

Consortium for
Educational
Research and
Evaluation–
North
Carolina

An Evaluation of the North Carolina Educator Evaluation System and the Student Achievement Growth Standard

2010-11 through 2013-14

Gary T. Henry and J. Edward Guthrie
Vanderbilt University

September 2015

Consortium for
Educational
Research and
Evaluation–
North
Carolina



Carolina Institute for Public Policy
THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Table of Contents

Executive Summary 2

Introduction..... 6

Findings..... 8

 NCEES Teacher Ratings, 2010-11 through 2013-14 8

 How many teachers needed improvement, according to the NCEES? 8

 To what extent did principals rate teachers below proficient (Not Demonstrated or Developing) or did teachers not meet expected growth according to their EVAAS scores? Have these changed over time? 8

 Summary of Key Findings 10

Principal Feedback 10

 Did principals provide teachers with information on their strengths and weaknesses by making distinctions in performance between the standards? 10

 Have principals’ ratings changed over time in terms of providing teachers with information about their strengths and weaknesses? 11

 Summary of Key Findings 11

Relationships between EVAAS Scores and Principal Ratings..... 12

 Were the EVAAS scores for teachers’ contributions to student achievement growth related to principals’ ratings of teachers’ performance, especially the ratings for facilitating student learning? 12

 Over time, did the performance measures converge, indicating greater alignment between EVAAS scores and principals’ ratings? 13

 Summary of Key Findings 13

Other Measures of Teacher Performance 13

 What instructional practices or teacher behaviors predicted gains or improvements in teachers’ EVAAS scores? 13

 Summary of Key Findings 15

Teacher Perceptions of NCEES..... 15

 What were teachers’ views about their evaluations and related topics during the period in which the NCEES evaluation with six standards has been implemented? 15

 Have teachers’ attitudes changed over time? 16

Summary and Conclusions 17

References..... 19

**AN EVALUATION OF THE NORTH CAROLINA EDUCATOR EVALUATION
SYSTEM AND THE STUDENT ACHIEVEMENT GROWTH STANDARD:
2010-11 THROUGH 2013-14**

Executive Summary

In 2011, as a part of the State Board of Education’s implementation of North Carolina’s Race to the Top (RttT) initiative, a sixth standard—a measure of student growth, the Educational Value-Added Assessment System (EVAAS)—was added to the existing five standards for evaluating teachers. The purpose of this report is to describe the outcomes of teacher evaluations that have occurred since the sixth standard was added and trends in those outcomes through 2013-14.

Evaluation Questions and Key Findings

1. *How many teachers needed improvement, according to the NCEES? To what extent did principals rate teachers below proficient (Not Demonstrated or Developing) or did teachers not meet expected growth according to their EVAAS scores?*

One of the most important purposes of teachers’ evaluations is to identify teachers who need improvement so leadership can intervene in ways that help ensure that students have access to high-quality teaching. Between 2010-11 and 2013-14, for all teachers with both individual EVAAS growth scores and principals’ ratings, 18.4 percent were found to need improvement, but this percentage varied between initially licensed teachers (21.4 percent) and fully licensed teachers (13.5 percent).

Key Findings

- Over 80 percent of teachers in each of the past two years who were assigned to a “Needs Improvement” category (*Not Demonstrated* or *Developing*) were assigned by their EVAAS score alone.
 - Approximately 5 percent of beginning teachers who received ratings on all five standards received one or more ratings below *Proficient* from their principals.
 - The percentages of teachers who received at least one rating by their principals below *Proficient* was 7.7 percent for teachers who did not meet expected growth according to EVAAS, 3.2 percent for teachers who met expected growth, and 1.2 percent for teachers who exceeded expected growth.
2. *Did principals provide teachers with information on their strengths and weaknesses by making distinctions in performance between the standards?*

Another important purpose for the NCEES evaluations was to provide teachers, especially initially licensed teachers who are, on average, less effective and more likely to turnover than more experienced teachers, with clear information about their strengths and weaknesses. Value-added scores, including but not limited to EVAAS scores, provide objective measures of the outcomes of teachers’ instructional practices but, unfortunately, do not provide information about which practices are strengths and weaknesses for individual teachers.

Key Findings

- Principals rated teachers either *Proficient* or *Accomplished* 90 percent of the time, which provided limited information on individual teachers' strengths and weaknesses.
 - Principals' ratings have not varied over time, indicating little refinement in using NCEES ratings to provide teachers with feedback on strengths and weaknesses.
 - Principals rate teachers globally rather than providing meaningful distinctions on teachers' performance on each standard.
3. *Were the EVAAS scores for teachers' contributions to student achievement growth related to principals' ratings of teachers' performance, especially the ratings for facilitating student learning?*

To enhance credibility of NCEES with teachers and to provide consistent information, the agreement between the principal ratings that teachers received and their student growth ratings (EVAAS) should be related. The objective measure of their students' achievement growth could reasonably be expected to relate to principals' ratings of teachers on the NCEES standards, especially the "teachers facilitate learning for their students" standard, which most closely has a bearing on student growth.

