



The View from the States

A Brief on Non-Tested Grades and Subjects

In designing multiple-measure evaluation systems for teachers and principals, one of the key questions to emerge is: how to obtain student growth data for teachers of academic subjects and grades not covered by statewide assessments. Race to the Top grantees that began early to implement new evaluation systems are well on their way to incorporating new measures of student growth into the evaluation process. This brief profiles how three Race to the Top grantees—Delaware, North Carolina and Tennessee—have addressed this question. Leaders from the Delaware Department of Education (DDOE), the North Carolina Department of Public Instruction (NCDPI) and the Tennessee Department of Education (TDOE) reported to their peers during a seminar on August 28, 2012 in Washington, DC, on how they have developed growth measures for teachers of non-tested grades and subjects. This brief, based on their presentations, offers a look at the paths each State is taking, but does not attempt a comprehensive overview of policy or practice, or an assessment of the quality of their initiatives.

Delaware: Internal Assessments to Measure Student Growth

At the beginning of the 2011–2012 school year, the DDOE engaged 600 Delaware educators in the construction of assessments from the ground up, for non-tested grades and subjects. This was not a new challenge: The DDOE had attempted to engage 400 educators in assessment identification and construction in January 2010 but the effort floundered, in part because of the absence of needed support from assessment specialists. Missteps in this first attempt, however, became an opportunity for learning and redesign.

Prior to Race to the Top, Delaware had in place a statewide teacher evaluation system based on the Charlotte Danielson framework: the Delaware Performance Appraisal System, or DPAS. What was new with the advent of Race to the Top was a requirement to integrate a student-growth component (“component 5”) valued at 20 percent of an educator’s final rating. Also new was an emphasis on multiple measures, to be rigorous and comparable across classrooms. As one outcome of the new policy, no teacher’s evaluation could depend solely on a school-wide score. This left the DDOE to determine a structure for multiple measures that would respect the policy and deal with the fact that two-thirds of the State’s educators did not have students who took the statewide test (the Delaware Comprehensive Assessment System or DCAS) that provides student growth data linked to teacher performance.

Linda Rogers, associate secretary for the Teaching and Learning Branch of DDOE, told the seminar how her agency engaged educators in the design of assessments that would provide teachers of non-tested grades and subjects with student-growth data. Most of these teachers could not obtain the growth data needed for their evaluations from either the DCAS or commercial assessments. In its new initiative, DDOE provided much more structure, carefully designed to provide teachers with preparatory professional development, engage the union leadership and define the assessment deliverables that the educators would be asked to create.

To get started, DDOE engaged a vendor to create a series of professional development **modules** for training that would total five days, spread over several months. The modules would train teachers to create standards-based assessments for those grades and courses whose students do not take the DCAS. The DDOE leadership team organized the teachers according to grade and subject into working groups of six to twelve—mathematics grades 9-12, for example. Facilitated by DDOE staff, the training began with instruction in the basics of assessment—distinctions between summative and formative assessment, for example—and moved on to the complexity of writing growth goals and test items. The modules covered the fundamentals of what constitutes a good rubric, explaining the eight strands in their model, which would pertain both to evaluating external measures and building internal measures.

Supported by professional development, the educator groups worked on the assessments for their assigned subject and grade range the remainder of the 2011–2012 school year. Each group was charged with producing a package of six deliverables that would constitute the complete set of materials necessary to provide data points for component 5 of DPAS: the purpose statement, test blueprint, test specifications, pre-test/post-test, administrative guidelines and scoring guidelines. Rogers noted at the seminar that the package was designed with these elements to ensure that the assessments met the expectations for rigor and comparability.

As the new school year approached, in July 2012, DDOE neared completion of an analysis of every item on every assessment, reconvening teachers to create an additional item bank in anticipation that some items would fail to perform. Although Rogers does not foresee that DDOE will bring teachers back again on the same scale, the leadership does intend to continue building the item bank, already funded through the State budget rather than Race to the Top grant funds, recognizing the long-term issue of having only one pre-test and one post-test per subject. By August 2012, 43 educator groups had produced assessment **packages** with growth goals specifying targets and evidence. (One example of these packages is the *DDOE Educator Growth Goals 2012 for Adapted Physical Education*.)

