UNC PEMBROKE | MAYNOR HONORS COLLEGE

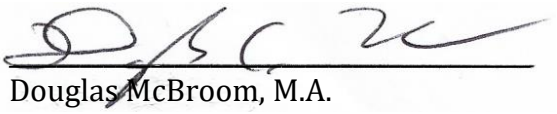
A Regional Comparison of Teacher Evaluations and Student Performance

Senior Project

In partial fulfillment of the requirements for
The Esther G. Maynor Honors College
University of North Carolina at Pembroke

By

Hannah Harrold
Mathematics and Computer Science
12/4/19

_____	_____
Hannah Harrold Honors College Scholar	Date
	_____
Douglas McBroom, M.A. Faculty Mentor	Date
_____	_____
Joshua Kalin Busman, Ph.D. Senior Project Coordinator	Date

Acknowledgements

I would like to express my deep gratitude to Professor Douglas McBroom and Professor Joshua Busman, my research advisors, for their guidance, encouragement and useful critiques of this research work.

Abstract

The following research will explore the teacher evaluation procedures of North Carolina and it's surrounding states. Following the analytic review of these procedures, the paper will then put focus on student performance in comparison to the type of teacher evaluation used in the region. What was found is that among the four states, Virginia had the highest level of student performance. The main difference founded was the inclusion of student evaluations of teachers and teacher-made portfolios with evidence of reaching professional standards. The paper ends with suggestions as to how to improve the evaluation of teachers in North Carolina to better improve student performance.

Keywords: teacher evaluation, comparison, student performance, region

A Regional Comparison of Teacher Evaluations and Student Performance

North Carolina education within recent years has made major changes that will affect the future of students in the NC education system. Beginning in May of 2018, teachers of North Carolina organized to strike in Raleigh to ensure that state lawmakers make room in the budget to give teachers raises as well as fund textbooks and classroom resources (Ball, 2018). Soon after, 40 employees from the North Carolina Department of Public Instruction (NCDPI) were laid off to meet an over five million dollar budget cut. Leaving NCDPI without support services for educators and technology services (Fain, Leslie, Dukes, & Browder, 2018; Hui, 2018).

Currently, North Carolina schools are undertaking big changes in standardized testing in both reading comprehension in elementary schools to the End of Course and final exams in middle and high schools (Hui, 2019;Pires, 2019) With the Testing Reduction Act of 2019 being passed and the confrontation between the NC House of Representatives and Governor Roy Cooper over 2019's state budget (Vaughan, Horsch, & Specht, 2019), it is time to assess how teacher evaluation and student performance in North Carolina compare to surrounding states with different evaluation systems. In consideration of this paper, student performance will be based on the students' overall growth on the state standardized tests. Based on the results, conjectures will be made on what does and does not work in North Carolina's teacher evaluation system.

Literature Review

Historical Context

In the 1700s, teachers in education were chosen by the clergy or other members of the local government. From then, the teachers were considered servants of the public that

was supervised by a selected person or group that had the power to manage the curriculum and give nay changes to employment. As industrialization grew, teachers with expertise in specific disciplines were in demand. These teachers were then supervised by a principal teacher who had knowledge of the disciplines rather than the knowledge of a clergyman. It was in the mid-1800s that teaching became more of a profession requiring training for improved instruction. It was then that the study of educational pedagogy began to rise (Marzano, Frontier, & Livingston, 2011).

By the early twentieth century, views on how education should be evaluated became split between a democratic and scientific view. The democratic method introduced by John Dewey involves practices being surrounded by student needs. Contrarily, the scientific method developed by Frederick Taylor focused on measuring the effectiveness of certain practices in education. Because the scientific method was focused on measuring effectiveness like factory workers, the industrial minded public endorsed the scientific measurement of education effectiveness. It was in 1916 when Ellwood Cubberly made a guideline for measuring teacher effectiveness for school administrators to use. The use of the guideline began the practice of administrators conducting observations in the classroom. In 1929, William Wetzel advocated using student performance to measure the effectiveness of teachers. To do measure student performance, students must be given aptitude tests that can reliably measure specific objectives of the course. From then, standardized testing became more common, however, students were being taught using a democratic philosophy focusing on community and being tested on content knowledge for effectiveness measures. After World War II, the focus of measuring effectiveness was switched to the focus on the

teacher as an individual. As a result of focusing on teacher needs, the effectiveness of teachers depended on teacher resources (Marzano et al., 2011).

