

NH PACE FLEXIBILITY WAIVER INSERT

A COMPREHENSIVE COMPETENCY-BASED ASSESSMENT SYSTEM

NH has recently been approved to implement and evaluate the **Performance Assessment of Competency Education (PACE)** pilot with four school districts during 2014-2015 and up to another four districts in 2015-2016.

NHDOE is committed to creating a balanced and robust system of assessments (formative, interim and summative) focused on supporting personalized learning models to evaluate students' competencies over rigorous academic content, adaptive skills, and critical work-study practices. For the 2014-2015, NH will administer only the Smarter Balanced summative assessments because the Smarter Balanced interim system will not be deployed until the 2015-2016 school year. However, NH schools have been taking advantage of the learning opportunities available through Smarter Balanced Digital Library of formative assessment probes and related supports. New Hampshire is firmly committed to the Smarter Balanced Assessment Consortium assessments and expresses this commitment in its role as a governing state. Further, Deputy Commissioner Paul Leather and NHDOE consultant, Scott Marion from the National Center for the Improvement of Educational Assessment, are two lead members of the Proficiency-Based Learning Task Force that is studying how the Smarter Balanced summative and interim assessments can support competency-based approaches to education (CBE). The Task Force is making recommendations to the Smarter Balanced executive committee for potential enhancements to the designs of both the summative and interim assessments to better support CBE assessment, monitoring/auditing, and reporting.

The New Hampshire State Board of Education rule (306.37) requiring all high school courses to be aligned to course-level competencies is one step toward fostering new practices of assessment that promote "deeper levels of understanding important academic content and skills." The NHDOE is also partnering with the Center for Collaborative Education (CCE) and the National Center for the Improvement of Educational Assessment (NCIEA) to develop a statewide performance assessment system that will balance local control with statewide accountability and comparability.

Performance Assessment to Support Meaningful Learning

New Hampshire Department of Education (NHDOE) is designing a coherent accountability system to foster deep understanding of learners. Many current educational accountability systems have stated goals of promoting deeper learning for students to, among other goals, improve college and career readiness. The NH initiative is based on the premise that performance-based and related assessment approaches must be meaningfully incorporated into accountability systems if we are to do more than pay lip service to these policy goals. We rely on the following definition for performance assessment:

Performance assessments are generally multi-step activities ranging from quite unstructured to fairly structured. The key feature of such assessments is that students are asked to produce a product or carry out a performance (e.g., a musical performance) that is scored according to pre-specified criteria, typically contained in a scoring guide or rubric.¹ In fact, the rubric is a critical component in establishing the validity of the score inferences since it is the bridge between the student work and the resulting score, the basis for the inference. Occasionally, performance assessments target key processes

¹ *Products* are sometimes thought of as a separate category of assessment form, but we argue that products are really one possible outcome or piece of evidence derived from a performance assessment.

or skills, such as communicating to diverse audiences, engaging in critical thinking, and listening to multiple viewpoints, that students employ when wrestling with a problem or participating in an event such as a debate or a mock presentation to a simulated (or real) city council. Like “authentic assessments,” performance assessments suffer from definitional problems in that this one term can encompass many different types of assessments. For example, performance assessment can range from 15-20 minute tasks (i.e., quite short) to multi-day activities with many scoreable units (Marion & Buckley, in press).

This definition does not distinguish among traditional academic and more cross-cutting (e.g., critical thinking, problem solving) knowledge and skills, because the principles for assessment design and validation apply to the multiple assessment targets. Shepard (2000) and others have argued that high quality tasks and assessments provide teachers and students the opportunity to learn more about the content being assessed than they could from selected-response items. Additionally, good assessments, especially performance tasks in which students have to generate solutions and reveal and/or explain their thinking can provide opportunities for teachers to develop sophisticated understandings about the nature of student learning (see also NRC, 2014). Although such insights are not impossible to obtain with selected response items, they are more likely to emerge from examining student work associated with complex performance tasks.

