NC Teacher Evaluations & Teacher Effectiveness

Exploring the relationship between value-added data and teacher evaluations

By Dayne Batten, Christopher Britt, Jennifer DeNeal, and Lauren Hales

Project 6.4 / August 2012

RESEARCH QUESTIONS:

1. **What is the relationship between teacher performance evaluation ratings and annual EVAAS student growth data?**
2. **Which regions and LEAs in North Carolina have the strongest connection between these two measures?**
3. **How can this research impact future teacher evaluation in North Carolina?**

INTRODUCTION

BACKGROUND

Understanding the best way to evaluate teachers has been a consistent challenge for policymakers. Traditionally, individual school administrators, such as principals and assistant principals, evaluate teachers using qualitative measures. In recent years, with the introduction of statistical value-added models that measure teachers' contributions to student growth, the national trend in education policy has moved towards adopting value-added metrics as an integral part of teacher evaluations.

Several states, including Tennessee, North Carolina, Georgia, Florida and Colorado, have begun to incorporate value-added metrics into their teacher evaluation systems through both Race to the Top grants and state legislation. Additionally, in November 2011, the National Governor's Association (NGA) awarded grants to Colorado, Guam, Nevada and North Carolina to assist with the redesign of their teacher evaluation systems (National Governors Association 2011). The North Carolina Department of Public Instruction (NCDPI), with the assistance of the NGA grant, is in the process of implementing the state's new evaluation tool.

In their NGA grant application, NCDPI (2011a) committed to implementing the teacher evaluation system in a way that "accurately identifies differences amongst teachers so as to help them grow substantially in the effectiveness of their practice." To reach this goal, NCDPI proposed to complete rigorous quantitative and qualitative analyses of teacher evaluation data in order to demonstrate how evaluation data correlates with student growth data. This paper addresses the quantitative analysis by exploring the relationship between teacher value-added scores and teacher performance evaluation scores in North Carolina on a state, regional and local level.

TEACHER EVALUATIONS IN NORTH CAROLINA

Since the 1980s, North Carolina has used a statewide teacher evaluation tool. The first iteration of the evaluation tool, the Teacher Performance Appraisal Instrument (TPAI), was used statewide until 2007. In February 2007, NCDPI partnered with Mid-continent Research for Education and Learning (McREL), a private nonprofit organization, to establish a new instrument for statewide evaluation known as the North Carolina Educator Evaluation System (NCEES) (Department of Public Instruction 2011b). The state rolled out NCEES to Local Educational Agencies (LEAs) across the state in waves. The final group of LEAs adopted the system in the 2010-2011 school year.

During the early stages of implementation and continuing through the 2010-2011 school year, beginning teachers (teachers with less than 3 years of experience) were evaluated yearly, while evaluations for career-status teachers were only required once every five years when their license required renewal. A new policy beginning in the 2011-2012 school year requires yearly evaluations for all North Carolina public school teachers using NCEES. Career-status teachers can use an abbreviated version of the evaluation instrument for the years they are not renewing their license.

Currently, North Carolina's Professional Teaching Standards (Department of Public Instruction 2008a) stipulate that each teacher be given performance evaluation scores, rating them on each of five standards. The North Carolina State Board of Education adopted these standards in 2007-2008 and established an evaluation rubric two years later. The 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
5. Teachers reflect on their practice

Teachers are assigned a value between one and five, depending on their level of proficiency, for each standard. According to the Professional Teaching Standards (Department of Public Instruction 2008b), scores should indicate the following:

1. Competency not demonstrated
2. Developing
3. Proficient
4. Accomplished
5. Distinguished

In July 2011, as part of North Carolina's Race to the Top effort, the North Carolina State Board of Education adopted a sixth standard for NCEES (Board of Education 2012). Implemented during the 2011-2012 school year, Standard 6 states that "teachers contribute to the academic success of students." Standard 6 will measure student growth as determined by a value-added metric. During the 2011-2012 school year, the student growth value will be weighted 70 percent based on the individual growth of students taught by the educator and 30 percent based on student growth for the entire school.

EVAAS & VALUE-ADDED DATA IN NORTH CAROLINA

North Carolina uses the Education Value Added Assessment System (EVAAS) from the SAS Institute as its value-added metric to measure student growth. EVAAS applies a combination of statistical models to assess LEA, school and individual teacher effectiveness based on student growth (SAS 2010). Because the EVAAS model uses standardized exam data to facilitate student growth calculations, value-added scores are only available for teachers whose courses contain End of Grade (EOG) or End of Course (EOC) assessments. EOG assessments are statewide summative exams given to students at the end of grades three through eight in reading and math, as well as science tests in grades five and eight. EOC assessments are summative exams administered to high school students in Algebra I, English I and Biology.