Between the period of 1960-1970, clinical supervision became widespread among the field of education. Clinical supervision continued to depend on observing the teacher in the classroom and promoted an increase in communication and reflection between the administrator and teacher. In the 1980s, clinical supervision was built upon by Madeline Hunter. In the clinical supervision model, the development of a common language among educators and administrators was supported so that observations could be focused on the development of a specific learning model. To improve on having a common language among teachers and administrators, many studies were produced to focus more on developmental evaluations rather than the goal-oriented method that was in use. Models similar to clinical supervision were later developed in the 1990s where teacher planning and preparation were included in the evaluation of teachers (Marzano et al., 2011).

In the current age, laws like the "No Child Left Behind Act" of 2001 and the "Race to the Top" initiative in 2009 pushed for the focus of student achievement on teacher evaluation (Dynarski, 2016; Marzano et al., 2011). As time went on, more studies such as the Widget study were conducted and the demand for evaluation changes has since then grown (Marzano et al., 2011). Currently, the question of whether student achievement can indicate teacher effectiveness is growing, leading education to reconsider how both teachers and student achievement are measured.

According to the National Assessment of Educational Progress (NAEP), math scores in the United States have grown since the No Child Left Behind Act, mostly among African American and Hispanic students. However, the scores did not continue to

grow and the current standing of students in the core subject areas show little to no growth. This lead to the Brown Center determining that the past scores on the NAEP may conclude that the educational system is providing stable results and there is little progress being made in student performance as education stands as of 2017 (Hansen, Mann, Valant, & Quintero, 2018).

In addition, two national test examinations, Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS), show that nation test scores can tell different stories. While the 2015 PISA rankings show the U.S. on an average level in comparison to other national systems, TIMSS shows the U.S. to be a competitor among the 34 systems present among the nations with the U.S. being 13th in 4th-grade math and 9th in 8th-grade math as of 2015. PISA shows the U.S. as being 37th among 69 countries, making the U.S. out to be in a rough place in education (Serino, 2017). However, TIMSS tells a different story, leading to the question of how the U.S. education system is doing in reality.

Evaluations

One of the models used today, known as the Value-added model, evaluates student test scores annually measure the quality of teachers. What the model fails to take into account is the other factors outside of the teacher's influence that affect student performance on tests. It has been shown that test scores of students range based on the type of test, whether it be on the same topic and the make-up of students assigned to the teacher. For example, teachers of English Second Language students have lower scores based on the VAM model. Factors that affect the student's scores are the language barrier between the teacher and student as well as other social factors that the student may be

facing. Test scores do not accurately measure the effectiveness of a teacher (American Educational Research Association, 2015; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). A written interview showed how a teacher who had high test scores for years and was considered an effective teacher, went to a teacher with a red flag within one year because the group students she received were ESL students. There are other cases where VAMs results in the inaccurate bias of teachers (Darling-Hammond et al., 2012; Koedel & Betts, 2011).

Characteristics of accurate systems for evaluation involve multiple observations of teachers that focus on the professional standards mandated by the National Board for Professional Teaching Standards (NBPTS). An example of such a program is the Teach Advancement Program, which involves six teacher evaluations throughout the year which are conducted by administrators that have been intensively trained in professional standards. This form of evaluation allows for teachers to gain extensive feedback about their planning and performance. Multiple studies have shown that one of the most considered characteristics of effective evaluations is the extensiveness of feedback (Darling-Hammond et al., 2012; Firestone, 2014). In consideration of student outcomes, the achievement of these professional standards is proven to meet desired student outcomes for the classroom (Darling-Hammond et al., 2012).

