Consortium for Educational Research and Evaluation– North Carolina

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

Table of Contents

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 policy[2] 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.

[1]CERE–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 efforts[9] 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 standard[10] 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.

Introduction

Research shows that the influence of a teacher is one of the most important and measurable influences on student academic performance (Darling-Hammond, 2000; Rivkin, Hanushek, and Kain, 2005; Rowan, Correnti, and Miller, 2002; Sanders and Rivers, 1996). Consequently, strengthening the education workforce is North Carolina's highest reform priority and a core pillar of its federally funded Race to the Top (RttT) plan. With an ultimate goal of high student achievement statewide, a wide-reaching focus of the NC RttT initiatives is to increase students' access to great teachers in every classroom and great leaders in every school.

Informed by North Carolina's rich network of administrative data, surveys, and partnerships, the state identified high-priority actions intended to develop strong teachers and effective principals ("North Carolina Race to the Top Proposal," 2010). Among those is the critical ability to evaluate educator effectiveness objectively, and to include within that evaluation accountability for the growth of students. North Carolina's RttT proposal included a three-part Educator Evaluation Plan comprised of:

1. Adding a requirement for an explicit student growth component to the Educator Evaluation process;
2. Fully implementing this new Educator Evaluation process for both teachers and principals statewide; and
3. Developing a uniform system for integrating student growth data into evaluations for all teachers and principals, based on the input and guidance of key constituents and stakeholders.

To meet these goals, the North Carolina Department of Public instruction (NCDPI) proposed to implement the following activities over the course of the four-year RttT grant period (2010-2014):

- Add a 6th standard to the Teacher Evaluation Process and an 8th standard to the Principal Evaluation Process that are based on measures of student growth;
- Provide professional development on the existing evaluation tool and on the changes to the evaluation components and process;
- Provide targeted professional development based on level of performance as measured by North Carolina Educator Evaluation System (NCEES);
- Establish teacher design groups to create assessments (pre- and post-) for untested courses;
- Engage accountability staff to test reliability and validity of assessments recommended by the workgroups;
- Develop and validate a student academic growth factor; and
- Develop longer-term systems for integrating student growth data into teacher and principal evaluations for all teachers and principals (including growth measured through the new assessments developed for untested courses).

The logic model in Appendix A provides additional details about the framework and sequence guiding this approach.

Brief Overview of the North Carolina Educator Evaluation System (NCEES)

The NCEES includes both the North Carolina Teacher Evaluation Process (NC TEP) and the Principal Evaluation Process (NC PEP). Prior to RttT funding, the NC TEP assessed teacher performance based on the North Carolina Professional Teaching Standards and was guided by an observation rubric and artifact review.[12] The original Professional Teaching Standards are as follows:

1. Teachers demonstrate leadership;
2. Teachers establish a respectful environment for a diverse population of students;
3. Teachers know the content they teach;
4. Teachers facilitate learning for their students; and
5. Teachers reflect on their practice.

In February 2012, the State Board of Education, in alignment with North Carolina's RttT plan, adopted a 6th standard that explicitly requires the use of student growth data as a measure of teacher performance:

6. Teachers contribute to the academic success of students.

Similar to the NC TEP, the standards for principals address multiple leadership attributes (strategic, instructional, cultural, human resources, managerial, external development, and micro-political). In addition to individual evaluations, the NC PEP includes data from the Teacher Working Conditions Survey to help principals prioritize areas of focus in order to improve the teaching and learning conditions of their schools. For principals, measures of student growth are added as an 8th standard. The North Carolina Standards for School Executives are as follows:

1. Strategic leadership;
2. Instructional leadership;
3. Cultural leadership;
4. Human resource leadership;
5. Managerial leadership;
6. External development leadership;
7. Micropolitical leadership; and
8. Academic achievement leadership.

[12] See http://www.ncpublicschools.org/docs/effectiveness-model/ncees/instruments/teach-eval-manual.pdf for the detailed rubrics used in the NC TEP ratings.

On Standards 1 through 5 of the NC TEP and Standards 1 through 7 of the NC PEP, an educator's observed level of competence for each standard is rated using a 5-point scale of *Not Demonstrated, Developing, Proficient, Accomplished,* or *Distinguished*. Combined with ratings on Standards 6 and 8 of the NC TEP and NC PEP, respectively, these ratings determine an educator's overall status, which falls into one of three categories: *Highly Effective, Effective,* or *In Need of Improvement.* For example, to earn a rating of "Effective," educators must meet expectations for each instructional practice standard, as well as for the student growth standard. "Highly Effective" educators must exceed expectations for each standard. All educators rated "In Need of Improvement" must complete a professional development plan. An educator receives an overall effectiveness status only once she or he has accrued three years of data for Standards 6 or 8 ("NC Educator Effectiveness Policies and Update," 2013*).*

Purpose of the Evaluation

The Consortium for Educational Research and Evaluation–North Carolina (CERE–NC)[13] is conducting an independent external evaluation of North Carolina's RttT initiatives. The roles of the RttT Evaluation Team are to (1) document the activities of the RttT initiatives; (2) provide timely, formative data, analyses, and recommendations to help the initiative teams improve their ongoing work; and (3) provide summative evaluation results toward the end of the grant period to determine whether the RttT initiatives met their goals and to inform future policy and program decisions to sustain, modify, or discontinue initiatives after the grant-funded period.

The state's RttT grant identified improving student learning by "enabling and ensuring great teaching" as a key goal of the initiatives. To help achieve this goal, Standards 6 and 8 are intended to establish measures that can identify effective teachers and leaders and ultimately contribute to a meaningful evaluation system that informs their professional development and growth.

The overriding evaluation goals for this initiative are to:

1. Ensure quality, consistency, and fairness of new and ongoing teacher and principal evaluation processes through examination of the validity and reliability of measures of student learning gains using multiple measures of teacher and teaching effectiveness; and
2. Examine educators' perspectives on new evaluation standards and the effect of these standards on educators' practices and attitudes, and compare practices and outcomes of the NCEES under the new system with those under the old system.

The evaluation will examine the implementation of the new evaluation standards to determine if the addition of student growth measures provides independent information that allows for meaningful assessment of teachers' and administrators' effectiveness. The evaluation has four primary foci: (1) the correlation of student growth measures with other measures of teacher quality, including the other standards in the Educator Evaluation Process; (2) the correlation

[13] CERE–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.

between Common Exams and other student assessments;[14] (3) educators' use of student growth data (past and present); and (4) educators' perceptions about the use of student growth data in evaluation (including the roll-out of the new Common Exams).

The overall evaluation approach is mixed-method. The Evaluation Team is conducting quantitative analyses that evaluate the relationship between different measures of teacher performance, how these vary by subgroups, and the characteristics of new student growth measures as provided by the Education Value-Added Assessment System (EVAAS; see *Initiative Implementation Activities to Date*, below, for description). The Team will conduct qualitative analyses, including independent teacher observations and teacher and principal interviews, to assess the impact of the new evaluation process on educators' attitudes and practices.

Purpose and Structure of this Report

The purpose of this report is to explore the relationship between the new standards and established standards as well as other teacher effectiveness indicators, and to assess the knowledge and perceptions of teachers and administrators in regard to the new evaluation system during initial stages of implementation. In doing so, the Evaluation Team will discuss the positive benefits of this initiative to date and the challenges to its implementation. Identifying potential barriers enables the Evaluation Team to provide recommendations for improving implementation.

The findings in this report address the following research questions[15] (see Appendix A for the full evaluation Scope of Work,[16] which includes additional research questions to be addressed in subsequent reports):

1. How does the 6th standard 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.)
   - 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?

[14] *Note*: The various Common Exams either are currently under development or were administered for the first time in school year 2012-13 and, therefore correlations are not included in this evaluation report. The correlation between Common Exams and other student assessments will be included in the final evaluation report.
[15] 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 NC PEP's 8th Standard.
[16] The current Scope of Work was approved October 2012. It was amended from the original scope (May 2011) to include a more mixed-method approach to best assess educators' perspectives and issues with implementation.

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

This report provides initial analysis of the association between alternate measures of teacher effectiveness currently being implemented. It also assesses educators' knowledge and perceptions of the new evaluation standards and Common Exam as of fall 2012.

The report begins with an overview of the policy initiatives currently in place and presents implementation activities to date. The report then uses quantitative data in the form of EVAAS scores and other teacher effectiveness measures to determine the extent to which EVAAS scores are predicted by these alternate measures. The report also uses qualitative data, obtained from teacher and principal interviews, to address current knowledge and perceptions of the new standards.

This report provides formative feedback for the initial phases of a multi-year initiative in order to inform opportunities for improvement; however, the timing of data collection for the report limited the Team's ability to assess the first full implementation of the Common Exams (assessments for grades 4 through 12 in English Language Arts, science, social studies, and mathematics for grades and courses currently without an End-of-Grade [EOG] or End-of-Course [EOC] test). Therefore, the report does not present conclusive statements related to the Common Exams.

Initiative Implementation Activities to Date

Addition of a Student Growth Measure to the NC Educator Evaluation Process

In 2008-09, NCDPI began a statewide roll-out of the NCEES for teachers and principals. During that school year, superintendents evaluated all principals using the NC PEP. The NC TEP was first used in 13 LEAs in 2008-09, and all LEAs participated by 2010-11. Every LEA in North Carolina currently implements the NCEES.

With funds from the NC RttT grant, NCDPI expanded its existing evaluation system to explicitly include student growth data. Before RttT, the system was comprised of five performance standards for teachers and seven for administrators. In 2012, the State Board of Education adopted the addition of student growth standards (Standard 6 for teachers and Standard 8 for principals) as a formal requirement for each teacher's and principal's evaluation. 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 receive equal weighting with 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 policy 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.

Educator 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, where student growth is defined as the change in student performance on a standardized test from one period to another.[17] EVAAS uses a Multivariate Response Model (MRM) or Univariate Response Model (URM), as appropriate.[18] This online utility provides NCDPI and its partners access to valuable implementation and outcome data regarding specific courses and grades[19] that receive common assessments. 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.

Measuring Student Growth

Beginning in the 2012-13 school year, NCDPI began adoption of a uniform, statewide set of acceptable measures of pre-approved student growth data. With over 60% percent of the teacher workforce employed in grades and subjects currently without a statewide EOG or EOC test ("Teacher Effectiveness and Support for Growth READY," 2012), the state had to develop additional long-term, unified measures of student growth. The state developed Common Exams (the first of several planned Measures of Student Learning [MSLs]) with input from over 800 educators across the state during the 2011-12 school year. These educators designed item specifications, reviewed items provided by an external vendor, and provided guidance on scoring rubrics and administration procedures. Their recommendations were then provided to the Center for Urban Affairs Technical Outreach for Public Schools (TOPS) at North Carolina State University in order to generate the assessment items.

[17] 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.
[18] 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.
[19] 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.

In the 2012-13 school year, NCDPI administered Common Exams, EOGs, EOCs, and Career and Technical Education Post-Assessments (as well as Pre-Assessments in some courses). Common Exams are MSLs for core subject areas (English Language Arts [ELA], science, social studies, and mathematics) in grades 4 through 12 that are not currently tested using EOCs or EOGs. All other MSLs, including K-2 literacy, Grade 3, and the Analysis of Student Work process (for the Arts, Healthful Living, World Languages, and other content areas)) are either being piloted in the 2012-13 school year or are scheduled to roll out in the 2013-14 school year.

NCDPI piloted the Common Exams in fall 2012 in 39 Local Education Agencies (LEAs) and one regional school, for a total of 86,546 completed exams. NCDPI elicited feedback from this initial implementation and responded with several substantive changes, including shortening exam lengths and adding more specificity, structure, and examples to the scoring rubrics (Garland & Quick, 2013).

Calculation of the 6th and 8th Standards

The relative weighting between teacher- and school-level EVAAS scores has evolved since initial implementation. In 2011, the state proposed a calculation of teacher effectiveness that balanced EVAAS measures of individual teacher effectiveness (70%) with measures of school-wide growth (30%), with the rationale that the inclusion of school-wide EVAAS estimates would encourage collaboration and collective ownership of overall outcomes. 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 a change to the teacher evaluation policy (TCP-C-006) for 2012-13 that replaces the weighted teacher effectiveness rating for teachers of tested subjects with a rating derived entirely from student growth values attributable to the individual teacher. 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. 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 efforts[20] addresses issues of content and quality.

[20] http://cerenc.org/rttt-evaluation/professional-development/

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 allows 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.

Data Sources and Methods

The evaluation questions for this initiative require both quantitative and qualitative data sources and analysis. In addition to the EVAAS and McREL data provided by NCDPI, this report also includes data from multiple surveys, data collected from independent classroom observations, as well as data gathered through teacher and principal interviews. These various data sources allow this report to address the more technical aspects and relationships of Standard 6 to other measures of teacher effectiveness, as well as its implementation and educator perceptions of its use. The various data sources and their level of collection are summarized in Table 1; Appendix A includes a more detailed summary of the research questions and data sources.

Table 1. Summary of Data Sources and Level of Data

Data Source	Teachers	Principals
Interview Data: Semi-structured interview data collected at up to six time-points over three years.	✓	✓
CLASS Observation Protocol: Independent observations of teaching quality at up to six time-points over three years	✓	
Tripod Student Survey: Student survey data collected during the 2011-12 academic year in pilot schools and, if continued, in future years as well.	✓	
Race to the Top Omnibus Teacher and Principal Survey: Teacher and principal survey collected four times in four years, including two pre-implementation years.	✓	✓
NCDPI observation rubric (NC TEP): Principal observations of teaching quality.	✓	
Administrative Data: Administrative data on students and teachers including EVAAS scores, McRel ratings, teacher mobility and retention, and teacher characteristics.	✓	✓

Sample Selection

Before discussing the particular data sources, as well as their collection and analysis, it is important to discuss the sample selection criteria for identifying the participating teachers.[21] The procedure for generating the sample was an iterative process that began in the 2011-12 school year. The first step was to generate a sample of schools for the Omnibus Survey. The sample of schools was selected using stratified random sampling. The strata consisted of, in order of priority: schools in Turnaround/District and School Transformation; STEM schools; rural schools; and low-performing but non-turnaround schools. With the exception of STEM schools (high school only) the sample pulled from all three levels of schooling. A fifth "general" stratum contained all remaining schools; charter schools were not included in the sample. The final count of schools included in the 2011-12 launch of the Omnibus Survey was 344. In the second step, a subsample of 100 schools was randomly selected from the stratified random sample. The distribution of schools in this sample reflected the distribution of the Omnibus Survey sample.

[21] This sample is used to provide data for CLASS observations, interviews, and Omnibus survey responses.

From that sample, a third step consisted of principals selecting two teachers from each school who met certain selection criteria (EOC/EOG teachers with at least three years of teaching experience). This step generated a sample of 200 teachers who taught subjects for which there was an EOC/EOG and who would then be observed using the CLASS observation protocol in the 2011-12 school year. Note that, at the time of this sample selection (2011-12), Common Exams were not yet being administered; while approximately 80 of the original 200 teachers selected taught in content areas that would receive both EOC/EOGs and Common Exams in 2012-13, the sample did not include teachers who taught in subject areas that would receive only Common Exams in 2012-13. Therefore, to properly evaluate research questions 4 and 5, the Evaluation Team conducted a final step in the sampling process. This final step involved a purposeful selection of a subset of schools from the current subsample of 100 schools where Common Exams will be administered in the 2012-13 school year. Teachers who taught Common Exam-only courses were identified and then randomly selected to replace approximately 35 teachers from the first wave of EOC/EOG teachers for whom data were incomplete in 2011-12. This step created the final sample of 100 principals and 200 teachers that was used for the 2012-13 CLASS observation and interview data collection.

It is important to note that this sample is primarily meant to follow teachers as opposed to schools. Accordingly, teachers who no longer teach at their original schools but who otherwise still teach in the North Carolina public education system will remain in the sample.[22]

EVAAS

The main component of EVAAS used for this report is the teacher value-added estimate calculated using either the URM or MRM approach.[23] Estimates for school years 2008-09 through 2010-11 were calculated by SAS, and the Evaluation Team obtained them from NCDPI. These estimates of the 6th Standard for teachers who teach EOC or EOG classes are used to estimate the relationships between the 6th Standard and the other five standards in the North Carolina Teacher Evaluation Process (NC TEP) for 2011-12. The 6th Standard also is used in analyses that determine the extent to which it correlates with other measures of teacher and teaching effectiveness, as well as to what extent those correlations vary by subgroup.

North Carolina Teacher Evaluation Process (NC TEP)

The North Carolina Teacher Evaluation Process supports teachers in their professional growth and assesses teacher performance in relationship to the North Carolina Professional Teaching Standards. These standards include: (1) Teachers demonstrate leadership, (2) Teachers establish a respectful environment for a diverse population of students, (3) Teachers know the content they teach, (4) Teachers facilitate learning for their students, and (5) Teachers reflect on their practice. School administrators lead the four step evaluation process with their teachers: (1) Training and orientation, (2) Self-assessment, goal setting and pre-conference, (3) Observation cycle, and (4) Summary evaluation and goal setting. Formal observations last for at least 45 minutes or for an

[22] Twelve teachers in the 2011-12 TLEE sample moved to different schools in 2012-13 and were followed for observation.
[23] For a detailed explanation of these two approaches, please 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.

entire class period. All probationary teachers receive three formal observations per year by their principal and a fourth formal observation by a peer. Career status teachers are evaluated annually by their principal and receive three observations, one of which is a formal observation, during their renewal years. In 2011-12, career status teachers who were not participating in a summative evaluation could be assessed using a reduced rubric that included only Standards 1, 4, and 6, but only teachers who received ratings on all five standards are included in this evaluation report[24]. On each standard, teachers can be rated as Not Demonstrated, Developing, Proficient, Accomplished, or Distinguished. The NC TEP data used for this report were obtained from NCDPI as reported by school administrators and are used to assess Research Question 1.