Key Findings

- Principals' ratings of teachers' overall performance and of their ability to facilitate student learning were loosely correlated with teachers' EVAAS scores. On average, principals' ratings of their teachers were not influenced by the state's official measure of the teachers' contributions to the achievement growth of their students.
 - The correlations between principal ratings and EVAAS scores varied across individual course subjects. Correlations with EVAAS were higher with science and substantially lower with reading/English/language arts, the mClass, and career and technical assessments.
4. *What instructional practices or teacher behaviors predicted gains or improvements in teachers' EVAAS scores?*

Teachers need actionable evidence about effective instructional practices that can improve their EVAAS scores. In interviews, most teachers reported receiving immediate and constructive feedback regarding their evaluations; however, some reported a lack of feedback from their administrators. Overall, teachers expressed an interest in receiving a higher quality and greater quantity of feedback.

In order to support principals and teachers in the development of more effective teachers, the Evaluation Team examined measures of teachers' instructional practices and behaviors that relate to improvements in their EVAAS scores—that is, related to higher EVAAS scores than predicted based on their prior EVAAS scores alone. Practices found to be effective and ineffective at improving EVAAS scores, as well as practices associated with teachers' capacity to improve their EVAAS scores over time, are listed in Table ES1 (following page).

Table ES1. Measures Associated with Impacting Teachers' Value-Added Scores More than Predicted by Prior Value-Added Scores

<i>Effective Practices</i>	<i>Practices Associated with Teachers' Capacity to Improve</i>
Facilitating Student Learning	Reflection on Practice
Collaborative Environment	Collaborative Environment
Higher-Order Instruction	Higher-Order Instruction
Classroom Management	Classroom Management
Positive Climate	

Ineffective Practices

- Busy Work
- Student-Led Environment

Key Findings

- Both teachers and principals want more information about how teachers can improve their practices, especially in ways that increase their value-added scores.
 - Using measures from principals' ratings, student surveys, and classroom observations we found several measures associated with higher value-added scores than were predicted by teachers' previous value-added scores alone.
 - These measures can be (1) directly incorporated into the feedback provided by principals to teachers, and (2) periodically tested for their relationship to value-added if student surveys and/or observation protocols for classrooms are conducted in the future.
5. *What were teachers' views about their evaluations and related topics during the period in which the NCEES evaluation with six standards has been implemented?*

When the student achievement growth measure was added to the original five NCEES standards, two concerns surfaced. The first concern was that the overall fairness of the evaluations would be eroded—that the protocols designed to promote teacher development would be undermined. Second, concerns were raised that the high-stakes evaluations would inhibit teachers from supporting one another and working together to improve student learning.

From 2011-12 to 2013-14, the favorability of teachers' views of their evaluations has declined significantly. Overall, teachers' rating of the evaluation process as measured by survey items declined from approximately 5.2, which indicated slight agreement with the items, to about 4.8, which moved them toward neither agreement nor disagreement with the items. Contrary to some expectations, teachers engaged in knowledge-sharing more in 2013-14 than in earlier years. Overall, their knowledge-sharing activities increased from once or twice per week to almost daily.

Conclusions

Along with approximately two-thirds of other states, North Carolina adopted value-added scores for individual teachers as an additional, sixth standard to supplement the pre-existing five-standard NCEES teacher evaluation system.

- Because almost all identifications of a need for improvement were based on the value-added score, value-added effectively acted alone to determine teachers' evaluation status, rendering the judgments of principals on all other aspects of teaching and teachers' performance much less important.
- Consideration should be given to systematically adding other direct measures of teaching performance into the NCEES in addition to the EVAAS scores that are currently included. In addition to providing more evidence for teachers who also have EVAAS scores, teachers who do not have EVAAS scores could then have direct measures of their individual performance incorporated into their evaluations.
- The second main conclusion from this evaluation is that teachers are being rated globally, and neither the ratings nor evaluation feedback are providing them with enough actionable information for them to improve.
- The NCEES process seems to have been accepted by teachers and principals. On the whole, they feel that the system is fair. Also, teachers' survey responses indicated that the implementation of the NCEES process has not produced negative side effects, such as decreasing teachers' willingness to share information. In fact, the opposite has occurred—teachers have engaged in more knowledge-sharing across the first four years of expanded NCEES implementation.

The main issue with the NCEES appears to be that the current system includes only one systematic data source—EVAAS—and, while EVAAS is an important objective measure of teachers' effectiveness, the inclusion of additional systematic measures may point out the strengths and weaknesses of individual teachers, increase the accuracy of identifying those who need improvement, increase the favorability of teachers' attitudes toward the evaluation system, and provide direct information about practices that can be used for improvement for all teachers.

Introduction

In 2011, as a part of the State Board of Education's implementation of North Carolina's Race to the Top (RttT) initiative, a sixth standard—a measure of student growth—was added to the existing five standards for evaluating teachers. On the recommendation of the North Carolina Department of Public Instruction (NCDPI), the State Board chose the Educational Value-Added Assessment System (EVAAS) that is operated by the SAS Institute to provide the measure of individual teachers' contributions to student achievement.

Prior to including the sixth standard, the North Carolina Educator Evaluation System (NCEES) for teachers was based on five standards that had been developed by the North Carolina Professional Teacher Standards Commission. The five standards, which are rated by teachers' principals (or other administrators), are: (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 (NCDPI, 2013). For each of these standards, principals rate teacher performance as *Not Demonstrated*, *Developing*, *Proficient*, *Accomplished*, or *Distinguished*, with associated numerical values from 1 through 5, respectively.

