

Consortium for
Educational
Research and
Evaluation—
North
Carolina

Executive Summary of

Measures of Student Growth in the North Carolina Educator Evaluation System

Formative Evaluation Report

Heather Lynn, Nate Barrett, Julie Marks, and Aubrey Comperatore
Carolina Institute for Public Policy, University of North Carolina
at Chapel Hill

Gary T. Henry and James Guthrie
Peabody College, Vanderbilt University

September 2013

Consortium for
Educational
Research and
Evaluation—
North
Carolina



MEASURES OF STUDENT GROWTH IN THE NORTH CAROLINA EDUCATOR EVALUATION SYSTEM: FORMATIVE EVALUATION REPORT

Executive Summary

The Consortium for Educational Research and Evaluation–North Carolina (CERE–NC)¹ is evaluating the Race to the Top (RttT) initiative to integrate and fully implement the addition of a student growth component into the educator evaluation process for teachers and principals. The goal of the Consortium’s evaluation is to examine this implementation, as well as perceptions of the initiative among teachers and principals and outcomes associated with the addition of the student growth measure to the evaluation process. The Evaluation Team is assessing correlations between measures of student growth and other measures of teacher performance and collecting baseline data for assessing the impact of the addition of student growth measures to the educator evaluation process on educator and student performance over time.

Expansion of the Teacher and Principal Evaluation Process

Addition of a Student Growth Measure to the Evaluation Process

The North Carolina Department of Public Instruction (NCDPI) developed the North Carolina Educator Evaluation System (NCEES), which consists of multiple rating categories, with the input of teachers and principals. Before RttT, the system was comprised of five performance standards for teachers and seven for administrators. The goal of this initiative is to adopt and implement a 6th and 8th standard, respectively, to the North Carolina Teacher Evaluation Process (NC TEP) and North Carolina Principal Evaluation Process (NC PEP) that formally integrates student growth data into assessments of educators’ effectiveness. Teachers and principals receive separate ratings on each of the standards that comprise their evaluations, as well as an overall effectiveness rating that takes into account their performance on all of the standards. Standards 6 and 8 are given the same weight as each of the other standards (Standards 1 through 5 on the NC TEP and Standards 1 through 7 on the NC PEP) when calculating an educator’s overall effectiveness status.

In April 2013, the North Carolina State Board of Education (SBE) adopted a policy² that requires educators to receive an overall effectiveness status (incorporating ratings from Standards 1 through 6) after three consecutive years of student growth data become available, in addition to the annual ratings received for each standard. Supervisors may continue to implement professional development plans (PDPs) for teachers and principals based upon annual feedback as needed during this time. The U.S. Department of Education approved an initial timeline in which school year 2012-13 was the first of the three years of data required for educators to receive an overall effectiveness rating; subsequently, school year 2014-2015 will be the first year for which such statuses will be provided.

¹CERE–NC is a partnership of the Carolina Institute for Public Policy at the University of North Carolina at Chapel Hill, the Friday Institute for Educational Innovation at North Carolina State University, and the SERVE Center at the University of North Carolina at Greensboro.

² TCP-C-006; <http://sbepolicy.dpi.state.nc.us/>

Measuring Student Growth

Expansion of Measurement Tools. With over 60% of the teacher workforce employed in grades and subjects currently without a statewide End-of-Grade (EOG) or End-of-Course (EOC) test (“Teacher Effectiveness and Support for Growth READY,” 2012), the state had to develop additional long-term, unified measures of student growth. These measures, hereafter referred to as Measures of Student Learning (MSLs), encompass measures of student growth in all subjects and grades, including previously untested subjects as well as EOC and EOG assessments. Implementation of newly developed MSLs will occur on a staggered schedule over the course of the grant period. The State developed Common Exams, the first of several planned Measures of Student Learning, with input from over 800 educators across the state for English Language Arts, science, social studies, mathematics in grades 4 through 12, and Career and Technical Education (CTE). NCDPI administered these Common Exams along with End-of-Course and End-of-Grade assessments in 2012-13.³

Education Value-Added Assessment System (EVAAS). In 2012, the NC State Board of Education selected the SAS Institute’s Education Value-Added Assessment System (EVAAS) to measure educators’ impact on student growth, with student growth defined as the change in student performance on a standardized test from one period of time to another.⁴ EVAAS uses a Multivariate Response Model (MRM) or Univariate Response Model (URM), as appropriate.⁵ In addition to calculating a teacher’s Standard 6 rating, the online EVAAS system includes an individualized dashboard for teachers to see evaluation ratings on Standards 1 through 5 integrated with Standard 6. For additional information, educators can access their self-assessments, principal ratings on Standards 1 through 5, and PDPs using the online North Carolina Educator Evaluation System, hosted by True North Logic. These online utilities provide the North Carolina Department of Public Instruction (NCDPI) and its partners access to valuable implementation and outcome data within Local Education Agencies (LEAs), across regions, and across different standards. The State intends to use these data to inform assessments of NCEES implementation.