Classroom Observations

The Evaluation Team conducted independent classroom observations of approximately 200 teachers in approximately 100 schools during the 2011-12 and 2012-13 school years to assess longitudinal changes in instructional practices related to the implementation of the 6th Standard. Data for this report reflect approximately three observations per teacher: fall 2011, spring 2012, and fall 2012[25]. The Team used Teachstone's Classroom Assessment Scoring System™ (CLASS™), which assesses teacher and student interactions that impact learning on a 7-point Likert scale across three major dimensions (Pianta, *et al*., 2008): Emotional Support, Classroom Support, and Organization (Appendix B). These dimensions provided the Team with additional measures of teachers and effective teaching that could be used to address Research Questions 1.1 and 1.2.

In order to receive certification to conduct observations using this instrument, all observers were required to participate in a two-day regional training session provided by Teachstone, Inc. and pass an online reliability test with 80% accuracy. Observers also participated in observation refresher sessions in fall 2012.

Interviews

In order to understand how student growth data inform teaching practices, as well as to assess perspectives about its use in evaluation, the Evaluation Team conducted interviews with school administrators and teachers. In 2011-12, the Evaluation Team interviewed 50 teachers who received an observation[26], and 22 of their principals. This sample was selected purposefully from the larger observation sample in order to capture the perspectives of teachers who administered various types of MSLs (EOGs, EOCs, and Common Exams) and who represent a variety of school contexts (schools with high average socioeconomic [SES] indicators, schools with low SES indicators, various performance rating levels, etc.). Participants were asked questions regarding their knowledge about Standard 6 and Standard 8. However, since the roll-out of Standard 6 did not officially begin until fall 2012, findings from these interviews were limited. As a result, the interview protocol was revised for 2012-13 data collection to more accurately

[24] Although this approach reduces the sample size from 29,227 to 10,616, the omission of partial evaluations using only Standards 1, 4, and 6, reduces the potential for those ratings to be influenced by raters' possible attempt to account for the non-rated Standards.
[25] Some teachers were not observed for three observations due to a staggered roll-out of observations in 2011-2012, teacher availability and/or attrition, or a 2012-13 resampling to include Common Exam-only teachers.
[26] For convenience, interviews and observations were conducted on the same date.

reflect changes in the ways in which NCDPI disseminated information about Standard 6 and Standard 8 and initial implementation of Measures of Student Learning (more specifically, the roll-out of the Common Exams). Since previous interview data could not be reconciled with the current Scope of Work and provided little insight into the revised evaluation questions, only the 2012-13 interview data informs this report.

The 2012-13 interview protocol followed a semi-structured format (Appendix C). The first half of the interview consisted of close-ended questions designed to gauge participants' prior and current use of EVAAS data, as well as their understanding of the new Common Exams. The second half of the interview consisted of open-ended questions that assessed participants' prior and current use of EVAAS data to inform instruction, knowledge about the implementation of the Common Exams, and perceptions of the use of student growth data in educator evaluation.

Tripod Student Survey (Student Perceptions Survey)

Previous research suggests that teaching effectiveness is multi-dimensional and that under appropriate conditions, student evaluations of teachers can provide measurement of these dimensions that are both reliable and stable (Marsh and Roche, 1997; Ferguson, 2008). This evaluation includes estimates of these additional measures of teacher effectiveness that are derived from the Tripod Student Survey (Appendix D). This survey was administered in 2011-12 to pilot schools in forty-one North Carolina LEAs. The primary intent of the Tripod Student Survey is to capture measures of teacher effectiveness based on the student experience in the classroom. The survey includes questions that define this experience that use what the survey refers to as the Seven C's. They include: Caring about students, Controlling behavior, Clarifying lessons, Challenging students, Captivating students, Conferring with students, and Consolidating knowledge. The student responses were obtained from NCDPI and are used to address Research Questions 1.1 and 1.2.

Omnibus Survey

The Evaluation Team's Omnibus Survey is based on a conceptual framework for effective schools. It was created to provide measures of key dimensions important for effective learning that could be used by several RttT evaluations. The survey is administered to all teachers and principals in a sample of selected schools. This report uses responses from teachers in the 100 schools in which the Team conducted observations in 2011-12. This report contains the responses from teachers to questions in the following survey dimensions: Evaluation Validity, Test Preparation, and Self-Efficacy (Appendix E). The Team selected these dimensions as the factors that might potentially relate to teacher value-added measures generated by EVAAS and could most help address Research Questions 1.1 and 1.2.

Findings

This section first provides a quantitative analysis of the extent to which alternate measures of teacher effectiveness predict EVAAS scores and the extent to which these predictions are related to other characteristics. A qualitative analysis of educators' reported use and perceptions of the new evaluation standards follows.

Quantitative Analysis: Comparing Teacher Value-Added Scores with other Measures of Teaching Quality

The Research Questions that guide this portion of the evaluation are:

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

This section of the report describes the relationships between the EVAAS scores (based on 2011-12 EOG and EOC test data) used for the 6th and 8th Standard ratings and other measures of teaching effectiveness, including: principals' ratings of the other five standards in the NC TEP; Evaluation Team observations of teachers (CLASS); student surveys (Tripod); and teacher survey responses to RttT Omnibus Survey constructs that have been shown to relate to effective teaching (constructs related to preparing students for the reading and mathematics objectives of the North Carolina Standard Course of Study and to teachers' beliefs in their own effectiveness).

The sections that follow address Research Questions 1 and 1.1; the report folds sub-group analyses (Question 1.2) into each section. To use data for the largest possible sample of teachers from 2011-12, each of the questions is addressed using different sub-samples. Each of the four sub-samples contain all of the teachers for whom the particular measure was available and who had a valid EVAAS score as described in the section on data sources. Table 2 (following page) provides summary statistics across the different data sources.

In evaluating Standard 6, the Team uses two criteria. First, because effective teaching is multi-dimensional, the new measure for Standard 6—EVAAS scores—should be correlated with other measures of teaching effectiveness. However, it should not be so highly correlated with other standards that the information could be considered redundant. In other words, Standard 6 should provide a summative measure that is related to the measures of various dimensions of effective teaching.

Table 2. Teacher Characteristics by Data Source

	EVAAS[27]		NC TEP[28]		TRIPOD		CLASS[29]		Omnibus	
Experience										
< 3 years	7,711	16.5%	1,686	15.9%	206	14.8%	6	7.9%	93	11.3%
3-5 years	7,695	16.5%	1,778	16.8%	204	14.7%	16	21.1%	126	15.3%
6-10 years	9,697	20.8%	2,203	20.8%	354	25.5%	16	21.1%	175	21.2%
11-25 years	15,380	32.9%	2,926	27.6%	516	37.2%	34	44.7%	260	31.5%
>25 years	3,851	8.2%	829	7.8%	108	7.8%	4	5.3%	90	10.9%
(Not Reported)	2,396	5.1%	1,194	11.3%	0	0.0%	0	0.0%	81	9.8%
Level										
Elementary	17,392	37.2%	—	—	484	34.9%	91	45.1%	118	14.3%
Middle	14,550	31.1%	—	—	—	—	88	43.6%	104	12.6%
Secondary	14,788	31.7%	—	—	904	65.1%	23	11.4%	603	73.1%
(Not Reported)	0	0.0%	—	—	0	0.0%	0	0.0%	—	—
Subject										
English-Language Arts	15,004	32.1%	—	—	413	29.8%	11	14.5%	85	10.3%
Mathematics	17,624	37.7%	—	—	431	31.2%	19	25.0%	124	15.0%
Science	10,500	22.5%	—	—	524	37.8%	46	60.5%	251	30.4%
Other/Contained	3,602	7.7%	—	—	—	—	—	—	365	44.2%
(Not Reported)	0	0.0%	—	—	20	1.4%	0	0.0%	0	0.0%
Gender										
Female	36,108	77.3%	7,564	71.3%	1,061	76.4%	63	82.3%	613	74.3%
Male	9,483	20.3%	2,690	25.3%	298	21.5%	13	17.1%	191	23.2%
(Not Reported)	1,139	2.4%	362	3.4%	29	2.1%	0	0.0%	21	2.3%
Race										
Black	5,905	12.6%	1,610	15.2%	207	14.9%	17	22.4%	168	20.4%
White	37,668	80.6%	8,152	76.8%	1,093	78.8%	57	75.0%	600	72.7%
Other Race	1,516	3.2%	409	39%	45	3.2%	2	2.6%	32	3.9%
(Not Reported)	1,641	3.5%	445	4.2%	43	3.1%	0	0.0%	25	3.0%
Total	46,730		10,616		1,388		76		825	

Second, concerns might be raised if any sub-groups of teachers seem to have systematically different rating on the 6th standard than other standards or other quality measures. This outcome could indicate that one of the measures is biased with respect to a certain sub-group; however, it is important to bear in mind that which of the two measures is biased cannot be definitively determined from this analysis. To assess these relationships, the Team uses an Ordinary Least Squares (OLS) approach. This empirical method allows for analysis of the relationships between

[27] Characteristics reported for EVAAS scores are by number of observations, since teachers may have multiple observations, depending on subjects taught and tests given.
[28] The NC TEP data provided do not allow the determination of the level or subject taught. Because some teachers have more than one EVAAS score, the Evaluation Team could not definitively establish the context of the observation and instead used a composite of the teacher's value-added score.
[29] The number of teachers who received a CLASS rating and a valid EVAAS score is lower than expected because of the reduction of End-of-Course tests pursuant to G.S. 115C-174.11. The original evaluation sample was drawn prior to the test reduction for the 2011-12 school year.

Standard 6 and other measures of teacher effectiveness while controlling for other factors including nesting of evaluations.

Comparing EVAAS Value-Added Scores with Principal Ratings of Teachers' Performance in the North Carolina Educator Evaluation System

This section reports the associations between the EVAAS teacher value-added score used for Standard 6 and the principal ratings on NC TEP teacher evaluation scores for the sample of teachers that have both EVAAS scores and principal ratings in 2011-12. The sample consists of 10,616 teachers, roughly one-tenth of the population of classroom teachers in North Carolina public schools.

Table 3 demonstrates that in each of the five other standards, teachers with higher ratings from their principals also have higher average teacher value-added scores. Also, more experienced teachers tend to receive higher ratings in all categories. Males are relatively under-represented—in terms of their population percentages—in the *Accomplished* and *Distinguished* categories in all five standards. White teachers are over-represented in the *Accomplished* and *Distinguished* categories for all five standards.

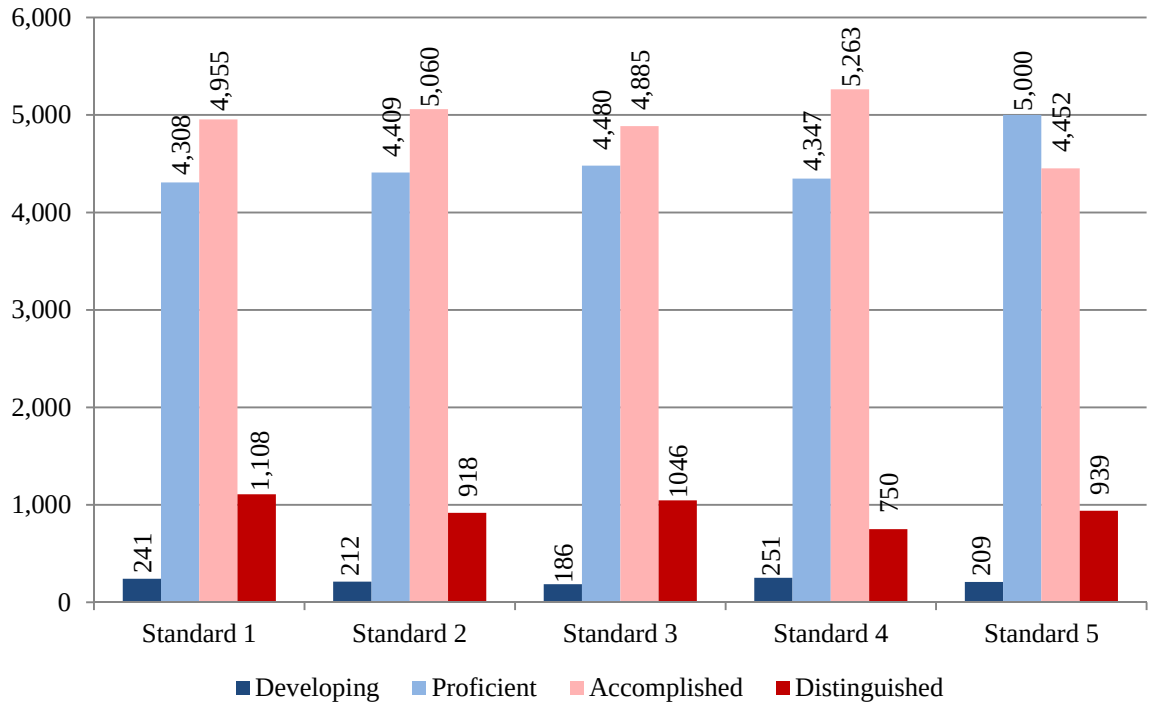
Table 3. Teacher Characteristics by NC TEP Ratings[30]

	Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard 1	4	241	4,308	4,955	1,108
White	75.0%	71.5%	78.2%	81.0%	85.1%
Male	25.0%	29.5%	30.3%	24.5%	18.4%
Avg. Experience	5.50	6.98	8.59	11.74	14.92
EVAAS	-2.94	-2.06	-0.53	0.31	0.82
Standard 2	17	212	4,409	5,060	918
White	88.2%	78.0%	77.9%	81.2%	84.8%
Male	35.3%	30.6%	30.9%	23.9%	17.2%
Avg. Experience	17.71	8.11	8.76	11.61	15.16
EVAAS	-0.10	-2.28	-0.50	0.35	0.67
Standard 3	19	186	4,480	4,885	1,046
White	89.5%	76.7%	77.2%	81.4%	86.5%
Male	26.3%	29.3%	29.3%	24.7%	20.2%
Avg. Experience	19.84	6.72	8.40	11.89	15.58
EVAAS	0.53	-1.99	-0.56	0.33	0.92
Standard 4	5	251	4,347	5,263	750
White	50.0%	74.6%	76.1%	82.6%	86.8%
Male	25.0%	31.1%	30.6%	23.7%	18.5%
Avg. Experience	9.40	7.41	8.83	11.64	15.94
EVAAS	-3.80	-1.98	-0.61	0.41	0.87
Standard 5	16	209	5,000	4,452	939
White	100.0%	73.5%	78.9%	80.5%	85.7%
Male	31.3%	37.0%	30.3%	23.1%	18.0%
Avg. Experience	17.06	6.35	9.04	11.89	14.42
EVAAS	0.48	-1.86	-0.43	0.35	0.71

[30] Evaluation standards defined as: 1) Teachers demonstrate leadership, 2) Teachers establish a respectful environment for a diverse population of students, 3) Teachers know the content they teach, 4) Teachers facilitate learning for their students, and 5) Teachers reflect on their practice.

One noticeable pattern across standards is the low number of teachers receiving a *Developing* rating, and a high number of teachers receiving a *Proficient* or *Accomplished* rating. Figure 1 presents a graphical illustration of the number of teachers within each rating category for each of the five standards. While Table 3 demonstrates expected patterns—most notably, the increase of mean EVAAS scores as the ratings scales increase—Figure 1 shows that most teachers are rated as either *Proficient* or *Accomplished*. This finding is consistent with research that suggests that administrators' ratings of teachers may be subject to potential positive and/or clustering bias (Popham, 1988; Danielson and McGreal, 2005).

Figure 1. Number of Teachers in Each Rating Category by Standard[31]



The Evaluation Team examined the correlations between the teacher value-added score and the other five standards by first looking at the extent to which being rated *Accomplished* or *Distinguished* on each of the five standards predicts teachers' value-added score. Table 4 (following page) summarizes these results. The rating for *Facilitate learning (Standard 4)* best predicts teachers' value-added scores. The other two ratings that are significantly related to teachers' value-added scores, in order of importance, are *Know the content they teach (Standard 3)* and *Demonstrate leadership (Standard 1)*. Neither the rating for *Establish a respectful environment for a diverse population of students (Standard 2)* nor for *Reflect on their practice (Standard 5)* significantly relates to teachers' value-added scores.

In addition, teachers with 11 to 25 years of experience who have ratings on Standards 1 through 5 exhibit significantly lower teacher value-added ratings compared to the reference group of teachers with 6 to 10 years of experience. Male teachers with ratings for Standards 1 through 5

[31] The *Not Demonstrated* category is omitted from the figure because of its infrequent use. See Table 3.

that are similar to those for female teachers also tend to receive lower teacher value-added scores. Similarly, Black teachers with similar ratings from principals receive lower teacher value-added scores than do White teachers. These findings suggest that male teachers, Black teachers, and teachers with 11 to 25 years of experience receive relatively higher principals' ratings than value-added scores.