The sixth standard—teachers contribute to the academic success of students—is based on a composite score that measures each teacher's contribution to the achievement score growth of her or his students who take state assessments. For teachers who do not teach tested grades or subjects, the school value-added achievement growth is taken as the measure of their contribution to the academic success of students. Currently, 35 states and the District of Columbia have added or are in the process of adding measures of teachers' effects on raising students' test scores (which will be referred to collectively as teacher value-added or EVAAS scores) to the evaluation of individual teachers (Doherty & Jacobs, 2013).

The purpose of this report is to describe the outcomes of teacher evaluations that have occurred since the sixth standard was added and trends in those outcomes through the 2013-14 school year. More specifically, the scope of this evaluation is limited to the teachers with individual measures of their contributions to student learning that are provided through EVAAS. The existing five standards, which were developed prior to the implementation of North Carolina's RttT initiatives, were not directly evaluated, but the relationships between EVAAS scores and principals' ratings for individual teachers are included. The report answers the following five sets of questions:

1. How many teachers needed improvement, according to the NCEES?
To what extent did principals rate teachers below proficient (*Not Demonstrated* or *Developing*) or did teachers not meet expected growth according to their EVAAS scores?
Have these frequencies changed over time?
2. Did principals provide teachers with information on their strengths and weaknesses by making distinctions in performance between the standards?

Have principals' ratings changed over time in terms of providing teachers with information about their strengths and weaknesses?

3. Were the EVAAS scores for teachers' contributions to student achievement growth related to principals' ratings of teachers' performance, especially the ratings for facilitating student learning?

Over time, did the performance measures converge, indicating greater alignment between EVAAS scores and principals' ratings?

4. Could other measures of teachers' performance have been used to provide information to teachers about their strengths and weaknesses, specifically in terms of improving their contributions to student learning?

What instructional practices or teacher behaviors predicted gains or improvements in teachers' EVAAS scores?

5. What were teachers' views about their evaluations and related topics during the period in which the NCEES evaluation with six standards has been implemented?

Have their attitudes changed over time?

To address these five sets of questions, the Evaluation Team assembled a dataset that included all NCEES teacher evaluation ratings recorded between 2010-11 and 2013-14 as well as EVAAS scores for teachers. These data were merged with datasets maintained by the Education Policy Initiative at Carolina that contain student, teacher, and school information and have been used in many of the prior RttT evaluations.¹ In addition, this evaluation used teachers' responses from the RttT omnibus teacher survey that was administered in 2011-12 through 2013-14 to a stratified random sample of North Carolina public schools. Also, the Team interviewed 101 teachers and 38 principals and collected classroom observation data using the CLASS instrument from diverse Local Education Agencies (LEAs)² across the state. These data were used in this report to describe the perspectives of teachers and principals on the measure of student achievement growth and classrooms and to analyze instructional practices associated with increases in EVAAS scores. Finally, the evaluation incorporated data from a pilot of the Tripod student survey in 38 school districts in the state collected during the 2011-2012 school year. For this study, descriptive statistics such as means and standard deviations, bivariate correlations, measurement test statistics (Cronbach's alpha), and multivariate regression were applied.

¹ RttT evaluations were conducted by the Consortium for Educational Research and Evaluation–North Carolina (CERE–NC), a partnership of: the SERVE Center, University of North Carolina at Greensboro; the Education Policy Initiative at Carolina, University of North Carolina at Chapel Hill; and the Friday Institute for Educational Innovation, North Carolina State University.

² LEA is North Carolina's term for traditional school districts and charter schools.

Findings

NCEES Teacher Ratings, 2010-11 through 2013-14

How many teachers needed improvement, according to the NCEES?

One of the most important purposes of teachers' evaluations is to identify teachers who need improvement so leadership can intervene in ways that help ensure that students have access to high-quality teaching. For all teachers with both individual EVAAS growth scores and principals' ratings, 18.4 percent were found to need improvement, but this percentage varied between initially licensed teachers (21.4 percent) and fully licensed teachers (13.5 percent).

To what extent did principals rate teachers below proficient (Not Demonstrated or Developing) or did teachers not meet expected growth according to their EVAAS scores? Have these changed over time?

Teachers receiving NCEES evaluations can be categorized as being in need of improvement if they are rated below *Proficient* by their principal on any of the first five standards or if they receive "Does Not Meet Expected Growth" as their EVAAS score. The type of rating that identifies a teacher as needing improvement is important. In the principals' ratings of the five standards, the ratings are directly connected to the principals' observations and other data sources that the principals can use to provide feedback on the practices, behaviors, or attitudes that the teacher should target. The benefit of the EVAAS score is that it is an objective measure of a teacher's contribution to her or his students' learning, but it does not provide any information about what instructional practices or other behaviors the teachers need to change in order to improve. If the EVAAS scores are the primary means by which teachers are identified to need improvement—that is, if the EVAAS score identifies teachers as in need of improvement more frequently than does the principal-rated standard—EVAAS may be over-weighted in the identification of teachers in need of improvement.

