Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

FairTest
National Center for Fair & Open Testing

Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

By Monty Neill
Executive Director, FairTest

A Report by the National Center for Fair & Open Testing

October 2016

FairTest
P.O. Box 300204
Boston, MA 02130
www.fairtest.org
fairtest@fairtest.org
617-477-9792

Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

Table of Contents

Note: The full report that this section is excerpted from is available online at:
http://www.fairtest.org/assessment-matters-constructing-model-state-system

Assessment Matters:
Constructing Model State Systems to Replace Testing Overkill

Part II

New Hampshire PACE

New Hampshire received a waiver in 2015 from No Child Left Behind (NCLB) to begin constructing a new statewide system, the Performance Assessment for Competency Education (PACE). Implementation started with four participating districts in the 2014-15, school year. It grew to eight in 2015-16 and includes nine in 2016-17, with 10 more preparing to join. The waiver has been extended through 2016-17. The state expects to be part of the Every Student Succeeds Act (ESSA) "Innovative Assessment" pilot.

New Hampshire describes its new system in these terms:

> "One of the motivating reasons why NH is piloting a new kind of accountability system is because the state wants to support meaningful learning and continuous improvement models, as well as promote shared accountability between districts and the state" (Evans, Lyons & Marion, 2016).

> "The PACE system is based on a rich system of local and common (across multiple districts) performance-based assessments that are necessary for supporting deeper learning as well as allowing students to demonstrate their competency through multiple performance assessment measures in a variety of contexts" (Marion and Leather, 2015).

For federal approval under ESSA, PACE must demonstrate that the new assessments are valid and reliable, and are comparable among themselves and with the state's current tests. PACE can be understood as a structure to unite two purposes: deeper learning and comparability for accountability. The first begins with locally designed and controlled performance tasks scored by teachers. These are intended to improve teaching and learning and provide key evidence for an "achievement level determination" (ALD). The determinations, placing every student's level of proficiency on a four-point scale, are made by classroom teachers using information gathered across the school year. New Hampshire has designed a structure to ensure these "competency determinations" are comparable across the districts and accord with state test results. The ALDs are also the basis for determining each school's level, an ESSA requirement. Thus, the ALDs are the core of the PACE system.

Description of PACE

PACE includes three kinds of assessments:

- State tests, which are the Smarter Balanced Assessment Consortium (SBAC) exams in English Language Arts (ELA) and math, given in one grade each in elementary and middle school, and the SAT college admissions test in grade 11.
- Teacher-made, PACE-wide common performance tasks in ELA and math in grades 3-8 that do not have state tests, as well as science tasks at grades 4, 8 and in high school; and
- District teacher-made tasks in all three subjects in grades 3-11 that do not use the state tests, with many districts also developing tasks for earlier grades and state-tested grades.

Common Tasks

Teachers from participating districts design 17 Common Tasks (CT) with task-specific scoring guides, one per required subject/grade. "There are three main purposes for the common tasks across districts: 1) to help measure the degree of cross-district comparability of scoring, 2) to serve as models of high quality tasks and build local capacity, and 3) to contribute to the long-term goal of building a large task bank from which districts can draw for local assessment purposes" (NH DoE, 2016a, p. 4).

As models, "The tasks are designed and reviewed specifically to allow for independent student inquiry, multi-step problem solving and argument building, and typically allow for multiple possible solutions" (NH DoE, 2016a, p. 5).

Tasks are designed to be accessible to students with disabilities and English language learners. The state uses Universal Design for Learning in writing the Common Tasks, to maximize accessibility, and also allows accommodations in line with those provided for SBAC. It has separate assessments for students with the most severe disabilities. It administers the WIDA test to English language learners, who do not take the CTs, as allowed under NCLB/ESSA.

Once CTs have been administered, the state posts them to a website for use by districts as local tasks and by teachers in their classrooms. The assessment bank is complemented with tasks from other teachers and states that the state has reviewed and approved (Marion-Leather, 2015, p 13).[i]

Local Tasks and the "Achievement Level Determination"

While Common Tasks are important, the heart of the new system is the use of locally made tasks included in local systems combined with each teacher's "achievement level determination" (ALD) for individual students.

In PACE, each participating district designs a system of performance-based assessments tied to the state's subject area "competencies" and standards (NH DoE, NDb). Districts submit their systems to the New Hampshire Department of Education (NH DoE) for peer review and approval (NHDoE, 2016a). Local assessments sometimes include gathering information on students' "work-study practices," but these are not part of any statewide data or subject competency determinations.