Table 4: EVAAS and NCTEP Rating (Accomplished or Distinguished Indicator), OLS Estimates

Variable	Coefficient (Std. Err.)
NCTEP Indicator	
Demonstrate Leadership	**0.26**** (0.08)
Respectful Environment	0.15 (0.08)
Content Knowledge	**0.32***** (0.08)
Facilitate Learning	**0.57***** (0.08)
Reflect on Practice	0.00 (0.08)
Teacher Experience	
< 3yrs	-0.03 (0.09)
3-5 years	0.12 (0.08)
11-25 years	**-0.18*** (0.07)
>25 years	-0.14 (0.11)
Controls	
Male	**-0.41***** (0.06)
Black	**-0.44***** (0.07)
Other Race	-0.02 (0.13)
Constant	**-0.51***** (0.07)
n	10,616
R^2	0.06

Note: Reference groups are: ***NC TEP Indicator***—developing rating level; ***Teacher Experience***— 6-10 years' experience; ***Controls***— female and White.
Note: * indicates statistically significant at the p<0.05 level; ** indicates statistically significant differences at the p<0.01 level; *** indicates statistically significant differences at the p<0.001 level.

In addition to assessing the relationship between being rated as either *Accomplished* or *Distinguished* on each standard, the Evaluation Team also examined the extent to which differences in a teacher's value-added score relate to her or his rating in each of the original five standards. Table 5 (following page) presents the results of the five models. Taking the results in total, there are several outcomes worthy of note. First, the *Not Demonstrated* rating is

inconsistently related to estimates of teacher value-added, though the significance of this relationship is questionable. One potential explanation is that administrators may be using the *Not Demonstrated* category differently. For example, some administrators may use this categorization to denote a complete absence of the behavior, while others may use it to denote the behavior was not looked for. Additionally, as Table 3 suggests, the use of the *Not Demonstrated* category is infrequent.

Second, recalling that the comparison group is teachers with a *Developing* rating, all remaining teacher ratings on all five standards are significantly and positively related to the teacher value-added measure as estimated by EVAAS.

Finally, the trend of the relationship between teachers' ratings on a given standard and their value-added scores is positive and significant for all standards. This suggests that higher ratings within a given standard are associated with higher value-added scores.

Table 5: EVAAS and NC TEP Standards, OLS Estimates

	Standard 1	**Standard 2**	**Standard 3**	**Standard 4**	**Standard 5**
Rating	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)
Not Demonstrated	-1.07 (1.32)	**2.04***** (0..66)	**2.37***** (0.63)	-1.71 (1.17)	**2.02***** (0.68)
Proficient	**1.41***** (0.17)	**1.65***** (0.18)	**1.29***** (0.20)	**1.27***** (0.17)	**1.20***** (0.19)
Accomplished	**2.18***** (0.18)	**2.42***** (0.19)	**2.12***** (0.20)	**2.21***** (0.17)	**1.90***** (0.19)
Distinguished	**2.67***** (0.19)	**2.71***** (0.20)	**2.71***** (0.21)	**2.66***** (0.19)	**2.23***** (0.21)
Teacher Experience					
< 3years	-0.09 (0.09)	-0.15 (0.09)	-0.09 (0.09)	-0.08 (0.09)	**-0.19*** (0.09)
3-5 years	0.08 (0.08)	0.07 (0.08)	0.09 (0.08)	0.09 (0.08)	0.07 (0.08)
11-25 years	**-0.18*** (-0.07)	**-0.16*** (-0.07)	**-0.21**** (0.07)	**-0.18*** (0.07)	**-0.17*** (0.07)
>25 years	-0.16 (0.11)	-0.11 (0.11)	-0.21 (0.11)	-0.13 (0.11)	-0.12 (0.11)
Controls					
Male	**-0.43***** (0.06)	**-0.44***** (0.06)	**-0.45***** (0.06)	**-0.42***** (0.06)	**-0.43***** (0.06)
Black	**-0.48***** (0.07)	**-0.52***** (0.07)	**-0.47***** (0.07)	**-0.43***** (0.07)	**-0.52***** (0.07)
Other Race	-0.05 (0.13)	-0.01 (-0.13)	-0.02 (0.13)	0.00 (0.13)	-0.07 (0.13)
Constant	**-1.60***** (0.18)	**-1.81***** (0.19)	**-1.51***** (0.20)	**-1.56***** (0.18)	**-1.28***** (0.20)
n	10,616	10,616	10,616	10,616	10,616
R^2	0.06	0.06	0.06	0.07	0.05

Note: Reference groups are: ***NC TEP Indicator***—developing rating level; ***Teacher Experience***— 6-10 years' experience; ***Controls***— female and White.
Note: * indicates statistically significant at the p<0.05 level; ** indicates statistically significant differences at the p<0.01 level; *** indicates statistically significant differences at the p<0.001 level.

While the ratings are correlated at some level with teacher value-added scores, the ratings along with experience, gender, and ethnicity explain only six percent of the variability of teachers' value-added scores. In other words, there is a considerable amount of variance in the value-added measure not explained by the model specified above, which may be attributable to several factors. One possible factor is that the EVAAS score, as used here, is measured as a point estimate and may have an associated variance that the model cannot account for. Another possible factor is that the information provided in the 6th standard adds information to the evaluation process, but that the ratings, the teacher value-added scores, or both may lack reliability. Finally, the variance also may be an artifact of the clustering of NC TEP ratings on the *Proficient* and *Accomplished* levels, as demonstrated above in Figure 1. More specifically, because there is little variance in the NC TEP ratings, the ability to explain a large amount of the variance of the value-added measure may be limited.

Comparing EVAAS Teacher Value-added Scores to other Measures of Teaching Effectiveness

The Evaluation Team also estimated correlations between teachers' value-added scores and two additional measures of teaching effectiveness: the Tripod Student Survey and CLASS observations. The Tripod Student Survey was developed to obtain students' responses about their teachers' instructional practices (Ferguson, 2008). The Tripod survey identifies seven specific constructs, including Captivate, Care, Challenge, Clarify, Confer, Consolidate, and Control, but a factor analysis conducted by the Evaluation Team of the 2011-12 survey responses showed that there were only three distinct constructs: *general pedagogy* (which combines Captivate, Care, Clarify, Confer, and Consolidate), *classroom management* (Control), and *challenge*, often referred to as "academic press" or setting high expectations for students (Challenge). The CLASS observations have three constructs that represent three distinct and positive attributes of interactions between teachers and students: *emotional support*, *classroom organization*, and *instructional support*.

Table 6 (following page) presents the average values of the ratings for *general pedagogy*, *classroom management*, and *challenge* for the sub-sample of teachers that had both EVAAS estimates and Student Tripod survey responses. It is interesting to see that *general pedagogy* scores seem unrelated to experience, except that teachers with more than 25 year of experience tend to have lower scores. *Classroom management* scores increase with experience. *Challenge* scores also increase with experience.

All three Tripod constructs correlate with teachers' value added scores, but, as with principal ratings of the teachers, the correlations vary, as shown in Table 7 (second page following). *General pedagogy* and *challenge* (academic press) most strongly predict teachers' value-added scores. *Classroom management* also predicts teachers' value-added scores but much less strongly than the other two constructs. Elementary teachers with student responses similar to secondary teachers have lower teacher value-added scores, suggesting that students in elementary schools give their teachers higher ratings than do secondary students when effectiveness (as measured by the three constructs) is held constant. However, this result could be an artifact of the different constructs used for elementary and secondary teachers. Male teachers with similar student responses also receive lower value-added scores.

Taken together, these results suggest that additional measures of teacher effectiveness, most notable *pedagogy* and *challenge*, and, to a lesser extent, *classroom management*, are related to a teacher's value-added score. However, as with the model estimating the relationship between teacher value-added and NC TEP ratings, several variables demonstrate a potential bias in these relationships.

Table 6: Tripod Student Survey, Descriptive Statistics

Variables	General Pedagogy	Classroom Management	Challenge
	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Average Factor Value	-0.01 (1.06)	0.47 (0.93)	0.00 (1.07)
Experience			
< 3years	0.14 (0.87)	0.26 (0.94)	-0.17 (0.97)
3-5 years	-0.09 (1.07)	0.38 (0.86)	-0.07 (1.04)
6-10 years	-0.03 (1.04)	0.50 (0.92)	-0.04 (1.12)
11-25 years	0.06 (1.10)	0.53 (0.90)	0.09 (1.05)
>25 years	-0.41 (1.05)	0.73 (1.13)	0.22 (1.18)
Level			
Elementary	-0.09 (1.14)	0.69 (0.69)	0.34 (1.06)
Secondary	0.03 (1.01)	0.36 (1.01)	-0.18 (1.03)
Subject			
English-Language Arts	0.12 (1.02)	0.48 (0.91)	0.16 (1.09)
Mathematics	-0.45 (1.00)	0.40 (0.91)	0.19 (0.99)
Science	-0.09 (1.13)	0.53 (0.96)	-0.28 (1.05)
Gender			
Female	0.01 (1.05)	0.48 (0.94)	-0.07 (1.08)
Male	-0.09 (1.08)	0.46 (0.91)	-0.23 (1.02)
Race			
Black	-0.02 (0.94)	0.52 (0.87)	-0.07 (1.02)
White	-0.01 (1.07)	0.46 (0.94)	0.02 (1.08)
Other	-0.12 (1.28)	0.52 (0.82)	-0.18 (0.98)

Table 7. EVAAS and TRIPOD Survey Factors, OLS Estimates

Variable	Coefficient (Std. Err.)
Factor	
Pedagogy	**0.49***** (0.05)
Classroom Management	**0.15*** (0.06)
Academic Press	**0.47***** (0.05)
Experience	
< 3years	-0.01 (0.18)
3-5 years	0.17 (0.18)
11-25 years	-0.15 (0.14)
>25 years	**-0.70**** (0.23)
Level	
Secondary	**0.51***** (0.12)
Subject	
Mathematics	0.07 (0.14)
Science	**0.10** (0.47)
Controls	
Male	**-0.44***** (0.14)
Black	0.05 (0.16)
Other Race	0.68* (0.31)
Constant	-0.19 (0.16)
n	1,388
R^2	0.13

Note: Reference groups are: ***Experience***—Teachers with 6-10 years' experience; ***Level***—elementary; ***Subject***—ELA; ***Controls***— female and White.
Note: * indicates statistically significant at the p<0.05 level; ** indicates statistically significant differences at the p<0.01 level; *** indicates statistically significant differences at the p<0.001 level.

Table 8 (following page) presents the average CLASS observation scores for each of the three constructs broken out by experience, level of schooling, subject, gender, and ethnicity. These data suggest that observed *classroom organization* does not vary substantially across any of the sub-categories.

Table 8: CLASS Observations, Descriptive Statistics

Variables	Emotional Support	Classroom Organization	Instructional Support
	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Average scale score (1-7)	4.78 (0.68)	5.78 (0.51)	4.22 (0.64)
Experience			
< 3years	4.97 (0.19)	5.78 (0.55)	4.53 (0.62)
3-5 years	4.77 (0.80)	5.84 (0.45)	4.12 (0.65)
6-10 years	4.83 (0.69)	5.80 (0.75)	4.36 (0.65)
11-25 years	4.80 (0.61)	5.75 (0.45)	4.20 (0.59)
>25 years	4.17 (1.08)	5.76 (0.20)	3.88 (0.94)
Level			
Elementary	5.00 (0.63)	5.90 (0.36)	4.46 (0.56)
Middle	4.68 (0.73)	5.71 (0.49)	4.12 (0.62)
Secondary	4.62 (0.58)	5.74 (0.73)	4.05 (0.72)
Subject			
English-Language Arts	4.81 (0.58)	5.76 (0.45)	3.97 (0.51)
Mathematics	4.54 (0.81)	5.73 (0.72)	4.11 (0.78)
Science	4.87 (0.63)	5.81 (0.43)	4.33 (0.59)
Gender			
Female	4.82 (0.63)	5.80 (0.52)	4.26 (0.63)
Male	4.59 (0.87)	5.72 (0.47)	4.04 (0.70)
Race			
Black	4.67 (0.65)	5.89 (0.40)	4.27 (0.44)
White	4.81 (0.70)	5.76 (0.55)	4.21 (0.70)
Other	4.94 (0.09)	5.56 (0.32)	4.39 (0.04)

While the observed levels of *emotional support* and *instructional support* are somewhat different, they tend to show similar patterns across descriptive categories. Females receive higher scores on both than do males. Mathematics teachers receive higher scores on both than do ELA teachers, and Science teachers are higher than both Mathematics and ELA teachers on these measures. The most prominent difference in these measures by experience is that teachers with greater than 25 years of experience tend to have lower scores on both *instructional* and *emotional support*.

Table 9 demonstrates that none of the three CLASS measures is significantly related to teachers' value-added scores. Collectively, with the controls for experience, gender, subject taught, level of schooling, and ethnicity, CLASS measures explain 25 percent of the variation in teachers' value-added scores. The sub-sample of teachers with both CLASS measures and EVAAS scores is the smallest of all of the sub-samples in this analysis, so it is possible that, with a larger sample of teachers, the relationship between CLASS measures and EVAAS scores would vary in significance, magnitude, and/or direction relative to the results in the present analysis.

Table 9. EVAAS and CLASS Observations, OLS Estimates

Variable	Coefficient (Std. Err.)
CLASS Domain	
Emotional Support	0.72 (0.67)
Classroom Organization	-0.81 (0.54)
Instructional Support	0.44 (0.76)
Experience	
< 3years	-0.83 (1.18)
3-5 years	-0.05 (0.80)
11-25 years	**-1.39*** (0.68)
>25 years	0.28 (1.25)
Level	
Middle	-0.19 (0.77)
Secondary	**1.65*** (0.82)
Subject	
Mathematics	0.31 (0.90)
Science	-0.08 (0.86)
Controls	
Male	-0.08 (0.70)
Black	-0.35 (0.67)
Other Race	-0.57 (1.63)
Constant	0.19 (3.46)
n	76
R^2	0.25

Note: Reference groups are: ***Experience***—Teachers with 6-10 years' experience; ***Level***—elementary; ***Subject***—ELA; ***Controls***— female and White.
Note: * indicates statistically significant at the p<0.05 level

Comparing EVAAS Teacher Value-Added Scores to Teachers' Survey Responses

In the previous sections, the Evaluation Team compared teachers' EVAAS scores with other measures of teaching quality, including the five other evaluation ratings provided by principals and other measures that were developed to measure teaching quality. This section highlights similar analyses for two measures from the Omnibus Survey that also are expected to relate to teachers' value-added scores and one that measures teachers' views of the teacher evaluation process.

The first of the constructs is teachers' attitudes about their own effectiveness in the classroom, or *teacher self-efficacy*. This construct is a composite measure of several items that relate to teachers' beliefs in their ability to motivate their students and make a difference in their learning. The second measure is also a composite measure, one that reflects teachers' beliefs that they have adequately covered the learning objectives for their grade and subject(s) and adequately prepared their students for success in the next grade, which is labeled here as *preparation*. Finally, this section examines teachers' views about the current teacher evaluation process and its outcomes, in terms of teacher development. Once again, this construct—*evaluation validity*—is a composite measure that incorporates teachers' beliefs in the fairness and clarity of the process and whether it contributes to professional growth and development.

Teachers in this sub-sample are those who responded to the survey and also had EVAAS scores, which results in a probability sample of teachers with an EVAAS score. Table 10 (following page) presents the descriptive statistics of the three constructs for the entire sample as well as by categories. On average, these teachers strongly believe they have prepared their students well with respect to the learning objectives that have been set for their grade and subject(s), as indicated by an overall average of 6.7 out of a possible 7.0 *preparation* score. These responses vary little across categories, except that teachers with less than three years of experience have slightly less confidence that their students have been well-prepared. On average, these teachers reflected high *self-efficacy* in the classroom, as indicated by an overall average of 5.7 on a 7-point scale. Teachers in this sample are somewhat less likely to believe in the fairness of the evaluation process and that it is developmentally oriented to enhance their own efficacy, as indicated by an overall average 5.3 *evaluation validity* score on a 7-point scale. Teachers with three to five years of experience and minority teachers seem less likely than others to view the evaluation process favorably, but the differences are relatively small.

Table 11 (second page following) presents the estimates of teacher value-added, as measured by the EVAAS score, in relation to the three Omnibus survey constructs and other control variables. The strongest relationship between value-added scores and survey responses was related to teachers' responses about *self-efficacy*. A one-point difference in *self-efficacy* is positively associated with a half-point difference in teachers' value-added score. Teachers' belief that they have adequately prepared their students is also significantly related to their value-added scores with a one-point increase in *preparation* associated with a .36-point increase in teachers' value-added scores. Teachers' sense of the fairness of the evaluation system is not related to teachers' value-added scores. Controlling for the other variables in the model, only males have significantly different value-added scores; their value added scores are .5 points lower than those of females.