METHODOLOGY

VALUE-ADDED DATA

To facilitate our research, SAS provided EVAAS data for all North Carolina public school teachers with value-added scores. The data included each teacher's school and LEA, as well as grade level, content area, number of students and a value-added report for each tested subject taught. The value-added report contained three data points: an individual Teacher Effect score, the standard error for that Teacher Effect score and whether student growth in a teacher's classes was "Above," "Below," or "Not Detectably Different" from the expected student growth (SAS 2010, Board of Education 2009). Teachers with scores falling within two standard errors of the state average were deemed "Not Detectably Different" from the state average (SAS 2010).

The scale of scores for each subject area and grade level differed greatly; therefore, we needed a way to make the EVAAS data comparable for all teachers. In order to normalize the scores, we calculated the number of standard errors[1] each Teacher Effect score fell from the state average, which allowed us to make comparisons between subjects and grade levels. We then averaged these figures across subjects and grades for each teacher, giving us a combined measure of that teacher's performance. Based on this measure, an elementary school teacher who excelled in teaching reading but was a comparatively poor math teacher would receive a moderate score, while a teacher who excelled at both would have a higher score. This measure provided a picture of a teacher's overall effectiveness that we could objectively compare to performance evaluation scores.

PERFORMANCE EVALUATION DATA

The Academic Services and Instructional Support Division of NCDPI provided us with evaluation data for all teachers receiving a performance evaluation in the 2010-2011 school year, providing data for approximately 46,000 teachers.

We used two different data points to compare performance evaluation scores to our EVAAS data. First, we selected Standard 4 (teachers facilitate learning for their students) of the North Carolina Professional Teaching Standards as the best point of comparison for EVAAS data. One can imagine

[1] *Standard error here refers to the statistical standard error calculation provided by SAS, not the standard deviation of the individual teacher effect scores in our dataset.*

teachers able to excel at Standards 1, 2 or 3 without contributing to student growth in the classroom. For example, some teachers may "know the content they teach" (Standard 3), but ineffectively communicate that knowledge to their students. Conversely, in a classroom where the teacher facilitates student learning, we would expect to see concurrent student growth. As such, we believe Standard 4 most closely relates to student growth and, therefore, best corresponds with EVAAS data at this time.

Next, we calculated an average score for teachers, as assessed by their administrators, across the five standards, giving us a more balanced perspective of their performance in all areas. When reporting one score per teacher, standard policy reports the median score. However, using the mean score across the standards provided us with the most precision for our calculations and analysis. We expected to find a significant correlation between performance evaluation scores and the EVAAS data.

DEMOGRAPHIC DATA

To understand the impact of various demographic factors on evaluation scores, we collected demographic information on North Carolina public school teachers from the North Carolina Window of Information on Student Education (NCWISE) database. The demographic information included race, gender and whether a teacher was classified as a beginning teacher.

FINAL DATASET

Our dataset included 11,430 teachers for which we had both EVAAS scores and performance evaluation ratings. (As previously mentioned, for the 2010-2011 school year, only beginning teachers and career-status teachers renewing their license were required to be evaluated. Moreover, only instructors teaching courses with an EOG or EOC assessment could receive an EVAAS value-added score.) Of those 11,430 teachers, we obtained demographic information on all but four teachers, leaving us with a dataset large enough to facilitate accurate state-level research as well as local analysis for North Carolina's largest LEAs. Though we used data on teachers from smaller LEAs in our regional and statewide statistics, we did not look at LEA level statistics from any LEA with less than 98 teachers in our data set. Selecting LEAs with at least 98 teachers ensured we were not reporting LEA-level results compromised by small sample size, while maintaining 35 LEAs for our LEA-level analysis.

STATISTICAL ANALYSIS

We used multiple analytical methods to examine trends in our data. We explored simple descriptive statistics to analyze the distribution of values on a number of measures. Additionally, we used Ordinary Least Squares (OLS) regression to look for relationships between teachers' EVAAS scores and performance evaluation ratings in all eight regions of the state, and in all LEAs for which we had at least 98 data points. We documented three key variables from the results of each of these regression models to facilitate comparison among LEAs and regions:

1. Regression coefficient – number representing the average impact of a one standard error increase in a teacher's EVAAS score on that teacher's performance rating. For example, a regression coefficient of 0.5 signifies a teacher's performance rating would go up by half a rating, on average, for every one standard error increase in their EVAAS Teacher Effect score.
2. P-value – number representing the probability that the observed relationship between Teacher Effect scores and performance ratings was spurious, rather than an actual significant relationship. Lower p-values indicate a greater likelihood that the observed correlation is real. Typically, a p-value of five percent or less is considered strong in statistical research.
3. R^2 – value representing the percent of the variation in teacher evaluation scores explained by the statistical model. High values indicate that the model explains most of the differences between teacher scores and low values indicate most of the variation is explained by unobserved or random factors.

