning Policy Institute and Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L., Bae, S., Cook-Harvey, C., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Stanford, CA: Learning Policy Institute and Stanford Center for Opportunity Policy in Education.

Wei, R.C., Pecheone, R.L. Pecheone, & Wilczak, K.L. (2014). *Performance assessment 2.0: Lessons from large-scale policy and practice*. Stanford, CA: Stanford University.

Wei, R. C., Pecheone, R. L., & Wilczak, K. L. (2015). Measuring what really matters. *Phi Delta Kappan*, 97(1), 8-13.

GreenTalks at Boston Green Academy: Student Reflections on Performance Assessment

by Christina Kuriacose

<http://vue.annenberginstitute.org/issues/47/greentalks-boston-green-academy-student-reflections-performance-assessment>

High school students reflect on their participation in GreenTalks, a public exhibition of their research on food justice issues in Boston. Interviews conducted by Christina Kuriacose, Program Associate, Quality Performance Assessment, Center for Collaborative Education



Christina Kuriacose

In spring 2017, for the third year running, 10th graders at Boston Green Academy (BGA) presented GreenTalks, a showcase of research on food justice issues. The day I visited the school, students were presenting the PowerPoints they had put together. All of them included a map plotting out the proximity of their neighborhood or BGA to grocery stores, bodegas, and fast food restaurants. This mapping was meant to show where there is limited access to healthy food in the areas most students were living in. Almost all the presenters commented on the lack of proximity to grocery stores, where it is easiest to get fresh produce. The presentations additionally included the data from student-conducted surveys. These surveys asked statements and questions ranging from "I know someone who has diabetes," to "If I had more access to healthy foods I'd eat it more often" to "How often do you eat fast food?" The students represented this data in pie charts or bar graphs. Some of the presenters spoke about nutrition programs in other countries and others about local healthy food programs, in order to highlight ways to increase access to healthy foods in less affluent areas.

Some of the calls to action at the end of presentations included:

- We need to lobby the government to subsidize healthy food.
- Local community gardens will provide fresh produce options to neighborhoods far away from grocery stores.
- Access to healthy food is a social justice issue that we need to rally around.

In a conversation with several participating students, I asked about the process of moving through this performance assessment. One student described her GreenTalk as “an end-of-the-year presentation that we do based on what we have been learning about in all our classes for the last few months. This year the focus was food in Boston. We could choose to research the availability of food in either Boston or our school.” Another student added, “We were talking about food scarcity in all of our classes. You could really see the connections, because we kept coming back to the same topic in math, English, social studies.”

For these students, this long-term research was the most extensive project they had been asked to take on by that point in their academic careers. When asked about what she thought when first encountering the GreenTalks performance assessment task, one student responded with trepidation, “You want us to do what? It was really overwhelming and looked like a lot of work. I didn’t really know how to get started. But then it was easier to see the connections between the different classes and how that built towards the project.”

Another student agreed that initial anxiety about the GreenTalk subsided as they delved into the work. “It’s a lot harder than what we are used to. At first, I was thinking, I’d rather just write an essay or something, like the sort of stuff we do in other classes. I wasn’t excited at all. But once I chose my topic, I began to get more into it. It’s connected to our lives; it’s closer to my real life than a lot of other stuff we do.”

Student got to choose to work on city-wide food issues or the quality of food at school. One student observed, “I chose to do food in Boston because it was easier. Doing the research was easy because I’m researching my life. I just talked with friends and my family about where they get food from and the quality of food they can get and I had all the information I needed.”

A student who focused on food at BGA said, “I chose food at BGA because it’s not very good. I wanted to show teachers what we think of food and what we would like to see.” This student offered the conclusion that it is inefficient to ship food for the Boston Public Schools from a production factory in Long Island, New York, noting also that the food isn’t fresh by the time it makes it to Boston, and that it would be better to invest in local companies to provide the food for schools.

When asked to reflect on the differences between the GreenTalks project and assessments they had done in the past, students described their own learning curves. “At first, I thought it was pointless,” said one student. “But I appreciate the independence of it all. I’m so used to having a right answer and right way of doing things, so when we were given this big open-ended project, it didn’t connect clearly to class. But as I got started, I saw how I was using what I learned in my own way. It was cool because I was using stuff from class in a way that was relevant to my life.” Another student agreed, observing, “I’d rather do this because it’s more fun. It’s easier because you have a lot of people doing the same thing and you can get more help. If there is something you don’t really get, you can ask another student. We helped each other out in doing the research and developing the projects.”

Offering final thoughts for educators in the process of developing assessments, one of BGA’s 10th graders offered sage advice. “Every student learns in a different way. Make projects that allow for that. There needs to be visual teaching and hands-on learning. Some people need that for things to really click.”