As shown in Figures 1 and 2 (following page), 14.8 percent of teachers (7,308 out of 38,263) were categorized as in need of improvement by EVAAS scores alone. These 7,308 represented 81.8 percent (7,308 out of 8,723) of all teachers placed in a "needs improvement" category. Of fully licensed teachers assigned to a "needs improvement" category, 96.1 percent were placed into that category based on their EVAAS score alone. Only 1.6 percent of the teachers were designated to need improvement by both EVAAS scores and principals' ratings. Principals' ratings alone assigned 2.3 percent of teachers to a "needs improvement" category—3.5 percent of initially licensed teachers and 0.3 percent of the fully licensed teachers.

From 2010-11 through 2013-14, principals' ratings assigned between 5.2 and 5.9 percent of teachers with full evaluations to the "Needs Improvement" category.

Figure 1. “Effective” and “Needs Improvement” Evaluation Ratings (Only Teachers with Individual EVAAS Scores)

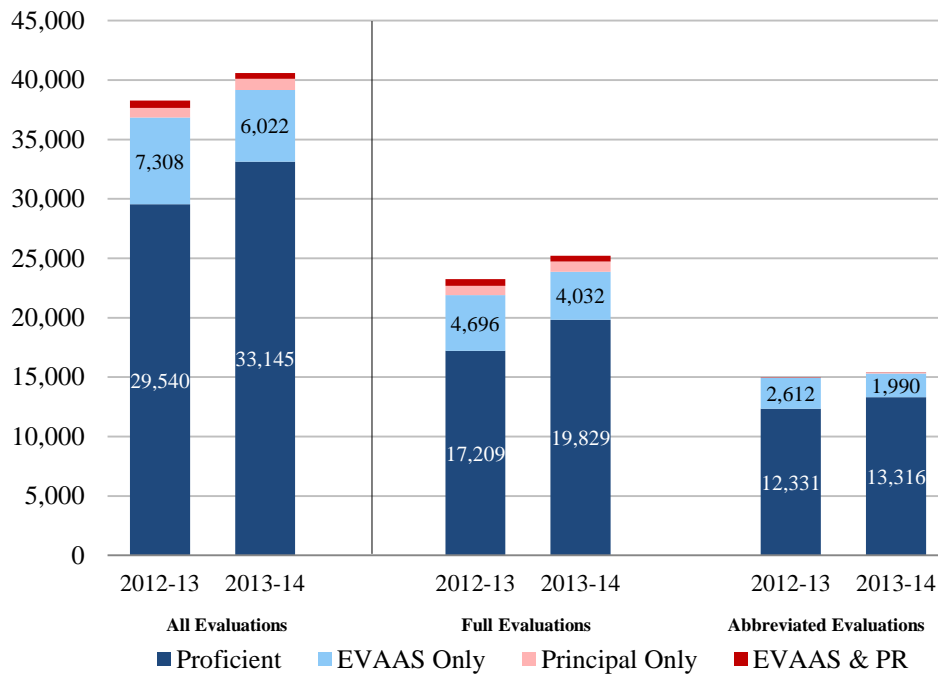
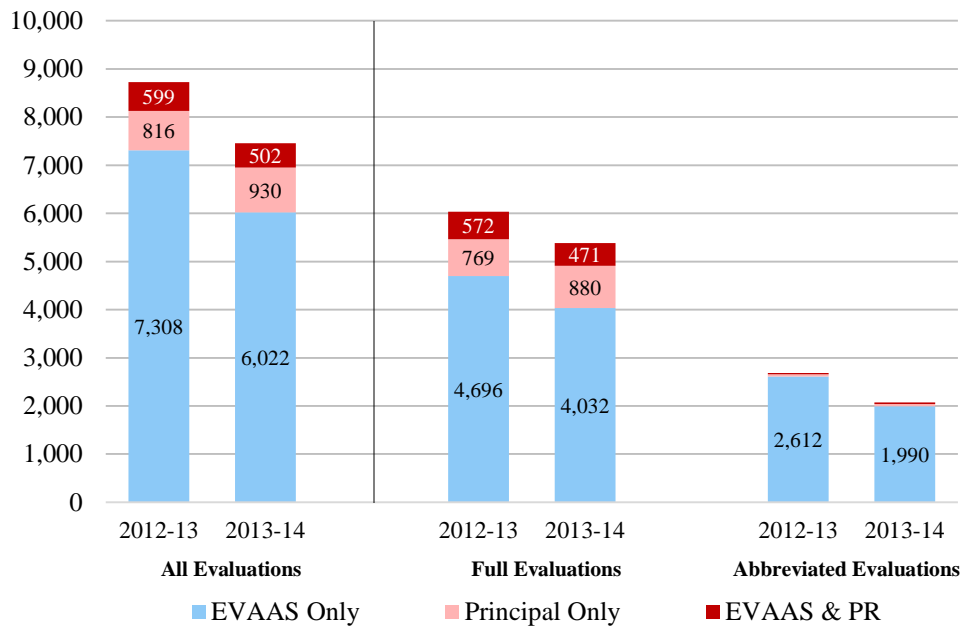


Figure 2. Teachers in Need of Improvement, by Source of Assignment



Summary of Key Findings

- Over 80 percent of teachers in each of the past two years who were assigned to a “Needs Improvement” category were assigned by their EVAAS score alone. That is, they received proficient or higher ratings on all five standards rated by their principals but were identified as not meeting expected growth by EVAAS.
- Approximately 5 percent of beginning teachers who received ratings on all five standards received one or more ratings below *Proficient* from their principals.
- For initially licensed teachers, the EVAAS scores and principal rating agreed that approximately 2.0 percent needed improvement. For the fully licensed teachers, the two types of ratings agreed that 0.2 percent of teachers need improvement.
- The percentages of teachers who received at least one rating by their principals below *Proficient* was 7.7 percent for teachers who did not meet expected growth according to EVAAS, 3.2 percent for teachers who met expected growth, and 1.2 percent for teachers who exceeded expected growth.