Rogers called DDOE’s involvement of teachers in assessment design “the best work that we have done.” The time and expense it took to engage and train teachers to prepare the assessments was extremely worthwhile, she noted, not only because of the value of the products, but because the teachers who took part helped in the rollout and implementation of the new assessments, commending them to colleagues. Had technical quality been the only concern, she continued, it might have been more efficient to hire assessment experts rather than train teachers to write items, but DDOE’s strategy had its reward: a great gain in “engagement and credibility” among teachers statewide for the new system of measures. The teachers themselves lauded the professional development and the opportunity to engage in the work.

North Carolina: Measures of Student Learning

The North Carolina State Board of Education passed a requirement early in 2011 that every teacher’s rating include student growth data, placing a new challenge in front of the NCDPI, the State education agency. The state assessment provided data for about 33 percent of the teaching workforce. With no change, most teachers would be rated on a school-wide growth value. NCDPI, as well as most teachers, would prefer to use another measure.

In 2011, the North Carolina State Board of Education added a sixth standard to the North Carolina Educator Evaluation System and required that student growth data determine every teacher's rating on the sixth standard. (School administrators determine ratings on the other five standards, using evidence from observations and other data.) For the 2011–2012 school year, the State board also set sixth-standard ratings. For teachers who had State assessment data, 70 percent of their rating would be based on a teacher's individual growth value and 30 percent on the school-wide growth value. For a teacher of a non-tested grade or subject, 100 percent of the rating would be based on the school-wide growth value.

Facing these circumstances—as Rebecca Garland, chief academic officer for North Carolina, explained at the August 28 seminar—NCDPI determined that in subsequent years, all North Carolina teachers would have an individual growth measure. To make this happen, NCDPI's leaders established a strategic plan for an intensive investment in assessment creation during the 2011–2012 school year. Focusing first on core subjects that lacked existing measures, they designed a strategy to create statewide “measures of student learning” (MSLs), common end-of-course assessments that local educational agencies (LEAs) and Race to the Top-funded charter schools could select as a source of student growth data. (The State holds that “these assessments belong to local school districts to administer and be responsible for,” Garland noted.) Second, NCDPI's leaders decided to partner with teachers to design the new assessments. Third, they carefully defined the role of the teachers as “setting specifications” or “blueprints” for the test items. Teachers would not write test items, but would review the items written by the NCDPI's contractor, the Center for Urban Affairs and Community Services at North Carolina State University.

To select the teachers to create the blueprints, late in 2011, NCDPI screened more than 1,500 online applications, seeking both experience and diversity. They chose 800 teachers. To prepare the teachers for their role, NCDPI leadership worked with a contractor to devise a package of training materials. First, they offered

the teachers background in educator effectiveness. Next, the NCDPI assessment development group walked the teachers through what Jennifer Preston, Race to the Top coordinator for educator effectiveness, described as “Assessment 101”—an explanation of item types, validity and reliability.

At this point, the NCDPI staff organized the teachers according to their content areas to meet with the NCDPI curriculum and instruction division, who briefed teachers on the **Common Core State Standards and the North Carolina Essential Standards**. They also discussed the State's educator effectiveness model and the role of MSLs in that model. Then the State staff provided templates for the teachers to reflect upon each standard—first individually and then in groups—and its connections with others, how much weight to give a standard and how to assess it. To create the blueprints for the test items, Garland explained, each content area group determined the standard to be assessed, the types of items to use and the item specifications. Preston noted that “our 800 teachers have been so critical to us, serving as that check to ensure that what we do is authentic” in designing an assessment that aligns with how teachers teach the standards.

Once the blueprints were completed and delivered to the Center for Urban Affairs and Community Services, the Center managed the production of test items to the blueprint specifications prepared by the teacher content area groups. NCDPI has chosen to work with the center in part because it hires North Carolina teachers to write test items, a critical plus for the credibility of the MSLs.

In August, NCDPI brought the teachers back in for two days to review each of the 6,000 items and ensure that each met specific state criteria. They also created rubrics and guidance for administering and scoring the assessments.

With this phase of the work completed, NCDPI approved and implemented **30 MSLs** at the beginning of 2012–2013 for grades 4–12 social studies, science, English language arts and mathematics, with the result that all teachers of these subjects in those grades now

have either a State exam or a common exam. Garland plans for NCDPI to revise the MSLs annually to permit the release of each year's items at the end of the school year to allow parents and others to see them.