A considered solution to motivating teachers to improve testing scores and the effectiveness of teaching is through financial incentives. Fierstone's (2014) study has shown that incentives are inefficient in motivating teachers and decreases the autonomy of teaching. According to a platform of teachers who are gaining their doctorate, they agree that the lack of trust in teachers as professionals has led to teachers being apathetic

to taking actions to improve instruction. This is because of evaluation methods measuring the wrong aspects of education and blaming the teacher for something that they are not in control of (Wallace, 2012).

In a study of emerging evaluation systems among eight states, most of the evaluations are now including multiple assessment models and weights in accordance with the significance of the content being evaluated. Methods that are used include classroom observations, multiple student assessment results, and student learning objectives. While some teachers are apprehensive about the use of student test scores in the evaluation, most schools have noticed a positive correlation with feedback on observations improving teacher performance. Most of the administrators and teachers in the schools are more motivated to become better in instruction (Anderson, Butler, Palmiter, & Arcaira, 2016).

Supporting these emerging evaluation systems, Close and Amrein-Beardsley (2018) concludes after reviewing multiple arguments presented by education associations, education systems should have multiple measures of assessment for their teachers to better evaluate if the teacher is effective. In addition, education systems should include teachers in their evaluation. As Wallace's (2012) multiple interviews stated, teachers should be given more autonomy in their profession. Lastly, evaluations should be served as a formative tool rather than be a consequential measure of effectiveness among educators. As in Anderson, Butler, Palmiter, & Arcaira's (2016) study, teachers and administrators are more motivated to become better instructors when they are given informative feedback from their observations.

Correlation of Evaluations to Student Performance

Based on several studies, teacher effectiveness shows to be correlated to student performance (Virginia Department of Education, 2018). Two factors that have been shown to affect student performance is the teacher and their previous academic achievement level from the previous year. While effective teachers rarely bring low performing students up to par with the expected achievement level, the student does show growth when place with an effective teacher (Sanders, Wright, & Horn, 1997). It has also been shown that teachers with high effectiveness based on evaluations are accurate in identifying students who have high academic growth levels. Along with this, teacher evaluations serve as an identifier of effective teaching practices that can be used to improve teacher with lover effectiveness levels (Milanowski, 2004)

North Carolina

North Carolina bases its evaluation system on the six professional standards established by the North Carolina Board of Public Instruction (NCDPI): Demonstrating leadership, establishing a respectful and diverse environment for students, content knowledge, facilitation of learning, reflection, and contribution to the academic success of students. Teachers are evaluated in four steps. In the first step, teachers, peers, and administration must complete training for the evaluation process. Teachers will then be given the rubric, policy, and schedule for the upcoming evaluation process. In step two, teachers will evaluate themselves and then meet with the principal to discuss the assessment as well as the growth plan and lesson(s) that will be observed during the process. Thirdly, the principal will observe the classroom. The number of evaluations and standards depends on the number of years employed. After, the teacher and principal will meet to discuss any strengths and weaknesses observed. Lastly, the principal will give the

teacher a summary of the evaluations conducted throughout the year based on the rubric and standards placed for evaluations. After summarizing the teacher's evaluation, the principal and teacher will make a growth plan based on their results (NC Department of Public Instruction, 2015).

In the case of North Carolina, 2,523 schools participated in state standardized testing. Based on the results, 27.5 percent of the schools exceeded their growth goals, 45.8 percent met their growth goals, and 26.7 did not meet goals (NC Department of Public Instruction, 2019). In comparison to last year's performance, there was a .5 percent increase in schools exceeding their goal as well as a .1 percent growth in schools meeting their goal. Overall, 1,843 schools met or exceeded growth goals which display a seven percent growth in comparison to the 2017-2018 year (NC Department of Public Instruction, 2019).