Principal Feedback

Did principals provide teachers with information on their strengths and weaknesses by making distinctions in performance between the standards?

Another important purpose for the NCEES evaluations was to provide teachers, especially initially licensed teachers who are, on average, less effective and more likely to turnover than more experienced teachers, with clear information about their strengths and weaknesses. Value-added scores, including but not limited to EVAAS scores, provide objective measures of the outcomes of teachers’ instructional practices but, unfortunately, do not provide information about which practices are strengths and weaknesses for individual teachers. To provide clear indications of teachers’ strengths and weaknesses, principals should be using the entire range of the rating scale, especially *Developing* through *Distinguished* (rating levels 2 through 5), and we expect that principals’ ratings may change over time as they develop expertise in rating their teachers.

In Table 1 (following page), we see that, on average, teachers were rated between *Proficient* and *Accomplished*, and that the ratings have not changed as principals and teachers have gained experience with the evaluation rubric. In both 2012-13 and 2013-14, 90 percent of the teachers that were rated on all five standards received a rating of either *Proficient* or *Accomplished*.

Table 1. Mean and Standard Deviations by Principal Rating Standards (Teachers with both EVAAS and Principals' Ratings)

	2010-2011		2011-2012		2012-2013		2013-2014	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Leadership	3.53	0.71	3.62	0.71	3.59	0.70	3.58	0.69
Respectful Environment	3.50	0.67	3.67	0.70	3.63	0.69	3.61	0.69
Content Knowledge	3.51	0.69	3.56	0.69	3.51	0.68	3.51	0.69
Facilitating Learning	3.64	0.70	3.57	0.67	3.53	0.67	3.52	0.66
Reflection on Practice	3.57	0.70	3.56	0.70	3.53	0.69	3.50	0.69

Have principals' ratings changed over time in terms of providing teachers with information about their strengths and weaknesses?

The stability of the average ratings for each standard could mask changes in ratings if principals had begun to use both higher ratings (such as *Distinguished*) and lower ratings (such as *Developing*) with nearly equal frequency. If this were the case, the higher and lower ratings would offset each other in the average, but the variability in the ratings, as measured by the standard deviation (S.D.) would change over time. We could determine from such a pattern that, as they gained experience with the rubric, principals were more discriminating in their ratings and thus were giving teachers more useful information about how they could develop and improve. In Table 1, we see the standard deviations have been stable across standards and over time. Thus, principals are not providing more information about teachers' strengths and weaknesses as they gain experience with using NCEES.

A potential concern arises from the limited range of the average ratings by category, something that has been noted in prior research on personnel evaluations. Even though the evaluation is divided into discrete categories and multiple dimensions, raters tend to provide global ratings of personnel rather than discrete ratings that reflect individual strengths and weaknesses. To examine the extent to which NCEES ratings tend to be global ratings of teachers rather than ratings of the teachers' strengths and weaknesses on each standard, the Evaluation Team conducted a test of the extent to which the five ratings were measuring the same thing (global ratings of the teacher) or different things (teachers' individual strengths and weaknesses). Using a standard statistic for examining the reliability of measures (Cronbach's alpha) that varies from 1 (global) to 0 (rating each standard independently), we found that the measure was 0.93 for each of the four years in which NCEES teacher evaluations have been conducted. The Team concluded that principals rated individual teachers' overall performance rather than making distinctions in teachers' performance on each of the five NCEES standards.

Summary of Key Findings

- Principals rated teachers either *Proficient* or *Accomplished* 90 percent of the time, which provided limited information on individual teachers' strengths and weaknesses. This raised concerns, especially for beginning teachers who are more likely to need this information for developmental purposes.
- Principals' ratings have not varied over time, indicating little refinement in using NCEES ratings to provide teachers with feedback on strengths and weaknesses.

- Principals rate teachers globally rather than providing meaningful distinctions on teachers’ performance on each standard.

Relationships between EVAAS Scores and Principal Ratings

Were the EVAAS scores for teachers’ contributions to student achievement growth related to principals’ ratings of teachers’ performance, especially the ratings for facilitating student learning?

To enhance credibility of NCEES with teachers and to provide consistent information, the agreement between the principal ratings that teachers received and their student growth ratings (EVAAS) should be related. The objective measure of their students’ achievement growth could reasonably be expected to relate to principals’ ratings of teachers on the NCEES standards, especially the “teachers facilitate learning for their students” standard, which most closely has a bearing on student growth. On the other hand, since principals had more direct evidence about teachers and their instructional practices than what could be captured by student performance on achievement tests, the relationship between principals’ ratings and EVAAS scores may have been weakened.