Seventy percent of the State's teachers now have assessments that yield student growth data, but 30 percent do not, including teachers of critical languages, the arts, physical education and health, and an array of local electives. NCDPI's leadership is considering SLOs, pre- and post-tests and rubrics that principals would be trained to use.

Looking back over the year, Garland emphasized that this was not only a wonderful professional development opportunity for teachers, but that it also led to "unintended positives" of great benefit to NCDPI. These teachers became leaders, noted Preston, persuasive ambassadors not only for the MSLs, but for the educator effectiveness system as a whole, helping their colleagues understand student growth assessment and its place in the new system.

Tennessee: The Fine Arts Portfolio

In 2010, TDOE leaders inaugurated an initiative to identify individual value-added growth measures for teachers not covered by the Tennessee Value-Added Assessment System (TVAAS), in order to fulfill the State's new policy that 35 percent of every teacher's rating be based on student growth data. Because TDOE committed to intense scrutiny of alternative measures before approving them, the 2011–2012 school year was a pilot year for many of the proposed alternate measures, as well as the inaugural year for the new statewide teacher evaluation system, the Tennessee Educator Acceleration Model (TEAM). As a stopgap that first year, nearly two-thirds of Tennessee's teachers received a school-wide or system-wide score because no State tests covered their subjects and grades.

At the August 28 seminar, Sara Heyburn, TDOE's assistant commissioner for teachers and leaders, framed the twin purposes of her State's policy of developing alternative measures: to ensure access for more teachers to an individual measure and to improve student performance. Tennessee leaders wanted to avoid

Tennessee's First to the Top Act, adopted in 2010 by the General Assembly with bipartisan support, established a statewide educator-evaluation system based on student outcomes. The act required basing 35 percent of the evaluation on student growth data from the TVAAS or a "comparable measure" for teachers of non-tested grades and subjects. (Classroom observation and other qualitative measures would account for 50 percent, and other measures of student achievement, to be established by the State Board of Education, for 15 percent.) Implementation of the new system—the Tennessee Educator Acceleration Model or TEAM—was set for the 2011–2012 school year. During that first year, about 36 percent of Tennessee's teachers received an individual value-added score through TVAAS, and the 64 percent who taught non-tested grades and subjects were assigned a school-wide or system-wide value-added score.

creating new State tests to measure student growth for the sole purpose of teacher evaluation. They also sought to maintain strict standards for design and approval of the new assessments, but eventually allow LEAs to choose whether or not to use the State-approved tools. This policy meant that TDOE would have to thoroughly vet the new alternative measures, which they understood to include time for pilots of each proposed alternative measure.

TDOE began recruiting teachers to create the new assessments, because State leaders believed that teachers know best how to measure growth in students. Teachers took part at different levels of engagement. Those on the educator work teams were recommended by their professional associations, supervisors and others knowledgeable about teacher leaders in the various subjects and grades. Beyond the work teams of 6-12, teachers throughout the state also reviewed and commented on proposals that the work teams developed.

At an early gathering of the educator work teams, teachers voiced dissatisfaction with some of the initial options presented by TDOE, recalled Dru Davison, chair

of arts education from Memphis City Schools, at the August 28 seminar. Many failed to meet one or more of the criteria that the teachers had set: Measures should help children, strengthen the teaching profession and provide an authentic range of teacher effectiveness scores. The group proposed using a portfolio to assess subjects outside the core academic areas. After conversation with the teachers, TDOE leaders decided to pilot a portfolio approach to fine arts assessment. The portfolio would consist of student work samples that documented growth in the arts, such as a performance or visual artwork.

The fine arts educator work team met throughout 2011–2012 under Davison’s leadership. The team wrestled with the inherent complexity of the assessment process and with the painstaking work of basing the portfolio on standards. According to State policy, all measures had to meet the same rigorous standards as the TVAAS. Measures had to be rigorous, taken across two points in time, comparable across classrooms, and aligned with college- and career-ready standards.

Because the fine arts have had formal standards for many years, educators in these disciplines are familiar with the idea of standards informing practice, Davison observed. Fine arts standards, however, tend to be broad and specify tasks, rather than set benchmarks for how well tasks should be done and at what age, which is necessary for portfolio assessment. The fine arts group had another challenge: to ensure that portfolios measured growth—as required by State policy—rather than achievement.