Table 10. Omnibus Teacher Survey, Descriptive Statistics

Variable	Evaluation Validity	Test Preparation	Self-Efficacy
	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Average Scale Score (1-7)	5.30 (1.29)	6.68 (0.64)	5.73 (0.78)
Experience			
< 3years	5.57 (1.11)	6.53 (0.62)	5.64 (0.87)
3-5 years	5.12 (1.28)	6.75 (0.40)	5.76 (0.76)
6-10 years	5.24 (1.46)	6.70 (0.50)	5.69 (0.76)
11-25 years	5.24 (1.30)	6.67 (0.76)	5.70 (0.77)
>25 years	5.50 (1.23)	6.69 (0.75)	5.96 (0.70)
Level			
Elementary	5.47 (1.18)	6.72 (0.41)	5.99 (0.62)
Middle	5.01 (1.51)	6.65 (0.85)	5.75 (0.73)
Secondary	5.31 (1.27)	6.67 (0.63)	5.67 (0.80)
Subject			
English-Language Arts	5.09 (1.45)	6.53 (0.58)	5.58 (0.81)
Mathematics	5.05 (1.40)	6.58 (0.78)	5.50 (0.77)
Science	5.29 (1.30)	6.71 (0.52)	5.78 (0.75)
Unknown/Contained	5.44 (1.20)	6.71 (0.61)	5.77 (0.78)
Gender			
Female	5.34 (1.27)	6.70 (0.61)	5.77 (0.75)
Male	5.12 (1.37)	6.58 (0.61)	5.58 (0.79)
Race			
Black	5.58 (1.12)	6.71 (0.61)	5.93 (0.69)
White	5.17 (1.33)	6.66 (0.63)	5.65 (0.78)
Other	5.85 (1.16)	6.70 (0.50)	5.88 (0.84)

Table 11. EVAAS and Omnibus Survey Scales, OLS Estimates

Variable	Coefficient (Std. Err.)
Omnibus Survey Scale	
Evaluation Validity	-0.13 (0.08)
Test Preparation	**0.36*** (0.16)
Self-Efficacy	**0.50***** (0.14)
Experience	
< 3years	-0.46 (0.36)
3-5 years	-0.47 (0.32)
11-25 years	0.12 (0.27)
>25 years	-0.37 (0.35)
Level	
Middle	0.18 (0.38)
Secondary	-0.17 (0.39)
Subject	
Mathematics	-0.25 (0.39)
Science	0.28 (0.40)
Unknown/Contained	-0.13 (0.34)
Controls	
Male	**-0.46*** (0.23)
Black	-0.38 (0.24)
Other Race	0.73 (0.49)
Constant	**-4.30***** (1.27)
n	825
R^2	0.06

Note: Reference groups are: ***Experience***—Teachers with 6-10 years' experience; ***Level***—elementary; ***Subject***—ELA; ***Controls***— female and White.
Note: * indicates statistically significant at the p<0.05 level; ** indicates statistically significant differences at the p<0.01 level; *** indicates statistically significant differences at the p<0.001 level.

Summary of the Relationships between the 6th Standard—Teachers' Value-Added Scores—and Other Measures

The quantitative findings above present correlates of Standard 6 with other measures that are expected to relate to teachers' effectiveness, as well as the variation of teachers' value-added scores across gender, ethnicity, and other characteristics.

Summary of the relationship between Standard 6 and the other five standards:

- Teachers' ratings as either *Accomplished* or *Distinguished* on Standard 3 (teachers know the content they teach) or Standard 4 (teachers facilitate learning for their students) are significantly and positively related to their EVAAS scores.
- There is also a correlation between teachers' *Accomplished* or *Distinguished* rating on Standard 1 (teachers demonstrate leadership) and their EVAAS score, though it is not as large.
- 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.

Summary of the relationship between Standard 6 and other measures of teacher effectiveness:

- Teachers' value-added scores are related to questions in the student survey that measure general pedagogy and setting high expectations.
- Omnibus Survey responses indicate that self-efficacy and the belief that their students are well prepared are also related to the Standard 6 measure.
- Measures of teacher effectiveness estimated by CLASS are not related to Standard 6.

Summary of sub-group findings:

- There are systematic differences in the relationship between Standard 6 and other measures of teacher effectiveness when controlling for certain sub-groups, such as gender and race.
- Males and minorities receive higher ratings from their principals on Standards 1 through 5 than would be consistent with their value-added scores.
- These tentative findings are based on a single sample and should be more thoroughly examined as new data become available; also, as new data become available, other potential biases should be examined.

In general, these findings indicate a consistency between measures expected to relate to a teachers' ability to raise student test scores and the value-added scores that measure a teacher's actual contributions to student test score growth. These findings should provide confidence that a teacher's value-added score is providing information that is consistent with other sources that measure teacher effectiveness. Although the other measures are not highly predictive of a teacher's value-added score, each seems to measure important and related concepts related to teachers' effects on student achievement. At the outset of this section, the Evaluation Team suggested that the value-added measure should relate to other measures of teacher effectiveness,

but not so closely that no additional information is provided. The significant and positive relationships between many of the additional teacher effectiveness measures and Standard 6 suggest that the teacher value-added measure accounts for many of the dimensions considered important to effective teaching; however, these relationships are not so highly correlated that they could be considered substitutes for one another.

Qualitative Analysis: Comparing Teacher Value-Added Scores with other Measures of Teaching Quality

The Research Questions that guide this portion of the evaluation are:

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?

The qualitative findings presented in this section are derived from 72 interviews in fall 2012 (22 principals and 50 teachers) and describe participants' views about the use of student growth data in evaluation, including the implementation of the Common Exams.

One focus of the NCEES is to provide access to EVAAS data for the purpose of informing instructional practices (Research Question 2). While EVAAS data ultimately will include data provided by the EOC, EOG, and Common Exam assessments, at the time of the interviews, teachers and administrators could access only EOC and EOG data from the 2011-12 school year. Therefore, the findings in this discussion related to Research Question 2 focus on participants' prior and current use of EVAAS data as it relates to EOC and EOG data. Participant perceptions about the use of student growth data in their evaluations (Research Question 3) and about the implementation of Common Exams (Research Question 4) should be interpreted within the context of early stages of implementation.

How are Principals Using EVAAS Data for Evaluation and to Inform Teaching Practices?

Principals were asked to report whether they had received prior access to EVAAS data (before the 2012-13 school year) and to provide information about methods for accessing the data and use of the data in evaluation. Table 12 (following page) shows that principals (n=22) reported having prior access to EVAAS data. Of those, 82% were able to access student, teacher, and school-level EVAAS data. Based on close-ended interview questions, a majority of principals (95%) were able to access EVAAS data through direct log-in; however, some (41%) of principals received EVAAS information, such as print-outs of scores, through their central office.

Table 12. Principal's Prior use of EVAAS

Prior to 2012-13, what level of data did you have access to? Check all that apply.	% of Respondents (*n*=22)	How did principal access data? Check all that apply.	% of Respondents (*n*=22)
Student-Level Data	82% (*n*=18)	Direct Access	95% (*n*=21)
Teacher-Level Data	82% (*n*=18)	Received Printouts/Scores from LEA	41% (*n*=9)
School-Level Data	82% (*n*=18)	Other	5% (*n*=1)

Principals also were asked about their current access (as of fall 2012) to EVAAS data. Table 13 demonstrates that 18 out of 22 principals (81%) had accessed EVAAS data. All of these principals were able to access these data directly and had student-, teacher-, and school-level data. It is important to note that the variability between the number of principals with prior access and current access may be attributable to the timing of the interviews.[32] Three out of 18 principals (17%) reported using these data for teacher evaluations, and two out of 22 principals (9%) reported that these data have been used for their own evaluations.

Table 13. Principal's Current Use EVAAS

In 2012-13, what level of data did you have access to? Check all that apply.	% of Respondents (*n*=18)	How did principal access data? Check all that apply.	% of Respondents (*n*=18)
Student-Level Data	83% (*n*=15)	Direct Access	100% (*n*=18)
Teacher-Level Data	78% (*n*=14)	Received Printouts/Scores from LEA	17% (*n*=3)
School-Level Data	83% (*n*=15)	Other	—

Data from the open-ended portion of the interviews reveal several key themes regarding how principals use EVAAS data to provide feedback to teachers and/or to inform the evaluation process. Findings reveal that principals tend to use EVAAS data to make school-based decisions and to establish a dialogue within the school community regarding instructional practices that bring about student growth. Some principals remarked that they used EVAAS scores as a talking point in evaluation conferences and planning meetings with their teachers.[33] For instance, principals used EVAAS data to drive conversations about ways to focus instructional practices that bring about student growth:

> [W]hen I met with that teacher my conversation with that teacher was, "Yeah, you're getting the proficiency but you're not getting the growth. This is an area that you need to be focusing on. You need to be looking at your instructional practices . . ." and then that

[32] Since principals were interviewed in the fall 2012, they were able to reflect on the full year (2011-2012) for prior access but were only able to reflect on current access as of September-October 2012.
[33] Student growth data will be used to calculate a teacher's Standard 6 effectiveness rating starting in the 2013-14 school year. Consequently, at the time of data collection, there were no reports indicating that principals use EVAAS scores in the direct determination of a teacher's effectiveness rating.

> will go right back to, "What are you doing in your PLCs? . . ." So, we really use EVAAS as a conversation about instruction.

Principals also use EVAAS data to make strategic staffing decisions, including hiring teachers specifically to work with students at risk for failure. One teacher noted:

> They hired me, as I said, to teach that class because of my past experience. . . . I specifically worked with gang-bangers and had been able to bring them up. . . . So, coming in with that background, they wanted me to teach this class. . . . EVAAS became a part of that in saying, "We're not just labeling these students. There is data to prove these students need this extra help."

In that same vein, principals remarked that EVAAS data assist in the identification of students for remediation and classroom placement. Principals reported that EVAAS scores also help teachers and principals track student growth throughout the year, as they can compare benchmark test scores to projected student scores and assess progress. Interventions for identified students include remediation, heterogeneous grouping, and re-teaching course content. One principal said:

> We basically did two things with it [EVAAS]. We would log in online, and we'd pull the teacher effectiveness reports . . . and we would give that to teachers. . . . The other thing we did is to pull student sheets to decide what classes we would place students in. For example, in our Algebra I class.

How are Teachers Using EVAAS Data for Evaluation and to Inform Teaching Practices?

Teachers were asked to report whether they had received prior access to EVAAS data (before the 2012-13 school year) and to provide information about methods for accessing the data and use of the data in evaluation. Table 14 reflects participants who teach a broad array of courses: 31 out of 50 (62%) indicated they taught an EOG- or EOC-tested course, 14 out of 50 (28%) indicated that they taught a course with a Common Exam, and 5 out of 50 (10%) taught a blend of courses with EOGs, EOCs, and/or Common Exams.

Table 14. Level and Type of Access to EVAAS Prior to 2012-13

Type of Test Administered	Respondents Having Access to EVAAS Data Prior to 2012-13	Of Those with Access . . .					
		Level of Access			*Method of Access*		
		Student	*Teacher*	*School*	*Direct*	*Printouts*	*Other*
EOC/EOG Only (*n*=31)	61% (*n*=19)	84% (*n*=16)	16% (*n*=3)	42% (*n*=8)	53% (*n*=10)	79% (*n*=15)	—
Common Exam Only (*n*=14)	86% (*n*=12)	58% (*n*=7)	58% (*n*=7)	33% (*n*=4)	67% (*n*=8)	75% (*n*=9)	8% (*n*=1)
Both EOC/EOG and Common Exam (*n*=5)	40% (*n*=2)	50% (*n*=1)	50% (*n*=1)	50% (*n*=1)	50% (*n*=1)	50% (*n*=1)	—
Total (*n*=50)	66% (*n*=33)	73% (*n*=24)	33% (*n*=11)	39% (*n*=13)	58% (*n*=19)	76% (*n*=25)	3% (*n*=1)

Table 15 indicates that 66% (33 out of 50) of teachers in this sample reported having access to EVAAS data prior to the 2012-13 school year. Of those teachers, 61% (*n*=19) of EOC/EOG teachers, 86% (*n*=12) Common Exam-only teachers, and 40% (*n*=2) of teachers who taught both EOC/EOG and Common Exam classes reported prior access of EVAAS. The majority of these teachers received access to EVAAS data via print-outs from school administrators (76%).[34]

When teachers were asked about their current EVAAS access, 16 out of 50 teachers (32%) reported accessing EVAAS as of fall 2012.[35] Of those teachers, 35% (*n*=11) of EOC/EOG teachers, 21% (*n*=3) of Common Exam teachers, and 40% of EOC/EOG and Common Exam teachers reported having current access. A majority of EOC/EOG teachers (91%) and all teachers who taught a combination of EOC/EOG and Common Exam courses (100%) were able to access student-level data. Participants accessed EVAAS data via a combination of direct log-in to EVAAS, print-outs, and other methods.[36]

Table 15. Level and Type of Access to EVAAS as of Fall 2012

Type of Test Administered	Respondents Having Access to EVAAS Data in 2012-13	Of Those with Access . . .					
		Level of Access			*Method of Access*		
		Student	*Teacher*	*School*	*Direct*	*Printouts*	*Other*
EOC/EOG only (*n*=31)	35% (*n*=11)	91% (*n*=10)	36% (*n*=4)	27% (*n*=3)	55% (*n*=6)	73% (*n*=8)	9% (*n*=1)
Common Exam only (*n*=14)	21% (*n*=3)	66% (*n*=2)	33% (*n*=1)	33% (*n*=1)	100% (*n*=3)	33% (*n*=1)	—
Both EOC/EOG and Common Exam (n=5)	40% (*n*=2)	100% (*n*=2)	—	—	100% (*n*=2)	—	—
Total (*n*=50)	32% (*n*=16)	88% (*n*=14)	31% (*n*=5)	25% (*n*=4)	69% (*n*=11)	56% (*n*=9)	6% (*n*=1)

Teachers provided mixed responses to open-ended interview questions about the use of EVAAS student growth data to inform instruction. Some teachers described other growth assessments they used to track student progress that are not currently used in EVAAS. For instance, one teacher remarked on using the ClassScape Assessment System, a North Carolina-based tool that helps educators monitor the academic progress of students on objectives aligned to the Common Core and North Carolina Essential Standards:

> I have not used EVAAS, but what we do here is analyze students weekly using ClassScape, for example, or other types of in-house assessments. . . .

[34] Teacher subgroups (EOC/EOG, Common Exam, and EOC/EOG and Common Exam) are based on self-reported designation as of fall 2012. It is possible that teachers in this sample taught a different set of courses prior to fall 2012.

[35] Some teachers in this sample were interviewed as early as September 2012. The timing of the interviews prevents direct comparison to findings regarding prior access of EVAAS.

[36] As these findings are based on participant self-reports, these tables should be interpreted with caution. All participants were asked, "Have you accessed EVAAS data for this current school year?" However, open-ended responses indicate that some participants had misconceptions about EVAAS data as compared with other student growth data.

Other teachers had received preliminary information via print-outs but did not have a clear concept of how to use the data to inform their instruction:

> Our principals had access to it, and we were shown some information through that, and we were given some print-outs, but I've not had any real training on how to use EVAAS to help me.
>
> We received a printout with our EVAAS data. It wasn't broken down by student or anything so there was not really much use of it except to know where we were, like on a chart.

Some teachers who had access to EVAAS data were using this information to identify students who were in need of remediation and to implement interventions or to customize individual instruction:

> We were given an account, PIN number, and a password. I would open up my account on my personal computer. . . . What I would go to is the reading scores. I would look at the fourth, sixth, and eighth grade EOG more than anything else, see how they have progressed from fourth to the eight, see where there were information lapses, if they fell off somewhere. I'd address those issues. And all in all, what this information allowed me to do was differentiate my instruction based on the students' needs.
>
> We used it to watch and kind of predict growth in students, especially once we went to some common formative assessments during the school year so we could see how students are mastering standards and knowledge so that we could identify deficiencies or lack of growth and look for interventions ideas and have news to support students.

What are Teachers' and School Leaders' Perceptions of the Use of Student Growth Data in the Evaluation Process?

The impetus for this evaluation question was to understand how principals perceive the addition of student growth data (as measured by EOCs, EOGs, and Common Exams) to Standard 6 of the NC TEP and Standard 8 of the NC PEP. As the timing of these interviews (fall 2012) corresponded with the roll-out of the Common Exams and the new standards, the Evaluation Team sought to understand participants' knowledge about the new standards and related components. Findings suggest that while participants had been informed about the new teacher and leader evaluation process, there were still numerous questions regarding the development and implementation of its various components. Interviews revealed respondents' perceptions about the potential benefits, concerns, and misperceptions about the use of student growth data in evaluation.

General knowledge about the use of student growth data in the teacher evaluation process (NC TEP). Principals reported on the extent to which they understood components of the teacher and leader evaluation process, including the addition of Standard 6, Standard 8, EVAAS data, and Common Exams. Table 16 (following page) shows that a majority of the 22 respondents indicated that they had at least a little to a lot of understanding of the components; however, few reported strong knowledge of the components. A majority of principals (73%) had little to no knowledge about the Common Exams at the time of these interviews.

Table 16. Principals' Understanding of Components of Standards 6 and Standard 8 in the NC TEP and NC PEP

	Understanding (*n*=22)				Source of Training on NCEES			
Components	**Not at All**	**A Little**	**A Lot**	**Fully**	**Yes**	**LEA[a]**	**State[a]**	**Both[a]**
Teacher Standard 6	—	50% (*n*=11)	50% (*n*=11)	—	95% (*n*=21)	62% (*n*=13)	29% (*n*=6)	10% (*n*=2)
Principal Standard 8	9% (*n*=2)	41% (*n*=9)	45% (*n*=10)	5% (*n*=1)	73% (*n*=16)	75% (*n*=12)	19% (*n*=3)	6% (*n*=1)
Common Exams	5% (*n*=1)	68% (*n*=15)	23% (*n*=5)	5% (*n*=1)	82% (*n*=18)	61% (*n*=11)	17% (*n*=3)	22% (*n*=4)
EVAAS	—	50% (*n*=11)	45% (*n*=10)	5% (*n*=1)	91% (*n*=20)	40% (*n*=8)	20% (*n*=4)	40% (*n*=8)

[a] Calculated as the percentage of those responding "Yes" to receiving professional development for the respective component.

