Consortium for Educational Research and Evaluation–North Carolina

An Evaluation of the North Carolina Educator Evaluation System for School Administrators: 2010-11 through 2013-14

Gary T. Henry and Samantha L. Viano
Vanderbilt University

February 2016

Table of Contents

AN EVALUATION OF THE NORTH CAROLINA EDUCATOR EVALUATION SYSTEM FOR SCHOOL ADMINISTRATORS: 2010-11 THROUGH 2013-14

Executive Summary

The purpose of this report is to evaluate the effects of adding an eighth standard, school-level Education Value-Added Assessment System (EVAAS) scores, to the evaluation of school principals that is based on the seven standards in the North Carolina Standards for School Executives (NCSSE). To that end, this report describes the relationship between the principal evaluation ratings and other measures of administrator effectiveness, as well as trends in the administrator evaluation between the 2010-11 and 2013-14 school years.

Major Findings

1. From 2010-11 through 2013-14, 27.8 percent of principals were found to need improvement based on their superintendents' ratings and/or school-level EVAAS results. In 2013-14, superintendents' ratings assigned 7.6 percent of principals (128 principals) to the "needs improvement" category.
2. In 2013-14, 25.1 percent of principals (515 principals) were rated as at least Proficient by their superintendents on all seven standards but had school EVAAS scores at the Did Not Meet Expected Growth level.
3. Principal Instructional Leadership measures (including clear communication, high standards and data use) were strongly correlated with superintendents' ratings of their principals, which indicates that superintendents are primarily rating principals based on their instructional leadership rather than differentiating between the seven standards.
4. Even though the Teacher Working Conditions survey is a suggested source of evidence for the principal evaluation process, key items on the survey are only loosely correlated with principals' scores, if at all, indicating the survey information is not being used systematically in the evaluation process.
5. The direct measures of principal effectiveness that were most highly correlated with the principal evaluation scores were measures that were not available for superintendents to use as artifacts in the principal evaluation process, indicating that the recommended measures of principal performance were not used by superintendents.
6. Objective measures of principals' performance, such as retention of effective teachers or school value-added scores, are not strongly correlated with superintendents' ratings of principals' performance, indicating that these measures do not systematically influence principal evaluation ratings.
7. Of the 20 objective measures of principal performance compiled for this evaluation, most are uncorrelated with principal ratings or, at best, loosely correlated, indicating that objective measures rarely influence superintendents' ratings of principals.
8. Most of the items from the Teacher Working Conditions survey that significantly correlate with composite principal evaluation scores are not items that would be expected to be important indicators of principal effectiveness.

9. Superintendents' ratings of principal effectiveness do not appear to be equally distributed across school context. As the percentage of black students or free/reduced price lunch students increases, the composite evaluation score of the schools' principal tends to decrease. It was not possible to test whether less effective principals were assigned to schools with concentrated poverty or black populations or if superintendents rate principals lower, regardless of actual principal quality, when they oversee those types of schools. Both explanations are plausible and both raise concerns for the evaluation of principals.
10. Superintendents rated principals either Proficient or Accomplished, on average, 75 percent of the time, which provided limited information on individual principals' specific strengths and weaknesses. Superintendents rate principals globally rather than providing meaningful distinctions on principals' performance on each standard.
11. Superintendents' ratings have not varied over time, indicating little refinement in using NCSSE ratings to provide principals with feedback on strengths and weaknesses.

Overall, it seems that, in spite of a strong theory that systematic evaluation of principals could lead to improving principals' performance through the NCSSE ratings, it is unlikely that the system as it is currently implemented will do so. The vast majority of principals receive ratings above Proficient for all standards, even though many schools are classified as performing below expectations. In addition, this evaluation provides some evidence that principals' rating may not be entirely fair—principals in schools with higher concentrations of African-American and economically disadvantaged students receive lower ratings. To overcome some of these issues, the State Board of Education may wish to incorporate other measures of principal performance—such as retention of effective teachers, teacher survey ratings of principals' instructional leadership, and teacher survey ratings of the fairness and feedback provided in teacher evaluations—into a composite quantitative rating of principals' overall performance.

Introduction

In December 2006, the North Carolina State Board of Education (SBE) approved the North Carolina Standards for School Executives (NCSSE). Assistant principals and principals are rated on these seven standards annually by district superintendents. The standards are (1) strategic leadership, (2) instructional leadership, (3) cultural leadership, (4) human resource leadership, (5) managerial leadership, (6) external development leadership, and (7) micropolitical leadership (NCDPI, 2013). For each of these standards, principals rate administrator performance as Not Demonstrated, Developing, Proficient, Accomplished, or Distinguished, with associated numerical values from 1 through 5, respectively.

In the 2011-12 school year, the SBE followed through with a commitment made in the state's application for Race to the Top (RttT) funds to add an eighth standard—academic achievement leadership. The new standard is based on school-wide student growth as measured by the Education Value-Added Assessment System (EVAAS), which also provides the student achievement growth measure for teachers. Principals receive a score of Does Not Meet Expected Growth, Meets Expected Growth, or Exceeds Expected Growth, with associated numerical values from 1 through 3, respectively.

The purpose of this report is to evaluate the effects of adding the eighth standard to the evaluation of school principals. To that end, the report describes the relationship between the principal evaluation ratings and other measures of administrator effectiveness as well as trends in the administrator evaluation ratings between the 2010-11 and 2013-14 school years. The scope of this evaluation is limited to principals with available evaluation scores. The report addresses the following five sets of research questions:

1. How many principals needed improvement according to the NCSEE? How frequently did superintendents rate principals as below proficient (Not Demonstrated or Developing)? How frequently were principals rated as not meeting expected growth according to the school-wide EVAAS? How frequently were principals designated as needing improvement by both the superintendent ratings and EVAAS scores? Have these frequencies changed over time?
2. Are the ratings principals receive on their evaluations correlated with other measures of principal performance? Are the correlations higher with measures that are recommended for use in rating principals than with other measures of performance? Do certain standards appear to be more closely related to other effectiveness measures than other evaluation standards?
3. Are there specific objective or subjective alternative measures of principal effectiveness that are good predictors of superintendents' composite principal evaluation scores?
4. Do the principal evaluation ratings appear to be higher or lower based on the context of the school? Is there evidence of either lower-performing principals concentrated at a certain type of school or of principals being rated lower when they oversee those types of schools?
5. Did superintendents provide principals with information on their strengths and weaknesses by making distinctions in performance between the standards? Has the value of superintendents' average ratings changed over time? Has the variation in ratings across

standards changed in terms of providing principals with information about their strengths and weaknesses?