A crucial element of Tennessee’s portfolio approach is its requirement for blind peer review by the best teachers in each field, Davison believes. The teacher and the peer reviewers each use the scoring guide to determine whether the portfolio demonstrates adequate growth. If the peer review and self-review ratings are within a point, the review is complete. Substantial differences trigger additional blind reviews.

Early in 2012, the proposed portfolio measure was reviewed by technical advisors and stakeholders, revised by the fine arts group and submitted to the

TDOE. Once the measure was conditionally approved by the TDOE, Memphis City Schools (MCS) piloted the portfolio, engaging 500 teachers. MCS is a Gates Foundation Partnership District, which provides funding for the MCS Teacher Effectiveness Initiative. The arts work is funded in part by the MCS Teacher Effectiveness Initiative, which enabled the working group to produce scoring guides, teacher professional development, the beta test, pilot test and data for the formal recommendation to TDOE. The pilot won support from the teachers and demonstrated a range of distribution of effectiveness scores.

In spring 2012, the working group submitted its report on the pilot and proposed the portfolio as an alternative growth model to the TDOE. Winning the Tennessee State Board of Education’s approval in August 2012, the path was open for adoption in the **2012–2013 school year**. In fact, 31 LEAs had voluntarily paid to attend the training in summer 2012 in anticipation of the board’s approval.

Looking back, Heyburn noted, “You can’t underestimate the importance of engaging educators in this work.” The TDOE plans to adopt the portfolio model to create growth measures for teachers of pre-kindergarten, kindergarten, physical education and health, and high school science courses.

Conclusion

The three very different assessment development initiatives conducted by Delaware, North Carolina and Tennessee do not represent the only approaches to devising measures of student growth for non-tested grades and subjects, but they are thought-provoking examples of efforts to address the challenge. All three states report that despite progress they have made, much work remains. They are approaching that work with a spirit of continuous improvement, using data—including teacher feedback—to evaluate their work, acknowledging missteps, making adjustments and finding new pathways for further advances.

Importantly, all of the States decided to engage teams of teachers to develop these student learning measures, although their specific roles varied in each State. The

State education agencies that collaborated with these teams voiced unanimous satisfaction with the technical quality and the commitment of the work, and the seriousness with which the teachers took part, but also recognized the importance of making sure the teacher teams got enough staffing and technical expertise.

In an unanticipated outcome for all three States, these teachers had a major impact upon returning to their home schools and LEAs. They became spokespeople not only for the new measures but for the educator effectiveness systems of which the measures were a single component. For the teachers who worked with the State agencies, the experience was an opportunity to learn about standards and assessments as well as to contribute to the new effectiveness system. State leaders believed that many teachers had concluded that the new system they were helping to build was a marriage of assessment and professional development that offered teachers new opportunities for professional growth and more of a role in their own evaluation than allowed by previous systems. In carrying their experiences back to their workplaces, these teachers gave other educators the keys to ownership of the new system. State agencies interested in building support for new educator evaluation systems might find this outcome instructive.

The presentations also demonstrated the differences in policy context and the ways the State education agencies adapted solutions appropriate to them. Tennessee's leaders described their approach as a "tight/loose coupling:"The State establishes rigorous assessments comparable across schools, but allows Tennessee's 136 LEAs discretion in choosing which measures to use. Similarly, North Carolina sought to "provide the tools and the process," but beyond that, leave it to LEAs to administer and be responsible for measures of student learning. In Delaware, teachers whose courses are not covered by DCAS (the statewide assessment), must choose one of the State-approved internal measures, set their own growth targets for a sample of their students and obtain their principal's approval for the targets.

Policy Background

Since 2008, many States have revised their statutes to require multiple measures for teacher evaluation. (See the National Council on Teacher Quality publication, *State of the States 2012: Teacher Effectiveness Policies*.) Race to the Top policy language requires that these evaluations include multiple measures, defined as "multiple rating categories that take into account data on student growth ... as a significant factor." States are asked to establish clear approaches to measuring student growth that account for individual students. Race to the Top policy language further requires that measures of student growth have the following qualities:

- Rigorous in reflecting high expectations for student progress toward college and career readiness
- Determined across two or more points in time
- Comparable across classrooms in predicting progress toward standards in the subject assessed
- As rigorous in non-tested subjects and grades as in tested subjects and grades

The language also distinguishes non-tested grades and subjects from tested grades and subjects. Tested grades and subjects are those covered by a State's assessment under the Elementary and Secondary Education Act (ESEA), usually math and reading, grades four through ten. Non-tested grades and subjects are any that fall outside of those fields of instruction. Although States are expected to use data from the statewide assessment for teacher evaluation, they are free to use additional measures.