Most principals commented that they received information and training related to the addition of Standard 6 to the NC TEP and Standard 8 to the NC PEP through a variety of sources, but the majority of training they received took place at the LEA level. Principals received information regarding the use of student growth data via a combination of LEA-level and state-level resources. Principals received a majority of information from their LEAs in the form of regular leadership meetings, staff development, emails, and direct coaching from LEA personnel (the Curriculum Specialist, the Testing Coordinator or the Associate Superintendent, etc.). Principals said:

> [O]ur Assistant Superintendent for Human Resources always updates us on that kind of information at Leadership Team Meetings so we'll know that kind of thing.
>
> [Student growth data] was delivered by our Director of Accountability and Testing, and we had a follow-up series that would come through our Associate Superintendent, who would come over and work with schools on student performance data, looking for trends and what types of intervention we could put in.

Principals cited webinars as the most utilized state-level resource, followed by state-led trainings, emails, and DPI representation at LEA and regional leadership meetings. Principals reported that they received information and updates about implementation of Common Exams and the use of student growth data in the evaluation on a rolling basis and that the state encouraged a "train-the-trainer" approach to disseminating this information to their teachers. Principals said:

> The district is just now rolling out some of the information. They actually sent an email out last week to provide training for teachers and administrators. . . .
>
> It's really been face-to-face…It was different because it was thoroughly explained, and I appreciate that our district took those steps, and then afterwards it was like [sic] "We will send you this information so now you'll be able to share it with your staff." [It was] a great approach, I thought.

While these data have been informative in many ways, principals also indicated that they were challenged by the ongoing roll-out of information regarding Standard 6 and addition of the Common Exams. Some principals indicated that they were purposeful in the timing of

information they disseminated to teachers. While many principals remarked that they passed information to their staff as they received it, some principals commented that they intentionally withheld information about the implementation of the Common Exams and EVAAS data from their teaching staff due to their own lack of understanding about how scores are calculated or because of lack of information:

> I think they [teachers] have just begun to grasp the ABCs model, and now it's changing, and I personally can't explain it to them. I could calculate growth and predicted scores for the ABCs, but I don't know how it's done with EVAAS.
>
> We've been informed about the way that it [Common Exams] will look, and that there are potential changes to it. So my style and my approach is "Well, when the changes are all done, let us know so I can properly inform my staff to avoid confusion."

Table 17 indicates that the level of understanding among teacher was similar to that of principals regarding Standard 6 of the NC TEP. Of the three new components, all respondents indicated that they had at least a little knowledge about Standard 6. Teacher understanding of EVAAS was variable; 14% reported having no knowledge of EVAAS, 50% had a little knowledge, and 26% indicated that they had a lot of knowledge. Similar to principal responses, 80% of teachers indicated that they had little (50%) to no knowledge (30%) of the Common Exams.[37] Findings related to participant knowledge about Common Exams and its implementation will be discussed in more detail in the following section of this report.

Table 17. Teachers' Understanding of Components of Standard 6 in the NC TEP

	Percentage of Respondents (*n*=50)			
Components of the NC TEP	**Not at All**	**A Little**	**A Lot**	**Fully**
Standard 6	—	54% (*n*=27)	34% (*n*=17)	12% (*n*=6)
Common Exams	30% (*n*=15)	50% (*n*=25)	12% (*n*=6)	8% (*n*=4)
EVAAS	14% (*n*=7)	54% (*n*=27)	26% (*n*=13)	6% (*n*=3)

Although there was variation in levels of teacher knowledge, overall, teachers expressed that they had a general awareness about the use of student growth data to calculate a Standard 6 rating as a result of staff meetings or discussions in their Professional Learning Communities (PLCs). Specifically, many teachers indicated they were aware that the student scores are based on a growth model that requires at least three years of student growth data in order to calculate an overall effectiveness rating:

> They've kind of touched on Standard 6. They haven't really gone fully into all the details about it. . . . My principal has talked to us about it and PLCs. . . . but it would be nice to have some more in-depth training on it.

[37]The variability in access to EVAAS data (as mentioned in findings for Research Question 2) and the timing of these interviews relative to Common Exam implementation (see findings for Research Question 4) may explain why some teachers had limited understanding.

> Well, I'm aware that there is Standard 6 and that it's going to be added to our evaluations this year and that it's the measure of our students' growth during the school year. That information was given to us in teachers' meetings and that's the extent of my knowledge of it.
>
> [T]hey let us know that [the rating] is more about growth. . . . It's more how did you grow them.
>
> [W]e were just informed that it is starting now, but . . . it's going to take three years for us to get to the rating.

Positive perceptions of student growth data in evaluation. When asked to provide their thoughts about the addition of Standards 6 and 8 to the NCEES, both teachers and school leaders acknowledged that their students' academic growth was a critical indicator of effectiveness. Teachers and principals said:

> I think to be an effective teacher that you need to demonstrate growth. And I'm not saying that it is or is not 100% fair. I do believe that all students can learn. And I would say that to adhere to the school's philosophy is that all students can learn at a high level. (Teacher)
>
> I think teachers are aware now that we're going to make sure that the students are learning, and that they're progressing from year to year. I think that's crucial. . . . (Principal)
>
> Ultimately, I'm responsible for what happens in the school and, you know, if I'm allowing people who are not getting the job done to stay in the classroom, and that's reflected in student data, whether it's through Common Exams or whether it's through EVAAS or EOCs, whatever the case may be, then that is a reflection of my leadership. (Principal)

Ratings for Standards 6 and 8 can augment other performance measures on the NCEES. If used formatively, teachers and principals hypothesize that student growth data in evaluations can be used to reflect upon, discuss, and modify their instructional practices. An accountability system would also encourage educators to align their instruction to learning targets set by the teaching standards. Teachers and administrators point out that an effective feature of student growth data in evaluation is the potential to focus instruction on student learning needs. Teachers and principals commented:

> To me, the benefit for the school and even for the district [is that] it makes the teachers accountable for what they have to do. They have to make sure that they are teaching the standards, they're doing their lesson plans, they are looking at individual data for their children, they're rearranging their flexible groups according to the data, because that's how I use the data in my classroom. (Teacher)
>
> So, I feel that as teachers see if their students are growing or see if their students are learning, then they will be able to [adjust] their instruction to meet the needs of the students. (Teacher)

> I think it's going to make [teachers] more aware and cause them to do greater reflection on their teaching, and look at their delivery of instruction, differentiation of instruction of students, and how they actually teach the children. (Principal)

Participants commented that the new standards will hold teachers and school leaders accountable for continuous improvement. Some participants pointed out that the new standard could be used to identify teachers who are implementing successful practices and provide an objective yardstick for teachers to measure student growth. Teachers and principals said:

> It will certainly motivate teachers if their students are making growth. It will motivate them to continue to do whatever they're doing or to enhance what they're doing. . . . (Teacher)
>
> I guess there are benefits because there are probably some principals who don't do what they're supposed to do and it makes them be responsible and keep a check on their teachers and make sure they're doing what they're supposed to be doing. (Principal)
>
> We [school administration] would give [our teachers] a copy of their [teacher value-added] report, explain the EVAAS to them, and that way they were able to look at their students and see what they did with the group of students as compared to students who had a similar testing history to their students throughout the state. That was powerful information for teachers because it removed a lot of excuses because teachers couldn't say that "Well, we aren't comparing apples to apples," because we were. (Principal)

Concerns about the use of student growth data in evaluation. Many teachers perceived the addition of Standard 6 as a source of stress in an already demanding career. While teachers and administrators appreciated the need for accountability for student learning and greater alignment between instruction and student outcomes, they admitted that the timing of Standard 6 amidst implementation of other RttT initiatives was stressful. Teachers and principals said:

> We've got so much stress on us as it is, and that will just add one more thing. (Teacher)
>
> I love what I do, and I'm in limbo right now. Do I continue and stay, or do I keep allowing them to put all this extra stress? And you can't handle but so much. (Teacher)
>
> I guess as a principal I feel like I'm doing what I should, it kind of puts an added stress on you that you may not [be doing enough]. It's kind of out of your control sometimes because you don't always get to hire the teachers you want to hire. They're there when you get there sometimes. (Principal)
>
> If done fairly and if presented in a way where teachers can grow from it, yes, but it's not really clear as to how they come up with these numbers, and if it's punitive in a way that money's given out but they don't get it or incentives are given out and they don't get it, I think it's going to destroy the spirit of a teacher if it's not presented in a way that we grow from it and not that, "You're a bad teacher." (Principal)

Related to this, participants expressed concern that the added pressure created by the addition of Standard 6 may influence the decision of some teachers to leave the profession. Some educators also hypothesized that Standard 6 could discourage potential teachers from entering the profession. Teachers and principals said:

> I'm also very concerned about morale in a county that has trouble hiring teachers already. What is this going to do? When you lose experience and depth of experience because of frustration over things like [the new standard] . . . then it's very hard to hire good people to come in. (Teacher)
>
> I've been teaching for 18 years and I have thoroughly enjoyed this job, but it's becoming to the point where now I'm thinking, "Hmm, is there an easier way to make money?" because I don't mind the hard work, but I don't like that additional stress, the stress that we already have on us that someone's going to lose their job over this. (Teacher)
>
> I think [the impact is] not going to be as immediate as we think, and I really believe that a lot of people are going to leave this profession in protest. (Principal)
>
> My job is to make sure everyone is learning and teachers are teaching, [and] teachers have the tools to teach. I think when you get to an evaluation instrument like [Standard 6] that could rip a good teacher away from me. . . . (Principal)

Negative perceptions about the use of student growth data in evaluation. Some teachers and principals expressed a belief that their Standard 6 ratings would not significantly influence their practices. While many participants noted that their main goal has been and will continue to be providing quality instruction that increases student learning, it was unclear whether they perceived student growth data as a tool to reach this goal. Many expressed a commitment to "keep the main thing the main thing" throughout the changes in the NCEES. This sentiment was further reinforced by some veteran administrators and teachers who had weathered similar changes in NCEES and student assessment in the past. Teachers and principals said:

> I really feel like an effective principal is an effective principal whether Standard 8 was there or not, just like an effective teacher is an effective teacher without Standard 6 being there. (Principal)
>
> I guess because of my mindset, I got into this to help kids, so that doesn't really make what I'm doing any different. . . . And I'm going to do it regardless [of the test] because they need it. (Teacher)
>
> Well, we dropped down from six EOC tests last year to only three. Now with the MSL we're jumping back up to 13 or 14 tests. I have been in the system long enough to understand in 25 to 30 years, these things go in cycles. We tested 20 years ago, and now we're testing again. So I don't see any difference. I just think more teachers are going to be paying attention to what they're supposed to in their Common Core and Essential Standards. (Principal)

Issues of clarification around the use of student growth data in evaluation. It was evident from interviews with teachers and principals that some of the concerns about the use of student growth data in evaluation stemmed from misconceptions and/or a lack of information. One reoccurring theme that emerged from interview data was that some teachers thought the calculation of a Standard 6 rating was based on student achievement rather than student growth. Many teachers cited that their student makeup, in terms of baseline ability, would influence their overall effectiveness rating and felt that having lower-achieving students in their class would unfairly disadvantage them in their Standard 6 ratings. Conversely, teachers in schools that were typically

high-achieving felt that their class make-up would predictably lead to effective Standard 6 ratings. Teachers said:

> It's a flat-line evaluation comparing you to everybody else, which I don't think is very fair, but it's never been very fair.
>
> My big concern is when you have a teacher who has been given a group of repeater students. . . . They've already failed once. . . . And if they get that same class year after year—if they're that person that the guidance department gives that class to year after year—that's going to hurt their average tremendously. . . .
>
> But if you have a student that enters your senior English class, as I do, who is maybe on the eighth grade reading level, and they take a test that's written for a senior reading level, I may have been able to help them jump one or even two grade levels in reading yet still they'll fall below in terms of their ability.
>
> I feel that since our students do so well at this school, this part of the evaluation process doesn't concern me at all because most of our students pass.

Some principal comments supported the notion that classroom makeup affects a teacher's effectiveness. As mention earlier, principals remarked that they made classroom placement decisions based on their EVAAS data:

> Not that it was unfair, but sometimes we've got teachers who work really well with, perhaps, a special education demographic and so perhaps, we might have put more of those types of students in a classroom before last year. But knowing that we needed to balance classrooms with a very diverse set of student, we wanted to make sure—knowing that 70/30 is populating someone's Standard 6— we wanted to be fairer that perhaps . . . we had to take that into consideration as we were building classrooms.

Related to this, teachers also expressed a misconception and concern about the idea that one measure, or one test, would dictate their Standard 6 effectiveness rating. Teacher comments revealed some misperceptions about the way that scoring data are aggregated to produce an effectiveness rating over time. Some felt that they had a lack of control over student behavior during the testing cycle and that students who have a "bad testing day" with this one measure could create an inaccurate representation of their contribution towards student growth:

> I've been teaching a number of years, and teacher evaluations are necessary for quality control, if you will. But student performance on one test given once a year doesn't always match what a teacher does in the room, both good and bad.
>
> I'm only being evaluated on the fact that my kids made this score on a test; no consideration to what kids you have, by what you're doing in school or how you do with the kids or what things you taught the kids.

Requests for information about the use of student growth data in evaluation. Throughout the interviews, teachers and administrators indicated interest in receiving additional information about the initial roll-out of the new standards. Of primary importance was a desire to understand how student growth data is calculated to form an effectiveness rating. Specific to this, teachers and principals want to know more about the factors that affect the rating, how the rating is

calculated (the formula), and how the formula accounts for teacher movement between grades and subjects:

> There are still questions I have about, you know, last year we had great numbers, the year before not so great, the year before more average. I guess specifically to see a formula would be helpful, but I probably don't know near as much about it as I should. . . . (Principal)
>
> I'd like to know more about how the effectiveness ratings come about, kind of I guess the math behind it, the formula behind it. . . . (Teacher)
>
> I guess more clarification on how the data will be used for Standard 6. If there will be just a baseline, "Your children did this well. Here it is," or if there's factors that affect it if it goes up and down. (Teacher)
>
> Using the Standard 6, I think there's been a lot of questions. [For instance], if a teacher went from 4th grade to 5th grade, does that still count? And a lot of my questions have been about the movement of teachers, and if a teacher came from another district in another county in North Carolina, how does that count if they had data, so there's a lot of questions asked. (Principal)
>
> And I'd like to see the research behind whether or not [predictors] are efficient and effective and [if there's] something to help us. (Teacher)

Teachers also want to understand more about how to use student growth data to tailor instruction and ultimately help their students:

> I think the better you understand these measures . . . and what they are using to predict, that knowledge is going to be a good thing. I would like to know how I can use to it to help the students. Sometimes it's just a lot of numbers, and you don't always know what's behind those numbers. . . . I would like to know more so that I can further help my kids.
>
> I would like to know how to use the predictive data. Like, I now understand how all the numbers work together to come up with that, but how do I take that back to the . . . individual students in front of me? How does that impact my instruction? . . . Really, what does Standard 6 mean, more than just sitting and reading it? Like, really, how does that affect all of us?

General knowledge about the Common Exams. At the time of these interviews (fall 2012), LEAs had not administered the first set of Common Exams. The Evaluation Team is currently collecting data regarding fall 2012 implementation of Common Exams, but this report reflects only participant knowledge about the implementation process and projected plans. Table 18 (following page) indicates that, as of fall 2012, teachers report a limited understanding about the Common Exam implementation process. Of the teachers interviewed, 38% (19 out of 50) indicated that they teach a course that would be tested by a Common Exam. Of those teachers, 74% reported that they know when implementation of Common Exams will occur for their grades, and 53% reported that they understand how Common Exams were developed and would be used as part of the NC TEP. They had less knowledge regarding items on the exam; only 11% had previewed sample Common Exam items. A majority of these respondents did not know

about the mode of Common Exams (58%), how they would be proctored (53%), or the scoring process for Common Exams (74%).

Table 18: Knowledge about Implementation of Common Exams

Mode of Common Exams?	% of Respondents (*n*=19)	Proctoring of Common Exams?	% of Respondents (*n*=19)	Scoring of Common Exams?	% of Respondents (*n*=19)
Online	5% (*n*=1)	Yes	37% (*n*=7)	Externally	5% (*n*=1)
Paper	21% (*n*=4)	No	11% (*n*=2)	By Teachers	11% (*n*=2)
Mixed	16% (*n*=3)	Don't Know	53% (*n*=10)	Others	11% (*n*=2)
Don't Know	58% (*n*=11)	—	—	Don't Know	74% (*n*=14)

Questions Regarding Common Exams	% of Respondents (*n*=19)
Knows When Implementation of Common Exam will Occur for Their Grade	74%
Previewed the Common Exams to be Used in Grade	11%
Understand How Common Exams developed and used as part of NCEES	53%

Teachers indicated that they learned about Common Exams through LEA and school resources. Open-ended interview data with *all* teacher participants (*n*=50) revealed that teachers received information about Common Exams from several sources, including school administrators, the NCDPI website, and second-hand knowledge or rumors from other colleagues. While teachers commented that they had received some information regarding the Common Exams, their impressions were that this information was, at times, incomplete or lacked coherence. Teachers said:

> I have seen some examples of it, some of the released questions . . . but having not had students sit for those things and getting feedback on the actual MSLs or Common Exams leaves a lot of uncertainty. . . .
>
> As I say, what we've been told, it has been good, but it just seems like we will find out one thing, and then next week, it might change a little bit.
>
> I know that at some point eventually it's going to be tested based on the Common Core standards that I'm teaching. I've heard it might be 2013. I've heard that it might be next year. I'm just not sure when it's going to start, but I know eventually it will.