To address these five sets of questions, the Evaluation Team assembled a dataset that included all NCSSE principal evaluations ratings conducted between 2010-11 and 2013-14. 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. The teacher evaluation scores and teacher value-added scores (North Carolina Educator Evaluation System scores and EVAAS scores) from the principal's school were merged in at the school level with the superintendent evaluation of principals. In addition, analyses included teachers' responses from the RttT Omnibus teacher survey that was administered between 2011-12 and 2013-14 to teachers in a stratified random sample of North Carolina public schools, as well as Teacher Working Conditions (TWC) survey results for the 2011-12 and 2013-14 waves of data collection. For this study, descriptive statistics such as means and standard deviations, bivariate correlations, measurement test statistics (Cronbach's alpha), and multivariate regression were applied.

Findings

The Distribution of Principal Ratings

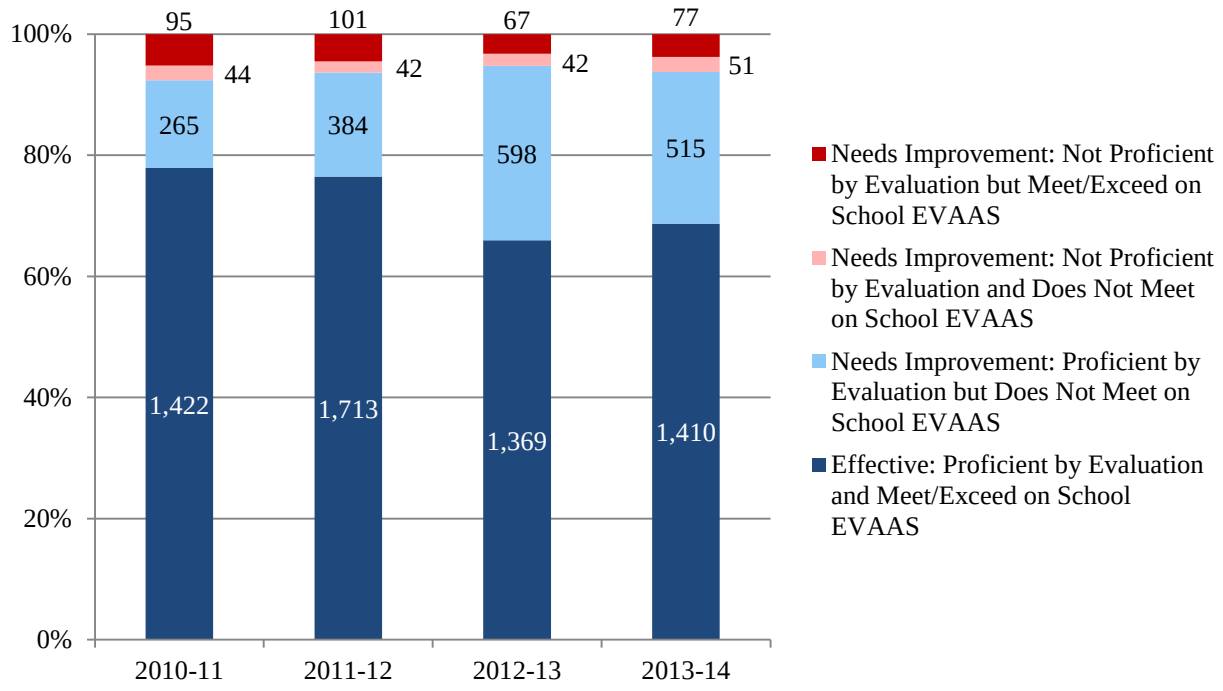
> *Research Question Set 1: How many principals needed improvement according to the NCSEE? How frequently did superintendents rate principals as below proficient (Not Demonstrated or Developing)? How frequently were principals rated as not meeting expected growth according to the school-wide EVAAS? How frequently were principals designated as needing improvement by both the superintendent ratings and EVAAS scores? Have these frequencies changed over time?*

One of the most important purposes of principals' evaluations is to identify areas in which principals need improvement so that they can develop their professional practice in ways that increase the quality of school leadership. Principals receiving NCSSE evaluations can be categorized as in need of improvement if they are rated Not Demonstrated or Developing by their superintendent on any of the first seven standards, or if they receive a Does Not Meet Expected Growth rating as their school-level overall EVAAS score. Among all principals with both school-level EVAAS growth scores and superintendents' ratings in the period covered by this evaluation, 27.8 percent were found to need improvement.

The type of rating that places a principal in the needs improvement group can affect the extent to which principals receive information they can use to develop more effective practices. In the superintendents' ratings of the seven standards, the ratings are intended to be directly connected to the principals' self-assessment, school performance data, and other data sources that the superintendents are encouraged to use to provide feedback on the practices, behaviors, or attitudes that the principals should target. The benefit of the school-level EVAAS score is that it is an objective measure of a school's contribution to student learning, but the score, like all other value-added measures, does not provide any information about what instructional practices or other behaviors the principal needs to focus on to improve school-wide achievement. If the EVAAS scores are the primary means by which principals are identified as being in need of improvement—that is, if the EVAAS scores more frequently identify principals as in need of improvement than the superintendent-rated standard—this may over-emphasize EVAAS in the identification of principals in need of improvement and minimize the amount of information that can be used by principals to improve.

As shown in Figure 1 (following page), 21.5 percent of principals were categorized as in need of improvement by school-level EVAAS scores alone. Only 2.2 percent of principals were designated as being in need of improvement by both school-level EVAAS scores and superintendents' ratings. Superintendents' ratings alone assigned 4.1 percent of principals to the needs improvement category.

Figure 1. Distribution of Effective *and* Needs Improvement *Ratings*



Major Findings

- From 2010-11 through 2013-14, superintendents' ratings assigned between 5.3 and 7.6 percent of principals (128 in 2013-14) to the needs improvement category.
- In the 2012-13 and 2013-14 school years, 28.8 percent and 25.1 percent of principals, respectively (598 in 2012-13 and 515 in 2013-14), were rated as Proficient or better by their superintendents but had school EVAAS scores at the Does Not Meet Expected Growth level.
- In those same school years, around three percent of principals were rated as below Proficient by their superintendents on at least one standard, but their school EVAAS level was Meets or Exceeds Expected Growth.