Evaluating Comparability in the Scoring of Performance Assessments for Accountability Purposes

by [Susan Lyons and Carla Evans](#)

<http://vue.annenberginstitute.org/issues/47/evaluating-comparability-scoring-performance-assessments-accountability-purposes>

Researchers report on their evaluation of comparability claims in local scoring of performance assessments across districts participating in New Hampshire's Performance Assessment of Competency Education pilot project.



Susan Lyons



Carla Evans

This brief summarizes “Comparability in Balanced Assessment Systems for State Accountability,” published in *Educational Measurement: Issues and Practice* (Evans & Lyons, 2017). This study evaluated comparability claims in local scoring of performance assessments across districts participating in New Hampshire’s Performance Assessment of Competency Education (PACE) pilot project.

With the passage of the Every Student Succeeds Act (ESSA), there has been increasing attention on how states could design innovative assessment and accountability systems to submit for approval under the law. The challenge lies in designing assessment and accountability systems that can support instructional uses while serving accountability purposes (Baker & Gordon, 2014; Gong, 2010; Marion & Leather, 2015). New Hampshire’s PACE project is an example of one such system. Within PACE, local and common performance assessments administered throughout the school year contribute to students’ overall competency scores, which are in turn used to make annual determinations of student proficiency for state and federal accountability. The challenge is using the information from multiple, local assessment sources to support comparable scoring across districts.

Comparability

Comparability is not an attribute of a test or test form, nor is it a yes/no decision. Instead, comparability is the degree to which scores resulting from different assessment conditions can support the same inferences about what students know and can do. In other words, can the scores resulting from different assessment conditions be used to support the same uses (e.g., school

evaluation)? Comparability becomes important when we make the claim that students and schools are being held to the same standard, particularly when those designations are used in a high-stakes accountability context.

METHODS AND RESULTS

There are many different methods for gathering evidence to support score comparability evaluations. Contrasting conceptions of comparability typically include statistical and judgmental approaches, or some combination of the two (Baird, 2007; Newton, 2010). The chosen approach is dependent upon the nature of the assessments and the intended interpretations and use of the test scores (Gong & DePascale, 2013). We apply two judgmental methods for estimating comparability that are used in international contexts: consensus scoring and rank-ordering.¹

Consensus Scoring

The consensus scoring method involves pairing teachers together, each representing different districts, to score student work samples from students outside of either of their districts. The student work samples are common performance tasks given across districts in particular grade or subject areas. Examining the work samples one at a time, the judges discuss their individual scores and then come to an agreement on a consensus score. The purpose of collecting consensus score data is to approximate “true scores” for the student work. To detect any systematic discrepancies in the relatively leniency and stringency of district scoring, we calculated averages differences in local teacher scores and the consensus scores (mean deviation index). Using this index, a negative deviation indicates an underestimation of student scores by classroom teachers (i.e., district stringency), and positive mean deviation indicates overestimation of student scores by classroom teachers (i.e., district leniency).

Across all districts, the consensus scoring yielded scores that were positive, meaning they were a bit lower than the teacher-given scores. This finding itself is not necessarily problematic from a comparability perspective, as long as the relative leniency of the teacher-given scores is even across districts. Analysis revealed uneven scoring across districts, suggesting that there remains a need for additional training on scoring and within-district calibration, as well as for increased cross-district calibration.

Rank-Order Method

High school biology presented a unique challenge in calibrating the cross-district scores because there was no common performance assessment administered across districts in this discipline; each district developed and implemented completely unique tasks. Typically, score calibration procedures require one of two conditions: 1) common persons across tasks, or 2) common tasks across persons. Because neither of these conditions was satisfied in the 2014-15 implementation of high school science in PACE, we looked to alternate methods of score calibration and modeled our method after the rank-ordering cross-moderation method used in England.

The seven participating judges, all high school science teachers, were given packets of student work that had been grouped by average rubric score and represented student work from biology performance tasks from all four districts. After training and an opportunity to familiarize

themselves with the different performance assessments from the four districts, the judges were instructed to rank papers within each packet based on merit, evidence of student understanding, demonstrated competence, and student knowledge of science, which are all different ways of saying “better,” as Bramley (2007) succinctly puts it. The rank orders from teacher judges were converted to scores, which were compared to original teacher scores.

The results revealed scoring differences across districts, most notably in one district in which teachers were scoring their student work a full standard deviation below (more rigorous) where the judges placed the same student work within the sample.

CONCLUSION

We found that applying the two methods in the context of the PACE system highlighted some strengths and limitations of each method. First, both methods provide comparability evidence in local scoring within a district. Both methods also involve teachers from multiple districts reviewing student work samples from other districts, which has the added benefit of providing a rich context for professional development. In previous research on the effects of high-stakes performance-based assessment systems on student performance (Borko, Elliott, & Uchiyama, 2002; Lane, Parke, & Stone, 2002; Parke, Lane, & Stone, 2006), professional development had a strong mediating effect on the relationship between the performance-based assessment system and changes in teacher instructional practices. Using a one of these methods not only provides the evidence necessary of comparability in local scoring, but also provides a built-in professional development opportunity for teachers.