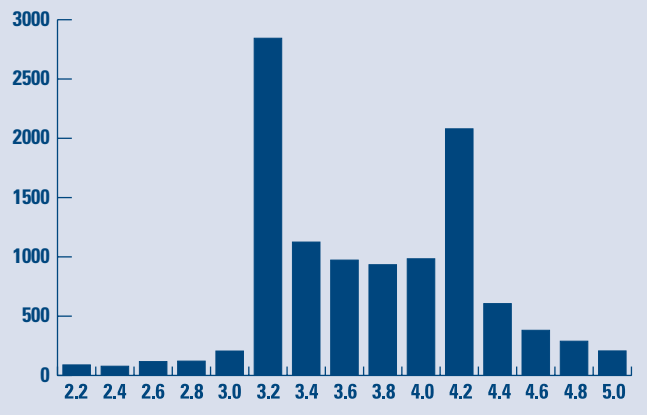
In addition to the individual LEA and regional relationships, we also looked for relationships on a statewide level. For this analysis, we created Generalized Linear Models using LEA as a classification variable to control for natural variations among LEAs and prevent statistical errors.

ANALYSIS

STATEWIDE DESCRIPTIVE STATISTICS

Our research showed that the vast majority of teacher performance evaluation scores were concentrated between a three ("Proficient") and just above a four ("Accomplished"). Averaging all standards, the mean teacher evaluation score was a 3.6, with more than 90 percent of teachers rated "Proficient" or higher and 40 percent of teachers as "Accomplished" or higher. There was a large concentration of teachers with an average of 3.2 or 4.2, indicating a tendency among administrators to give teachers four ratings of "Proficient" and one "Accomplished" or four ratings of "Accomplished" and one "Distinguished". The distribution of scores was very similar for each of the individual standards.

AVERAGE EVALUATION SCORES (STATEWIDE)



We also noted that teachers tended, on average, to be rated slightly higher on certain standards of performance. Standard 4, in particular, tended to be the standard on which teachers were rated the highest. The following table shows the average scores for each performance indicator.

AVERAGE SCORE BY STATE STANDARD

Demonstrate Leadership	3.56
Respect Diversity	3.54
Know Content	3.57
Facilitate Learning	3.61
Reflect on Practice	3.60

Finally, we looked at the degree of fluctuation in performance evaluation scores statewide. Standard deviation for performance scores was only six tenths of a point, indicating that the vast majority of scores were very close to the mean.

LEA DESCRIPTIVE STATISTICS

We also looked at descriptive statistics for each of the 35 LEAs for which we had at least 98 data points. The data show a fairly wide range of mean scores and standard deviations for the state. Mean scores for performance evaluations across all five standards ranged from a low of 3.24 to a high of 4.01 within the 35 LEAs in our study.[2] Standard deviations ranged from a low of 0.4 to a high of 0.71. These patterns held constant when we considered individual standards.

LEA LEVEL MODELS

We next analyzed the relationship between teacher evaluation ratings and EVAAS Teacher Effect scores in each of the 35 LEAs considered in our study. The results indicated a varied relationship between teacher evaluation ratings and Teacher Effect scores across the 35 LEAs. In certain LEAs, the connection was so weak that we could not determine a statistically significant trend. However, in the LEAs with a significant relationship, the relationship between Teacher Effect scores and teacher evaluation results was as great as six times higher in certain LEAs than in others for Standard 4 ratings and almost five times higher in certain LEAs for mean evaluation ratings.

Teacher Effect scores had the strongest relationship with teachers' Standard 4 performance evaluation ratings in the following six LEAs:[3]

1. Rowan-Salisbury
2. Moore County
3. Burke County
4. Craven County
5. Buncombe County
6. Wilson County

[2] *A full summary of our descriptive statistics for each of the 35 LEAs is available in Appendix A.*

[3] *A full summary of our regression results for each of the 35 LEAs is available in Appendix B.*

However, even in these LEAs, the relationship between Teacher Effect scores and Standard 4 ratings was considerably small. Our results showed, on average, that Teacher Effect scores predicted only about nine percent of the variation in Standard 4 performance evaluation ratings in these six LEAs. The other 91 percent of the variation resulted from unexplained or random factors. Of the 35 LEAs, our model found that six LEAs had no statistically significant relationship between Teacher Effect scores and Standard 4 ratings.