Correlation of Principal Ratings with Other Measures of Principal Performance

> *Research Question Set 2: Are the ratings principals receive on their evaluations correlated with other measures of principal performance? Are the correlations higher with measures that are recommended for use in rating principals than with other measures of performance? Do certain standards appear to be more closely related to other effectiveness measures than other evaluation standards?*

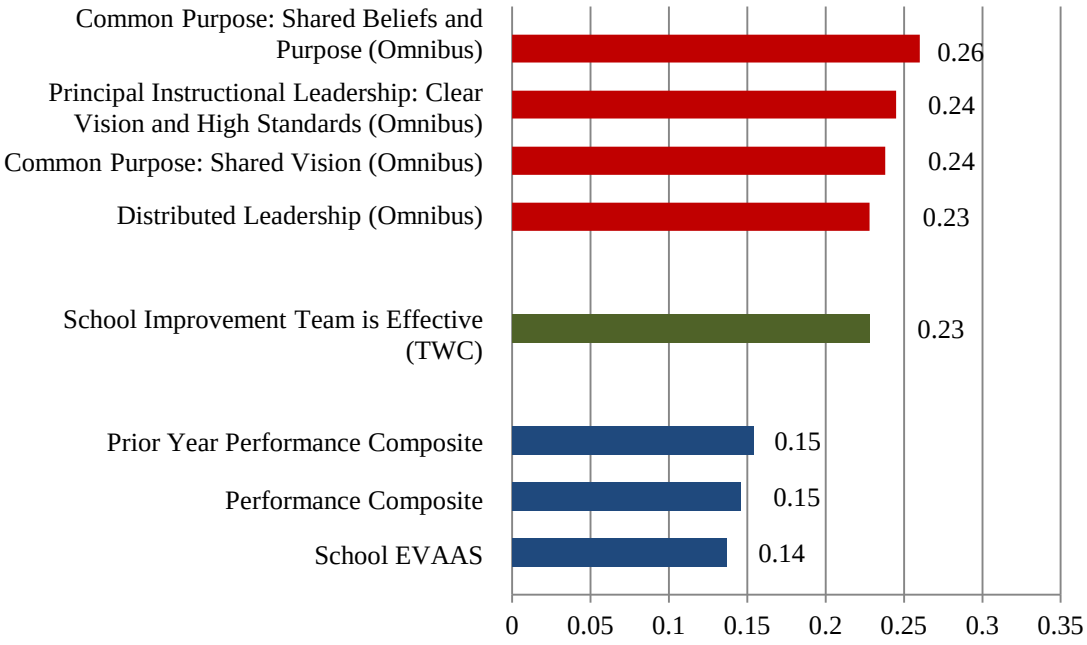
As part of the principal evaluation process, superintendents and principals agree on a list of artifacts that the principals will then include in their consolidated performance assessments that the superintendents review to determine final ratings. These artifacts refer to the evidence the principals use to substantiate their ratings and can include any piece of data or information they wish. Recommended artifacts include measures from the TWC survey, student achievement and testing data, documents (e.g., stated mission statements, school improvement plans, etc.), and

teacher retention data. Superintendents and principals are given a list of suggested artifacts for each standard and are permitted to use additional evidence not explicitly listed.

The dataset compiled for this evaluation includes many of the recommended artifacts that superintendents are asked to reference when determining principals' evaluation ratings. The Evaluation Team calculated the correlations between ratings on individual standards and the artifacts that superintendents would be expected to use, and examined the correlations to see if there is evidence that principal ratings are based on these measures of principal effectiveness. While we investigate dozens of different possible artifacts, this report focuses on a limited number of them—primarily, items that aligned well with the particular standard and were specifically recommended as evidence for the ratings.

The bars in Figure 2 represent the correlation between measures of principal effectiveness and principals' scores on the Strategic Leadership standard. These correlations range from one, perfect correlation, to negative one, perfect negative correlation, with zero indicating no correlation. The colors of the bars reflect the data sources: red bars refer to the scale-level variables from the Omnibus survey, green bars are from the TWC survey items, and the blue bars represent objective measures of principal performance. The Omnibus survey was not available to principals or superintendents during the evaluation, but including it in the analysis allows us to examine correlates of the superintendents' ratings of principals' effectiveness with a measure separate from the evaluation process.

Figure 2. Selected Correlates with Strategic Leadership Scores[1]

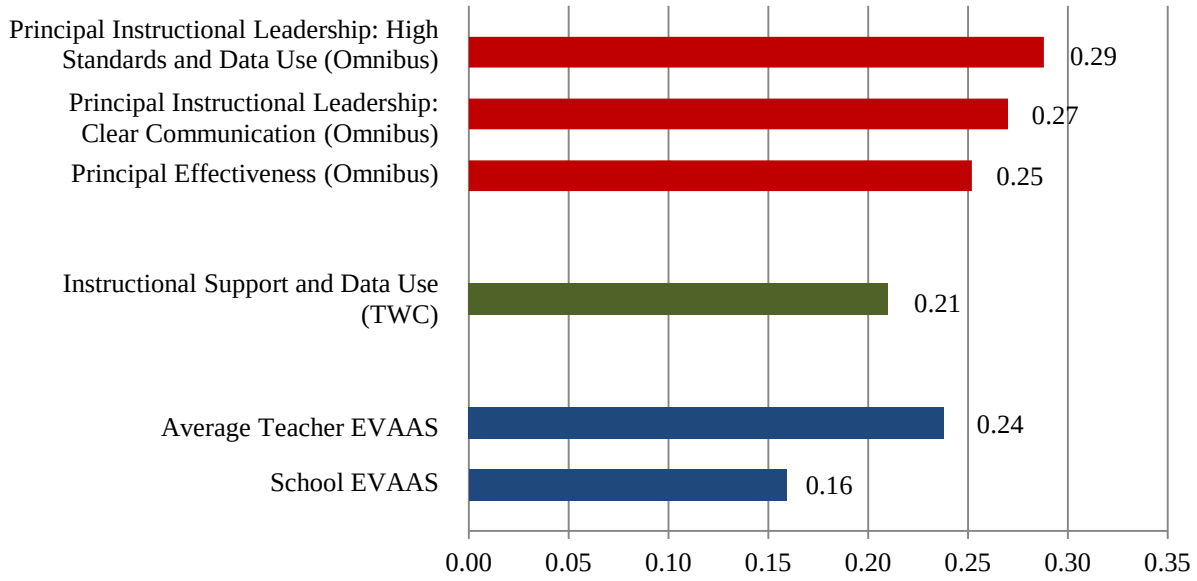


[1] Correlations with items from the TWC on data use, school leadership, and shared vision as well as the change in performance composite from the previous year to the next were included in analysis but not in the figure because of low correlations.

