



Reciprocal Accountability for Transformative Change: New Hampshire's Performance Assessment of Competency Education

SCOTT F. MARION, JONATHAN VANDER ELS, AND PAUL LEATHER

In New Hampshire, a new performance assessment system focuses on reciprocal accountability and shared leadership among teachers and leaders at the school, district, and state levels.

For every increment of performance I demand from you, I have an equal responsibility to provide you with the capacity to meet that expectation. Likewise, for every investment you make in my skill and knowledge, I have a reciprocal responsibility to demonstrate some new increment in performance. (Elmore 2002, p. 5)

Scott F. Marion is the executive director of the National Center for the Improvement of Educational Assessment in Dover, New Hampshire. Jonathan Vander Els is the executive director of the New Hampshire Learning Initiative. Paul Leather is deputy commissioner of the New Hampshire Department of Education.

This concept of *reciprocal accountability*, developed by school improvement expert Richard Elmore, is at the core of New Hampshire’s Performance Assessment of Competency Education (PACE), a competency-based educational approach designed to ensure that students have meaningful opportunities to achieve critical knowledge and skills (see Marion & Leather 2015; Rothman & Marion 2016; New Hampshire Department of Education 2016). For PACE, reciprocal accountability means that local educational leaders are involved in designing and implementing the assessment and accountability systems and receive intense technical, policy, and practical support and guidance from the New Hampshire Department of Education (NHDOE) and other experts in the field. PACE attempts to foster organizational learning and change by appealing to the intrinsic motivation of adults to improve their work rather than relying on top-down accountability and compliance strategies.

Beginning in 2012, New Hampshire worked with the Center for Collaborative Education (CCE) to implement performance assessment literacy training, using professional development and capacity building to lay the groundwork for moving forward. In March 2015, the U.S. Department of Education granted permission to New Hampshire and their advisors from the National Center for Improvement of Educational Assessment (Center for Assessment) to pilot PACE, a new assessment and accountability system with significantly greater levels of local design and agency, with an overall goal to facilitate transformational change in performance that best supports the goal of significant improvements in college and career readiness.

As part of this shift in orientation, the state is supporting a competency-based approach to instruction, learning, and assessment within an internally oriented accountability model, in which those being held accountable have responsibility for co-developing the standards,

measures, and bars set for proficiency. Assessment of competency-based learning almost always requires performance-based assessment, and the information learned through this process will continue to inform the design of the accountability system and, hopefully, better inform school improvement (Hargreaves & Braun 2013).

PACE involves multiple lines of work and multiple players. Here, we use three specific perspectives to provide tangible examples of reciprocal accountability in action:

- The first example – of *shared leadership* – is presented by Paul Leather, New Hampshire’s deputy commissioner of education, who as the official leader of the project had to build a structure based on shared decision making among the state, districts, and external partners.
- The second story – of *building local capacity* and expertise – is told by Jonathan Vander Els, the current executive director of the New Hampshire Learning Initiative and former principal of Memorial Elementary School in Sanborn Regional School District, one of the original PACE districts.
- The last example is presented by Scott Marion, executive director of the Center for Assessment and the lead technical advisor to PACE. He discusses the ways in which the evaluation of *technical quality* of the PACE assessment system is based on the reciprocal notion of supporting expertise among local educators while meeting rigorous psychometric requirements.

THE VIEW FROM THE STATE: SHARED LEADERSHIP AND RECIPROCAL ACCOUNTABILITY (PAUL LEATHER)

Under former Commissioner Virginia Barry’s leadership, the NHDOE has long practiced reciprocal or “shared leadership” for the major decisions in our state’s public education. Barry met with the district superintendents and

other educational leadership groups monthly to discuss major issues such as educator effectiveness, educational innovative practices, and the opioid crises. In particular, shared leadership discussions have addressed assessment and accountability for many years, from the adoption of the Smarter Balanced Assessment Consortium¹ in 2014 to the design of state accountability systems since the onset of No Child Left Behind in 2002. It was at just such a discussion, held within the confines of the state's Accountability Task Force in 2014, where the idea for PACE was born.

The task force, made up of superintendents, curriculum supervisors, teachers, and association chapter directors, discussed the idea of moving to a new kind of accountability system more in keeping with competency-based education. Chris Rath, then superintendent of the Concord School District, said in no uncertain terms, "We can't take on something this innovative without you providing us some space to innovate. With the Common Core, Smarter Balanced, and other efforts all being implemented this year [2014-2015], our educators are overburdened as it is." After some discussion, the group agreed with the idea of advancing a pilot to include volunteer districts, where Smarter Balanced would be implemented only once each in elementary, middle, and high school, and a bank of complex performance tasks would be used in grades and subjects where Smarter Balanced was not administered. In this way, the idea of "space to innovate" was integrated into New Hampshire's accountability system.

This model of shared decision making became the operational norm for PACE. A roundtable was created, made up of field representatives from the original

four participating districts, two external partners (Scott Marion of the Center for Assessment and Dan French of CCE), and NHDOE staff (Deputy Commissioner Paul Leather and PACE State Director Mariane Gfroerer). Originally, this group met at least monthly to address all of the issues of design, planning, professional development, implementation, reporting, and technical quality. Nothing moved forward without the full consensus of the group.