That said, reviewing student work samples across districts is costly and time-consuming. The practicality and feasibility of scaling up the proposed methods in a large-scale performance assessment program is a real concern, particularly within a state that has many more districts or other units with a large number of different local assessment systems. One way New Hampshire has attempted to address scale issues is through improved technology. As this project continues to scale, New Hampshire is undergoing an intensive research and development process to procure additional software that will support the scaling of this effort through virtual task design and scoring.

States awarded flexibility under ESSA’s Innovative Assessment Demonstration Authority will have to demonstrate that all students have the same opportunity to learn and are held to the same performance expectations. In so doing, accountability systems based on school-based assessments or other innovative assessment systems permitted under the Innovative Assessment Demonstration Authority must provide evidence to support comparability claims. The methods presented in this brief provide tools to strengthen the body of evidence related to the comparability of scores across districts implementing an innovative system of assessments.

FOOTNOTES

1 Consensus scoring is a version of external moderation used in Queensland, Australia (Queensland Studies Authority, 2014). Rank-ordering is a version of cross-moderation used in England (Bramley, 2005, 2007).

REFERENCES

- Baird, J. A. (2007). Alternative conceptions of comparability. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 124-165). London: Qualifications and Curriculum Authority. Retrieved from <http://hdl.handle.net/1983/1004>
- Baker, E. L., & Gordon, E. W. (2014). From the assessment OF education to the assessment FOR education: Policy and futures. *Teachers College Record*, 116(11).
- Borko, H., Elliott, R., & Uchiyama, K. (2002). Professional development: A key to Kentucky's educational reform effort. *Teaching and Teacher Education*, 18, 969-987. [http://doi.org/10.1016/S0742-051X\(02\)00054-9](http://doi.org/10.1016/S0742-051X(02)00054-9)
- Bramley, T. (2005). A rank-ordering method for equating tests by expert judgment. *Journal of Applied Measurement*, 6(2), 202-223.
- Bramley, T. (2007). Paired comparison methods. In P. E. Newton, J. A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 246-300). London: Qualifications and Curriculum Authority.
- Evans, C., & Lyons, S. (2017). Comparability in innovative assessment systems for state accountability. *Educational Measurement: Issues and Practice*. <http://doi:10.1111/emip.12152>
- Gong, B. (2010). *Using balanced assessment systems to improve student learning and school capacity: An introduction*. Washington, DC: Council of Chief State School Officers and Renaissance Learning.
- Gong, B., & DePascale, C. (2013). *Different but the same: Assessment "comparability" in the era of the Common Core State Standards*. Washington, DC: The Council of Chief State School Officers.
- Lane, S., Parke, C. S., & Stone, C. A. (2002). The impact of a state performance-based assessment and accountability program on mathematics instruction and student learning: Evidence from survey data and school performance. *Educational Assessment*, 8(4), 279-315. <http://doi.org/10.1207/S15326977EA0804>
- Marion, S., & Leather, P. (2015). Assessment and accountability to support meaningful learning. *Education Policy Analysis Archives*, 23(9). Retrieved from <http://dx.doi.org/10.14507/epaa.v23.1984>
- Newton, P. E. (2010). Contrasting conceptions of comparability. *Research Papers in Education*, 25(3), 285-292. <http://doi.org/10.1080/02671522.2010.498144>
- Parke, C. S., Lane, S., & Stone, C. A. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation*, 12(3), 239-269. <http://doi.org/10.1080/13803610600696957>
- Queensland Studies Authority. (2014). *School-based assessment: The Queensland assessment*. Queensland, Australia: The State of Queensland (Queensland Studies Authority).

Voices from the Field: Performance Assessments in State Accountability as discussed at the National Conference on Student Assessment

by Susan Lyons and Yuxi Qiu

<http://vue.annenberginstitute.org/issues/47/voices-field-performance-assessments-state-accountability-discussed-national-conference>

This field report from 2017's National Conference on Student Assessment shares possibilities for flexibility and innovation in assessment and accountability made possible by the Every Student Succeeds Act.



Susan Lyons

Yuxi Qiu

I

In June 2017, the National Conference on Student Assessment offered opportunities to delve into the possibilities for flexibility and innovation in assessment and accountability presented under the federal Every Student Succeeds Act (ESSA). The conference, sponsored by the Council of Chief State School Officers, provided deep dives into a range of topics, including discussions related to the use of performance assessments for school accountability.