For Strategic Leadership, the highest correlations, between 0.23 and 0.26, are with teacher responses on the Omnibus scales measuring Common Purpose, Distributed Leadership, and Principal Instructional Leadership. The TWC item on if the school improvement team is effective has the next-highest correlation. Comparatively, the objective measures of principal performance, the performance composites and school level EVAAS, have much lower correlations than the survey measures.

Figure 3 shows the correlations between measures from the Omnibus survey and EVAAS measures with the Instructional Leadership scores. While the Omnibus scales are still more highly correlated with the evaluation rating, the Average Teacher EVAAS correlation with the Instructional Leadership score is closer to the Omnibus correlations than the objective measures of performance shown in Figure 2. Of note is the relatively high correlation between teachers' ratings of principal effectiveness (Omnibus) and principals' Instructional Leadership ratings. This seems to indicate that superintendents may have a sense of how teachers feel about their principals' effectiveness when rating their Instructional Leadership.

Figure 3. Correlates with Instructional Leadership Scores[2]

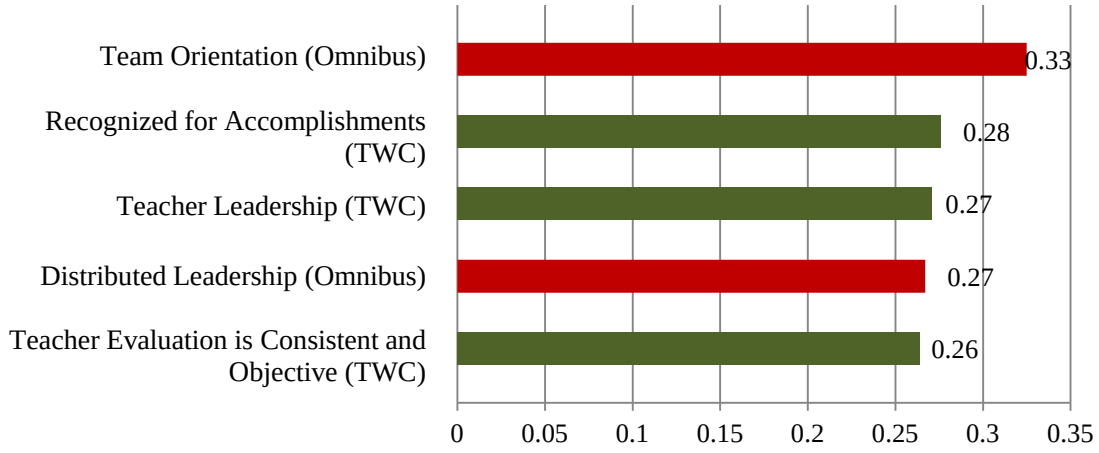


The highest correlation with an aligned TWC measure is 0.21 (with the Instructional Support and Data Use scale), which may indicate that the recommendation to use the TWC as evidence for this rating is not influencing the rating. This interpretation is further supported by the fact that the correlation between the Instructional Leadership score and survey results from related Omnibus survey items (which are not available to superintendents or principals) are larger than the correlations between the TWC and the Instructional Leadership score.

[2] Other scales from Omnibus and TWC were correlated with this standard but are not shown because of space constraints.

Because of the nature of the Cultural Leadership standard, the analyses did not include any objective measures of performance with these correlations (Figure 4). The highest correlation is between teachers' ratings of principals' team orientation (Omnibus) and the Cultural Leadership score. From the TWC, teachers' responses to items on teachers being recognized for their accomplishments, opportunities for teacher leadership, and the consistency and effectiveness of principals' teacher evaluations were correlated with principals' Cultural Leadership rating.

Figure 4. Correlates with Cultural Leadership Scores[3]

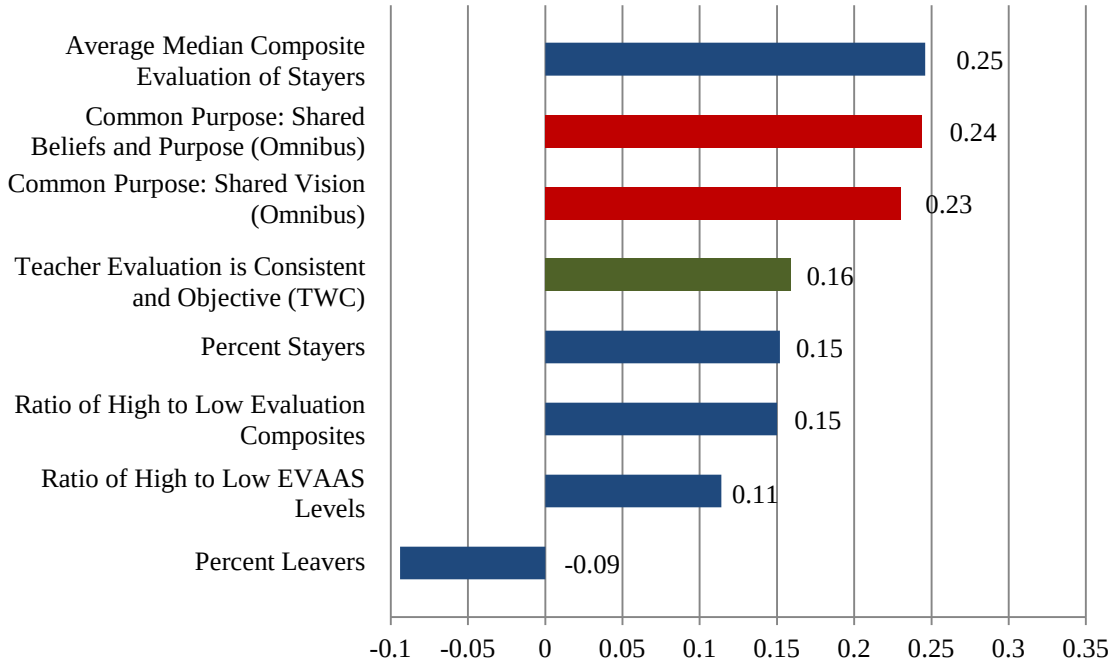


As noted above, Omnibus scales have consistently higher correlations with the scores as compared to the TWC correlations, even though the superintendents and principals are asked to use the TWC results as part of the principal evaluation process and are unaware of the results of the Omnibus. The correlations between the survey measures and the score on Cultural Leadership remain moderately high, around 0.30.