Table 2 shows that the correlations between principal ratings were significantly—albeit loosely—correlated with EVAAS scores. Correlations of EVAAS with principals’ ratings for Standard 4 only (“teachers facilitate learning for their students”) were not different in level or trend from correlation with the overall composite ratings. These findings reinforced that teachers were rated based on a global concept of their ability rather than according to their performance on each standard, and that the state’s official objective measure of student performance did not strongly influence principals’ ratings of their teachers.

Table 2. Correlations between Principal Ratings and EVAAS (Concurrent and Lagged EVAAS).

	2010-2011		2011-2012		2012-2013		2013-2014	
	<i>Conc.</i>	<i>Lag</i>	<i>Conc.</i>	<i>Lag</i>	<i>Conc.</i>	<i>Lag</i>	<i>Conc.</i>	<i>Lag</i>
Composite Standards 1–5 All Evaluations	0.23	0.23	0.22	0.24	0.22	0.24	0.24	0.26
Facilitating Student Learning All Evaluations	0.19	0.19	0.21	0.24	0.21	0.23	0.23	0.25

In addition to the examination of all teachers who received both individual EVAAS scores and NCEES ratings on two or more standards, we analyzed the correlations by subject, grade, and type of test. The Team found higher correlations between EVAAS and science subjects (roughly 0.32), somewhat lower correlations with math (roughly 0.26), and the lowest correlations with reading/English/language arts (always below 0.20). Associations with Microsoft Word and PowerPoint, the only Career and Technical Education assessment courses for which there were a sufficient number of observations to analyze, also were low (roughly 0.16). Compared to other types of assessments, the correlations of the mClass assessments of Text and Reading Comprehension with principals’ ratings were substantially lower (0.11 in 2013-14). Across each grade, 3rd through 8th, and in high school, the correlations between principals’ ratings and teachers’ EVAAS scores were very similar, ranging from 0.22 to 0.27 in 2013-14.

Over time, did the performance measures converge, indicating greater alignment between EVAAS scores and principals' ratings?

As principals gained experience with the EVAAS measures, it could have been expected that they would have more closely aligned their ratings with the objective student growth measure. It is possible, however, that the timing of the availability of the EVAAS scores, usually during the summer following the school year, would have caused principals to rely on the teachers' score from the prior year (which we refer to as the lag score) when they conduct teachers' annual evaluations.

As indicated in Table 2, the correlations with teachers' composite score remained stable over time. In addition, the correlations with the lagged EVAAS score, which principals had access to when their ratings were completed, were nearly the same as with the concurrent EVAAS scores, indicating principals did not align their ratings with the growth measures that they had available.

Summary of Key Findings

- Principals' ratings of teachers' overall performance and of their ability to facilitate student learning were loosely correlated with teachers' EVAAS scores.
- The correlations did not increase over time, indicating that, on average, principals' ratings of their teachers were not influenced by the state's official measure of the teachers' contributions to the achievement growth of their students.
- The correlations between principal ratings and EVAAS scores varied across individual course subjects. Correlations with EVAAS were higher with science and substantially lower with reading/English/language arts, the mClass, and career and technical assessments.

Other Measures of Teacher Performance

Existing research has shown teachers' value-added scores influence principals less than direct observations of the teachers' instructional practices (Goldring et al., 2015), which may explain the loose correlations between EVAAS scores and principals' ratings of teachers' performance. When combined with the finding above that principals rate teachers on their overall performance and tend not to provide meaningful distinctions based on performance on each standard, this suggests that principals and teachers may need more *actionable* information to support the development of teacher performance.

What instructional practices or teacher behaviors predicted gains or improvements in teachers' EVAAS scores?

In order to support principals and teachers in the development of more effective teachers, the Evaluation Team examined measures of teachers' instructional practices and behaviors that relate to improvements in their EVAAS scores—that is, related to higher EVAAS scores than predicted based on their prior EVAAS scores alone.

For this evaluation, we used regression analysis in which we included measures of teachers' practices and behaviors from three sources to see whether any of these measures were related to the teachers' concurrent value-added scores after controlling for prior value-added scores. These

three sources were the NCEES principals’ ratings of teachers on the original five standards, a pilot of the Tripod student survey administered in 38 school districts, and the CLASS observation instrument.

In terms of improving value-added scores, the Team found both effective and ineffective practices, which we list in Table 3. Teachers rated higher by their principals on Standard 4—“teachers facilitate learning for their students”—increased their EVAAS scores more than those rated lower. In elementary schools, students’ report of both the presence of a collaborative classroom environment and evidence of classroom management predicted higher EVAAS scores. In secondary schools, student reports of both higher-order instruction and classroom management predicted higher EVAAS scores. A positive climate, as noted in classroom observations by independent observers, also predicted higher EVAAS scores.

Table 3. Measures Associated with Impacting Teachers’ Value-Added Scores

	Source	β
Effective Practices		
Facilitating Student Learning	Principal Ratings	0.09
Collaborative Environment	Student Survey – Elementary	0.22
Classroom Management	Student Survey – Elementary	0.17
Higher-Order Instruction	Student Survey – Secondary	0.14
Classroom Management	Student Survey – Secondary	0.12
Positive Climate	Classroom Observation	0.29
Ineffective Practices		
Busy Work	Student Survey – Secondary	-0.15
Student-Led Environment	Student Survey – Secondary	-0.20
Teachers’ Capacity to Improve		
Reflection on Practice	Principal Ratings	0.02
Collaborative Environment	Student Survey – Elementary	0.17
Higher-Order Instruction	Student Survey – Secondary	0.20
Classroom Management	Student Survey – Secondary	0.18

Betas (β) represent standardized regression coefficients from models in which a value-added score is regressed on two lagged values of value-added and significant measures identified through backward stepwise regression by data source. For Effective and Ineffective Practices, the target criterion (DV) is value-added from the concurrent year of the observational measures. For Capacity to Improve, the target criterion (DV) is value-added from the year following the observational measure.