Stuart Kahl of Measured Progress outlined a rich history of state-led initiatives to adopt performance assessments in the 1980s and '90s, such as the Connecticut Assessment of Educational Progress – Science.¹ In response, Susan Tave Zelman of the Ohio Department of Education offered that despite great interest in successfully maintaining a performance-based system of assessments and a long history of attempting to do so, “states are still struggling with how to integrate performance tasks in a system with standardized tests, . . . and no state has yet been successful in creating a state-wide system integrated into state accountability.” Zelman identified three sources of obstacles to implementing performance assessments for this purpose:

1. **Design challenges**, such as a lack of consensus on the role of states, alignment of performance assessments to states’ standards, integration into state accountability systems, and opportunities to learn;

2. **Resource challenges**, such as limits on money, time, state capacity, vendors, technology platforms, and political will; and
3. **Measurement challenges**, such as validity, reliability, timeliness of results, comparability, and scalability.

John Weiss of the Pennsylvania Department of Education and Joyce Zurkowski from the Colorado Department of Education echoed the challenges related to political will.² Zurkowski recommended introducing performance assessments into a balanced assessment system, but avoiding immediate association of results with accountability. Following this model, Colorado has had early success experimenting with performance assessments locally to improve instruction and student learning before making the transition to accountability.

Though implementing performance assessments at the classroom level may be a pathway for states to build towards innovative systems of accountability, this type of work at the district level is not without its own challenges. Giselle Martin-Kniep of Learner-Centered Initiatives works with teachers to design and implement performance assessments.³ Martin-Kniep explained that designing authentic assessment tasks is a challenge for teachers in every subject and grade. Teachers also need guidance on how to provide formative assessment opportunities prior to administering summative tasks. However, early evidence suggests that once successfully implemented, school districts value the use of performance-based assessments to inform instruction. Jacob Mishook of Achieve, Inc. commented that his work with districts to inventory the suites of assessments administered to students throughout the year revealed that “tests that are the most valued [by districts] measure real learning [and] assess problem solving and critical thinking.”

Alternate assessments for students with significant cognitive disabilities may offer a more readily-accessible entry point for the practice of using performance tasks for accountability. Jan Sheinker of Sheinker Educational Services warned states that “portfolio assessments are not going to fly anymore.... States need to start thinking about complex performance tasks that can be developed with integrated accommodations.”⁴ Sheinker suggested that performance assessments can be designed to measure student proficiency with scaffolding that could be added or removed based on student individual needs. While Sheinker touted consistent scoring and comparability as benefits of performance assessments over portfolio-based alternate assessment systems, performance assessment design for accountability still presents significant technical challenges, as Zelman previously noted.

The tenor of this conference suggested great interest shared among education stakeholders in using performance assessments as part of state accountability systems. That being said, overcoming the real technical and political barriers will require additional research and evaluation – such as those results presented by Art Thacker of HumRRO at the session titled, “Formative Evaluation of New Hampshire’s Performance Assessment of Competency Education (PACE)”⁵ – and a widespread effort to improve assessment literacy, not only among teachers, but also for policy makers, parents, and even students.

FOOTNOTES

1 Pecheone, R., Wei, R. C., Kahl, S. & Zelman, S. T. (2017). Bringing performance-based assessment into the summative space. Symposium presented at CCSSO's National Conference on Student Assessment, Austin, TX.

2 Wylie, C., Zurkowski, J., & Weiss, J. (2017). Performance assessment for accountability: Measuring what matters of distracting from the classroom. Symposium presented at CCSSO's National Conference on Student Assessment, Austin, TX.

3 Martin-Kniep, G., Mishook, J., & Hildreth, B. (2017). Why less is more: Working with districts and states to reduce redundant and unnecessary assessments. Symposium presented at CCSSO's National Conference on Student Assessment, Austin, TX.

4 Sheinker, J., Michaels, H., & Pence, N. (2017). Alternate assessment alignments: What best practice and lessons learned have taught us. Symposium presented at CCSSO's National Conference on Student Assessment, Austin, TX.

5 Leather, P., Wilhoit, G., Tacker, A., & Lyons, S. (2017). Formative evaluation of New Hampshire's Performance Assessment of Competency Education (PACE). Symposium presented at CCSSO's National Conference on Student Assessment, Austin, TX. For more information on PACE visit <https://www.education.nh.gov/assessment-systems/pace.htm>

Exploring the Role of Performance Assessment in Competency-Based Education: Recommendations from the National Summit on K-12 Competency-based Education

by [Laurie Gagnon](#)

Based on reports created for the National Summit on K-12 Competency-based Education, this article explores how educators can take action to build momentum for and implement competency education and performance assessment.