We saw very similar results for the relationship between Teacher Effect scores and teachers' averaged performance ratings. In fact, many of the LEAs with the strongest relationships across mean evaluation ratings were the same as those for the Standard 4. Teacher Effect scores shared the greatest relationship with teachers' average performance evaluation ratings in the following six LEAs:[4]

1. Rowan-Salisbury
2. Moore County
3. Craven County
4. Buncombe County
5. Pender County
6. Wilson County

In these six LEAs, the relationship between Teacher Effect scores and mean evaluation ratings was still small, though larger than for Standard 4 ratings. On average, in the six LEAs with the strongest relationships, Teacher Effect scores explained 13 percent of the variation in mean performance ratings. Additionally, we found four LEAs where Teacher Effect scores had no statistically significant relationship with mean performance ratings.

REGIONAL LEVEL MODELS

We also determined relative levels of connection between EVAAS data and teacher evaluation scores for the state's eight regions. The table below summarizes regression results for the relationships between Teacher Effect scores and Standard 4 data in each region. All values are highly statistically significant.

Region	Coefficient	R^2
1	0.068	7.5%
2	0.058	5.5%
5	0.047	4.5%
7	0.046	3.2%
8	0.046	3.2%
4	0.042	4.5%
3	0.037	2.5%
6	0.036	2.3%

Regions 1 and 2 showed the highest levels of connection, with an average of 6.5 percent of the variation in Standard 4 evaluation scores explained by Teacher Effect results. Regions 3 and 6 had the weakest relationship, with an average of 2.4 percent of the variation in Standard 4 scores explained by the Teacher Effect data.

The results for the model predicting the connection between Teacher Effect scores and mean evaluation scores, shown in the table below, were very similar.

Region	Coefficient	R^2
1	0.067	8.9%
2	0.064	8.5%
8	0.057	6.0%
5	0.050	6.4%
7	0.048	4.3%
4	0.045	6.5%
3	0.040	3.8%
6	0.040	3.7%

Once again, Regions 1 and 2 had the strongest relationship, with the lowest levels of connection seen in Regions 3 and 6. As with the LEA findings, the relative levels of interrelation were slightly higher for average evaluation scores than for Standard 4 scores.

STATEWIDE MODELS

Our final analysis examined statewide data on teacher evaluations and Teacher Effect scores, including teacher demographic data using a Generalized Linear Model. Again, we used LEA as a classification variable, enabling the model to control for the differences between LEAs. The table below reports the results for our regression model predicting teachers' Standard 4 evaluation ratings.

STANDARD 4 REGRESSION MODEL

	Estimate	P – Value
Teacher Effect	0.04	.000
African American?	-0.06	.001
Beginning?	-0.33	.000
Female?	0.11	.000

The results of the regression model highlighted several interesting statewide trends. Most notably, Teacher Effect data had a relatively weak relationship with teachers' Standard 4 evaluation ratings. In fact, for each standard error that a teacher's individual teacher effect lay above zero, that teacher's Standard 4

[4] *A full summary of our regression results for each of the 35 LEAs is available in Appendix C.*

rating would increase by about four hundredths of a point on average. To put this in perspective, for the 2010-2011 school year, a teacher rated "Above Average" according to EVAAS data had an 8 percent chance of having a higher Standard 4 rating than a teacher who was comparable in every other way, except for an EVAAS score of "Not Detectably Different."

Also noteworthy is the role of demographic factors in a teacher's Standard 4 evaluation rating. Our analysis found that beginning teachers and African Americans tended to have lower Standard 4 ratings, even while holding their level of effectiveness as determined by EVAAS constant. We did not report results for other races as they were not statistically significant. Female teachers tended to have higher Standard 4 evaluation ratings than male teachers, even at the same levels of effectiveness.

We also analyzed the relationship between our available data and teachers' mean evaluation scores across all standards. The table below summarizes the results of this regression model.

AVERAGE SCORE REGRESSION MODEL

	Estimate	P – Value
Teacher Effect	0.04	.000
African-American?	-0.09	.000
Beginning?	-0.35	.000
Female?	0.11	.000

As with our LEA and regional analysis, the results for the mean evaluation scores were very similar to the results for Standard 4. In the statewide model, a one standard error increase in Teacher Effect score tended to increase a teacher's average performance evaluation score by four tenths of a point. These findings imply that, during the 2010-2011 school year, a teacher with an "Above Average" Teacher Effect score would have at least a 40 percent chance of having a rating higher on any of the standards than a comparable teacher, except for an EVAAS score of "Not Detectably Different.