The analysis revealed two practices that consistently predicted lower EVAAS scores, both from student surveys: busy work and a student-led environment. The classrooms with high scores on busy work were characterized by evidence of student boredom and lack of student attention and by students finding the lessons neither interesting nor enjoyable (though students in these classes also report having tasks at all times). Student-led environments were characterized as ones in which students made decisions about activities and were asked to share their thoughts; students in these classes also indicated that they did not believe that they learned much. While having students exercise some levels of control and decision-making can have academic benefits for

students, it may take greater teaching expertise than many teachers can master to use these practices effectively.

Also, as shown in Table 3, four measures were associated with teachers' capacity to improve. In other words, teachers with higher scores on these measures had higher EVAAS scores in the following year than was predicted by their concurrent EVAAS scores alone. Teachers who received higher ratings from their principals on their ability to reflect on their practice raised their EVAAS scores more than was predicted by their EVAAS score in that year. As was the case with explaining improvement over prior years, a collaborative environment in elementary classrooms was associated with higher EVAAS scores in the following year. A collaborative environment was defined as one in which students are asked to share their thoughts and to explain the reasoning behind their work, an important characteristic distinguishing it from a student-led environment.

In secondary classrooms, students' report of both classroom management and higher-order instruction are indicators of teachers' capacity to improve their EVAAS scores. Classrooms with better classroom management are characterized by students who behaved as the teacher wanted them to, stayed busy, and did not waste time. Also, in classes rated highly on this measure, bad behaviors did not distract the students from their work. Classes with high ratings for higher-order instruction were characterized by students who explained their answers and shared their thoughts. These findings suggest that, to be effective, student interactions should be directly related to their classwork; sharing thoughts can be effective when connected to the curriculum but perhaps not as an end in itself.

Summary of Key Findings

- Both teachers and principals want more information about how teachers can improve their practices, especially in ways that increase their value-added scores.
- Using measures from principals' ratings, student surveys, and classroom observations, we found several measures associated with higher value-added scores than were predicted by teachers' previous value-added scores alone.
- These measures can be (1) directly incorporated into the feedback provided by principals to teachers, and (2) periodically tested for their relationship to value-added if student surveys and/or observation protocols for classrooms are conducted in the future.

Teacher Perceptions of NCEES

What were teachers' views about their evaluations and related topics during the period in which the NCEES evaluation with six standards has been implemented?

When the student achievement growth measure was added to the original five NCEES standards, two concerns surfaced. The first concern was that the overall fairness of the evaluations would be eroded—that the protocols designed to promote teacher development would be undermined. Second, concerns were raised that the high-stakes evaluations would inhibit teachers from supporting one another and working together to improve student learning.

Have teachers' attitudes changed over time?

From 2011-12 to 2013-14, the favorability of teachers' views of their evaluations has declined significantly. Overall, teachers' rating of the evaluation process as measured by survey items declined from approximately 5.2, which indicated slight agreement with the items, to about 4.8, which moved them toward neither agreement nor disagreement with the items. The largest declines in favorability related to the developmental use of the evaluation. For example, the largest decline was on the item, "The teacher evaluation process encourages professional growth", closely followed by "The evaluation process encourages teachers to reflect on their instructional practice." The lowest declines were in the responses to items, "Teacher evaluation is fair." and "The criteria on which I am evaluated are clear."

Contrary to some expectations, teachers engaged in knowledge-sharing more in 2013-14 than in earlier years. Overall, their knowledge-sharing activities increased from once or twice per week to almost daily. The largest increases were on items concerning sharing information about their students and instructional practices, such as "Teachers at my school share ideas on teaching," or "Teachers at my school share and discuss student work." Both practices are highly encouraged to improve student learning. The two items that showed the least increase were "Teachers at my school discuss what you/they learned at a workshop or conference," and "Teachers at my school share and discuss research on effective instructional practices for English language learners."

Overall it appears that the favorability of teachers' views about the evaluation system has declined. This decline seems to reflect disappointment in the developmental purposes for which the evaluations were conducted, rather than in concerns about fairness or the specific standards. It also seems that the types of teacher knowledge-sharing that were job-embedded and focused on student learning are flourishing, in spite of concerns about a higher-stakes evaluation that includes a measure of student achievement growth. Traditional professional development workshops do not seem to promote knowledge-sharing, and the sharing of information about working effectively with English learners is not increasing, which may be a concern since North Carolina ranked among the top 10 states in growth of its foreign-born population between 1990 and 2011 (United States Census Bureau, 1990, 2000, 2011).

Summary and Conclusions

Along with approximately two-thirds of other states, North Carolina adopted value-added scores for individual teachers as an additional, sixth standard to supplement the pre-existing five-standard NCEES teacher evaluation system. Adopting a student achievement measure for teachers' evaluations was included as a part of the state's application for federal RttT funds. This evaluation report focused on the sixth standard and the trends in the ability of the NCEES to both identify lower-performing teachers and provide information about teachers' individual strengths and weaknesses.