Performance Assessment of Competency Education (PACE)

New Hampshire is committed to raising the bar for all students by defining college and career-readiness to encompass the knowledge, skills, and work-study practices that students need for post-secondary success including deeper learning skills such as critical thinking, problem-solving, persistence, communication, collaboration, academic mindset, and learning to learn. Therefore, the state is piloting an accountability system with significantly greater levels of local design and agency to facilitate transformational change in performance. As part of this shift in orientation, the state is supporting a competency-based approach to instruction, learning, and assessment contextualized within an internally-oriented approach to accountability to best support the goal of significant improvements in college and career readiness. The information learned through competency-based assessments would then be used to support accountability determinations and, hopefully, better inform school improvement (e.g., Hargreaves & Braun, 2013).

A competency-based system relies on a well-articulated set of learning targets that helps connect content standards and critical skills leading to domain proficiency. Such a system requires careful tracking of student progress and ensures that students have mastered key content and skills before moving to the next logical set of knowledge and skills along locally-defined learning trajectories. Current systems that rely on compensatory systems (e.g. averaging) for grading and related record-keeping may allow students to slip through the cracks in terms of possessing necessary knowledge for building deep understandings in the focal disciplines.

The PACE system is designed to foster deeper learning on the parts of students than is capable under current systems. This requires timely assessments linked closely with curriculum and instruction. 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. Thus, the accountability option was established to enable schools and districts to demonstrate student achievement and learning growth through means other than or in addition to standardized tests, with an emphasis on performance assessment.

In the PACE option, the New Hampshire Department of Education (NHDOE) has created a route for districts and schools to demonstrate quality not solely or primarily dependent upon state standardized tests. The creation of the PACE accountability option reflects NHDOE's belief that school accountability works best if the responsibility for design and implementation is shared by districts and the state, rather than top-down mandates. Known as "reciprocal accountability," districts and schools are responsible for determining and reporting on local accountability measures, while the state is responsible for support and oversight in helping districts establish strong accountability systems.

As noted above, NHDOE received approval from USED to limit state (or consortium) standardized testing to select grade levels (e.g., 4, 8, and 11). NHDOE is a strong supporter and governing member of the Smarter Balanced Assessment Consortium, but it argues that once per year assessments, as good as Smarter Balanced may turn out to be, are not enough to drive and support deeper learning. Further, NHDOE is concerned that having external, large-scale assessments at almost every grade will control the conversation and not allow the space for the competency-based reform to take hold. The current PACE model, described here, is not necessarily a fully realized competency-based accountability system. Rather, we are presenting a "transitional system" that incorporates expected requirements of federal/state accountability, but points the way to what a fully realized system would look like with a possible change in ESEA or other policy changes on the federal level.

PACE Implementation Plan

It is one thing to put forth a proposal for a richer approach to education, but it is another thing to create the conditions necessary for successful implementation. NHDOE is engaged in a multi-faceted implementation plan to ensure the success of the PACE option that includes requirements for participating districts, technical and professional learning support, including task development and scorer calibration activities, and wrestling with complex technical issues. We describe below key aspects of PACE implementation to provide an understanding of how NHDOE will eventually scale-up this pilot to all NH school districts.

Requirements for participating districts ("guardrails")

Districts participating in the 2014-2015 pilot have all adopted the state graduation competencies in ELA, mathematics, and science (the arts competencies will be approved this year and social studies competencies during the 2015-2016 school year) and developed a coherent and high quality set of K-12 course and grade competencies mapped to the state graduation competencies. These competencies were developed by teams of NH educators and approved by the NH State Board of Education. These districts must have demonstrated the leadership and educator capacity to participate effectively in the pilot. In addition to having a well-articulated set of competencies, these districts must have developed or be close to completing the development of a comprehensive assessment system tied to these competencies. Districts considered for the 2015-2016 pilot must have adopted graduation competencies and have a commitment during 2014-2015 to fully build out their course and grade competency systems in K-12 as well as their comprehensive assessment systems.

Participating districts must be willing to participate in a peer and expert review process where they submit their systems of performance-based assessments for evaluation based on clear and rigorous criteria including alignment with state standards and competencies, consistency and accuracy of

scoring, and fairness to all test takers. Further, PACE districts will be required to administer the state summative assessments (Smarter Balanced) in at least three grades, one at each level (e.g., 4, 8, and 11), which will serve as both an internal and external audit regarding school and district performance (see Table 1 below). Both local and Smarter Balanced assessments will be components in the PACE accountability system.