Our model for mean evaluation scores also showed very similar results for demographic factors. Beginning teachers and African Americans tended to have lower evaluation scores, while females tended to be rated relatively higher.

DISCUSSION

SCORE DISTRIBUTIONS

One of the most significant trends we observed was the comparatively small distribution of actual performance evaluation scores. Statewide, standard deviations for evaluation scores tended to be remarkably low, with the vast majority of scores clustered near the mean. We classified teachers based on EVAAS Teacher Effect scores from highest to lowest and found that, out of more than 11,000 teachers, the average performance evaluation rating of the 100 least effective teachers was a 3.2, while the average evaluation score for the 100 most effective teachers was only a 3.8. In other words, out of more than 11,000 teachers, the 100 whose instruction contributed the most to student growth and the 100 whose instruction contributed the least to student growth were all rated somewhere between "Proficient" and "Accomplished" on their performance evaluation ratings.

We suspect the low degree of variation in evaluation scores may partially explain the weak relationship between student growth data and teacher evaluation ratings. If administrators give similar scores to all their teachers, a high performing teacher would not likely stand out in terms of their performance evaluation. Conversely, if an administrator gives a wide range of scores, he or she would likely be particularly careful to ensure that high scores go to the most effective teachers.

One of the most striking findings of our research was the relationship between gender, race, years of experience and mean performance evaluation ratings. We observed that male, African American, and beginning teachers tended to have lower performance evaluation scores than their peers. However, variation in actual student growth data can explain much of that trend. When we compared demographic factors to student growth numbers, we found that Teacher Effect scores tended to be slightly lower among African Americans, male teachers, and beginning teachers. Thus, the lower student growth outcomes partially explain the lower performance evaluation scores for these groups. Nevertheless, we found that the influence of

demographics persisted, even when we controlled for student growth scores. For instance, if we selected two teachers who were the same in every respect (including EVAAS Teacher Effect score), except that one was male and the other female, we would expect the male teacher to have a lower performance evaluation score. The same held true for race and experience.

We believe there are three factors that could possibly explain this trend:[5]

1. There may be other aspects of teacher performance (aside from student growth) not addressed by our model. In this case, it would simply mean administrators are assigning performance evaluation scores using additional information we were unable to include in our model.
2. It is possible that racial or gender bias is influencing administrators' performance evaluation decisions.
3. Administrators could be using certain factors as "mental short cuts" to calculate performance evaluation scores. For instance, if administrators know their beginning teachers are generally less effective, they may tend to give all beginning teachers slightly lower performance evaluation scores, regardless of whether those scores are justified in every individual instance.

Ultimately, we believe the possibility exists that each of these trends is contributing to the observed patterns to a greater or lesser degree. We recommend further analysis of these relationships as a promising area for future study.

"MENTAL SHORT CUTS"

Several aspects of our results indicated administrators may use certain characteristics as "mental short cuts" in calculating performance evaluation scores. We first noticed this trend as we observed that beginning teachers tend to receive performance evaluation scores lower than warranted by their student growth numbers.

To understand how these "mental short cuts" work, consider an evaluator who knows that beginning teachers, on average, perform less effectively. With that knowledge, the evaluator may automatically give slightly lower scores to all her beginning teachers. While appropriate in many cases, an exceptionally effective beginning teacher may receive a lower score than his performance deserved. Identifying beginning teachers, in other words, has become a "mental short cut" to assigning performance evaluation ratings.

RELATIONSHIP AMONG PROFESSIONAL TEACHING STANDARDS

We also found evidence for a different type of "mental short cut" in our analysis of the relationship between Teacher Effect scores, mean evaluation ratings and Standard 4 ratings. As discussed above, there was at least a 40 percent chance that a teacher rated "above average" in terms of student growth would get one additional point on one of their performance evaluation standards. Yet, there was only an 8 percent chance they would get an extra point on Standard 4, the standard which most closely reflects an assessment of student learning at the time of our study. In other words, above average student growth was likely to be associated with an increase in a teacher's score on one standard, but it was unlikely that the extra point would come on Standard 4.