Perceptions about the Common Exams. One primary sentiment expressed by teachers and principals about the new Common Exams was the need for additional time to learn more about the implementation process and to learn about the test itself. For instance, a perceived lack of information about the content that would be covered on the Common Exams has contributed to rising concerns about the ability to prepare their students. One teacher noted:

> We don't have any guides to go by, and that's probably one of the hardest things, just a little bit, because even the things that you can purchase, we're not sure that that's going to help us until we get past through this first round of testing.

Related to this, some participants noted that the simultaneous roll-out of Standard 6 assessments with implementation of the new Standard Course of Study creates heightened pressure for aligning the new curriculum with the assessment. Many teachers and administrators expressed concern that they were still adjusting class instruction to the Common Core and Essential Standards and that they lacked sufficient time to successfully align their curriculum while preparing their students for a new high-stakes assessment. One principal noted:

> It's just all new, and I think especially if you've been around for a little while, you're just so caught up in what you know, and right now I think most people are just . . . wrapping [their] heads around Common Core and trying to learn the new curriculum, so you can help support your teachers in that, that this [Standard 6] really has taken a back burner, and I haven't paid as much attention [to it] . . . and I really feel like I need a lot more help with all of those things.

Additionally, teachers and principals expressed concerns about the scoring process for the Common Exams. Common Exams contain performance tasks (i.e., short-answer) that must be scored by two readers using a rubric. LEAs and schools are given flexibility to determine who scores the exams. "Scorers may be the teacher of the class of students tested and another teacher at the base school or scorers may be teachers located at another school" ("Measures of Student Learning: NC's Common Exams, 2013).

Teachers and principals felt that it could be a conflict of interest for teachers to grade the performance tasks for their own students, particularly when the scores would be used as a component of their evaluation. Teachers and principals said:

> I would have concerns about the quality and about the grading and how they're going to be scored, the rubric, and whether they're going to be scored fairly and who's going to score them. (Teacher)

> To me, it seems very specious reasoning to have me grade your students' papers, MSLs, when I know that the results will have a result on you. I don't see how that inherent bias is going to be accounted for. (Principal)

Furthermore, some teachers felt that scoring the performance items on the Common Exams using a rubric is too subjective. For example, teachers with an EOG or EOC commented on the fact that their assessments consisted of multiple-choice items and that the scoring was "cut and dry. With it all being multiple choice, you either say it was right or it was wrong." Conversely, they felt the Common Exams employed performance items and a different scoring process than the EOCs and EOGs, which made the tests too different. An EOG teacher commented:

> I like the EOG personally because it's more like "This is the question, this is the information you're taught, and this is how we're going to basically understand the information." So, with rubrics, there's more leeway, versus with a math test, you either

> got it right or wrong. With the rubric, it's like, there's a range. . . . [T]he rubric seems more broad.

The use of performance items on the Common Exams also prompted many respondents to make comparisons between Common Exams and EOG/EOC assessments. Some participants expressed that they believed that performance items could be a more authentic way of assessing student learning. One EOG teacher said:

> I feel that it gives the students a better opportunity to apply those 21st Century learning skills. . . . [I]t might be a little more fair instead of the paper/pencil setting for a long time . . . that we normally complete for the EOG. So, I do feel that . . . the rubric may be fairer to the student.

On the other hand, many participants felt that the performance items on the Common Exams made the tests too different from the EOGs and EOCs to serve as comparable measures for a teacher's effectiveness rating. One teacher said:

> I don't think it's fair to give me one set of standards and someone else another set of standards by which they are graded. It just does not work, and it's still fluctuating, too.

A principal hypothesized that the perceptions of fairness regarding the different tests will factor in to teachers' grade-level and content area decisions:

> You're going to have a hard time finding [grade] 3-5 teachers if they feel like the K-2 assessment or a 6th grade social studies assessments isn't weighted the same. If they feel like it's any easier, or if they feel like you get graded differently, then absolutely [they will change grades].

A teacher of a course with a Common Exam supported this sentiment, stating that although the basic rationale for performance items on student assessment was positive, the subjectivity in grading could be unfair for students and teachers:

> I think it's a step in the right direction. I have concerns about the equity of one test versus another test. You know, if you have subjective reviewers of the test data, how are we going to have quality control on that, and how exactly is that going to be adjusted to be fair for the students and for the teachers?

Summary of the Use of EVAAS Data to Inform Instruction and Perceptions of Student Growth Data in Evaluation

This report's qualitative findings include participant views on the use of EVAAS data to inform instruction, perceptions of the use of student growth data in evaluation, and information on how Common Exams are being implemented in LEAs.

<u>Summary of Using EVAAS Data for Evaluation and to Inform Teaching Practices</u>:

- A majority of principals were able to access EVAAS data through direct log-in; however, some principals received EVAAS information (such as print-outs of scores) through their

central offices. Teacher access to EVAAS data varied; only 16 out of the 50 teachers (32%) interviewed in this study reported accessing EVAAS data as of fall 2012.[38]

- Principals tend to use EVAAS data to make school-based decisions and to establish a dialogue within the school community regarding instructional practices that bring about student growth. Teacher use of EVAAS was mixed. Some teachers did not understand the role of EVAAS data to inform instruction and misclassified EVAAS data with other student growth data. Other teachers used EVAAS data to identify students for remediation or to customize instruction.

Summary of Teachers' and School Leaders' Perceptions of the Use of Student Growth Data in the Evaluation Process:

- Participants had some knowledge of Standard 6 and EVAAS, but a majority had little knowledge regarding the Common Exams.
- Teachers and principals acknowledged that their students' academic growth was a critical indicator of educator effectiveness. If used formatively, teachers and principals hypothesized that student growth data in evaluations could be used to reflect upon, discuss, and modify their instructional practices.
- Some teachers and principals perceived the addition of Standard 6 as a source of stress in an already-demanding career and raised concerns about potential attrition in the profession due to added stress.
- Some teachers thought the calculation of a Standard 6 rating was based on student achievement rather than student growth. Related to this, teachers also expressed a misconception and concern about the idea that one measure, or one test, would dictate their Standard 6 effectiveness rating
- Both teachers and principals expressed a desire to understand how student growth data is calculated to form an effectiveness rating.

Summary of Implementation of Common Exams across LEAs:

At the time of these interviews (fall 2012), LEAs had not administered the first set of Common Exams. Therefore this report addresses pre-implementation perceptions only.

- Many teachers and principals expressed concern that they were still adjusting class instruction to the Common Core and Essential Standards and that they lacked sufficient time to successfully align their curriculum while preparing their students for a new high-stakes assessment.
- Some teachers and principals felt that the scoring process of the performance items on the Common Exams could be too subjective. They believed that it could be a conflict of interest for teachers to grade the performance tasks for their own students, particularly when the scores would be used as a component of their evaluations.

[38] Some teachers in this sample were interviewed as early as September 2012. The timing of the interviews prevents direct comparison to findings regarding prior access of EVAAS.

- Participants expressed concern and confusion about the comparability of the Common Exams with the EOGs and EOCs.

Teachers and principals demonstrated awareness that student growth data will be used to calculate their effectiveness ratings. They were also aware that the Common Exams are intended to provide a measure of student learning for all teachers, regardless of their grade or subject area. They knew less about how the effectiveness rating is calculated and how Common Exams will be implemented and scored. Because they knew these ratings have implications for their evaluations, teachers and principals expressed concerns about the fairness of using tests for their evaluation ratings without adequate preparation. Coupled with the rollout of other initiatives, such as the new Standard Course of Study, some reflected that this assessment may negatively affect teacher morale. While teachers and principals believed that educators should be held accountable for student growth, they desired more information about how to use the data to inform their instruction.

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 the Tripod student survey. The Tripod constructs *general pedagogy* and *challenge* most strongly predict teachers' value-added scores. The *classroom management* construct predicts teachers' value-added scores but much less strongly than the other two constructs. Measures of teaching effectiveness provided by 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 (as measured using the RttT Evaluation Omnibus Survey) 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 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. Reflecting

findings for Research Question 2, 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.

Conclusion

The utility of performance evaluations is predicated upon not only how well performance is measured but also what actions are taken with the resulting information (Taylor and Tyler, 2011). The findings of this report suggest that North Carolina's measurement of teacher performance by means of the added 6th Standard is consistent with several other key measures of dimensions of teacher effectiveness. While these other dimensions are important on their own, a teacher's value-added, as measured by her or his EVAAS score, provides valuable insight into how those dimensions translate into student achievement gains. What remains to be seen, however, is whether statewide implementation of this new component of the evaluation process will result in a clear understanding of its measurement and consistent use of the resulting information by educators, so that the potential benefits of this new standard are not eclipsed by errors and misconceptions. Addressing uncertainties and possible misconceptions in a timely manner may help to buffer these possible hindrances.

The addition of the student growth standards to the NCEES comes at a time of great change in North Carolina public schools. Common Core and Essential Standards are replacing the previous Standard Course of Study with which teachers and principals are familiar, and new Measures of Student Learning are being added in previously untested grades and subjects. These changes are an important and necessary part of a large-scale reform effort, but as qualitative data from this report demonstrates, it can be difficult in the midst of so much change to clearly communicate updates to teachers and principals. Consequently, participant perceptions shared in the qualitative findings of this formative report should be interpreted within the context of all of the challenges associated with this initial phase of the incorporation of student growth measures into the NCEES.

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.[39] 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

[39] 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, including 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 elminiate entirely individual rater bias. Accordingly, the expansion of the scale also should be accompanied by evaluator training.

Next Steps in the Evaluation

The Evaluation Team will:

- Continue qualitative data collection to track changes in perceptions of the NCEES and use of student growth data to inform instruction over time;
- Follow implementation of the Common Exams and track the use of student growth data to inform instruction;
- Provide an up-close view of best practices in data-driven instruction through case studies of early adopters;
- Continue to observe the teacher sample in 2013-14 using the CLASS instrument;
- As additional data become available, conduct analyses in response to the following question: What characteristics of Common Exam implementation are correlated with score distributions necessary to calculate Value-Added scores?; and
- As additional data become available, conduct additional teacher and leader subgroup analyses that are related to key evaluation questions.

References

Albaugh, D., McBroom, R., & Stewart, H. (n.d.) Using EVAAS to improve student performance (PowerPoint slides]. Retrieved from North Carolina Department of Public Instruction website http://www.ncpublicschools.org/ready/resources/

Danielson, C., & McGreal, T. (2005) *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*. 8(1).

Ferguson, R.F. (2008). *The TRIPOD Project framework*. Cambridge, MA: Harvard University.

Garland, R. & Quick, A. (2013). *MSL/Common Exam spring 2013 update*. Retrieved from http://www.ncpublicschools.org/docs/effectiveness-model/resources/2013/02-memo.pdf

Garland, R., Johnson, L., & Preston, J. (2013). *Educator Effectiveness policies and update*. Retrieved from http://www.dpi.state.nc.us/docs/effectiveness-model/measures/201303-sbe-update.pdf

Marsh, H.W., & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist* 52.11: 1187-1197.

Measures of Student Learning: NC's Common Exams. (2013) Retrieved from http://www.dpi.state.nc.us/docs/effectiveness-model/measures/implementation-guide.pdf

North Carolina Race to the Top Proposal (2010) Retrieved from http://www.ncpublicschools.org/rttt/state/

Pianta, R.C., La Paro, K., & Hamre, B. K. (2008) *Classroom Assessment Scoring System*. Baltimore: Paul H. Brookes.

Popham, J.W. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education, 1,* 269-273.

Rivkin, S., Hanushek, E. & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrica*. *73*(2), 417-458.

Rose, R., Henry, G., & Lauen, D. (2012). *Comparing value-added models for estimating teacher effectiveness*. Retrieved from http://cerenc.org/wpontent/uploads/2011/10/EE_VAM_Briefing_2-7-12.pdf

Rowan, B., Correnti, R., & Miller, R.J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects student of elementary schools. *Teachers College Record*. *108* (2), 1525-1567.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Research progress report. Knoxville: University of Tennessee Value-Added Assessment Center.

Taylor, E. S., & Tyler, J. H. (2011). *The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers* (No. w16877). National Bureau of Economic Research.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect. *Brooklyn, NY: The New Teacher Project*.

WestEd (2010) *Options for Incorporating Student Academic Growth as One Measure of the effectiveness of Teachers in Tested Grades and Subject;* A Report to the North Carolina Department of Public Instruction

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

Appendix A. Scope of Work: Measures of Student Growth in the NC Educator Evaluation System

Overview

This evaluation will examine the perceptions, implementation, and outcomes of the addition of student growth data to the Teacher and Principal Evaluation Process (TEP & PEP). Correlations between measures of student growth and other measures of teacher performance will be assessed, as well as variability in these outcomes between different LEAs and schools in the state.

The implementation of the new evaluation standards will be examined to determine if the addition of student growth measures, including Common Exams, provides independent information that allows for meaningful assessment of teacher and administrator effectiveness. Qualitative analyses, including independent teacher observations and teacher and principal interviews, will be conducted to assess the impact of the new evaluation process on educators' attitudes and practices.

Quantitative analyses will be conducted to evaluate the relationship between different measures of teacher performance, how these vary by subgroups, and the characteristics of new student growth measures (i.e. Common Exams) that provide score distributions necessary to calculate value-added scores.

RttT Initiative Context

Policy Objective(s)/Purpose(s) of the Initiatives

- Fully implement the new teacher and principal evaluation processes statewide;
- Add requirement for explicit student growth data component in teacher and principal evaluation processes; and
- Develop with the advice of the Educator Effectiveness Workgroup, a long-term, uniform system for integrating student growth data into evaluations for all teachers and principals.

Initiative Activities

- Add a 6^{th} standard to the Teacher Evaluation Process and an 8^{th} standard to the Principal Evaluation Process for measures of student growth;
- Provide professional development on the existing evaluation tool and changes to the evaluation components and process;
- Provide targeted professional development based on level of performance as measured by NC Educator Evaluation System (NCEES);
- Establish teacher design groups to create assessments (pre- and post-) for untested courses;
- Engage accountability staff to test reliability and validity of assessments recommended by the workgroups;

- Develop and validate a student academic growth factor; and
- Develop longer-term systems for integrating student growth data into teacher and principal evaluations for all teachers and principals (including growth measured through the new assessments developed for untested courses).

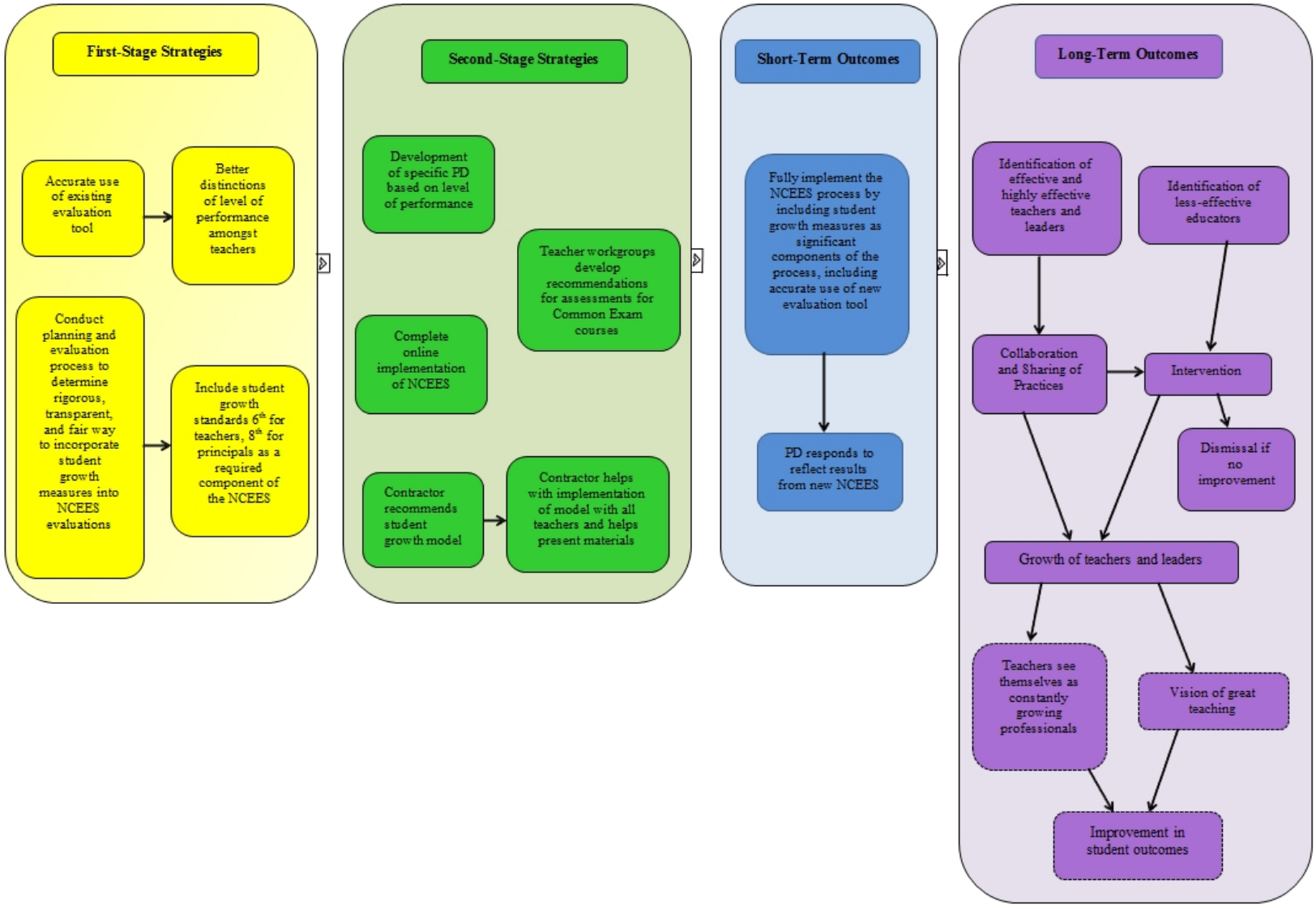
Evaluation Goal(s)/Purpose(s) of the Evaluation

- Ensure quality, consistency, and fairness of new and ongoing teacher and principal evaluation processes through examination of validity and reliability of measures of student learning gains using multiple measures of teacher and teaching effectiveness.
- Examine educators' perspectives on new evaluation standards and the effect of these standards on educators' practices and attitudes. Compare practices and outcomes of the NCEES under the new system with those under the old system.