Over the past two years, approximately 80 percent of teachers who received an individual EVAAS score were judged to be *Proficient* or above on all six standards. The vast majority of the teachers who were identified as in need of improvement were assigned solely on the basis of their value-added scores. Because almost all identifications of a need for improvement were based on the value-added score, value-added effectively acted alone to determine teachers' evaluation status, rendering the judgments of principals on all other aspects of teaching and teachers' performance much less important.

Many states have multiple sources of data that are used directly in teacher evaluations, such as student surveys and systematic, direct observations of classrooms. The scores from these alternative sources can be systematically translated into ratings, thereby providing direct evidence about specific valuable practices and removing some of the individual discretion in the final evaluation ratings. Consideration should be given to systematically adding other direct measures of teaching performance into the NCEES in addition to the EVAAS scores that are currently included. In addition to providing more evidence for teachers who also have EVAAS scores, teachers who do not have EVAAS scores could then have direct measures of their individual performance incorporated into their evaluations. Having direct evidence of the individual performance of all teachers also may increase the perceived fairness of the evaluation system. The weighting for the other evidence in the evaluation could be set using the same techniques used in this study to identify and highlight practices that relate to increasing student achievement.

The second main conclusion from this evaluation is that teachers are being rated globally, and that neither the ratings nor evaluation feedback are providing them with enough actionable information for them to improve. Once again, data from other sources, such as student surveys, systematic, direct observations of classrooms, and teacher surveys, could be analyzed (as was done in this evaluation) to identify measures that are associated with improving student learning. One procedure that could be used for providing teachers with evidence-based feedback and support is described in Figure 3 (following page).

Figure 3. Procedure for Targeting Teacher Improvement using Multiple Performance Measures

1. Identify measures of intrinsically valuable practices
2. Identify measures of practices empirically related to EVAAS scores by calculating correlations between multiple measures of teacher performance (e.g., ratings, surveys, and direct observation) and EVAAS scores at the state levels
3. Identify measures of practices empirically related to improvements in EVAAS scores by estimating two regressions: (1) to identify practices that increase teachers' capacity using concurrent measures of practice and EVAAS scores to predict future EVAAS scores; and (2) to identify practices associated with higher EVAAS scores than would be predicted by their prior scores at the state or LEA levels
4. Group teachers into quintiles of performance at the highest level of data collection (state or LEA) for each of the multiple measures: measures associated with higher EVAAS scores and gains in EVAAS; and measures of intrinsically valuable practices
5. For teachers with EVAAS scores, identify those not meeting expected growth on EVAAS scores, and for teachers without those scores, identify those in the lowest quintiles of measures associated with gains in EVAAS
6. Determine the lowest scores on the practices identified in steps 1 through 3 for each teacher not meeting expected growth on EVAAS scores and for teachers without EVAAS scores
7. Focus teachers in the lowest quintile(s) of practices that are deemed to have intrinsic importance (Step 1) to improve in these areas
8. Focus teachers not meeting expected growth or teachers without EVAAS scores on improving practices associated with increases in EVAAS scores in Step 3 to improve these practices
9. Support teachers with professional development for the focal areas identified in Steps 7 and 8 with activities that may include paired mentorship with teachers identified as highest-performing in these areas of weakness, coaching, and/or formal training
10. Re-analyze relationships between measures periodically (Steps 2 and 3) to allow for changes in associations as standards, assessments, demographics, and policy contexts change

NCEES seems to have been accepted by teachers and principals. On the whole, they feel that the system is fair. Also, teachers' survey responses indicated that the implementation of the NCEES process has not produced negative side effects, such as decreasing teachers' willingness to share information. In fact, the opposite has occurred—teachers have engaged in more knowledge-sharing across the first four years of expanded NCEES implementation.

The main issue appears to be that the current system includes only one systematic data source—EVAAS—and, while EVAAS is an important objective measure of teachers' effectiveness, the inclusion of additional systematic measures may point out the strengths and weaknesses of individual teachers, increase the accuracy of identifying those who need improvement, increase the favorability of teachers' attitudes toward the evaluation system, and provide direct information about practices that can be used for improvement for all teachers.

References

- Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104.
<http://doi.org/10.3102/0013189X15575031>
- Doherty, K. M., & Jacobs, S. (2013). *State of the States 2013: Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice*. Washington, D. C.: National Council on Teacher Quality.
http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report
- North Carolina Department of Public Instruction. (2013). *North Carolina Professional Teaching Standards*. Raleigh, NC: Author. <http://www.ncpublicschools.org/docs/effectiveness-model/ncees/standards/prof-teach-standards.pdf>
- United States Census Bureau (1990). Census of Population and Housing.
<http://www.census.gov/prod/www/decennial.html>
- United States Census Bureau (2000). Census of Population and Housing.
<http://www.census.gov/prod/www/decennial.html>
- United States Census Bureau. (2011). American Community Survey.
<http://factfinder2.census.gov>

Contact Information:
Please direct all inquiries to Gary T. Henry
Gary.henry@vanderbilt.edu

© 2015 Consortium for Educational Research and Evaluation–North Carolina



Carolina Institute
for Public Policy



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