Figure 5 (following page) shows the correlation between the Human Resource Leadership scores and several important objective indicators of principal performance. The terms *Stayers* and *Leavers* that are referenced in Figure 5 refer to teachers who stayed at a principal's school between the previous school year and the year the principal was evaluated and teachers who left a principal's school and are not teaching at another North Carolina public school after the previous school year, respectively. The percent of *Leavers* is negatively correlated with the Human Resource Leadership scores, which we would expect since higher percentages of teachers leaving North Carolina public schools may indicate less effective Human Resource management by the principal.

[3] Omnibus scales on teacher knowledge sharing and teacher self-efficacy as well as TWC items on time use and the school improvement team had lower correlations with this standard than the ones shown.

Figure 5. Correlates with Human Resource Leadership Scores[4]



The EVAAS and composite scores of *Stayers* refer to the scores they received the previous school year. The highest of these correlations is with the average median composite evaluation of *Stayers*. This variable is created by identifying the group of *Stayer* teachers, taking the median score these teachers received on their performance evaluations the prior year, and averaging those median scores (i.e., the composite score).

Another relatively high correlation is with the ratio of high to low evaluation composites. This variable is created by comparing the number of teachers whose median teacher evaluation score is a 4 or a 5 to the number of teachers whose median teacher evaluation score is a 1 or a 2 within the principal's school. As the ratio of high to low evaluation composites increases, the number of teachers at a school who are rated as effective increases in comparison to the number of teachers at a school with lower ratings. We would expect higher-performing principals to be at schools with more higher-performing teachers than lower-performing teachers, or higher ratios of high- to low-performing teachers.

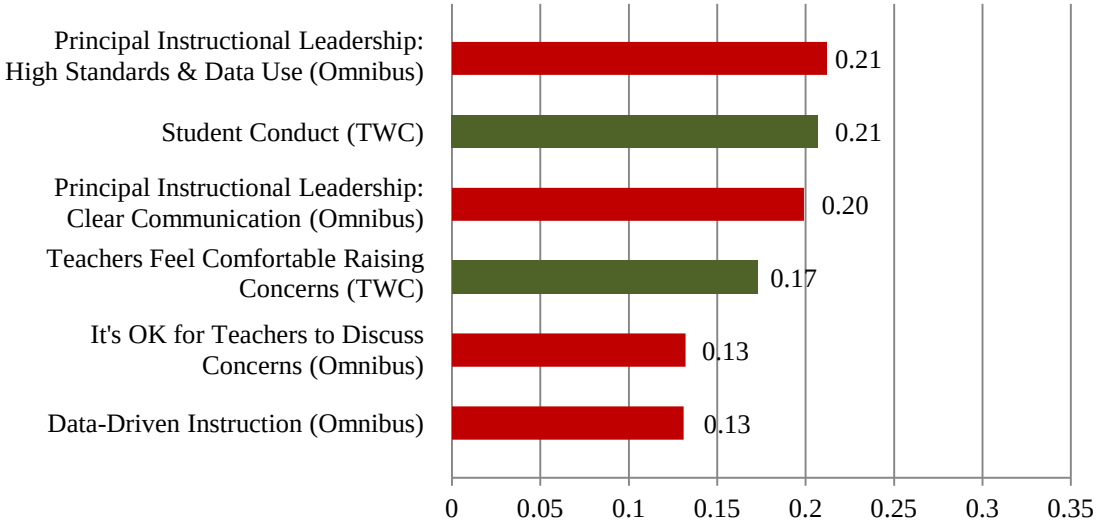
The ratio of high to low EVAAS levels is calculated in the same way as the ratio of evaluation composite scores. This ratio compares the number of teachers who receive an EVAAS level of Exceeds Expected Growth to the number of teachers who receive an EVAAS level of Does Not Meet Expected Growth. Similarly, we would expect that higher-performing principals would have a teaching staff at their school with more higher-performing teachers, in terms of EVAAS growth levels, than lower-performing teachers. We see that this correlation is not as strong as the

[4] Other Omnibus scales had lower correlations with this standard. As well, many other indicators of teacher mobility and quality were correlated with this standard but were excluded due to space constraints.

correlation was with the ratio of high to low evaluation scores, which might be expected since the principals themselves are likely to determine the teachers' evaluations scores but have less of a direct impact on the EVAAS growth level.

For the Managerial Leadership scores (Figure 6), there were no available objective measures that aligned with the performance expectations. Therefore, the evaluation only looked at correlations with survey measures of principal effectiveness. Overall, the highest correlation is with the Omnibus survey measures of high standards and data use and clear communication. These same measures also were significantly correlated with the Strategic Leadership and Instructional Leadership scores (Figures 2 and 3), indicating that principals' performance in these areas influence superintendents' ratings. Only modest correlations were found with two items on the TWC—student conduct and teachers' comfort with raising concerns—again suggesting that recommended evidence for this rating did not greatly influence actual ratings.

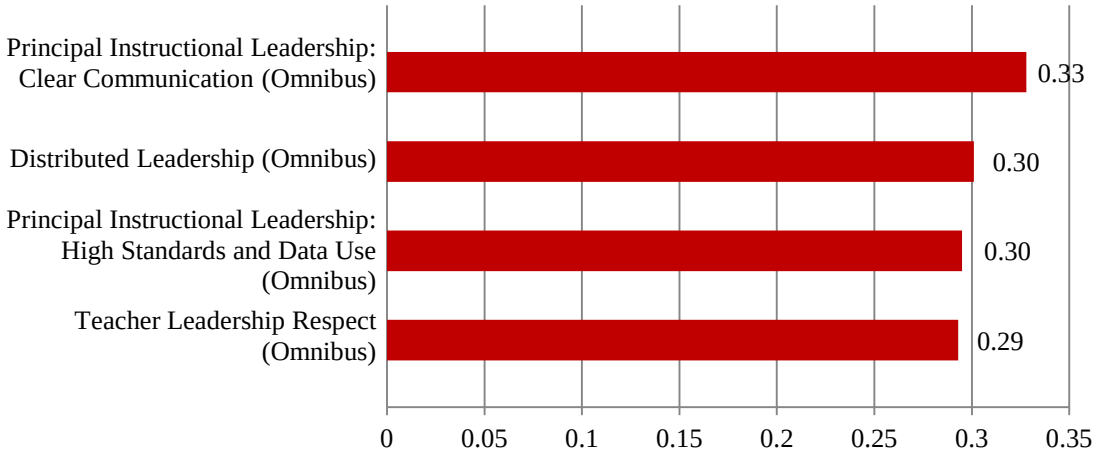
Figure 6. Correlates with Managerial Leadership Scores