Laurie Gagnon

In June 2017, CompetencyWorks and iNACOL gathered “100 leading innovators to move the field of competency-based education through the next generation of ideas and actionable outcomes” at the [National Summit on K-12 Competency-based Education](#). Such a summit had not been convened since 2011, when the Competency-Based Learning Summit resulted in the current [working definition of competency-based education](#).

The Summit sought to tackle the major issues that have emerged as the field has developed. Prior to the Summit, organizers wrote four framing reports to prepare participants to generate shared solutions. The themes of equity and diversity were interwoven both in the content and the process. Technical advisory groups (TAGs) were used to foster broader participation in the crafting of the four reports and particular attention was paid to the diversity of participants at the Summit.

USING TEXT-BASED DISCUSSION TO RELATE SUMMIT TAKEAWAYS TO YOUR OWN WORK

This article explores how we can apply the resources and takeaways of the Summit to our own work to build momentum for and fully implement competency education and its essential corollary, performance assessment. As a guide to taking action, in the next section we have provided:

- [links to the four framing reports from the Summit](#)

- six key issues in performance assessment that emerged from the discussions of the framing reports at the conference, listed under the relevant report
- suggestions for relating these concepts to your own work through close reading of parts of the text and sample discussion questions, in a methodology often referred to as a [text-based protocol](#)

Text-based discussions can be used as an entry point to analyze our own practices. Consider the goals and needs of the group as you select a protocol and text excerpts. Here are a couple of starting points:

- The 4-As protocol: engage with a text through identifying assumptions, what you agree with, what you might argue with, and what you aspire to in your own work?
- Basic text-based discussion with a framing question: The framing question plays a key role in focusing the discussion on the connections between the ideas in the text and the reality of your work in practice.

FRAMING REPORTS, KEY TAKEAWAYS, AND DISCUSSION QUESTIONS

Framing Report 1:

[In Pursuit of Equality: A Framework for Equity Strategies in Personalized, Competency-Based Education](#)

By Chris Sturgis with contributions from Ashley Jones

1. Performance assessment can help transform systems that enhance equity and access. Read pages 2-4 and one of the VUE 46 Case Studies about the [New York Performance Standards Consortium](#), [International High School at Langley Park](#), or [Chelsea High School](#). How do the assessment systems in the case studies generate a different kind of data than a standardized accountability test? Where might the school in the case study still want to explore further transformation?

2. Performance assessment allows students to take responsibility for their own learning by emphasizing feedback and practice and creating multiple opportunities to engage in assessment for learning. Explore the chart unpacking the working definition of competency-based education on pages 8-9. Where do you see connections between performance assessment and the competency education definition?

Framing Report 2:

[Meeting Students Where They Are](#)

By Antonia Rudenstine, Sydney Schaefer, and Dixie Bacallao

3. Instructional practices that are student-centered and meet students where they are drive the student experience in a competency-based system. Which performance assessment designs exhibit the features and strategies discussed in the report? As you examine specific performance assessment examples, consider how you might incorporate

the practices discussed – from fostering engagement, access, and rigor to offering timely, differentiating supports. Opening up our practice for analysis ensures that the performance assessment leads to instructional shifts at the heart of competency-based education.

Framing Report 3:

[Fit for Purpose: Taking the Long View on Systems Change and Policy to Support Competency Education](#)

By Susan Patrick, Maria Worthen, Natalie Truong, and Dale Frost

4. How we define success shapes how and what we teach, learn, and assess. Pages 3-4 discuss how a [Profile of the Graduate](#) can define a more holistic vision of what we care about from our education system. Beyond academics, what helped many of us be successful also included transferable skills, habits and dispositions for perseverance, empathy, and collaboration, and social-emotional intelligence. What Vision of the Graduate do your promotion and graduation requirements create? Start or revisit the process of building a community vision using the [Quality Performance Assessment Vision of the Graduate protocol \(free resource; registration required\)](#).

5. Assessment literacy and teacher capacity building are essential elements in the process of implementing personalized, competency-based practices and systems. Performance assessment, in the hands of students and educators, is specifically designed to foster their learning. For example, in VUE 46, Dan French and Barnett Berry look at [how micro-credentials build educators' performance assessment literacy](#) through a competency-based, personalized learning process.

Framing Report 4:

[In Search of Efficacy: Defining the Elements of Quality in a Competency-Based Education System](#)

By Chris Sturgis with contributions from Natalie Abel

6. Full implementation is key and we need to focus on pathways for implementing with fidelity and sustaining (re)designed systems. Select a compelling section of the framework to explore, such as structural domain 3 on pages 12-13, Continuum of Learning Objectives, Student Performance, Growth, and Progress are Transparent. Ask yourself the key questions from that section and discuss what to look for and exemplars from your own system or a model system you are examining. As you explore what success could look like in this domain, ask yourself how you are addressing the lessons.