Calculation of the 6th and 8th Standards. In 2011, the State proposed a calculation of teacher effectiveness that balanced EVAAS measures of individual teacher value-added with measures of school-wide growth,⁶ with the rationale that the inclusion of school-wide EVAAS estimates

³ NCDPI piloted the Common Exams in fall 2012 and made several substantive changes based on feedback, including shortening exam lengths and adding more specificity, structure, and examples to the scoring rubrics (Garland & Quick, 2013). All other Measures of Student Learning, including K-2 literacy, Career and Technical Education (CTE), Grade 3, and Analysis of Student Work assessments (Arts and Healthful Living) are either being piloted in the 2012-13 school year or are schedule to roll out in the 2013-14 school year.

⁴ As of the 2012-13 school year common assessments used to calculate a value-added score were End-of-Course and End-of-Grade Assessments.

⁵ The primary factor necessitating the need for the two different models is the test being used for the estimation of growth. In cases where the test uses the same scale from year to year, the MRM model is used. This model reports value-added in terms of gains, and works well with end-of-grade tests. Growth is determined by comparison to a population in a base year. The URM model is used when there is *not* an identical scale, as is the case for End-of-Course tests. In these cases, a prediction model is used which allows value-added to be reported as deviations from the prediction. Growth is determined by comparison to grade-level peers in the same year. For a more technical summary see: Wright, S.P., White, J. T., Sanders, W. L. & Rivers, J.C. (2010).SAS EVAAS statistical models. Retrieved from: <http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf>.

⁶ 70% individual teacher effectiveness; 30% school growth.

would encourage collaboration and collective ownership of overall outcomes.⁷ Standard 6 for teachers without individual EVAAS growth values would be based entirely on school-wide growth data. Analysis of this approach by NCDPI revealed that inclusion of school-wide growth unfairly lowered ratings of high-performing teachers who worked in low-performing schools and also raised the performance ratings of low-performing teachers who worked in high-performing schools (“NC Educator Effectiveness Policies and Update,” 2013).

As a result, in May 2013, the State Board of Education approved an amendment to the teacher evaluation policy (TCP-C-006) for 2012-13 that replaces the weighted teacher value-added rating for teachers of tested subjects with a rating derived entirely from student growth values attributable to the individual teacher.⁸ The amendment also includes incorporation of Common Exam data into calculations of principal effectiveness.

NCEES Professional Development

A significant implementation activity of this initiative involved developing and administering statewide professional development specific to the new NCEES. This included presentations at the State’s READY meetings and Summer Institutes, hosting webinars, and regional trainings on educator effectiveness held in spring 2012 for approximately 600 teachers and leaders. A separate evaluation of NCDPI’s statewide professional development efforts⁹ addresses issues of content and quality.

A final implementation milestone was the launch of Teacher Dashboards at the beginning of 2013. The dashboards are accessible via EVAAS and provide every North Carolina educator with a customized view of his or her ratings on Standards 1 through 6 of the NCEES. The dashboard also will allow educators to track their progress toward an overall educator effectiveness status. Professional development focused on this progress is currently implemented with the support of web-based EVAAS learning modules.

Evaluation Questions

The findings in this report address the following research questions:

1. How does the 6th standard¹⁰ correlate with the other five standards in the North Carolina Teacher Evaluation Process (NC TEP)?
 - 1.1 How does the 6th standard correlate with other measures of teacher and teaching effectiveness? (e.g., CLASS, Tripod, etc.)

⁷ Growth scores do not count toward an educator’s three years of data until it is attributable to a teacher’s specific content standards and students.

⁸ Estimates of teacher effectiveness for teachers without individual EVAAS growth values will remain the same; however, those estimates will not count toward overall educator effectiveness status.

⁹ <http://cerenc.org/rttt-evaluation/professional-development/>

¹⁰ The Evaluation Team collected and analyzed qualitative data for both teachers and principals regarding Standards 6 and 8, but for this report, the Team’s quantitative data analysis focuses only on the 6th Standard; future reports will incorporate analysis of quantitative data related to the North Carolina Principal Evaluation Process’s 8th standard.

- 1.2 How do the correlations between student growth and other teacher evaluation measures vary by subgroups? (e.g., beginning versus experienced teachers and principals; within high- vs. low-performing schools)
2. How are teachers and principals using EVAAS data for evaluation purposes and to inform teaching practices?
 - 2.1 How did school leaders use growth measures in teacher evaluation before implementation of the new standards?
3. What are teachers' and school leaders' perceptions around the use of growth data in the evaluation?

Summary of Findings

1. *How does the 6th standard correlate with the other five standards in the Teacher Evaluation Process?*

The EVAAS measures seem to provide an objective measure of teachers' contributions to student learning as indicated by correlations with the other five standards. The relationships between NC TEP ratings on the first five standards and Standard 6 are positive and significant and become larger as the NC TEP rating increases.