For the External Development Leadership scores, we identified only one measure we would expect to be correlated with these scores. The measure is a scale from the TWC survey that measures Community Support. The correlation between the Community Support scale the External Development Leadership scores is 0.24, which, judging by the correlations with other ratings, is relatively high.

For Micropolitical Leadership scores (Figure 7, following page), only Omnibus measures were analyzed for potential correlations. The correlations are all above 0.25 (relatively high) with teachers' rating of the clarity of their principals' communication having a 0.33 correlation with the Micropolitical Leadership scores. Once again, this measure and teachers' rating of principals' high standards and data use are highly correlated with superintendents' ratings.

Figure 7. Correlations with Micropolitical Leadership Scores



Major Findings

- The most highly correlated measures of principal effectiveness with the principal evaluation scores are also pieces of information that could not have been used as artifacts in the principal evaluation process: the Omnibus survey results.
- The correlations suggest that objective measures of principals' performance do not strongly influence superintendents' ratings of principals' performance.
- Even though the TWC survey is suggested for use in the principal evaluation process, the correlations between key items on the survey are only very loosely correlated with principals' scores, perhaps indicating the survey information is not being used systematically in the evaluation process.
- The Omnibus survey measures of Principal Instructional Leadership (including clear communication and high standards and data use) were especially highly correlated with superintendents' ratings of their principals.

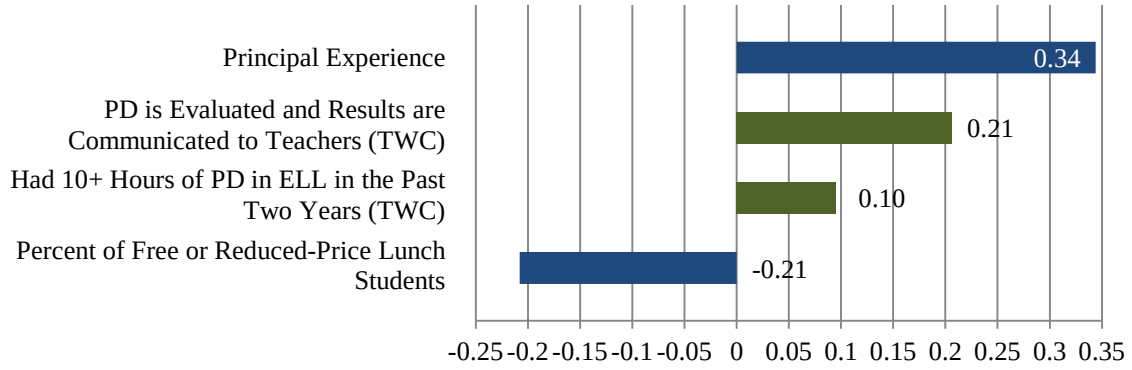
Other Measures of Effectiveness Related to Principal Ratings

Research Question Set 3: Are there specific objective or subjective alternative measures of principal effectiveness that are good predictors of superintendents' composite principal evaluation scores?

The analyses for this section used multivariate regression techniques to investigate which of the objective and survey-based measures that are indicative of principal performance significantly predict principals' composite evaluation scores. We perform this method twice, for the data from the 2013-14 school year and then the 2011-12 school year, as these are the school years during which the TWC was administered. The Omnibus survey measures are not included in these analyses because a limited sample of schools (a stratified random sample) participated in the Omnibus survey administration so we are unable to consider the Omnibus and other survey and objective measures concurrently.

Figure 8 displays the coefficients from the final model predicting median principal evaluation scores in the 2013-14 school year. As with the previous bar graphs, TWC items are displayed in green and the objective measures in blue. The TWC item with the largest coefficient is on the item that measures perceptions of whether professional development (PD) is evaluated and results are communicated to teachers. The standardized coefficient indicates that, for every standard deviation increase in the agreement with this item, the evaluation composite is predicted to increase by a fifth of a standard deviation. The other TWC item that predicts higher composite principal evaluation scores is the indicator of having had ten or more hours of PD in English language learner (ELL) instruction in the past two years. These seemingly anomalous higher correlations, along with the analysis in the prior section of this report, indicate that the TWC is not being used systematically in evaluations of principal performance.

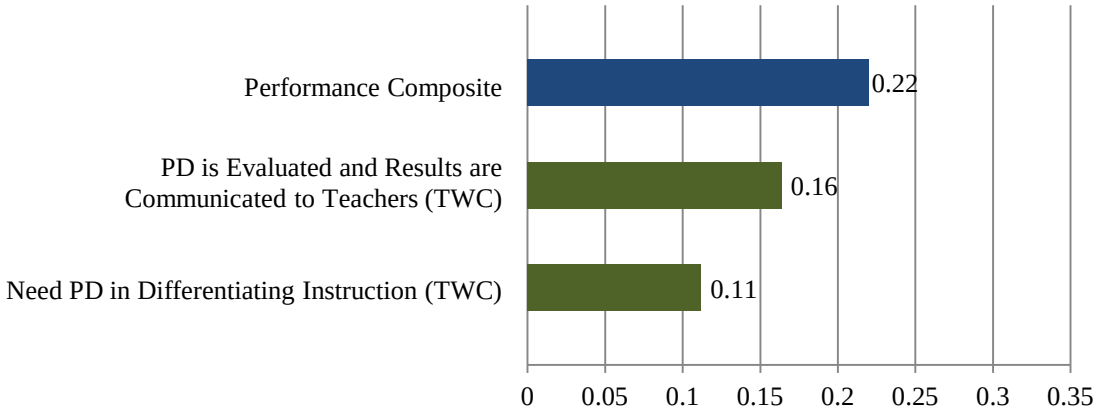
Figure 8. Significant Predictors of Principal Evaluation Composite Scores, 2013-14 School Year



Two objective indicators also significantly predict composite principal evaluation scores. Principal experience is a positive predictor of composite principal evaluation scores, with a one standard deviation increase in principal experience (about five years) leading to a composite evaluation score that is over a third of a standard deviation higher. The percentage of students eligible for free or reduced price lunch is a negative predictor of composite principal evaluation scores, with a one standard deviation increase (about 24%) in the percentage of students eligible for free or reduced price lunch leading to a decrease in composite principal evaluation scores of about a fifth of a standard deviation.