This leaves significant portions of the teacher workforce without a source for student growth data based on State assessments. Most States collect assessment data for only 25–35 percent of their teachers, leaving 65–75 percent without a statewide measure that can be used to calculate a value-added score (Goe 2010)—thousands of students and teachers in most States.

Because so many teachers lack student growth data from State assessments, States implementing teacher evaluation systems face the challenge of identifying other sources of student growth data. States and LEAs may use a variety of methods, depending upon the State context, the assessments already available, and the costs and benefits of different methods. At this writing, there are three broad approaches emerging, with others under exploration:

- **Student Learning Objectives (SLOs)** are a participatory method of setting measurable goals, or objectives, based on a specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of students and the measurable gain in student performance during the course of instruction. SLOs can be based on standardized assessments or on teacher-developed or other classroom assessments if they are “rigorous and comparable across classrooms.”
- **Other measures of student growth** involve developing or adapting other ways to quantify student learning over time for non-tested grades and subjects used by schools or LEAs. These measures may include early reading measures; standardized end-of-course assessments; formative assessments; benchmark, interim or unit assessments; and standardized measures of English language proficiency.
- **Measures of collective performance** are measures required by ESEA or other standardized assessments used to gauge the collective performance of groups of teachers. Measures of collective performance may assess the performance of the school, grade level, instructional department, teams or other groups of teachers. These measures can take a variety of forms including school-wide student growth measures, team-based collaborative achievement projects and shared value-added scores for co-teachers.

References and Resources

General

1. Laura Goe. (2010). *Teacher Evaluation in Transition: Using Evaluation to Improve Teacher Effectiveness*. The National Comprehensive Center for Teacher Quality. Published in Katie Buckley and Scott Marion. (June 2, 2011). *A Survey of Approaches Used to Evaluate Educators in Non-Tested /Grades and Subjects*. Harvard University and National Center for the Improvement of Educational Assessment.
2. Reform Support Network. (2011) *Measuring Student Growth for Teachers in Non-Tested Grades and Subjects*. Washington, DC.
3. August 28, 2012 seminar materials at <http://www2.ed.gov/programs/racetothetop/communities/teacher-leader-effectiveness-cop.html> (all materials listed below.)

Delaware

- Delaware Department of Education DPAS II—Component 5-Part III: External Measures PDF (135K) Provides a list of external, pre-existing measures that have been approved for use in the Core Content Areas and includes a description of each measure.
- Delaware Department of Education Measure B—Internal Assessments PDF (291K) Provides a list of internal assessments approved in Delaware by grade level or course.
- Delaware Department of Education Algebra I Post-test PDF (422K) Provides a sample of Delaware’s new Algebra 1 Post-test.
- Delaware Department of Education Measure C—Growth Goals PDF (188K) Provides a list of subjects for which there are Delaware Measure C growth goals.
- Delaware Department of Education Educator Growth Goals 2012: Adapted Physical Education PDF (606K) Indicators for Adapted Physical Education educators.

North Carolina

- North Carolina Design of the Measures of Student Learning: NC’s Common Exams—Summer 2012 PDF (492K) Provides a PowerPoint to inform individuals throughout the State about the measures of student learning for non-tested subjects for districts to use to populate Standard 6.
- Implementation Timeline—Measures of Student Learning NC’s Common Exams PDF (57K) Provides a list of subjects that will be tested in each semester for the 2012-13 and 2013-14 school years.

Tennessee

- Teacher Evaluation in Tennessee: A Report on Year 1 Implementation PDF (958K) Summarizes the work Tennessee did during the first year of implementation of teacher evaluation.
- Tennessee Arts Growth Measures System PDF (299K) Describes the portfolio system process that can be used by schools as an evaluation model for Fine Arts Teachers.

This publication features information from public and private organizations and links to additional information created by those organizations. Inclusion of this information does not constitute an endorsement by the U.S. Department of Education of any products or services offered or views expressed, nor does the Department of Education control its accuracy, relevance, timeliness or completeness.