The relationship between student growth and other evaluation standards could also explain part of the trend. For instance, we would expect teachers who "reflect on their practice" (Standard 5) to have higher rates of student growth. A strong relationship between the standards may also partially explain this trend, as administrators possibly view teachers who excel at one aspect of their jobs as more likely to excel in other aspects as well. Nevertheless, it is counter intuitive that an increased Teacher Effect score, in general, did not occur on the evaluation standard that loosely relates to student growth. This may again be the result of a "mental short cut" phenomenon. Evaluators may generally know who their most effective instructors are, but not necessarily in which standards those teachers excel. Thus, evaluators reward student growth on performance evaluations, but not always by increasing Standard 4 scores.

To address the "mental shortcut" issue, we recommend encouraging evaluators to spend a sufficient amount of time with teachers to gather a thorough understanding of their performance in all areas. It may also be helpful to gather information on teacher performance from other sources (e.g., other teachers), to better inform performance reviews.

BEST PRACTICES

As shown above, our results indicated a varied relationship between EVAAS Teacher Effect data and teacher performance evaluation data on both an LEA and regional level. Therefore, we suggest that the state examine performance evaluation practices in LEAs where evaluation scores have the strongest relationship with student growth data in order to identify potential best practices for performance evaluations. Although NCDPI has provided training statewide on NCEES, we recommend looking at how certain LEAs have implemented the evaluation system and promoting best practices through qualitative analysis of these LEAs. For example, we acknowledge that several LEAs have taken initiatives to understand teacher perspectives on the NCEES and teacher effectiveness including Charlotte-Mecklenburg's research in relation to their pay for performance system. We would also recommend increasing differentiated training for individual standards to assist administrators in understanding the criteria for each standard.

[5] *We also considered the possibility that these results were a result of multi-collinearity in our model. However, the maximum Variance Inflation Factor for any of our coefficients was a 1.02, well within safe limits. Moreover, the large sample size increased our confidence that our estimates were accurate.*

OTHER FACTORS FOR CONSIDERATION

Finally, we believe it is important to discuss the role of administrators in the performance evaluation process. With recent budget constraints, both principals and assistant principals are taking on more duties. Proper evaluations are necessary for accountability, but require administrators to take an appropriate amount of time for each evaluation. Yet administrators' time is becoming more and more constrained. This could become a critical issue as policy shifts for administrators to evaluate all teachers during each school year. Administrators must view performance evaluations as a priority and be given both the time and the support to devote to this task.

Nevertheless, we must take into account that evaluations are a subjective tool and vulnerable to human factors. Relationships between administrators and teachers will consistently factor into performance evaluations, and thus should be taken into consideration as well.

CONCLUSION

In conclusion, contrary to our hypothesis, we did not see a strong relationship between performance evaluation ratings and EVAAS Teacher Effect scores for North Carolina public school teachers. Overall, the relationship between these two measures was weak with a high amount of the variation resulting from unexplained or random factors. However, we did find that some regions and more specifically, certain LEAs, had a stronger relationship between performance evaluation ratings and Teacher Effect scores. Rowan-Salisbury had the strongest relationship of the 35 LEAs considered in our study, with an average 100% chance that an "Above Average" teacher would receive a higher rating on one standard in his or her performance evaluation.

In a statewide analysis, we also found that demographic factors demonstrated a role in teacher evaluation ratings. Holding all factors in our model constant, including Teacher Effect scores, we found female teachers tended to have higher evaluation ratings, while male teachers and African American teachers tended to have lower evaluation ratings. Additionally, beginning teachers tended to receive lower evaluation ratings than career-status teachers with the same Teacher Effect score.

We believe additional research into the relationship between teacher performance evaluation and EVAAS Teacher Effect scores is critical to strengthen the connection between these two measures. We recommend this study be repeated using the 2011-2012 school year data. This data will contain performance evaluations for all teachers throughout the state, establishing a much larger data set, as well as including a Standard 6 rating in the performance evaluation data. A future study will be able to look closer at the relationship between performance evaluations and Teacher Effect scores and allow North Carolina to continue progressing towards its goal of teacher effectiveness.