The Summit asked us to chart a course to an education system that meets the needs of each child in order to find a pathway to a successful and productive life. The variety of contexts and experiences represented at the Summit underscore that high-quality, equitable, competency-based education need not look exactly the same, but rather that within a shared framework, there are multiple pathways and approaches to a transformed system. Performance assessment offers a high-leverage entry point to one of the pathways for schools, districts, and states to design assessment, teaching, and learning practices that are aligned with and supportive of each other.

A New Age of Implementation: Guiding Principles for Implementing Performance Assessment Systems

by [Gary Chapin, Laurie Gagnon and Virgel Hammonds](#)

In an examination of the conditions required for the successful implementation of performance assessment, the authors draw on a range of personal experience and other insights to guide practitioners and policymakers.



Gary Chapin



Laurie Gagnon

During the 2000s, some practitioners and education researchers aimed to produce pedagogy that embodied the ideals of equity of opportunity and outcome, including initiatives such as personalized learning, competency-based learning, flexible pathways, attention to habits and dispositions, and authenticity in learning and assessment. Taken together, these initiatives provide ways for students to learn and demonstrate their learning, directly address the transferrable skills that are the foundation for all learning regardless of content, shape learning so that it is genuinely relevant to each student, and elevate student agency as a value. At the center of all of these practices is *performance assessment*, which can be defined as “multi-step assignments with clear criteria, expectations, and processes that measure how well a student transfers knowledge or applies complex skills to create or refine an original product” (CCE 2012).¹

Performance assessments come in many forms – artistic performances, labs, exhibitions of research, internships, and portfolios. Students show us not only that they know something, but also that they know how to use that knowledge (or skill). Brian Stecher (2010), writing for the Stanford Center for Opportunity Policy in Education, expresses a definition that is broader and somewhat more elegant. He writes that performance assessment is “judging student achievement on the basis of relatively unconstrained responses to relatively rich stimulus materials.” A lot is unsaid in Stecher’s definition, but he captures the ethical imperative, pointing not only to the technical aspects of performance assessment, but also to the values and cultural changes it implies.

Building on the authentic assessment work of the Boston Pilot Schools (CCE 2004), in 2008 the Center for Collaborative Education (CCE) began developing a system to design and implement performance assessment that was research-based and educator-driven and that achieved a high

level of technical quality. We conducted research and worked with a cohort of educators to bridge research and practice. The initiative culminated in the assessment model Quality Performance Assessment (QPA).²

WHAT DO WE KNOW ABOUT SUCCESSFUL IMPLEMENTATION?

Gregg Palmer, principal of Falmouth High School in Maine and an early champion of standards-based reform in that state, once said, “The most dangerous time for any innovation is when you try to scale it.”³ For the purposes of this conversation, “scaling” and “implementing” are synonymous, defined as enacting an innovation in a new space. Our experience lends credence to Palmer’s view that, “When an innovation fails, it is most often the implementation – rather than the innovation – that goes awry.”

In the early 2000s, a number of states attempted to foster Comprehensive Local Assessment Systems using standards-based criteria and reporting. In Maine, the effort (in which two of the authors, Gary and Virgel, participated) was indeed comprehensive, but not sustainable. The strict demands of validity and reliability took far too much time, and the recording requirements created so much documentation that it was common in Maine to quip that the effort “died under the weight of its own paper.” Other efforts, as reported by Tung and Stazesky (2010), have ended because funding ended, because of leadership changes, or because of political pressure at the local and state level.

Since then, we have learned more about implementing performance assessment in schools. Innovation funders, partners, state agencies, and curriculum leaders have been paying attention to which methods of implementation seem to produce the best outcomes. We are not as far developed as our colleagues in the health field, who have seen the birth of a subsidiary field of study, *Implementation Science*, which looks at the uptake of research findings into routine healthcare practice.⁴ Still, qualitative research,⁵ along with the experience of the authors, suggests that four key considerations, if taken into account, will increase the probability of a successful implementation effort:

- Shared moral vision and leadership driving policy and practice
- Abundant, informative, and compassionate communication to build understanding and harness public will
- An insistence on technical quality, fostered by sustained professional development
- A collaborative culture within which to build capacity

MORAL VISION AND LEADERSHIP

Any change process involves necessary loss and disruption, and a huge amount of effort and learning by educators and stakeholders. Effective implementation requires that everyone in the system comprehend its necessity on rational, emotional, and moral levels. Student achievement, graduation, and college placement rates, while necessary to cite, are never enough on their own to drive change. We must also surface the moral and ethical commitments to equity that will move our schools toward implementing performance assessment. These commitments must become the foundation for supporting policy and decisions around practice. As leaders, we will

know that we've succeeded not when people tell us they want to transform schools, but when they tell us they would be appalled if the change process went off the rails.