Overall Approach to Evaluation

Mixed-method: Evaluation questions to be addressed by applying analyses from multiple qualitative and quantitative sources.

Logic Model of Initiative



Research Questions & Anticipated Data Sources

Projected/Proposed Analysis Tool	**Document/ Course Review**	**Educator Evaluation Tool Results**	**Observations** (Classroom/ Institute/ Workshop/ Other)	**Interviews** (Teacher/ Admin/ Other)	**Surveys** (Student/ Teacher/ Other)	**Quant. Analysis**	**Admin. Data Review**	**Accounting Data Review**
Evaluation Question								
Primary Research Questions								
How does the 6th standard correlate with the other five standards?		X	X		X	X	X	
How are teachers and principals using value-added and Common Exam data for evaluation and to inform teaching practices?				X	X	X		
What are teachers' and school leaders' perceptions around the use of growth data in the evaluation?				X	X			
How are Common Exams being implemented across LEAs?		X			X	X	X	
What characteristics of Common Exam implementation are correlated with score distributions necessary to calculate Value-Added scores?		X			X	X	X	
Secondary Research Questions								
How does the 6th standard correlate with other measures of teacher and teaching effectiveness? (e.g., CLASS, Tripod, etc.)		X	X		X	X	X	
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, etc.)		X	X			X	X	
How did school leaders use growth measures in teacher evaluations before implementation of the 6th and 8th standards?				X	X	X		

Projected/Proposed Analysis Tool	Document/ Course Review	Educator Evaluation Tool Results	Observations (Classroom/ Institute/ Workshop/ Other)	Interviews (Teacher/ Admin/ Other)	Surveys (Student/ Teacher/ Other)	Quant. Analysis	Admin. Data Review	Accounting Data Review
Secondary Research Questions (cont.)								
Are there any identifiable correlates between past use of growth data and teacher, leader, school, or LEA characteristics?		X	X		X	X	X	
How do teachers' and leaders' perceptions of the use of student growth data in evaluation vary by subgroup?		X		X	X		X	
Do educator's feel differently about the use of student growth data for EOGs/EOCs teachers compared with Common Exam teachers?		X		X	X		X	
How do the distributions of COMMON EXAM scores vary across testing approaches as well as Common Exam subjects?		X				X	X	
If Common Exams can be used for VAMs, how does it correlate with other measures of teacher effectiveness?		X		X	X	X	X	

Evaluation Activities

Approach

The evaluation will have four primary foci: educators' use of student growth data (past and present); educators' perceptions around the use of student growth data in evaluation; the correlation of student growth data with other measures of teacher quality, including the other standards in the Educator Evaluation Process; and the correlation between implementation characteristics of Common Exams with student score distributions.

The overall approach is mixed method. This approach allows us to provide formative feedback before the quantitative data will be available, as well as to provide a richer context once those outcome data are eventually reported. Additionally, longitudinal tracking of respondents' beliefs and practices can demonstrate change over time and qualitative analysis of open-ended items may provide insights into key issues that emerge once the new standard is implemented.

Evaluation questions will be addressed by applying analyses from both qualitative and quantitative sources, as shown in the following table.

Data Sources

Data Source	Teachers	Principals
Semi-structured interview data collected at up to six time-points over three years	✓	✓
Independent observations of teaching quality[(1)] at up to six time-points over three years	✓	
Student survey data[(2)] collected during the 2011-12 academic year in pilot schools and, if continued, in future years as well.	✓	
Teacher and principal survey[(3)] collected four times in four years, including two pre-implementation years (approx. 350 schools).	✓	✓
Principal observations of teaching quality.[(4)]	✓	
Administrative data on students and teachers including EVAAS scores, teacher mobility and retention, and teacher certification and professional development.[(5)]	✓	✓

(1) CLASS observation protocol; (2) Tripod student survey; (3) RttT Teacher & Principal Omnibus Survey; (4) NCDPI observation rubric (NC TEP); (5) CIPP administrative database

Sample

The procedure for generating the TLEE sample has been an iterative process beginning in the 2011-12 school year. The first step selected 100 schools randomly from the stratified random sample generated by the Omnibus Survey sampling procedure. This generates a sample such that the underlying distribution of schools is represented in the "TLEE 100". From that sample the second step consisted of principals selecting two teachers that met certain selection criteria (EOC/EOG teachers with at least three years of teaching experience). At the time of this selection, only EOC/EOG teachers with at least three years of experience were selected into the sample, and approximately 80 of the 200 teachers taught in content areas that would receive both EOC/EOGs as well as Common Exams. However, teachers that only teach Common Exams are not presently in the sample and will need to be added. The final step will involve a purposeful selection of a subset of schools from the current TLEE 100 where Common Exams will be administered in the 2012-13 school year, and identify the teachers that teach only Common Exam courses. We will then replace approximately 35 teachers from the first wave of EOC/EOG teachers who had incomplete data with these Common Exam-only teachers, to keep a total sample size of approximately 200 teachers to follow with CLASS observations and a sample of 75 interviews with teachers and principals.

It is important to note that this sample is primarily meant to follow teachers as opposed to schools. Accordingly, teachers who are no longer within their original school but otherwise still within the NC public education system will remain in the sample.

The qualitative study will provide formative and summative data on the implementation of the new student growth components' progress and impacts. Data collection will incorporate a variety of methods conducted during two annual visits to each of 100 schools chosen from a probability sample, including classroom observations of two teachers within each school, and face-to-face

interviews with a purposeful sample of 50 teachers and 25 administrators within the schools. In total, over 300 interviews will be coded and analyzed for key patterns and changes between pre-implementation and post-implementation responses linked to the evaluation questions.

The *quantitative* study will provide formative and summative assessments of the implementation of the Educator Evaluation System. These assessments will provide analysis of the correlations between the added 6th standard and the other five standards used in evaluating teacher effectiveness. Correlations will also be estimated between the 6th standard and other teacher evaluation tools such as CLASS observations and Tripod surveys. These relationships will be estimated across various partitions of the sample including school and teacher performance as well as school and teacher characteristics. Implementation characteristics of the new Common Exam administration will be assessed and correlated with the distribution of Common Exam scores for use in value-added modeling. The distribution of scores will also be examined for consistencies across testing approaches and subjects.

Anticipated Schedule:

- First stage (July 2012-Sept 2012)
    - Preliminary analysis of existing interview data.
    - Develop plans for waves 3 and 4 (2012-13) observation and interview data collection; develop new interview scripts.
- Second stage (Oct 2012-June 2013)
    - Analysis and reporting for formative feedback report.
    - Conduct teacher observations (*n*=~200) and teacher and principal interviews (*n*=75).
- Third stage (July 2013-September 2013)
    - Outline of final report developed and provided to DPI.
    - Analysis of waves 3 and 4 interview data.
    - Develop plans for waves 5 and 6 (2013-14) observation and interview data collection.
- Fourth stage (October 2013-September 2014)
    - Conduct teacher observations (*n*=~200) and teacher and principal interviews (*n*=75).
    - Analyze all current and previous observation, interview and survey data collections towards developing the final evaluation report.
    - Analyze administrative data.
    - Prepare 2014 evaluation report.

Major Evaluation Deliverables

- Formative feedback evaluation report — April 30, 2013
- Final evaluation report — September 30, 2014

Appendix B. CLASS Observation Score Sheet and Protocol

Score Sheet

Comprehensive Class Observation Score Sheet		
School: _____ **Grade/Subject**: _____ **Teacher**: _____ **Lesson Content**: _____	**Instructional Format** (circle all that apply): Whole Class Small-Group Individual **# of Adults**: _____ **# of Students**: _____ **Interruptions (**tally/0=none**)**:_____	**Observer**: _____ **Observation Date**: _____ **Day of the Week**:_____ **Start Time**: _____ **End Time**: _____ **Session #** (circle): 1 2 3 4

Domain	Dimension/Indicators	Observations	Score
Emotional Support	**Positive Climate (PC)** • Relationships • Positive Affect • Positive Communications • Respect		
	Negative Climate (NC) • Negative Affect • Punitive Control • Disrespect		
	Teacher Sensitivity (TS) • Awareness • Responsiveness to academic and social/emotional needs • Effectiveness in addressing problems • Student Comfort		
	Regard for Student Perspectives (RSP) • Flexibility and adolescent focus • Connections to current life • Support for autonomy and leadership • Meaningful peer interactions		
Classroom Organization	**Behavior Management (BM)** • Clear expectations • Proactive • Effective redirection of behavior • Student behavior		
	Productivity (P) • Maximizing learning time • Routines • Transitions • Preparation		
	Instructional Learning Formats (ILF) • Learning targets/organization • Variety of modalities, strategies, and materials • Active facilitation • Effective engagement		

Domain	Dimension/Indicators	Observations	Score
Instructional Support	**Content Understanding (CU)** • Depth of understanding • Communications of concepts and procedures • Background knowledge and misconceptions • Transmission of content knowledge and procedures • Opportunity for practice of procedures and skills		
	Analysis and Problem Solving (APS) • Inquiry and analysis • Opportunities for novel application • Metacognition		
	Quality of Feedback (QF) • Feedback loops • Scaffolding • Building on student responses • Encouragement and affirmation		
	Instructional Dialogue (ID) • Cumulative content-driven exchanges • Distributed talk • Facilitation strategies		
	Student Engagement (SE) • Active Engagement		

Protocol

Observation Sessions: Timing

- CLASS observations officially begin at the start of the class period
  - Begin at the bell
  - If no bell, begin at the time the class officially starts
- CLASS observation sessions include 15 minutes of observing and 10 minutes of scoring
- Observers complete four observation sessions for a teacher
- Goal of two teacher observations per day for the researcher; one morning and one afternoon

Observation Sessions: Classroom Placement

- CLASS observers situate themselves on the side of the classroom (profile view of the students) or in a front corner of the classroom
  - Observers do not situate themselves behind students, or directly in front of students
- CLASS observers remain stationary in the classroom. If the ability to observe is compromised by remaining stationary (observers cannot hear student-student or student-

teacher dialogues), researchers may reposition themselves with as little disruption to the class as possible.

Observation Sessions: Focus of Observation

- CLASS is a protocol designed to gauge the classroom experience for the average student. During whole group instruction, this means the researcher should primarily focus on the teacher and his/her interactions with students. If students break into small groups, the researcher should focus their attention across each of the groups to gauge the average student experience.
- CLASS observers should NOT take into account materials in the classroom (daily agenda or objectives) or what is written on the board, unless the teacher specifically references such materials during instruction.

Observation Sessions: Interruptions

- If an observation is interrupted by an event (e.g., a fire drill), the observation can be scored if 8 or more minutes of time have passed. If less than 8 minutes have elapsed, the observation session should be scrapped and started again.

Observation Sessions: Evaluation Considerations

- CLASS observations must be scheduled (not a surprise) to ensure the day/lesson is appropriate for evaluation.
- All raw scores for all dimensions and observations should be reported.
- CLASS observers should observe the same teacher in the fall and spring and year to year if possible
- CLASS observers must observe in a mix of regions and treatment conditions

Appendix C. 2012-13 Interview Protocols

Measures of Student Growth in the NC Educator Evaluation System
Interview Protocol
(Principals)
FALL 2012

SCRIPT: *I am going to ask you some questions about the educator evaluation process at your school. Your answers will be kept confidential and will not be shared with anyone at your school or in your district. Your responses will help inform both program and policy decisions surrounding the implementation of the NC Educator Evaluation System. You may choose not to answer any question for any reason, and you may stop the interview at any time.*

This interview has two parts: The first half will consist of "close-ended" questions, meaning we won't probe for additional information outside of the pre-assigned response categories.

In the second half of the interview we will explore some of these topics in more depth. However, if any points of discussion arise during this first portion please ask me to take note of the topic and we can be sure to address in Part B.

We will only tape record the second half of the interview, so I may ask some follow-up questions or ask you to repeat some of your thoughts from the first half if necessary.

Do you have any questions before we begin?

<<NOTE to interviewer: If respondent begins to supplement the categorical answers with a anecdotal or deeper response, please ask them if you can take a note of the key ideas so they can be discussed when we have the tape recorder on in the second half. Even if they don't want to be recorded, it should be saved for the second half protocol. >>

PART A: QUANTITATIVE QUESTIONS

NOTE to interviewer: Please read this close-ended portion aloud to the participant and fill in the information. This is a guided interview designed to be read aloud.

(1) Demographic Information:
a. Interviewer Name: _____
b. Date: _____
c. Principal Last Name: _____
d. Principal First Name: _____
e. School Name: _____
f. County/LEA: _____

(2) Please indicate for which of the following topics you have participated in formal professional development, and at what level (District or State):

a) Standard 6, or the addition of student growth to the Teacher Evaluation Process

- ❑ YES → District Level? State Level?
- ❑ NO

b) Standard 8, or the addition of student growth to the Principal Evaluation Process

- ❑ YES
  - o District Level
  - o State Level
  - o Both
- ❑ NO

c) The new Measures of Student Learning, or "MSL" assessments (also referred to as Common Exam)

- ❑ YES
  - o District Level
  - o State Level
  - o Both
- ❑ NO

d) The use of EVAAS (software) to access and make use of data

- ❑ YES
  - o District Level
  - o State Level
  - o Both
- ❑ NO

(3) How well do you feel you understand each of the following components of the teacher and leader evaluation process? Please answer either: No understanding, some, quite a bit, substantial understanding

a) Standard 6, or the addition of student growth to the Teacher Evaluation Process

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

b) Standard 8, or the addition of student growth to the Principal Evaluation Process

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

c) The new Measures of Student Learning (MSL) assessments or the Common Exam

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

d) The use of EVAAS to access and make use of data

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

> The following questions refer to access and use of EVAAS data <u>prior</u> to this current 2012-13 school year….

(4) Have you had access to EVAAS data before this current 2012-13 school year?

- ❑ YES → Since when? ____yrs
- ❑ NO

a) (<u>If yes</u>) What level of data (*check all that apply*)?

- ❑ Student
- ❑ Teacher
- ❑ School

b) (<u>If yes</u>) How did you access it (*check all that apply*)?

- ❑ Direct access through logging on to EVAAS
- ❑ Received printouts / scores from LEAs
- ❑ Other

(5) Have any of your <u>teachers</u> had access to EVAAS data before the 2012-13 school year?

- ❑ YES → Since when? ____yrs
- ❑ NO

a) (<u>If yes</u>) What teachers?

- ❑ All
- ❑ Only teachers in tested grades/subjects
- ❑ Other subset_____

b) (<u>If yes</u>) What level of data? *(Check all that applies)*

- ❑ Student
- ❑ Teacher
- ❑ School

c) (If yes) How did they access it? *(Check all that applies)*
- ❑ Direct access through logging on to EVAAS
- ❑ Received printouts / scores from LEAs
- ❑ Other:_____

(6) Was EVAAS data used as part of your principal evaluation (prior to 2012-13)?
- ❑ YES
- ❑ NO

<<Now, we'd like to know about this current 2012-13 school year to date…>>

(7) For this current school year (2012-13), have you accessed EVAAS data?
- ❑ YES
- ❑ NO

a) (If yes) What level data? *(Check all that applies)*
- ❑ Student
- ❑ Teacher
- ❑ School

b) (If yes) How did you access it? *(Check all that applies)*
- ❑ Directly through logging into EVAAS system
- ❑ Received printouts / scores from principal
- ❑ Other

c) (If yes) Have you used these data in any kind of evaluative process to date for your teachers?
- ❑ YES
- ❑ NO

d) Has EVAAS data been used in any kind of evaluative process to date for you in this current school year?
- ❑ YES
- ❑ NO

<<Now we will ask you a few questions about the new MSLs, or Measures of Student Learning also referred to as Common Exams>>

(8) Will your school be administering Measures of Student Learning or Common Exams this year?
- ❑ YES
- ❑ NO

a) (If yes) Have you received guidance from your district on implementation plans at this point?

- ❑ YES
- ❑ NO

b) (If yes) Is the implementation guidance the same district-wide, or does it vary at the school or class level?

- ❑ Same across the district
- ❑ Varies within districts by schools
- ❑ Varies within schools by classes
- ❑ Don't know

(9) Please indicate how you've implemented or plan to implement so far.

Circle the applicable selection for each category below:

Mode: Online, Paper, Mixed, or Don't Know
Proctored: Yes, No, Don't know
Scoring: Scored Internally, Scored Externally, Don't know

(10) Were any of these types of implementation plans informed by the district?

- ❑ YES
- ❑ NO

This concludes Part A of the interview. We will now move on to the open-ended portion, where we'd like to hear about your experiences in more detail surrounding these topics. We will be tape recording this half of the interview.

PART B: QUALITATIVE QUESTIONS (Begin tape recording here)

Speak into the recorder: Your name, the school name, name of principal being interviewed, and the date prior to proceeding.