All pilot districts are expected to have the same general assessment requirements in the same courses and grades. As noted above, the Smarter Balanced summative assessment will be administered in select grades. The current plan involves staggering the Smarter Balanced subject areas according to when the results will be most useful for informing programs and auditing the local and common performance assessments. The current state science assessment (NECAP) will be phased out as these districts play a lead role in beginning to pilot “next generation” science assessment tasks. In fact, the National Research Council advocated, in a recent report, that moving to assessments of the Next Generation Science Standards must be led by classroom-based assessments rather than trying this complex endeavor with large-scale assessments first (NRC, 2014). The PACE districts will be particularly suited to pilot this new approach, given their intensive efforts in implementing complex performance assessments. Further, the NH graduation competencies in science are based on cross-cutting concepts drawn from the Next Generation Science Standards so this assessment approach is entirely consistent with the competencies and the NGSS.

Importantly, local performance assessment, used for competency determinations, will be administered in all subjects and grades. In certain grades and subjects, they will be “anchored” by Smarter Balanced assessment results, but in many others, they will be tied to performance assessments common to all participating districts (PACE Common Assessments). The competency determinations for all grades and subjects depicted above will include local (to each district perhaps) performance and other assessments designed to represent the full range and depth of the target competencies at each grade level. They were not depicted in Table 1 simply to avoid cluttering the chart. These common performance assessments (PACE) are intentionally limited to just one or two major tasks in most grade levels and content areas because NHDOE does not intend to simply replace one state assessment with another. Rather, these common performance assessments will be used to help calibrate performance expectations across participating districts and will be incorporated into local competency determinations.

Grade	Competency Grading	English Language Arts	Mathematics ²	Science
K-2	☑			
3	☑	Smarter Balanced	PACE	
4	☑	PACE	Smarter Balanced	PACE
5	☑	PACE	PACE	
6	☑	PACE	PACE	
7	☑	PACE	PACE	
8	☑	Smarter Balanced	Smarter Balanced	PACE
9	☑	PACE	PACE	PACE
10	☑	PACE	PACE	PACE
11	☑	Smarter Balanced	Smarter Balanced	PACE
12	☑	**CAPSTONE PERFORMANCE ASSESSMENT		

Table 1. Common summative performance-based assessments (PACE) and Smarter Balanced assessments administered by grade and content areas in all PACE districts.

The Task Bank

An ultimate goal of the PACE pilot is to enhance the capacity of educators to develop and use their own classroom assessments. However, creating a set of tasks for common administration and scoring purposes as well as helping to jumpstart local capacity is critical to the success of this project. The NH Task Bank is a repository of quality performance tasks that have been designed specifically to assess student attainment of the New Hampshire State Model Competencies. Additionally, the tasks in the NH Task Bank serve as models that teachers can use in their own assessment design work.

One of two key sources for performance tasks are those designed and submitted by New Hampshire teachers, most of who have participated in New Hampshire’s Quality Performance Assessment Initiative over the past three years. These teachers received training in task design, quality assurance, analysis of student work and calibration. Tasks that are submitted to the NH Task Bank undergo a rigorous vetting and revision process. The NH task bank is organized according to content-specific competencies arranged along a developmental trajectory. The second key source of performance tasks is through the CCSSO’s ILN Performance Assessment Project. The ILN project is collecting and curating a set of quality performance tasks that will populate an open-source, vetted task bank accessible to teachers. The emphasis of the work is on the type of performance-based measures that support assessment of deeper learning.

Professional learning support

The professional learning opportunities associated with PACE are embedded in the actual work of PACE, including task development, scorer calibration activities, system design, and peer review. The implementing schools established work groups, creating common developmental competencies in the key content areas aligned to the state graduation competencies as well continuing to build the state task bank. Sharing and analyzing student work is the core of any meaningful professional learning activity, therefore a key aspect of such learning opportunities for PACE teachers involves

² High school mathematics and science PACE assessments are tied to courses (e.g., algebra, physics) and not to grade level.

learning how to carefully analyze student work using established protocols to engage in common scoring sessions designed to foster consistent and accurate scoring of complex tasks.