Now in its third year, the pilot has grown to eight districts and one charter school, and the makeup of the leadership team remains the same, with each district or charter school represented at the table. Meanwhile, consistent with the principles of reciprocal accountability, the field leaders and teachers have taken on more and more of the ongoing work of PACE. Eighteen teacher content leaders now facilitate the construction of new common PACE performance assessment tasks in English language arts, math, and science for grades 3–7 and 9–10.

With the NHDOE's support, a new organization has been constructed: the NH Learning Initiative, which serves as an intermediary entity supporting the work of both the field and the Department. Also, the New Hampshire chapter of the National Education Association is supporting another group of teacher leaders to facilitate PACE implementation with fellow educators within and across districts. All of this work is overseen by the PACE leadership team, which continues to meet monthly. Members demonstrate their shared ownership and commitment to the success of the pilot in many ways, including through presentations at district, state, and national conferences and to state government officials.

RECIPROCAL ACCOUNTABILITY AT THE SCHOOL LEVEL (JONATHAN VANDER ELS)

When I served as a principal in one of the original implementing PACE schools, reciprocal accountability was at the core

.....

1 Smarter Balanced and the Partnership for Assessment of Readiness for College and Careers (PARCC) are assessment systems that 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://www.smarterbalanced.org/> and <http://www.parcconline.org/>.

of our vision ensuring that all students achieve at high levels. I and my teachers subscribed to a shared leadership model in which we were together responsible for the success of our students, and we needed to work collaboratively to truly maximize the strength of the whole school.

In order for PACE to be effective, the capacity of all educators in each of the implementing schools must be developed to the fullest extent possible. Teachers must possess deep understanding of content, discipline-specific pedagogy, and well-developed assessment literacy to teach and assess a rigorous curriculum using complex performance tasks. Teachers must also be willing and able to work collaboratively in and across schools to develop shared expectations and vision.

We worked hard to develop a culture in which it was safe to innovate. Teachers were used to (and comfortable with) working either individually or within their school-based team. PACE required teachers across schools and districts to function in a professional learning community, through which they learned how to work together most effectively, how to look at student work, understand data, and most importantly, make changes to their instruction to meet the needs of all learners. Our teachers' role was to embrace the uncertainty that comes with stepping out of their comfort zones, committing to working collaboratively with colleagues, and sharing our learning to benefit all.

PACE came along at the right time for our school and our district. We had transitioned to "competency-based learning" a few years earlier, but our teachers really began to develop their assessment literacy by creating, administering, and refining Quality Performance Assessments, a professional development opportunity provided by CCE and initially made available over the summer by the NHDOE. Because we were already engaged in developing high-quality performance assessments, PACE was a logical and timely opportunity to participate in an assessment and

accountability effort that was not based on a single, standardized measure to evaluate students and schools.

Teachers' capacity and professionalism are at the heart of PACE. Relying on teacher leadership and autonomy to be "in charge" of the project has put teachers back into the driver's seat, determining students' competency and utilizing the data from the performance assessments to provide support, intervention, and extension, as appropriate, in a timely manner. For teachers, the essence of reciprocal accountability is a sense of "being heard." As one of our lead PACE teachers explained:

I think PACE has been successful so far because the people working on the initiative believe in the work. The people in charge listen to teacher feedback and are adaptable. We all understand the importance of the work and want it to be successful because it's what is best for kids.

We all have a role to play in the success of PACE, and all clearly understand the need to work with, and for, each other to support our students.

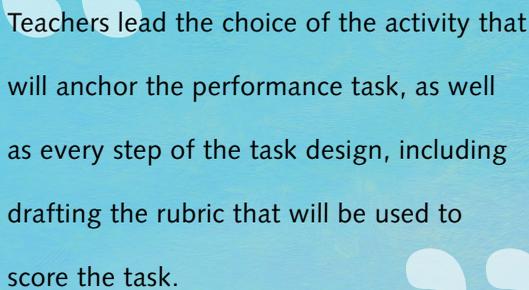
A RECIPROCAL ACCOUNTABILITY APPROACH TO EVALUATING TECHNICAL QUALITY (SCOTT MARION)

PACE has been recognized for its multifaceted approach to the evaluation of technical quality. (See, for example, Lyons & Evans, forthcoming; Rothman & Marion 2016.) In most cases, technical quality evaluations are the purview of highly trained psychometricians like those of us who work at the Center for Assessment. PACE leadership has always had a goal of ensuring that only high-quality assessments were used in participating schools, but we insisted from the beginning of the project that technical quality had to be a participatory sport. In other words, the evaluations of technical quality had to both gauge the quality of the assessments used and to increase the

assessment expertise of participating educators. While there are many aspects of our shared approach to evaluate assessment system quality, we highlight three key components here.