(11) Please tell us more about the types of information you have received to date on the addition of student growth data to the *Teacher Evaluation Process?*

→ *PROBES:*

- Describe the information sources: (emails, webinars, district/state trainings etc)
- Describe what has or has not been useful to you
- Describe what you still feel you need/want to know

(12) What kind of information have you received to date regarding the addition of student growth data to the *Principal Evaluation Process:*

→ *PROBES:*
- Describe the information sources: (emails, webinars, district/state trainings etc)
- Describe what has or has not been useful to you
- Describe what you still feel you need/want to know

(13) What is your current understanding of the newly developed Measures of Student Learning (MSLs) or Common Exam?

→ *PROBES:*
- How they were developed?'
- When and for whom they will be implemented?
- How they will be used as part of the Teacher Evaluation Process (TEP)?
- Describe what you still feel you need/want to know about MSLs.

(14) What are your thoughts regarding the implications of rubric-based scoring for the new MSL assessments or Common Exam compared with the more traditional EOC/EOC scoring?

→ *PROBES*:
- Any concerns around the quality of the data for MSLs or Common Exam?
- Any concerns about perceptions surrounding a lack of equity in the scoring system?

(15) <*If they've accessed EVAAS prior to 2012-13*>…We'd like to understand the extent to which ***principals*** were using EVAAS data prior to the 2012-13 school year. Can you please walk us through how you accessed and used the data in the past?

→ *PROBES*:
- To what extent were you informed about understanding and using these data? (i.e. school or district PD, web-based training at individual level, etc)
- To what extent was it used to provide feedback to and/or evaluation of teachers?
- To what extent was the data used as a part of *your* evaluation process?
- In what other ways have you used this data? (i.e. PD, class assignments, etc.)

(16) <*If their teachers have accessed EVAAS prior to 2012-13*> We'd also like to understand how ***teachers*** were using EVAAS data prior to the 2012-13 school year. Can you please walk us through who (which teachers) would have access to it, how they would access it, and what would be done with the data?

→ *PROBES:*
- To what extent were they informed about understanding and using this data?
- Please provide any examples you are aware of about how teachers have used EVAAS data in the past.

(17) Describe your experience with EVAAS student growth data, if any, ***in this current school year.***

→ *PROBES:*

- To what extent has it been it used to provide feedback to and/or evaluation of teachers?
- To what extent has the data used as a part of *your* evaluation process?
- In what other ways have you used this data? (i.e. PD, class assignments, etc.)

(18) What are your current feelings about the use of student growth data (EVAAS and MSLs/ Common Exam) in *teacher* evaluations?

→ *PROBES:*

- How does this differ between implementation with EOG/EOC teachers versus MSL/ Common Exam teachers?
- What potential challenges do you see for the addition of this new standard to the Teacher Evaluation Process?
- What do you see as the potential benefits?
- To what extent do you feel the addition of these measures will impact/improve teacher effectiveness?

(19) What are your current feelings about the use of student growth data in *principal* evaluations?

→ *PROBES:*

- What potential challenges do you see for the addition of this new standard to the Principal Evaluation Process?
- What do you see as the potential benefits?
- To what extent do you feel the addition of these measures will impact/improve principal effectiveness?

(20) Describe any areas that you would appreciate further professional development related to the new evaluation standards (6 & 8), the use of EVAAS, and the addition of MSLs/Common Exam, both for yourself and for your teachers.

(21) What other questions, concerns, or comments would you like to share with us at this time?

THANK YOU FOR YOUR TIME & PARTICIPATION!!

Measures of Student Growth in the NC Educator Evaluation System
Interview Protocols
(Teachers)
FALL 2012

SCRIPT: *I am going to ask you some questions about the teacher evaluation process at your school. Your answers will be kept confidential and will not be shared with anyone at your school or in your district. Your responses will help inform both program and policy decisions surrounding the implementation of the NC Educator Evaluation System. You may choose not to answer any question for any reason, and you may stop the interview at any time.*

This interview has two parts: The first half will consist of "close-ended" questions, meaning we won't probe for additional information outside of the pre-assigned response categories.

In the second half of the interview we will explore some of these topics in more depth. However, if any points of discussion arise during this first portion please ask me to take note of the topic and we can be sure to address in Part B.

We will only tape record the second half of the interview, so I may ask some follow-up questions or ask you to repeat some of your thoughts from the first half if necessary.

Do you have any questions before we begin?

<<NOTE to interviewer: Please read this close-ended portion aloud to the participant and fill in the information. This is a guided interview designed to be read aloud.

If respondent begins to supplement the categorical answers with a broader or deeper response, please ask them if you can take a note of the key ideas so they can be discussed when we have the tape recorder on in the second half. Even if they don't want to be recorded, it should be saved for that second half. We want to capture anecdotal information and valuable insights during the recorded portion of this interview.>>

PART A: QUANTITATIVE QUESTIONS (No tape recorder)

(1) Demographic Information:
a. Interviewer Name: _____
b. Date: _____
c. Teacher Last Name: _____
d. Teacher First Name: _____
e. School Name: _____
f. County/LEA: _____
g. Courses currently teaching: _____

(2) Which of the following assessments will be administered in the class(es) you teach this year? (Check all that apply.)

- ❑ EOC
- ❑ EOG
- ❑ Course with a Measure of Student Learning (MSL) or Common Exam

(3) Please indicate for which of the following topics you have participated in formal professional development, and at what level (school, district, or state):

a) Standard 6, or the addition of student growth to the Teacher Evaluation Process (TEP)

- ❑ YES
  - o School Level
  - o District Level
  - o State Level
- ❑ NO

b) The new Measures of Student Learning, or "MSL" assessments (also referred to as Common Exam)

- ❑ YES
  - o School Level
  - o District Level
  - o State Level
- ❑ NO

c) The use of EVAAS (software) to access and make use of data

- ❑ YES
  - o School Level
  - o District Level
  - o State Level
- ❑ NO

(4) How well do you feel you understand each of the following components of the Teacher Evaluation Process? Please answer either: Not at all, a little, a lot, or fully understand.

a) Standard 6, or the addition of student growth to the Teacher Evaluation Process

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

b) The new Measures of Student Learning (MSL) assessments (or Common Exam)

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

c) The use of EVAAS to access and make use of data

- ❑ Not at all
- ❑ A little
- ❑ A lot
- ❑ Fully

> The following questions refer to your access to and use of EVAAS data <u>prior</u> to this current 2012-13 school year.

(5) Have you had access to EVAAS data <u>before</u> the 2012-13 school year?

- ❑ YES→ Since when? ____yrs
- ❑ NO

a) (<u>If yes</u>) What level data? *(Check all that applies)*

- ❑ Student
- ❑ Teacher
- ❑ School

b) (<u>If yes</u>) How did you access it? *(Check all that applies)*

- ❑ Directly through logging into EVAAS system
- ❑ Received printouts / scores from administrator or colleague
- ❑ Other _____

>Now, we'd like to know about this current 2012-13 school year to date.

(6) ***(<u>EOC/EOG Teacher</u>s)*** For <u>this current school year</u> (2012-13), have you received your predicted student scores?

- ❑ YES
- ❑ NO

(7) Have you accessed EVAAS data (for example, school level data)?

- ❑ YES
- ❑ NO

a) (<u>If yes</u>) What level data? *(Check all that applies)*

- ❑ Student
- ❑ Teacher
- ❑ School

b) (If yes) How did you access it? *(Check all that applies)*
- ❑ Directly through logging into EVAAS system
- ❑ Received printouts / scores from principal
- ❑ Other

c) Has EVAAS data been used by your principal as part of your evaluations to date?
- ❑ Yes
- ❑ No

(8) If you are teaching a course this year that receives a Measure of Student Learning (MSL) or Common Exam, please answer the following:<<***NOTE:*** *If the participant is* ***not*** *teaching a MSL course, please move on to Question 11*>>

a) Do you know when implementation of the MSLs or the Common Exam will occur for your grades/courses?
- ❑ YES
- ❑ NO

b) Have you previewed the MSL or Common Exam to be used in your grade/course?
- ❑ YES
- ❑ NO

c) Do you understand how the MSLs or Common Exam were developed and how they will be used as part of the Teacher Evaluation Process?
- ❑ YES
- ❑ NO

d) Have you received any guidance on how the MSLs or Common Exam will be implemented this year?
- ❑ YES
- ❑ NO

e) (If yes) Is the guidance from your school or district?
- ❑ School
- ❑ District
- ❑ Both
- ❑ Neither

(9) Please indicate your MSL or Common Exam implementation plans to date.

Circle the selection for each category below:
1. Mode/What is the type?: Online, Paper, Mixed, or Don't Know
2. Proctored/Were they or will they be proctored?: Yes, No, Don't know
3. Scoring/How is or will scoring done?: By external graders, by teacher, other _____, Don't know

Were any of the implementation plans mentioned above informed by the district If so, which ones (mode, proctoring, scoring)?

(10) To your knowledge, how will (not just for your course) the MSLs or Common Exam be administered in the (as in the question above) in your school?

- ❑ Same across the school
- ❑ Varies between MSL classes
- ❑ Don't know

This concludes Part A of the interview. We will now move on to the open-ended portion, where we'd like to hear about your experiences in more detail surrounding these topics. We will be tape recording this half of the interview.

PART B: QUALITATIVE QUESTIONS (<u>Begin tape recording here</u>)

Speak into the recorder: Your name, the school name, name of teacher being interviewed, and the date prior to proceeding.

(11) Please tell us more about the types of information you have received to date on the addition of student growth data to the <u>*Teacher Evaluation Process?*</u>

→ *PROBES:*
- Describe the information sources: (emails, webinars, district/state trainings etc)
- Describe what information has or has not been useful to you
- Describe what you still feel you need/want to know

(12) *(If they have accessed EVAAS)…* We'd like to better understand the extent to which teachers were using EVAAS data ***<u>prior to the 2012-13 school year</u>***. Can you please walk us through how you've accessed and used the data in the past?

→ *PROBES*:
- To what extent were you informed about understanding and using this data?
- To what extent was it used to provide feedback to or evaluation of your instruction?
- Provide an example of how you have used EVAAS data to inform your instructional practices in your classroom and/or school (For example, did you change lesson plans or individualize instruction for students based on your data?)

(13) Describe your experience with EVAAS student growth data, <u>if any</u>, ***<u>in this current school year</u>***.

→ *PROBES:*
- To what extent has it been used to provide feedback to and/or evaluation of your teaching?
- To what extent has it been used to inform instruction?

*(*14) What is your current understanding of the newly developed Measures of Student Learning (MSLs) or Common Exam?

→ *PROBES (for example…)*

- how they were developed?
- when and for whom they will be implemented?
- how they will be used as part of the Teacher Evaluation Process (TEP)?
- describe what you still feel you need/want to know

(15) What are your thoughts regarding the implications of rubric-based scoring for the new MSL or Common Exam assessments compared with the more traditional EOC/EOC scoring?

→ *PROBES*:

- Any concerns around the quality of the data for MSLs?
- Any concerns about the scoring system?

(16) What are your current feelings about the use of student growth data (EVAAS and MSLs/ Common Exam) in *teacher* evaluations?

→ *PROBES*:

- How does this differ between implementation with EOG/EOC teachers versus MSL teachers?
- What potential challenges do you see for the addition of this new standard to the Teacher Evaluation Process?
- What do you see as the potential benefits?
- To what extent do you feel the addition of these measures will impact/improve teacher effectiveness?

(17) Describe any areas that you would appreciate further professional development related to the new teacher evaluation standard, the use of EVAAS, and the addition of MSLs/ Common Exam.

(18) What other questions, concerns, or comments would you like to share with us at this time?

THANK YOU FOR YOUR TIME & PARTICIPATION!!

Appendix D. Tripod Student Survey

Secondary Grades Tripod Student Survey Items

Care

1) My teacher in this class makes me feel that she/he really cares about me.
2) My teacher seems to know if something is bothering me.
3) My teacher really tries to understand how students feel about things.

Control

1) Student behavior in this class is under control.
2) I hate the way that students behave in this class.
3) Student behavior in this class makes the teacher angry.
4) Student behavior in this class is a problem.
5) My classmates behave the way my teacher wants them to.
6) Students in this class treat the teacher with respect.
7) Our class stays busy and doesn't waste time.

Clarify

1) If you don't understand something, my teacher explains in another way.
2) My teacher knows when the class understands and when we do not.
3) When she/he is teaching us, my teacher thinks we understand even when we don't.
4) My teacher has several good ways to explain each topic that we cover in this class.
5) My teacher explains difficult things clearly.

Challenge

1) My teacher asks questions to be sure we are following along when she/he is teaching.
2) My teacher asks students to explain more about answers they give.
3) In this class, my teacher accepts nothing less than our full effort.
4) My teacher doesn't let people give up when the work gets hard.
5) My teacher wants us to use our thinking skills, not just memorize things.
6) My teacher wants me to explain my answers—why I think what I think.
7) In this class, we learn a lot almost every day.
8) In this class we learn to correct our mistakes.

Captivate

1) This class does not keep my attention—I get bored.
2) My teacher makes learning enjoyable.
3) My teacher makes lessons interesting.
4) I like the ways we learn in this class.

Confer

1) My teacher wants us to share our thoughts.
2) Students get to decide how activities are done in this class.
3) My teacher gives us time to explain our ideas.
4) Students speak up and share their ideas about class work.
5) My teacher respects my ideas and suggestions.

Consolidate

1) My teacher takes the time to summarize what we learn each day.
2) My teacher checks to make sure we understand what he/she is teaching.
3) We get helpful comments to let us know what we did wrong on assignments.
4) The comments that I get on my work in this class help me understand how to improve.

Appendix E. RttT Omnibus Survey Items in Analysis

All of the following items were scales as follows: Strongly Disagree, Disagree, Slightly Disagree, Neither Agree nor Disagree, Slightly Agree, Agree, Strongly Agree, Not Applicable.

Evaluation Validity: Items were obtained from The Chicago Consortium on School Research.

At my school...

1. The evaluation process encourages teachers to reflect on their instructional practice.
2. Teachers use feedback from the teacher evaluation system to improve their teaching.
3. Teacher evaluation is fair.
4. The criteria on which I am evaluated are clear.
5. The teacher evaluation process encourages professional growth.

This school year, overall…

1. I am satisfied with the teacher evaluation process.

Self-Efficacy: Items were obtained from the Pattern of Adaptive Learning Scales and The Teacher's Sense of Efficacy Scale.

1. If I try really hard, I can get through to even the most difficult student.
2. I am good at helping all the students in my classes make significant improvement.
3. I am certain that I am making a difference in the lives of my students.
4. I can deal with almost any learning problem.
5. I can get students to believe they can do well in school work.
6. I can assist families in helping their children do well in school.
7. I can motivate students who have low interest in school work.
8. I can help students value learning.

Test Preparation: Items were created by the Omnibus survey development team.

This school year…

1. I covered the material required by the State Standard Course of Study.
2. I prepared my students for their EOG/EOC exams.
3. I prepared my students to move to the next level of schooling.

Appendix F. Common Exam Implementation Timeline

Implementation Timeline
Measures of Student Learning
NC's Common Exams

First Semester, 2012 - 2013 School Year

Earth/Environmental Science	Physics
Chemistry	Physical Science
English Language Arts I	English Language Arts III
English Language Arts IV	Pre-Calculus
Advanced Functions and Modeling	Geometry*
Algebra II/Integrated Math III	World History
Civics and Economics	U.S. History (2003 standards)
American History I	American History II
OCS English Language Arts I	OCS English Language Arts III
OCS English Language Arts IV	OCS Financial Management
OCS Applied Science	OCS Introductory Math

*Students enrolled in Integrated Math II in the 2012 - 2013 school year will take the Algebra I/Math I assessment.

Measures of Student Learning replace final exams in high school courses.

Second Semester, 2012 - 2013 School Year

Grade Four Science	Grade Six Science
Grade Seven Science	Grade Four Social Studies
Grade Five Social Studies	Grade Six Social Studies
Grade Seven Social Studies	Grade Eight Social Studies

Districts select the Measures of Student Learning that must be administered to gather growth data for all teachers. The administration of each common exam is not required.

August 2, 2012

First Semester, 2013 - 2014 School Year

21st Century Global Geography	Psychology
Sociology	World Humanities
American Humanities	The Cold War
21st Century Civil Liberties and Rights	Turning Points in American History
Grade Nine Health Education	Grade Nine Physical Education

Second Semester, 2013 - 2014 School Year

Grade Six Health Education	Grade Six Physical Education
Grade Seven Health Education	Grade Seven Physical Education
Grade Eight Health Education	Grade Eight Physical Education

More Information Coming Soon

K-3 Measures of Student Learning - focus on literacy instruction

World Languages

The Arts

Extended Content Standards

Important Notes

Measures of Student Learning will be administered to determine teacher ratings on the sixth standard of the NC Educator Evaluation System. The Department of Public Instruction will provide a guide to districts that outlines under what circumstances a Measure of Student Learning must be administered. For example, a fourth grade teacher with Math and English Language Arts End-of-Grade assessment results for students is not required to administer a science or social studies Measure of Student Learning. The sixth standard rating will be based on the End-of-Grade assessment results.

A Measure of Student Learning should not be administered unless a teacher has instructed students for 70 days (for schools using a block schedule) or 140 days (for schools on the traditional schedule).

Please email educatoreffectiveness@dpi.nc.gov with questions

August 2, 2012

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

SERVECenter • Carolina Institute for Public Policy • The University of North Carolina at Chapel Hill • The William & Ida Friday Institute for Educational Innovation, NC State University College of Education