Technical Issues and Considerations

In order for this reform initiative to be credible to New Hampshire stakeholders and to satisfy USED requirements, NHDOE is focused on ensuring the technical quality of the PACE system. Some of the key technical challenges include: creating comparable annual determinations, documenting longitudinal student progress (growth), measuring and reporting the performance of key student groups (equity), and establishing systems for the effective use of assessment and accountability results (utility).

Comparability of annual determinations

One of the major challenges with the PACE pilot accountability system is ensuring that students from all NH schools receive meaningful opportunities to learn the required knowledge and skills. One of the ways to evaluate these opportunities is to require all students to participate in the same assessment of the same knowledge and skills. But it is not the only way. There are many examples, both with educational programs and outside of education, where we recognize that the “same” is not the only way to define comparability. For example, consider students applying for a competitive music program. Students will play different songs, perhaps using different instruments, but judges will have to determine who should be admitted to the program. We accept that judges are able to weigh the different types of evidence to make “comparable judgments.” Why do we accept this? Because we have great trust in expert judges and their shared criteria. When the criteria are not explicit and applied systematically, then people have concerns.

True psychometric comparability (i.e., “interchangeability”) across districts administering different systems of assessment cannot be assured. In fact, it is not expected. However, NHDOE is taking important steps to ensure that students in pilot districts receive a high-quality education that meets or exceeds the expectations for non-pilot districts held to the same high expectations. For example, students deemed proficient in a particular grade or content area likely should be considered proficient regardless of the type of assessment.

Comparability efforts should not be focused on individual assessments administered throughout the year, rather the focus of comparability must be on the annual determinations of “proficient,” “on-track,” “competent,” or any other label. NHDOE has proposed an approach to do just that. The Smarter Balanced achievement level descriptors (ALDs) are the basis for establishing cutscores on the Smarter Balanced assessments (this process was recently completed). The ALDs serve as the narrative descriptions of performance and the role of the standard setting panelists is to match the narrative descriptions with actual performance on the test. Therefore, NHDOE has decided to require all PACE districts to anchor their annual determinations of proficiency (competency) to the Smarter Balanced ALDs for the respective grade level and subject area.

Of course, it is one thing to use common descriptors, but having assessment evidence to evaluate against these descriptors is another critical component of comparability. Therefore, all PACE districts have agreed to participate in a common standard setting process based on thoughtfully-identified set of summative competency assessments administered throughout the year along with the common summative PACE performance assessment. Participating in a common standard setting process, where student work is compared with the ALDs will allow for comparably rigorous achievement standards to be established in all PACE districts.

To audit the extent to which the intended comparability has been achieved, NHDOE will rely on the results of the Smarter Balanced assessments in math and ELA in at least three grades. Further, as noted above, NHDOE is closely examining the Smarter Balanced interim assessments to replace or augment current local benchmark assessments to support comparability while raising the level of performance expectations. These common state assessments provide both an internal and external audit for locally-designed systems of assessment, evaluating the degree to which student performance on the local performance assessment system relates to performance on the statewide assessments. 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.

All districts participating in the PACE pilot will participate in a peer review process during the first two years of implementation in order to examine their system design, assessment results, and annual determinations. Peer review will be structured to provide support and technical assistance to districts to ensure that local systems maintain high quality.

Lastly, NHDOE is taking steps to ensure scoring comparability by promoting accurate and consistent scoring of performance assessment tasks across classrooms, schools, and districts. NHDOE will sponsor Professional Development Institutes, including summer and school-year Quality Performance Assessment institutes on assessment literacy, competencies and designs for teaching them (knowledge, skills, and work-study practices), assessment task design and validation, scoring calibration, and data analysis to track student progress and inform instruction. Regional task validation sessions will be conducted to assist districts in fine-tuning assessment tasks to ensure they measure target knowledge, skills, and work-study practices. Regional calibration scoring sessions will be conducted to build inter-rater reliability and consistency in scoring across districts. These sessions are designed to build expertise among a core group of participants who can then lead task validation and calibration scoring sessions at the local level.

Equity

The competency-based educational system at the foundation of this pilot is, by design, more equitable than traditional educational approaches because educators focus on the learning needs of every student and do not allow any students to fall through the cracks. That said, the state will continue to aggressively monitor and report the performance of student groups as outlined in this waiver. In addition, districts participating in the PACE pilot will be subject to additional examination of student group performance through their required participation in a peer review process to evaluate aggregate and student group performance results.