- 1.1 *How does the 6th standard correlate with other measures of teaching and teaching effectiveness?*

The EVAAS measures are significantly and positively related to teaching effectiveness measures provided by a piloted survey that measures student perceptions of their classroom environment—the Tripod Student Survey—but measures of teaching effectiveness using the CLASS observation tool are not related to teacher value-added as measured by EVAAS. Teachers' views of their own efficacy and of the degree to which they believe they have prepared their students are significantly and positively related to their value-added measure, while their sense of the fairness of the evaluation process has no statistical relationship with their value-added measure.

- 1.2 *How do the correlations between student growth and other teacher evaluation measures vary by subgroups? (e.g., beginning versus experienced teachers and principals; within high- vs. low-performing schools)*

Most teacher scores on Standards 1 through 5 were clustered around the “proficient” and “accomplished” categories, thus limiting the ability to provide detailed sub-group comparisons. However, males and minorities receive higher ratings on Standards 1 through 5 than would be consistent with their value-added scores. These biases persist in other measures of teacher effectiveness as well.

2. *How are teachers and principals using EVAAS data for evaluation purposes and to inform teaching practices?*

Educators' current use of EVAAS indicates limited but promising use of student growth data to inform instruction. Teachers and principals use a variety of assessments, including

EVAAS and the state's online ClassScape Reading 3D Assessment System software, to measure growth and to identify students for intervention. However, there is evidence to suggest that not all teachers and administrators are using student growth data to inform instruction at this time.

2.1 *How did school leaders use growth measures in teacher evaluation before implementation of the new standards?*

Prior to the 2012-13 school year, all principals in the evaluation sample reported having access to EVAAS data, with all but one reporting direct access. Principals suggested that growth measures were used to establish a dialogue with teachers about their instructional practices.

3. *What are teachers' and school leaders' perceptions around the use of growth data in the evaluation?*

Educators acknowledge that student growth is an important indicator of effectiveness, but their perspectives around the use of student growth data in evaluation reflects some confusion about the measures. Their misperceptions related to Standard 6 as a growth measure, coupled with uncertainties about the formulas used to calculate an effectiveness rating, raise concerns about the ability to effectively use student growth data to inform instruction. EVAAS values are not displayed like the ABC growth model, and principals cannot calculate teachers' scores. The data collection timeline did not allow the Evaluation Team to fully assess implementation and perceptions around Common Exams. As of fall 2012, teachers had limited knowledge about the Common Exams and expressed concerns about their ability to adapt their curriculum to the new Standard Course of Study and to adequately prepare their students for a new assessment.

Recommendations

- *Broaden communications strategies.* When rolling out new assessments that will contribute to Standards 6 and 8 ratings, clearly label those that are trials or pilot efforts and follow up with communications directly to teachers via email and other media that clearly communicate the purpose of such assessments. Furthermore, anticipate that some principals or LEA administrators may strategically withhold information from teachers or delay communication until they feel that they are sufficiently prepared to respond to questions and implement the reforms.
- *Expand training related to Standard 6.* Leverage LEA- and school-based staff, including identification of teachers at schools who can serve as resident trainers, to lead additional face-to-face trainings regarding variables that inform a Standard 6 rating, how Standard 6 reflects student growth, and how to use EVAAS data to inform instruction.¹¹ Also, expand promotion of webinars offered by the EVAAS vendor (SAS), and consider implementation of face-to-face trainings with additional vendors for teachers.
- *Continue to seek out teacher input.* Offer additional opportunities for teachers to provide feedback regarding the administration of MSLs. Many teachers will have more experience

¹¹ Since fall 2012 data collection, NCDPI has responded to many information requests related to the NCEES.

with the Common Core and Essential Standards following the 2012-13 administration of MSLs and Common Exams, and thus will have the opportunity to focus the alignment of these items to their curriculum. Furthermore, opportunities for feedback may help teachers to develop greater ownership of, and therefore buy-in for, the assessments.

- *Consider revision to the NC TEP ratings and evaluation system.* The analysis demonstrates that a majority of teachers (89%-91%) were rated as “proficient” or “accomplished” in each of the five Standards. These findings are consistent with research that has found that subjective measures of teacher performance may be upwardly biased or benchmarked to minimum requirements when they are used in summative evaluation (Weisburg et al., 2009). As a result of this possible upward bias, the realized measurement scale of Standards 1 through 5 (with most ratings at “proficient” or “accomplished”) may limit the potential of the evaluation system to provide a full range of measurement and subsequent formative assessment and feedback. The expansion of the scale above the “proficient” benchmark (e.g., through the inclusion of an additional rating level) may afford more differentiation in teacher effectiveness ratings. However, it is important to note that this likely will not eliminate entirely the tendency of evaluators to benchmark their teacher ratings, nor will it eliminate entirely individual rater bias. Accordingly, the expansion of the scales also should be accompanied by evaluator training.

Contact Information:
Please direct all inquiries to Julie Marks
Julie_Marks@unc.edu

© 2013 Consortium for Educational Research and Evaluation–North Carolina



Carolina Institute
for Public Policy



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