The number of local tasks varies by district, grade and competency. These tasks are "curriculum-embedded and administered in local districts." Deputy Commissioner Paul Leather (2016) explained this means teachers base the tasks on their curricula and administer them at an appropriate time.

Among directives to districts regarding the local assessments are (NH DoE, 2016a, p. 10):

"3. Students must be allowed multiple opportunities to demonstrate evidence of achieving a competency over the course of a year.

"4. Districts must use a mixture of locally-designed performance assessments and assessments drawn from validated state/multi-state task banks."

"Leaders and teachers in each district determine how to score their local competency-based assessments" (NHDoE, 2016b). Each district task creates its own rubric, though in some cases a general rubric can be used across tasks (e.g., different writing samples). Local scoring, the state recommends, begins with selecting 10-20 previously completed tasks from across the range of student achievement (NHDoE, 2016, p 36 ff). Teachers sort these into four performance levels. They then use the anchors with the local task rubrics to score new student work. The results of the local tasks are included by the teachers as they render their ALDs.

Photo from Rollinsford Grade School.

Teachers decide each of their students' proficiency levels – the ALD – on a 1-4 scale in the subjects and grades mandated for NCLB/ESSA accountability. These determinations take the form of a summary judgment by each PACE teacher about each student. It includes all the academic information gathered from the student's work over the year. The

judgments are based on SBAC's "Achievement Level Descriptors," modified to fit PACE. These Descriptors also serve as a scoring guide that enables the analysis of comparability (NHDoE, 2016, p 10 and pp. 41-46).

Establishing Comparability

PACE has designed a complex system for establishing comparability. It begins with the requirements districts face to join PACE. It includes tools for re-scoring ("moderating") local and common assessments, and protocols to determine the degree of comparability across SBAC or SAT, PACE common tasks, district tasks, and the ALD. Disaggregated group scores are also examined. New Hampshire points out that comparability does not require psychometric exactness, such as the comparison of student scores on a single standardized test. PACE's methods also serve to demonstrate validity and reliability (required by ESSA), as well as assist teacher professional development.

First, the design, validation and approval of district systems. Staffs in participating districts go through extensive training and two years of peer review. Local systems must meet a set of criteria for quality, have well-prepared core staff, demonstrate validity, reliability and comparability, and adequately address such issues as bias/fairness.

Second, determining consistency in scoring common tasks across districts. The Common Tasks serve as reference points for comparing results across districts. Students take one CT per subject in each grade that has no state test. New Hampshire officials did not want a larger set of such tasks as they do not want a new form of state exam.

Each PACE CT has a scoring guide teachers use to rate student work. To help teachers do this well, the state sets up sessions for teachers to collectively re-score a sample of completed tasks and discuss the results. Teachers bring this knowledge back to their districts to apply to local tasks.

NH Department of Education staff worked with measurement experts from the National Center for the Improvement of Educational Assessment (NCIEA) to design a procedure to ensure accurate scoring. The process builds on models from Australia and Britain (Evans, Lyons & Marion, 2016). In the first year of the pilot, local teachers' common task scoring was adequately comparable across the four districts (NHDoE, 2016, p 18). Time and experience should strengthen consistency, though the influx of new districts and teachers will make this an ongoing project.

Third, and most central, ensuring accuracy and consistency of the local competency determinations. As the state says, "Comparability in scoring performance assessment tasks is important but the ultimate goal is that 'annual determinations' are comparable across school districts" (NH DoE, 2016b). The state devised procedures for linking district teacher-determined ALDs to common task scores and SBAC results.[ii] Teams of teachers compared the 2015 local ALDs with samples of completed Common Tasks from each district in each subject across the

four achievement levels. Teachers did not score work from their own districts. They found no systematic variation. That means overall scoring was adequately consistent across the grades and districts (Evans, Lyons & Marion, 2016).

PACE also compared the proportion of students across the state scoring at levels 3 and 4 on SBAC in the tested grades with ALDs from pilot districts in the other grades. The study found the scores across districts "were quite similar, indicating a high degree of comparability between PACE and non-PACE districts" (NH DoE, 2016b, p 11; NH DoE, 2016a, p. 11). In addition, "the differences in performance among major subgroups and the all students group were similar for both PACE and Smarter Balanced annual determinations" (NH DoE, 2016a, p. 22).