In 2006, when Gary was a curriculum teacher-leader at Hall-Dale High School in Farmingdale, Maine, the superintendent and principal moved to implement a standards-based reporting system. Their first step was not to change policy, but to begin working with teachers. First, the most enthusiastic were given access to professional development, and then word began spreading. More went for trainings. Book groups formed. Teachers, leaders, students, and parents traveled to districts engaged in similar work. At the forefront of all discussions was the question: Why is this necessary? When the time came to throw the switch and change the grading policy for the school, the school committee met with a group of 60 teachers. During the two-hour meeting, one teacher summed it up for all when he said, "Let us do the right thing."

COMMUNICATION

When Virgel arrived as superintendent of Regional School Unit 2 (RSU 2), which comprises five towns and to which Hall-Dale High School belongs, he began his tenure with a conversation tour of the large and far-flung district. As he recalled in an interview, "I needed to get to know the communities, the schools. What has made us successful? What hasn't? What do we need to target? I made a commitment to go to every community. Homes. Patios. Barns. Town fairs" (Center for Best Practice, 2012). The frequent gatherings tended to be small, and listening happened on both sides. For Virgel, change happened one conversation at a time.

The district had learned early on that gathering parents into an auditorium and speaking at them from the stage was decidedly not the way to invite cultural change to a school district. Smaller meetings, in which parents could talk with students and teachers – rather than administrators – and, not incidentally, eat lasagna, turned out to be much more successful. Importantly, the leaders and teachers at RSU 2 realized that people aren't afraid of change. Rather, as Michael Fullan (2006) has pointed out, they are afraid of loss.⁶

Faced with the shift to a performance assessment system, educators may feel loss of a sense of competence as the teachers move from masters of content to facilitators of learning. Or they may fear losing the precision (and the illusion of accuracy) that traditional grades can convey, and the resulting uncertainty. High-performing students may fear loss of hierarchy based on those grades, and the GPA-based honors that attach to them. Parents of high-performing students may similarly fear the possible loss of advantage that their children accrue in the current system (Kohn 1998), especially when it comes to college admissions and scholarships. Some may feel fear because even if they agree that the current system is flawed, they aren't sure of the advantages of a new system, and don't want to "experiment" on their children. And administrators and innovators may fear the reactions of their communities and possible resulting pushback.

None of this fear is baseless, though it is insufficient cause to withdraw from change. In the many conversations leading up to a transformation, everything is to be gained by avoiding treating anyone as if they are the enemy. Every stakeholder should be treated with compassion. In RSU 2, curriculum leaders met with individual parents for dozens of hours, discussing the foundations and subtleties of the system, and trying to assuage concerns. When RSU 2 formed an implementation committee, they not only invited parents on board, they invited the two most vocal skeptics of the new system, representatives of a much larger group. The committee met

every two weeks for the school year, and set the conditions for the initial launch the following September.

TECHNICAL QUALITY

Our advocacy for QPA is inspired by its potential to support equity. QPA also has at its core a commitment to technical quality – that is, an assurance that the assessments deployed are valid and reliable. A performance assessment system must have an effect on student learning that is considerable, measurable, and demonstrable. The mechanism for ensuring this level of technical quality is a sustained professional development effort such as the Performance Assessment for Competency Education (PACE)⁷ pilot, which New Hampshire launched in 2014.

In PACE, both local and cross-district collaborative processes are the foundation of technical quality evidence and are guided by the National Center for the Improvement of Educational Assessment (Center for Assessment). Key processes to ensure the technical quality of the assessments, scoring, and the overall annual determinations that result from the local assessment systems include:

- Content Area Leads, selected from among PACE district educators, lead their peers in developing the PACE Common Performance Tasks. The Content Area Leads work closely with Center for Assessment staff, who also conduct technical reviews of the tasks.
- Local calibration and double-scoring of the PACE common performance tasks occurs during the school year. Educators further analyze results and student work samples during summer cross-district calibration sessions.
- Annual determinations are made based on achievement level descriptors (ADLs) written by teachers and then applied in a teacher survey matching each student's body of work to the appropriate ALD, and through the analysis of samples from a body of student work from the local assessment system.

PACE teacher involvement is a cornerstone of technical quality, along with the support and additional psychometric analysis of the resulting data by the Center for Assessment, redefining traditional psychometrics for a locally-driven reciprocal accountability system (NHDOE 2016).

COLLABORATIVE CULTURE

While Virgel was superintendent of RSU 2, the district joined with nearly two dozen others to form the Maine Cohort for Customized Learning (MCCL). Their goal was to bring together energy, expertise, and resources. Pooling together funds in order to share professional development costs proved to be vital. The MCCL facilitated a years-long professional development agreement with the Re-inventing Schools Coalition to provide professional development to their collective faculties. Similarly, when looking for standards-based reporting software, they reached out to the designers of the Empower software package, and worked with them to customize the software the MCCL needs. Finally, in designing proficiencies and standards, the cohort created content area committees, drawing expertise from all member districts.