South Carolina

South Carolina's system for teacher evaluation strives to measure professional growth and development. The system has been used since 1998 and is known as Assisting, Developing, and Evaluating Professional Teaching (ADEPT). ADEPT bases its evaluation rubrics on the four domains of teacher standards: Planning, instruction, environment, and professionalism (Spearman, 2018). Before the evaluation process begins, teachers are notified of being formally evaluated before they renew their contracts for the new school year. From then, administrators will conduct a series of observations in the teacher's classroom. Each teacher has an evaluation team with an administrator and someone else with the same subject content knowledge as the teacher being evaluated. Before being observed, the teacher being evaluated must attend an orientation for the

evaluation process. Teachers must provide lesson plans to the evaluation team before their observations as well as a Student Learning Objective (SLO) report. Afterward, the teacher will present a self-reflection report for the period of observation and the evaluation team will develop a professional growth and development plan with areas for refinement and/or areas of continued improvement (Spearman, 2018).

The South Carolina Department of Education provides the public with school report cards for each school year presenting academic achievement, school improvement, and other data used to determine rates of student success based on the school year. On average, 50.43 percent of students met or exceeded growth goals for the 2018-2019 year (SC School Report Card, 2019). These scores are based on the SC Ready and EOC state standardized tests for reading and mathematics.

Tennessee

As of 2011, Tennesse uses the Tennesse Educator Acceleration Model (TEAM) evaluation model to give teachers feedback based on multiple observations that are both announced and unannounced (Tennessee Department of Education, 2018). During the observations, administrators look for the four domains that indicate teaching skills, knowledge, and professional performance standards. These four domains include instruction, environment, planning, and professionalism. All to which have their own set of standards and expectations described in the Tennessee Educator Acceleration Model Teacher Evaluation handbook (2018). Teachers are given scores based on the specific standards under each domain on a scale of one to five. Five being exemplary. Before the announced observations, administrators conduct a pre-conference to discuss the teacher's planning for the lesson and the overall unit, as well as to get any significant information

about students who will be in the observed class. After the observation, a post-conference is conducted for the teacher to self-reflect along with the administrator's guidance and recommendations for growth. From there a reinforcement or refinement plan is made (Tennessee Department of Education, 2018). Reinforcement plans are used to make clear what the teacher is doing well and how to remain constant in their performance. Refinement plans are used to identify an area where the teacher needs work and identify a plan to improve on the area. Both plans are primarily developed by the teacher's responses during their self-evaluation interview (Tennessee Department of Education, 2018).

The state of Tennessee measures the districts' performance using the Tennessee Value-Added Assessment System (TVAAS). The system takes the scores from the Tennessee Comprehensive Assessment Program (TCAP) for 4th to 8th-grade students and End of Course (EOC) for high school students to evaluate the school's performance and growth. According to the 2019 composite analysis of the TCAP and EOC results, the schools of Tennessee were the least effective in reaching their growth goals for their schools in 2018-2019.

Virginia

The Virginia Department of Education's (2015) Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers measures teacher effectiveness to provide professional development to the teachers of Virginia. Standards that teachers must follow include professional knowledge, instructional planning, instructional delivery, assess of and for student learning, learning environment, professionalism, and student academic progress (Virginia Department of Education,

2015). Administrators conduct formal and informal observations throughout the school year. Each observation includes a post-observation conference to review the evaluator's comments on teacher performance. Other data sources for teacher performance include student surveys, teacher-made portfolios of teaching standards, and self-evaluations (Virginia Department of Education, 2015).

VDOE measure their state performance by a school quality indicator. Each school is rated on a scale of one to three. One is that the school exceeds or meets school improvement goals. The scores are based on standardized testing, dropout rates, attendance, and civic readiness. On average, 93.3 percent of schools were at level one. Meaning out of 1,825 schools, approximately 1,703 schools met improvement goals (Virginia Department of Education, 2019).