Student Progress

Student Learning Objectives (SLO) continue to be the main component of NH's educator evaluation system for all NH districts. This was the clear intention of the NH Task Force on Effective Teaching (NHDOE, 2013). The state believes that it can successfully document changes in student learning while supporting positive changes in local assessment and instruction. Pilot districts, because of the improvements in their assessment capacity, will be able to produce higher quality SLOs than most NH schools and districts. Therefore, the question should focus more on can pilot districts produce valid educator evaluation results and less on specific (and distal) approaches for calculating current achievement conditioned on prior achievement (e.g., SGPs, VAM).

NH has been using Student Growth Percentiles (SGP, see Betebenner, 2009) for school accountability purposes for many years and plans to support districts in incorporating aggregate SGP results into educator evaluations starting in the 2015-2016 school year. The NH Task Force on Effective Teaching recommended not attributing SGP results to individual teachers, unless the district's specific evaluation plan requires such use. The Task Force recommended, and NHDOE agreed, that aggregate SGPs must be used at least as part of a "shared attribution" approach according to a district's (or school's) theory of improvement (e.g., grade-level or content area teams). This is an important distinction because a similar—but not exactly the same—model can be applied in the PACE schools. In other words, NH proposes to use Smarter Balanced assessments at select grades to calculate SGPs and use the results aggregated at the school level. These school-level results can be used to audit the individual SLO results and compare the "growth" of students in the pilot schools with other schools in the state.

Utility

Henry Braun stated that utility is the most important technical criterion by which we should judge the quality of accountability systems (Braun, 2012). Utility refers to the degree to which the policy/accountability system is able to support its intended aims. In the case of PACE, this would mean that the accountability system provides structure and information to help transform educator practices and deepen student learning. Focusing on utility changes the accountability conversation from one of labeling and sorting to one focused on using the results to bring about desired improvements in schools and student learning (Hargreaves & Braun, 2013).

Scaling Up

The state is not blind to well-known challenges with implementing performance assessments as part of accountability systems as well as with the challenges of building the local capacity necessary for raising the level of student learning, improving local performance assessments, and supporting local accountability determinations. The state is not attempting to meet the levels of standardization and psychometric specifications associated with a state-controlled assessment and accountability system (e.g., AERA, APA, NCME, 2014). NH argues that the theories of action for such systems are impoverished and have little evidence that such state-led systems bring about the levels of student and organizational learning the NHDOE would like to see. Rather, NHDOE is willing to engage in the challenge of supporting local capacity and agency in order to bring about transformational changes in student learning.

The state is currently developing plans for scaling such efforts to all NH schools. The current PACE accountability system, even if fully successful, is based on a voluntary proof of concept pilot with high-capacity schools. Improving chronically low-performing schools will be an enormous challenge. The state is committed to supporting the development of local leadership and capacity to help low performing schools implement the PACE system with fidelity. However, there are no illusions that this will happen overnight. In fact, the networked approach supported through PACE and other NH reform initiatives is likely the only viable strategy for bringing PACE to scale. This would involve growing this reform at a rate that can be managed and supported, while continuing to focus on building local expertise as part of regional and statewide networks. Additionally, NHDOE is working closely with and supporting NH's 18 Institutions of Higher Education (IHE) in their efforts to transform teacher education and to require performance-based evidence of acceptable pre-service candidacy. Again, NHDOE does not assume that implementing a reciprocal accountability will be easy or smooth, but is committed to employing an approach couched in research on

individual and organization learning to realize the deeper learning for students envisioned by many NH stakeholders.

Evaluation

As was discussed in the approval letter sent by Assistant Secretary Deborah Delisle of March 5, 2015, the NH Department of Education is arranging for a third-party evaluation of the outcomes of the pilot to be completed after the 2015-2016 school year. The evaluation will address the efficacies of this multiple assessment system accountability system, through comparability studies, an analysis of the effectiveness of teacher calibration and large-scale moderation, as well as a review of Smarter Balanced Assessment results and PACE Common Assessment results by content area.

Next Steps

Upon completion of the evaluation and an analysis of the pilot overall, the NH Department of Education will determine whether to seek a change to this waiver, once approved, in order to make project amendments and to expand the implementation of the PACE system for the 2016-2017 school year.