High-quality assessment design

Assessment quality starts with principled and high-quality assessment design. The assessment design templates were drafted by staff at the Center for Assessment, but revised based on feedback and interaction with participating teachers. The Center for Assessment team provides technical support and some oversight to the teacher-led task development teams, but the decisions about which assessments are used in the project are made collaboratively among the teacher leaders, project staff, and the technical consultants. The teachers lead the choice of the activity that will anchor the performance task, as well as every step of the task design, including



Teachers lead the choice of the activity that will anchor the performance task, as well as every step of the task design, including drafting the rubric that will be used to score the task.

drafting the rubric that will be used to score the task. Teachers suggest ways in which the task or tasks will work best within their instructional programs and together with the technical advisors negotiate among district content experts and the technical advisors to design tasks that can serve both instructional and accountability purposes.

Reliable and accurate scoring

Performance assessments must be scored accurately and consistently in order to support their uses to inform instruction and to serve as accountability measures. Further, a key tenet of PACE is that inferences regarding student achievement must be *comparable* across participating districts and between pilot and non-pilot districts, meaning that given a certain set of student work, a student rated as “proficient” in one district would be rated similarly by educators in a different district.

Ensuring scoring quality and comparability starts at the school and district levels, where participating PACE schools engage in calibration exercises to develop a shared understanding of student work quality. The PACE calibration protocol was developed and tested collaboratively among my staff, PACE teachers, and PACE district leads. This process was another example where more top-down technical quality approaches had to be negotiated with the practical realities of doing this work with teachers who have many other responsibilities. For example, we would have liked to have larger samples of student work for our calibration work, but that would have been a burden on the teachers, so we negotiated a sample size that is manageable for the teachers but still provides enough data for us to conduct the necessary technical analyses. In addition to the internal calibration work, each district collects data on the degree to which teachers score the performance tasks consistently with other teachers in the district. The Center for Assessment uses these data to compute inter-rater consistency statistics and then reports back to districts so they can use the information to improve their scoring quality.

Comparability of assessment results across participating districts

The key activity in evaluating cross-district comparability involves a massive collaborative effort led by my psychometric staff and involving hundreds of

educators and project leaders with the main event taking place over the course of two days each summer. Anonymized student papers are distributed to randomly arranged teams of teachers to produce “consensus scores.” These consensus scores serve as benchmarks by which local district scoring is evaluated. (Out of more than 400 papers scored, fewer than five each year required a third rater to help the original raters come to consensus.) Ideally, there should be only small differences between the consensus scores and the scores provided by the original teacher. This alignment would indicate a high degree of scoring accuracy. The more immediate concern is to ensure that the average differences between each district’s local scores and the consensus scoring are similar across districts. The extent to which a district deviates from other districts is a measure of leniency or stringency in local scoring (see Queensland Curriculum & Assessment Authority 2014).

We could have chosen to employ a more typical statistically based approach to comparability, but that would have been more top-down and would have done little to build the skills of participating teachers. The approach we designed allows teachers to collaboratively interrogate student work and to have their consensus judgments play a crucial role in the comparability evaluations. Further, this close examination of student work allows teachers to build their assessment literacy and understanding of student learning.

CONCLUSION

An innovative assessment and accountability project like PACE is unique and important for many reasons. The extensive use of performance assessments helps support learning (Shepard 2000) and increases teacher assessment literacy. The focus on high-quality performance tasks is something we have not seen on a large-scale since initiatives in several states in the 1990s. PACE seeks to demonstrate that some of the past technical concerns with the use of

performance assessments for accountability can be satisfactorily addressed (Evans & Lyons 2017). PACE provides a vivid example of reciprocal accountability in action, framing the ways in which PACE operates at all levels – from the NHDOE, to the approaches for evaluating and improving technical quality of performance assessments, to the collaboration among teachers, to the interactions between teachers and students.

REFERENCES

- Elmore, R. 2002. *Bridging the Gap between Standards and Achievement: The Imperative for Professional Development in Education*. Washington, DC: Albert Shanker Institute.
- Evans, C., and S. Lyons. 2017. “Comparability in Innovative Assessment Systems for State Accountability,” *Educational Measurement: Issues and Practice*. 36, no. 1:1–11.
- Hargreaves, A., and H. Braun. 2013. *Data-Driven Improvement and Accountability*. Boulder, CO: National Education Policy Center. <https://goo.gl/gRYXhk>.
- Marion, S., and P. Leather. 2015. “Assessment and Accountability to Support Meaningful Learning,” *Education Policy Analysis Archives* 23, no. 9. <https://goo.gl/bvASfq>.
- New Hampshire Department of Education. 2016. *Moving from Good to Great in New Hampshire: Performance Assessment of Competency Education (PACE)*. Concord: New Hampshire Department of Education. <https://goo.gl/fPU3hP>.
- Queensland Curriculum & Assessment Authority. 2014. *School-Based Assessment: The Queensland System*. Queensland, Australia: The State of Queensland (Queensland Curriculum & Assessment Authority).
- Rothman, R., and S. F. Marion. 2016. “The Next Generation of State Assessment and Accountability,” *Phi Delta Kappan* 97, no. 8:34–7.
- Shepard, L. A. 2000. “The Role of Assessment in a Learning Culture,” *Educational Researcher* 29, no. 7:4–14.