Consequences

The NH DoE says, "Discrepancies between local and state/consortium assessment results do not mean that the local results are wrong. Rather, it should lead to conversations and inquiries to try to understand the reason for any large differences between the two sets of results" (NH DoE, 2014). In any event, the analysis of the results from the first year concluded that no district systematically scored more stringently or more leniently, indicating there would have been no need to modify any district's scores.

New Hampshire began by including districts the state education department viewed as well-prepared for the new program. It recognizes that capacity will need to be built to expand PACE to districts that are less ready for performance assessment. Under ESSA, NH will have five years to expand PACE statewide.

Benefits and Concerns

There are significant benefits from the work being done by PACE:

- First, the performance assessments offer students a range of ways to show their knowledge and skills, many of which are not adequately covered by SBAC or SAT. Since PACE local assessments are tied to the curriculum, students are being taught content they may not have covered in the past due to pressure to raise standardized test scores. That includes higher-order thinking, applications of knowledge, problem solving, communicating, and connecting learning across subject areas.
- Second, designing, administering and scoring the assessments provides a vehicle for professional learning. Teachers deepen their knowledge about assessment, curriculum and instruction, and strengthen their ability to work and learn together.
- Third, PACE is creating a valuable model for the nation. Several states were developing performance assessment systems in the 1990s, but NCLB halted most of that work. While a knowledge base exists, for the most part states and educators have to start

nearly from scratch. They also have to address accountability and comparability issues that were generally not concerns in the 1990s.

- Fourth, the moderation and comparability systems developed by PACE and its primary technical partner, the NCIEA, can be useful to other states.

PACE is still very much a learning process. For example, beginning teachers typically have limited capacity to craft good tasks. However, evidence shows educators quickly learn how to make tasks better by trying them out in classrooms, sharing with other teachers, reflecting and discussing.

Photo from Rollinsford Grade School.

New Hampshire Principal Jonathan Vander Els (2015) observed, "Each performance assessment that I see being constructed is of higher and higher quality. This is due not only to our teachers' overall increased understanding of assessment in general, but also to their increased understanding of the nuances within each assessment. Considerations such as specific wording of a question, students' background experiences, ability to provide appropriate accommodations, and the level of the depth of knowledge are intuitively included."

Despite progress, there are concerns. Some observers fear a system like PACE will end up allowing low-performers to skate by, allowing some students to slip through the cracks. These critics want to retain a system rooted in standardized tests, despite evidence of damage to educational quality and student learning. While most have acceded to giving "innovative assessments" a try, they may fight for constraints that end up undermining assessment quality and thus its benefits for teaching and learning.

Other concerns emerge from those who have been engaged in performance assessments. These include:

- Some question the quality of tasks, though they recognize they will improve. Tasks may be artificial, not exemplifying real-world problems. Other issues may reflect the history, in New Hampshire and elsewhere, of using narrow rubrics to judge limited forms of writing. This includes the "five-paragraph essay," which lacks real world applicability.[iii]
- The performance tasks are administered as tests. They do not evolve out of ongoing student work in the curriculum, as they do at Rollinsford NH Grade School, the New York Performance Standards Consortium, Big Picture Learning and elsewhere (see Part III). In

those instances, students have strong say over their tasks and projects – a key principle for high-quality assessment. A related concern is that tasks may not be particularly engaging for many students. The famous Coleman report (1966) found that next to parent's socio-economic status, the most significant predictor of academic success was a sense of control over learning, which can be strengthened when students choose their work.

- The common tasks must fit into a traditional curriculum. The PACE tasks can demonstrate knowledge, problem-solving skills and communications, but they are not part of an evolving inquiry. At Rollinsford, the NY Consortium, etc., the projects and tasks are themselves learning experiences that can be assessed during the process and at completion.

- Due to how they are constructed, says Rollinsford Principal Kate Lucas, "The current tasks offer a *prescribed way to assess students.* For example, the ELA grade 6 PA was to write a persuasive essay (rain forests). We question if an essay is the best way for all students to demonstrate their ability to synthesize and analyze information and then persuade others. Is it possible *to offer multiple modalities* for demonstration that require the same skills? This opens up the door to success for *all* students. It also requires deeper thinking and decision making" (Lucas, 2016, emphases in original). Of course, writing is a highly valued skill that schools do need to teach and assess – but the "essay" form is not the only possible mode for demonstrating content knowledge.

- The process demands a large commitment of teacher time from participating schools. Certainly it is a learning experience for many, but Rollinsford fears that the time commitment would detract from its own labor-intensive efforts to improve. More generally, the question of teacher time is a significant issue that designers of new systems will have to address.

Can There Be a "System of Systems?"