APPENDIX A – DISTRICT DESCRIPTIVE STATISTICS

MEAN AND STANDARD DEVIATION, ALPHABETICAL, DISTRICTS WITH AT LEAST 98 DATA POINTS

LEA Name	Mean Evaluation Score (All Standards)	Standard Deviation	Sample Size
(All Standards)	Standard Deviation	Sample Size	0.1877
Alamance-Burlington	3.56	0.62	187
Brunswick County	3.58	0.60	106
Buncombe County	3.75	0.61	150
Burke County	3.78	0.63	98
Cabarrus County	3.57	0.50	217
Catawba County	3.30	0.40	117
Charlotte-Mecklenburg	3.30	0.54	1094
Cleveland County	3.78	0.50	134
Columbus County	3.43	0.49	110
Craven County	3.65	0.65	164
Cumberland County	3.46	0.61	339
Davidson County	3.41	0.55	114
Durham County	3.56	0.59	210
Forsyth County	3.41	0.48	396
Gaston County	3.66	0.71	247
Granville County	3.52	0.52	100
Guilford County	3.46	0.59	520
Harnett County	3.33	0.45	155
Hoke County	3.66	0.62	99
Iredell-Statesville	3.97	0.60	161
Johnston County	3.74	0.59	298
Lincoln County	3.66	0.64	98
Moore County	3.49	0.58	104
Nash-Rocky Mount	3.39	0.53	113
New Hanover County	3.76	0.58	187
Onslow County	3.24	0.48	196
Pender County	3.81	0.61	143
Pitt County	3.42	0.56	187
Randolph County	3.31	0.43	145
Robeson County	3.42	0.56	138
Rowan-Salisbury	3.69	0.66	163
Union County	4.01	0.58	374
Wake County	3.62	0.55	942
Wayne County	3.30	0.44	137
Wilson County	3.67	0.65	177

APPENDIX B – DISTRICT REGRESSION RESULTS FOR STANDARD 4

STANDARD 4 RESULTS, HIGHEST COEFFICIENT FIRST, DISTRICTS WITH AT LEAST 98 DATA POINTS

LEA Name[6]	Sample Size	Coefficient[7]	P-Value	R^2
Rowan-Salisbury	163	0.103	0.000	15.8%
Moore County	104	0.089	0.001	9.9%
Burke County	98	0.079	0.004	8.5%
Craven County	164	0.069	0.003	5.4%
Buncombe County	150	0.068	0.003	5.8%
Wilson County	177	0.068	0.001	6.1%
Granville County	100	0.067	0.001	10.6%
Lincoln County	98	0.063	0.030	4.8%
Alamance-Burlington	187	0.063	0.000	7.3%
Cabarrus County	217	0.063	0.000	5.9%
Pender County	143	0.059	0.004	5.8%
Brunswick County	106	0.059	0.002	8.6%
Onslow County	196	0.055	0.002	5.1%
Forsyth County	396	0.053	0.000	6.0%
Johnston County	298	0.053	0.000	5.4%
Nash-Rocky Mount	113	0.053	0.012	5.5%
Cleveland County	134	0.051	0.001	8.2%
Durham County	210	0.046	0.005	3.7%
Cumberland County	339	0.045	0.000	4.7%
New Hanover County	187	0.044	0.017	3.0%
Iredell-Statesville	161	0.043	0.014	3.8%
Union County	374	0.041	0.000	3.6%
Charlotte-Mecklenburg	1094	0.039	0.000	3.6%
Guilford County	520	0.037	0.000	3.4%
Gaston County	247	0.035	0.015	2.4%
Columbus County	110	0.033	0.032	4.2%
Harnett County	155	0.033	0.026	3.2%
Randolph County	145	0.028	0.043	2.8%
Wake County	942	0.017	0.027	0.5%
Pitt County*	187	0.038	0.052	2.0%
Catawba County*	117	0.034	0.053	3.2%
Wayne County*	137	0.026	0.141	1.6%
Davidson County*	114	0.022	0.369	0.7%
Robeson County*	138	0.011	0.415	0.5%
Hoke County*	99	0.005	0.821	0.1%

[6] ** denotes LEAs for which data was not statistically significant*

[7] *Coefficients are statistical best estimates of the relationship between the two variables. The actual relationship for a given LEA may be marginally stronger or weaker than the relationship observed in the current data set.*