The analyses found fewer predictors of principals' composite evaluation scores in the 2011-12 school year than in the 2013-14 school year, and all of the predictors identified were positive (Figure 9, following page). From the TWC, needing PD in differentiating instruction and evaluating and communicating PD quality both predicted higher principal evaluation scores, with the latter item also predicting scores in 2013-14 (Figure 8). The school performance composite was a positive predictor of composite principal evaluation scores, and a standard deviation increase in performance composite (about 14%) predicted an increase in composite principal evaluation scores of over a fifth of a standard deviation.

Figure 9. Significant Predictors of Principal Evaluation Composite Scores, 2011-12 School Year



Major Findings

- Most of the items from the TWC that significantly correlate with composite principal evaluation scores are not items that would be expected to be important indicators of principal effectiveness.
- We find a general lack of consistency in which factors predict composite principal evaluation scores over time.
- While some objective measures do significantly predict principal evaluation scores, relatively few do so considering the large quantity of items—20 in total—that were examined but were not found to be significant (e.g., teacher turnover, teacher EVAAS, school EVAAS, etc.).

School Context and Principal Ratings

> *Research Question Set 4: Do the principal evaluation ratings appear to be higher or lower based on the context of the school? Is there evidence of either lower-performing principals concentrated at a certain type of school or of principals being rated lower when they oversee those types of schools?*

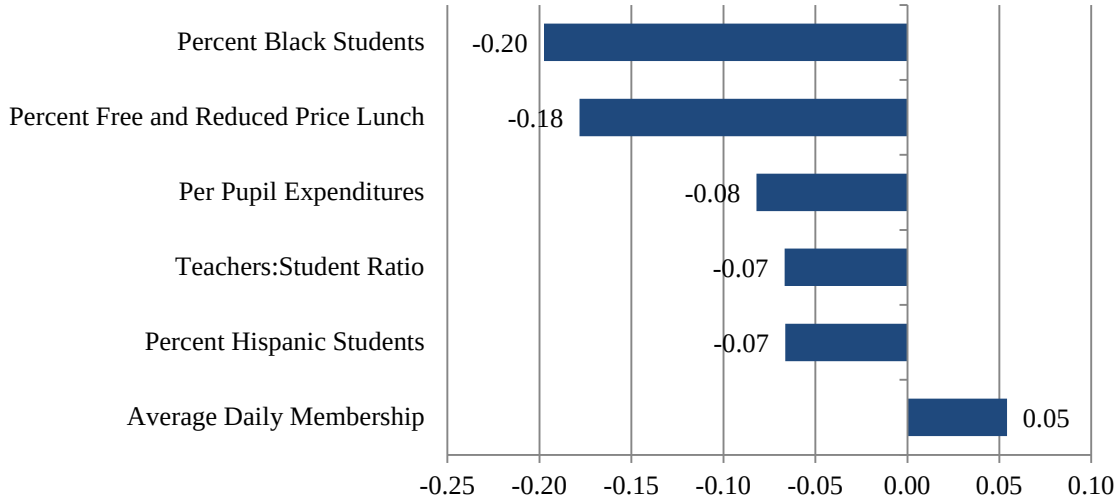
If superintendents' ratings of principal performance were completely independent from the context of the school the principal oversees, then we would not expect the evaluation ratings to be correlated with characteristics of schools. In other words, if principal quality was evenly distributed across schools randomly, then we would expect that schools would have a mix of effective and less effective leadership across the spectrum of school context factors like socio-economic composition, proportion of the student body in each racial category, and urbanicity. However, there are at least two scenarios that could lead to an uneven distribution of principals across schools. First, because more effective principals may be given more latitude to choose to work in certain schools, less effective principals (i.e., principals with fewer professional options) could be disproportionately assigned to schools that are perceived to be less desirable work placements (e.g., lower test scores, higher local crime rate). If this occurs systematically, indicators of challenging school environments will be correlated with superintendents' ratings. Second, if principal effectiveness *is* distributed evenly across schools, but principals at certain

types of schools receive lower ratings due to the challenges of the school context and not due to the principals' performance, these correlations will occur.

While this evaluation is unable to pinpoint which of these mechanisms explains an unequal distribution of superintendents' ratings of principal effectiveness across school contexts, it is at least able to investigate the extent to which superintendents' ratings are correlated with indicators of school context. For indicators of school context, the analysis focuses on factors that are part of the school itself and are unlikely to be affected by the principal. These variables include the size of the student population, the racial makeup of the school, and the poverty rate (as measured by free or reduced price lunch eligibility).

The correlations in Figure 10 are between the listed school context indicators and the composite superintendent rating, which is calculated by taking the median of the scores on the seven standards of the principal evaluation. A positive correlation indicates that as that variable increases in value, the composite score of the principal evaluation tends to increase as well. Conversely, negative correlations show that as the value of the variable increases, the composite score of the principal evaluation tends to decrease.

Figure 10. Correlation Coefficients between School Context Variables and the Composite Principal Evaluation Score



From Figure 10, all of the variables have negative correlations except for average daily membership, indicating that principals at larger schools tend to have higher composite evaluation scores, although at 0.05 this is a very low correlation. The two largest correlations are between the percent of students eligible for free or reduced price lunch and the percent of black students and the principals' composite evaluation score. Principals at schools with a higher percentage of black students or free/reduced price lunch students tend to have lower composite evaluation scores echoing a finding from an earlier section of this report (Figure 8).

Major Findings

- Superintendents' ratings of principal effectiveness do not appear to be equally distributed across school context. As the percentage of black students or free/reduced price lunch students increases, principals' composite evaluation scores tend to decrease.
- It was not possible to test whether less-effective principals were assigned to schools with concentrated poverty or black populations or if instead superintendents rate principals lower, regardless of actual principal quality, when they oversee those types of schools. Either explanation remains plausible and both raise concerns for the evaluation of principals.