The broader issue raised by Rollinsford concerns the vision of education, for example, whether it should be inquiry-driven (project-based) or more traditional. It also concerns whether a state developing a new system can allow a Rollinsford (or a Big Picture, a NY Consortium, or a school using the Learning Record) to join as a partner despite its different approach to assessments. NH Deputy Commissioner Paul Leather thinks it could not under the current NCLB waiver.

This could change. ESSA allows differing local assessments (as in New Hampshire) provided there is a vehicle to establish comparability. PACE does this by linking local competency determinations to Common Tasks and the SBAC tests. The determinations are made by teachers based on the evidence gathered over the year. Rollinsford does exactly that. If Rollinsford joined PACE under ESSA requirements rather than the current NCLB waiver, it could continue to use its own assessment processes rather than design local tasks. They would be

something of an outlier among PACE partners but could be a powerful example to districts interested in moving toward inquiry-based schooling.

What Rollinsford, the NY Consortium and other examples lead to is a "system of systems" in which local assessment systems may vary but all must provide evidence of comparability. A mix of anchor tasks and moderation can do that.

PACE has opened the door toward creating a state system of performance assessing that is significantly decentralized, places teachers at the heart of the process, ensures significant professional development, and directs students toward deeper learning. As such, it is a strong model for the nation to consider. FairTest calls on states to take account of the even richer possibilities of allowing a range of local systems, including those that are inquiry driven and build on student work as it evolves out of the curriculum.

References

Carla M. Evans, Susan Lyons, & Scott Marion. 2016. "Comparability in Balanced Assessment Systems for State Accountability." Paper Presented at the National Council for Measurement in Education (NCME) Coordinated Session "Advances in Balanced Assessment Systems," April, Washington, DC

Coleman, J., *et al*. 1966. *Equality of educational opportunity*. Washington, DC: United States Department of Health, Education, and Welfare, Office of Education. U.S. Government Printing Office.

Leather, P. 2016. Personal email communications, July 26.

Lucas, K. 2016. Personal email communications, July 26.

Marion, S. 2015, Feb. Two sides of the same coin: Competency based education and Student Learning Objectives. Published by Competency Works. http://www.competencyworks.org/resources/two-sides-of-the-same-coin-competency-based-education-and-student-learning-objectives/

Marion, S., & Leather, P. 2015. Assessment and accountability to support meaningful learning. Education Policy Analysis Archives, 23(9). http://dx.doi.org/10.14507/epaa.v23.1984

NH DoE. 2014. New Hampshire Performance Assessment of Competency Education: An Accountability Pilot Proposal to The United States Department of Education, November 21. http://education.nh.gov/assessment-systems/documents/pilot-proposal.pdf

NH DoE. 2016a. New Hampshire Performance Assessment of Competency Education: Progress Report to the United States Department of Education, March 1.

NH DoE. 2016b. Moving from Good to Great in New Hampshire: Performance Assessment of Competency Education (PACE), revised, January. http://education.nh.gov/assessment-systems/documents/overview.pdf

NH DoE. NDa. New Hampshire Accountability Pilot Overview Performance Assessment of Competency Education (PACE). http://education.nh.gov/assessment-systems/documents/pilot-overview.pdf

NH DoE. NDb. State Model Competencies. http://education.nh.gov/innovations/hs_redesign/competencies.htm

Vander Els, J. 2015. "Setting the PACE: Teacher Assessment Practices in a Competency-Based Education System." Competency Works. http://www.competencyworks.org/insights-into-implementation/setting-the-pace-teacher-assessment-practices-in-a-competency-based-education-system

---

[i] There is a list of 2015 tasks at http://tinyurl.com/alltaskslist, but it contains only task titles. For an example, see Water Tower, http://www.ewa.org/blog-educated-reporter/building-better-student-assessments, which this report summarizes in part I. See also Algebra task at Appendix I of NH DoE Report to USDoE, March 2016. The state plans to make tasks available online in 2016-17.

[ii] See chart, NHDoE, 2016b, p 10, and related discussion; NHDoE, 2016a, p, 6 ff, also Appendix J, on details developing ALDs, and appendix F, on producing the ALDs.

[iii] For some brief "rules" about rubrics, see "Gail's Axioms" in Neill, *et al., N.D., Implementing Performance Assessment*, p. 30*.* FairTest, Cambridge, MA. For examples of high-quality rubrics, see New York Performance Standards Consortium, *Education for the 21st Century*, http://performanceassessment.org/articles/DataReport_NY_PSC.pdf.