As a central part of its work supporting performance assessment, the Center for Collaborative Education has fostered collaborative networks in Rhode Island, Vermont, Oregon, Massachusetts, and New Hampshire (through the aforementioned PACE). CCE itself works as part of a national collaborative cohort of 12 organizations called the Assessment for Learning Project (ALP). Each ALP organization is engaged in a learning pilot around some aspect of Assessment for Learning. One, for example, is looking into the power and quality of feedback (or feedforward) for kids. Another is examining a place-based, culturally responsive approach to habits and dispositions. CCE has developed a system of micro-credentials around performance assessment, and is piloting them as part of a system of professional development with districts in Rhode Island, Kentucky, and Georgia. ALP's participating organizations gather online, and occasionally in person, to provide feedback, insights, encouragement, and support. It is a genuine learning community.

Districts must also build within themselves a culture that supports educator, student, and community collaboration. In many districts this has taken the form of some sort of formal model, for example Professional Learning Communities (PLCs) or Critical Friends Groups (CFGs). Through practice in these models, educators engage in intentional, structured conversations that center on student work, data, and practice. The protocols of these models allow for a safe place in which that level of vulnerability is possible. Over time, PLC practice stops being novel, stops being something teachers do, and becomes the culture of the school.

A NEW AGE OF IMPLEMENTATION

In implementing QPA across New England and beyond, we have learned how to support schools, districts, and states as they create local systems of performance assessment. We have seen that curriculum-embedded performance assessment operates as a leverage point for many of the practices that comprise competency-based and personalized learning. The vision of performance assessment, with its relatively unrestrained responses and rich materials, is one of broad possibility and permission for students to take agency over their learning. The challenge is embodying this vision in all of our schools, in the widest variety of contexts, and in ways that ensure equity of opportunity and outcome for every student.

All students need, deserve, and have a right to our care, attention, and best efforts. Our nation requires a citizenry with the capacity to thrive in (and hold on to) the democratic republic of the 21st century. The broader range of possibilities allowed by performance assessment, and the associated careful use of data required, mean that our biases and assumptions will be less likely to close our eyes to the ways students can be successful. The commitment to authenticity and the call to allow students to become co-conspirators in the construction of their learning mean that they will be better prepared to move about in the world and shape it for their own futures.

FOOTNOTES

1 See also Brown & Mevs 2012.

2 See CCE 2012. For more information on QPA, see <http://cce.org/work/instruction-assessment/quality-performance-assessment>.

3 Personal conversation with Gary Chapin, 2007.

4 For more on implementation science, see <https://www.fic.nih.gov/researchtopics/pages/implementationsscience.aspx>

5 See Tung & Stazesky (2010) and Center for Best Practice (2013)

6 In a keynote at a 2007 conference in Maine, discussing the book, *Breakthrough* (co-written by Fullan and Crévola, see References)

7 For more information on PACE visit <https://www.education.nh.gov/assessment-systems/pace.htm>

REFERENCES

Brown, C., and P. Mevs. 2012. *Quality Performance Assessment: Harnessing the Power of Teacher and Student Learning*. Boston, MA: Center for Collaborative Education.

Center for Best Practice, 2012. [The long conversation, or ‘It’s hard, but worth it. Did I mention that it’s hard?’: RSU 2 student-centered learning implementation case study](#). Maine Department of Education.

Center for Best Practice, 2013. [Threads of implementation: A thematic review of six case studies of Maine school districts implementing proficiency-based systems](#). Maine Department of Education.

Center for Collaborative Education. 2004. [How pilot schools authentically assess student mastery](#). Boston, MA: Center for Collaborative Education.

Center for Collaborative Education. 2012. *Quality Performance Assessment: A Guide for Schools and Districts*. Boston, MA: Center for Collaborative Education.

Fullan, M., Hill, P., & Crévola, C. 2006. *Breakthrough*. Thousand Oaks, CA: Corwin.

New Hampshire Department of Education. 2016. [Moving from good to great in New Hampshire: Performance assessment of competency education \(PACE\)](#).

Kohn, A. 1998. [Only for my kid: How privileged parents undermine school reform](#). *Phi Delta Kappan* (April).

Rush, B. 1987. [Of the mode of education proper in a republic](#).” *The Founders’ Constitution*. Chicago, IL: University of Chicago Press.

Stecher, B. 2010. *Performance Assessment in an Era of Standards-Based Educational Accountability*. Stanford Center for Opportunity Policy in Education. Stanford, CA: Stanford University.

Tung, R. and Stazesky, P. (2010). [Including performance assessments in accountability systems: A review of scale-up efforts](#). Boston, MA: Center for Collaborative Education.