The Principal Rating Process as a Tool for Professional Growth

> *Research Question Set 5: Did superintendents provide principals with information on their strengths and weaknesses by making distinctions in performance between the standards? Has the value of superintendents' average ratings changed over time? Has the variation in ratings across standards changed in terms of providing principals with information about their strengths and weaknesses?*

Another important purpose of the NCSSE evaluations is to provide principals with clear information about their strengths and weaknesses in order to help principals reflect upon and continually improve their effectiveness throughout their career as school administrators. The eighth standard, school-level EVAAS, provides an objective measure of the contribution of the school to student achievement, but it does not provide information either about which practices are strengths or weaknesses, or about whether school-level growth is attributable to school leadership or other school-level factors unrelated to administration. To determine whether superintendents use NCSSE to provide clear indications of principals' strengths and weaknesses in these areas, the analysis looked for evidence that superintendents are using the entire range of the rating scale, especially the Developing through Distinguished (2-5) sub-range. The analysis also considers whether, in aggregate, superintendents' ratings changed over time as they develop expertise in rating their principals.

Table 1 shows that, on average, principals were rated between Proficient (3) and Accomplished (4), and the ratings have not changed as superintendents and principals have gained experience with the evaluation rubric. Each year, between 72 and 80 percent of the principals with evaluation ratings received a rating of either Proficient or Accomplished on each standard.

Table 1. Mean and Standard Deviations by Principal Rating Standards

	2010-11		**2011-12**		**2012-13**		**2013-14**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Strategic Leadership	3.8	0.8	3.9	0.8	3.9	0.8	3.8	0.8
Instructional Leadership	3.8	0.8	3.9	0.8	3.9	0.8	3.9	0.8
Cultural Leadership	3.9	0.8	3.9	0.8	3.9	0.8	3.8	0.8
Human Resource Leadership	3.8	0.8	3.9	0.8	3.9	0.7	3.8	0.7
Managerial Leadership	3.8	0.8	3.9	0.8	3.9	0.7	3.9	0.8
External Development Leadership	3.9	0.7	3.9	0.8	3.9	0.8	3.9	0.8
Micropolitical Leadership	3.9	0.8	3.9	0.8	3.9	0.8	3.8	0.8

The stability of the average ratings for each standard could mask wider variability in ratings if superintendents had begun to use both higher ratings (e.g., Distinguished) and lower ratings (e.g., 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 (SD), would change over time. Such a pattern would indicate that, as they gained experience with the rubric, superintendents were more discriminating in their ratings and thus gave principals more useful information through the NCSSE about how they could develop and improve. Table 1 shows, however, that the standard deviations have been stable across standards and over time. Thus, superintendents are not providing more information about principals' strengths and weaknesses as they gain experience with NCSSE.

A potential concern arises from the limited range of the average ratings by category that has been noted in prior research on personnel evaluations. Even though the principal evaluation is divided into discrete categories and multiple dimensions, raters tend to provide global ratings of personnel rather than ratings that reflect individual strengths and weaknesses. To examine the extent to which NCSSE ratings tend to be global ratings of principals rather than ratings of the principals' strengths and weaknesses on each standard, the Evaluation Team conducted a test of the extent to which the seven ratings measured the same thing (global ratings of the principal) or different things (principals' 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), the Team found that the measure was 0.9 for each of the four years in which NCSSE principal evaluations have been conducted. This value indicates that most superintendents rated individual principals' overall performance rather than making distinctions in principals' performance on each of the seven NCSSE standards.

Major Findings

- Superintendents rated principals either Proficient or Accomplished, on average, 75 percent of the time, which provided limited information for principals via the NCSSE about their specific strengths and weaknesses.
- Superintendents' ratings have not varied over time, indicating little refinement in using NCSSE ratings to provide principals with feedback on strengths and weaknesses.
- Superintendents rate principals globally rather than providing meaningful distinctions on principals' performance on each standard.

Conclusions and Recommendations

Currently, it does not appear that the NCSSE is likely to lead to improved performance of principals.

1. Principals' performance as instructional leaders seems to be the primary influence on many of their ratings, rather than measures more directly related to each standard.
2. In addition to school EVAAS, measures of principals' performance—both objective and subjective—should affect evaluation ratings systematically, but few measures are closely related to ratings of relevant standards. Directly incorporating these measures systematically into an overall composite rating should be considered, rather than just recommending that they be used, as is the current policy. Using a composite score from several measures of principals' performance also will decrease the weight placed on the single objective measure of performance that is currently used—the school EVAAS score.
3. Consistent and relatively strong negative relationships between principals' scores and school context variables (such as percentage of free and reduced price lunch students and percentage of black students) indicates that either ratings are not fair or that less-proficient principals are systematically assigned to those schools. Either explanation indicates a problem. Superintendents should be informed of this and attention through professional development or principal transfer policies should be given to reducing or eliminating these correlations.

Overall, it seems that, in spite of a strong theory that systematic evaluation of principals could lead to improving principals' performance through the NCSSE ratings, it is unlikely that the system as it is currently implemented will do so. The vast majority of principals receive ratings above Proficient for all standards, even though many schools are classified as performing below expectations. In addition, this evaluation provides some evidence that principals' rating may not be entirely fair. This may occur due to the addition of school achievement growth measures into NCSSE and higher stakes associated with ratings below Proficient. The North Carolina Department of Public Instruction may wish to convene a group of principals and district leaders to further investigate alternatives for improving the NCSSE ratings. Finally, in addition to the school-level EVAAS scores, the SBE may wish to directly incorporate other measures of principal performance into the principal evaluation process, including retention of effective teachers, teacher survey ratings of principals' instructional leadership, and teacher survey ratings of the fairness and feedback provided in teacher evaluations. Several of these measures could be combined into a composite quantitative rating of principals' overall performance. In addition, the single objective measure of principals' performance, the school EVAAS score, would not be as strong an influence on the principals' overall rating. To accomplish this, the Teachers Working Condition survey may need to be revised to include more questions relevant to principal performance.

Reference

North Carolina Department of Public Instruction (2013). *North Carolina Standards for School Executives*. Retrieved from http://www.ncpublicschools.org/docs/effectiveness-model/ncees/standards/princ-asst-princ-standards.pdf

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

SERVECenter THE UNIVERSITY OF NORTH CAROLINA AT GREENSBORO People · Possibilities · Promise
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