APPENDIX C – DISTRICT REGRESSION RESULTS FOR EVALUATION AVERAGE

AVERAGE RESULTS, HIGHEST COEFFICIENT FIRST, DISTRICTS WITH AT LEAST 98 DATA POINTS

LEA Name[8]	Sample Size	Coefficient[9]	P-Value	R^2
Rowan-Salisbury	163	0.107	0.000	19.9%
Moore County	104	0.099	0.000	13.8%
Craven County	164	0.097	0.000	13.0%
Buncombe County	150	0.083	0.000	11.1%
Pender County	143	0.072	0.000	12.4%
Wilson County	177	0.065	0.000	7.9%
New Hanover County	187	0.063	0.000	7.8%
Cabarrus County	217	0.062	0.000	9.6%
Johnston County	298	0.060	0.000	8.7%
Forsyth County	396	0.058	0.000	10.5%
Iredell-Statesville	161	0.057	0.000	7.6%
Burke County	98	0.056	0.017	5.8%
Brunswick County	106	0.055	0.001	9.7%
Durham County	210	0.054	0.000	6.0%
Alamance-Burlington	187	0.054	0.000	6.8%
Onslow County	196	0.051	0.000	6.4%
Granville County	100	0.050	0.003	8.5%
Cumberland County	339	0.050	0.000	7.3%
Union County	374	0.049	0.000	6.4%
Davidson County	114	0.048	0.032	4.0%
Cleveland County	134	0.047	0.000	9.1%
Charlotte-Mecklenburg	1094	0.043	0.000	6.3%
Gaston County	247	0.043	0.001	4.2%
Harnett County	155	0.042	0.000	8.5%
Guilford County	520	0.041	0.000	5.3%
Pitt County	187	0.039	0.017	3.0%
Catawba County	117	0.035	0.011	5.4%
Columbus County	110	0.034	0.018	5.0%
Randolph County	145	0.027	0.017	4.0%
Wayne County	137	0.027	0.046	2.9%
Wake County	942	0.023	0.001	1.2%
Nash-Rocky Mount*	113	0.034	0.067	3.0%
Lincoln County*	98	0.043	0.100	2.8%
Hoke County*	99	0.020	0.318	1.0%
Robeson County*	138	0.010	0.429	0.5%

[8] ** denotes LEAs for which data was not statistically significant*

[9] *Coefficients are statistical best estimates of the relationship between the two variables. The actual relationship for a given LEA may be marginally stronger or weaker than the relationship observed in the current data set.*

Works Cited

- Board of Education. 2009. "Educational Value Added Assessment System (EVAAS) Teacher Module." North Carolina State Board of Education Policy Manual. Retrieved July 30, 2012 (http://sbepolicy.dpi.state.nc.us/policies/TCS-C-021.asp?pri=04&cat=C&pol=021&acr=TCS).
- Board of Education. 2012. "North Carolina State Board of Education Policy Manual." North Carolina State Board of Education. Retrieved July 30, 2012 (http://sbepolicy.dpi.state.nc.us/).
- Department of Public Instruction. 2008a. "North Carolina Professional Teaching Standards." North Carolina Department of Public Instruction. Retrieved July 30, 2012 (http://www.ncpublicschools.org/docs/profdev/standards/teachingstandards.pdf).
- Department of Public Instruction. 2008b. "North Carolina Teacher Evaluation Process." North Carolina Department of Public Instruction. Retrieved July 30, 2012 (http://www.ncpublicschools.org/docs/profdev/training/teacher/required/rubricassessmentform.pdf).
- Department of Public Instruction. 2011a. "Grant Proposal – National Governors Association State Strategies to Evaluate Teacher Effectiveness." Unpublished document.
- Department of Public Instruction. 2011b. "Validation Study, North Carolina Educator Evaluation System." Unpublished document.
- National Governors Association. 2011. "Four States and Territories Selected to Redesign Teacher Evaluation Systems." National Governors Association. Retrieved July 30, 2012 (http://www.nga.org/cms/home/news-room/news-releases/page_2011/col2-content/main-content-list/four-states-and-territories-sele.html).
- SAS. 2010. "SAS EVAAS Statistical Models." SAS Institute. Retrieved July 30, 2012 (http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf).

By Dayne Batten, Christopher Britt, Jennifer DeNeal, and Lauren Hales

The Financial and Business Services Area is in its sixth year of the Research Intern Program. The Program is designed to help build a quality research program within NCDPI to supplement and supply data for discussions related to procedural, process, and policy changes. This year's program included students from Duke University's Master of Public Policy program, The University of North Carolina at Chapel Hill's Master of Public Administration program, and North Carolina State University's Master of Public Administration program. The intern program is managed by Eric Moore (919-807-3731) and Kayla Siler (919-807-3824) | intern_research@dpi.nc.gov.

NC DEPARTMENT OF PUBLIC INSTRUCTION :: **June St. Clair Atkinson, Ed.D., State Superintendent :: 301 N. Wilmington Street :: Raleigh, NC 27601-2825**

In compliance with federal law, NC Public Schools administers all state-operated educational programs, employment activities and admissions without discrimination because of race, religion, national or ethnic origin, color, age, military service, disability, or gender, except where exemption is appropriate and allowed by law.

Inquiries or complaints regarding discrimination issues should be directed to: Dr. Rebecca Garland, Chief Academic Officer
Academic Services and Instructional Support :: 6368 Mail Service Center, Raleigh, NC 27699-6368 :: Telephone: (919) 807-3200 :: Fax: (919) 807-4065