Analysis

Based on the literature review, all four states are using teacher evaluations to enhance teacher development rather than an indicator of whether a teacher should keep employment. In addition, all four states sought for teachers to self evaluate and be involved in the process of their development. While each state had characteristics of an effective evaluation as reviewed previously, there was a range of student performance rates that give rise to questioning what each state did differently to get different results.

While all states placed professional development dependently on observations, each state had different standards to measure. However, the states have a common theme of looking at how teachers plan, perform, and assess students during the class period. Beyond that, the states also look at how teachers create a learning environment and reflect on their teaching. There were no obvious differences among the standards that

teachers are being held at. Also, each state gave teachers feedback for professional growth and development based on their observations.

When looking at student performance levels of each state, there is a stark contrast in Tenessee and Virginia's scores, from Tennessee not meeting growth expectations, to Virginia having 93 percent of their schools meet their goals. Looking at North Carolina and South Carolina, both states have at least half of their schools meet expectations. An evident difference among the evaluation processes of the states is Virginia's inclusion of student evaluations as well as teacher portfolios of evidence meeting teacher standards for the state. Another difference seen among the evaluation process is Tenessee's reinforcement and refinement plans being produced based on teacher reflection rather than the administrators. Both South Carolina and North Carolina have similar processes.

Discussion

Implications

Based on the analysis of North Carolina, South Carolina, Tennessee, and Virginia's teacher evaluation systems and student performance, classroom observations are only effective when they are implemented in professional growth correctly. In the case of North Carolina, the current system is effective in identifying practices that improve student growth, however, there is room for improvement. Based on Virginia's student performance and differences in evaluation practices, it is recommended that North Carolina add student evaluations to their evaluation process. In addition to teacher reflection, North Carolina should also have teachers build portfolios with evidence of reaching professional standards.

Further research

The present research bases student performance on state standardized testing. Given the 2019 Testing Reduction Act that was recently passed in North Carolina, it is recommended to further research how the act will affect the measures of student performance. The reliability of state standardized test scores representing student performance is debated. Therefore, it is unkown if students have progressed in their learning. In addition, the measures of student growth given by the states are based on different scales and could effect the data presented in the paper. Another area to possibly look into are the state's priorities regarding teacher performance and state goals. In addition, the paper is restricted to the surrounding states of North Carolina. Further research could expand their analysis to states throughout the U.S. or in high performing nations.

Conclusion

North Carolina is proven to be effective in measuring teacher effectiveness based on student performance. Through their evaluation system, North Carolina can identify practices that are effective in teaching and encourage the use of these practices in low performing schools. Based on the analysis of North Carolina's evaluation system to surrounding states, a recommended procedure to improve on NC's evaluation method is to include student feedback as well as include teacher evidence of reaching professional standards. Through this, North Carolina's student performance may be improved upon and their evaluation system may be spread to other states of low student performance. It is recommended that further research be done to other states in the country or other high performing nations.

References

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*(8), 448-452.

Anderson, L. M., Butler, A., Palmiter, A., & Arcaira, E. (2016). *Study of emerging teacher evaluation systems* [pdf]. Washington D.C., VA: U.S. Department of Education.

Ball, B. (2018, April 2). As teachers strike across the U.S., North Carolina teachers plan advocacy day in May. Retrieved from http://pulse.ncpolicywatch.org/2018/04/02/as-teachers-strike-across-the-u-s-north-carolina-teachers-plan-advocacy-day-in-may/

Close, K., & Amrein-Beardsley, A. (2018). Learning from what doesn't work in teacher evaluation. Retrieved from: https://www.kappanonline.org/learning-from-what-doesnt-work-in-teacher-evaluation/

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15.

Dynarski, M. (2016, December 6). Teacher observations have been a waste of time and money. Retrieved from https://www.brookings.edu/research/teacher-observations-have-been-a-waste-of-time-and-money/

Fain, T., Leslie, L., Dukes, T., & Browder, C. (2018, June 29). NC superintendent: Lawmakers should delay $5.1M cut to education agency. Retrieved from https://www.wral.com/40-state-education-staffers-laid-off-21-vacant-positions-eliminated/17665067/.

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43*(2), 100-107.

Hansen, M., Mann, E., Valant, J., & Quintero, D. (2018, July 5). 2018 Brown Center report on American education: Trends in NAEP math, reading, and civics scores. Retrieved from https://www.brookings.edu/research/2018-brown-center-report-on-american-education-trends-in-naep-math-reading-and-civics-scores/

Hui, T. K. (2018). 40 NC education employees laid off, including some who help low-performing schools. Retrieved from https://www.newsobserver.com/news/local/article214065504.html

Hui, T. K. (2019, September 5). NC Governor Roy Cooper signs law eliminating more than 20 tests given to students. Retrieved from https://www.newsobserver.com/news/local/education/article234749352.html

Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, *6*(1), 18-42.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). Chapter 2. *A brief history of supervision and evaluation*. Alexandria, VA: Association for Supervision & Curriculum Development.

Milanowski, A. (2004) The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33-53. DOI: 10.1207/s153279320pje7904_3

NC Department of Public Instruction. (2015). Evaluation instruments and processes. Retrieved from: http://www.dpi.state.nc.us/effectiveness-model/ncees/instruments/

NC Department of Public Instruction. (2019). 2018–19 performance and growth of North Carolina public schools excecutive summary. Retrieved from: http://www.dpi.state.nc.us/docs/accountability/reporting/2019/documentation/exsumm19.pdf

Pires, R. (2019, August 27). Questions, concerns swirl around changes to statewide reading test. Retrieved from https://wlos.com/news/local/questions-concerns-swirl-around-changes-to-statewide-reading-test

SAS Institutes. (2019) 2018-2019 composite trends. Retrieved from: https://tvaas.sas.com/evalComposite.html?as=v&aj=u&w4=106&ww=174506&ab=dA&x9=15

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57-67.

SC School Report Card. (2019). Academic achievement. Retrieved from https://screportcards.ed.sc.gov/overview/academics/academic-achievement/?q=eT0yMDE5JnQ9UyZzaWQ9MDAwMA.

Serino, L. (2017, April 6). What international test scores reveal about American education. Retrieved from https://www.brookings.edu/blog/brown-center-chalkboard/2017/04/07/what-international-test-scores-reveal-about-american-education/

Spearman, M. M. (2018, April). Expanded ADEPT support and evaluation system guidelines for classroom-based teachers. Retrieved from https://ed.sc.gov/educators/educator-effectiveness/expanded-adept-resources/https-ed-sc-gov-educators-educator-effectiveness-expanded-adept-resources-educator-evaluation-guidance-2018-19/2018-19-expanded-adept-guidelines-april-2018/

Tennessee Department of Education. (2018). Tennessee educator acceleration model team teacher evaluation evaluator handbook. Retrieved from https://team-tn.org/wp-content/uploads/2013/08/TEAM-Teacher-Evaluator-Handbook-2017-18_Add-Gifted-Doc1.pdf

Vaughan, D. B., Horsch, L., & Specht, P. A. (2019, September 11). NC house overrides budget veto in surprise vote with almost half of lawmakers absent. Retrieved from https://www.newsobserver.com/news/politics-government/article234962017.html.

Virginia Department of Education. (2015). Guidelines for uniform performance standards and evaluation criteria for teachers. Retrieved from http://www.doe.virginia.gov/teaching/performance_evaluation/teacher/index.shtml

Virginia Department of Education. (2019). School accreditation ratings. Retrieved from http://www.doe.virginia.gov/statistics_reports/accreditation_federal_reports/accreditation/index.shtml.

Wallace, J. D. (2012). Teacher evaluation: A conversation among educators. *Phi Delta Kappan, 94*(3), 44-46.