


1-2015

Teacher effectiveness and student achievement on STAAR: Implications for school leaders

Yanira Oliveras Ortiz

The University of Texas at Tyler, yoliverasortiz@uttyler.edu

Follow this and additional works at: http://scholarworks.uttyler.edu/edulead_fac

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Leadership Commons](#), [Elementary and Middle and Secondary Education Administration Commons](#), and the [Elementary Education Commons](#)

Recommended Citation

Oliveras Ortiz, Yanira, "Teacher effectiveness and student achievement on STAAR: Implications for school leaders" (2015). *Educational Leadership and Policy Studies Faculty Publications and Presentations*. Paper 4. <http://hdl.handle.net/10950/532>

This Article is brought to you for free and open access by the Educational Leadership and Policy Studies at Scholar Works at UT Tyler. It has been accepted for inclusion in Educational Leadership and Policy Studies Faculty Publications and Presentations by an authorized administrator of Scholar Works at UT Tyler. For more information, please contact tbianchi@uttyler.edu.



“Educational leaders must protect their children’s instructional time; they must ensure that our students are engaged in learning that’s meaningful, relevant and will prepare them for the real future, not just an upcoming test.”

-Y. Olivera-Ortiz



Teacher Effectiveness and Students Achievement on STAAR: Implications for School

Leaders

Dr. Yanira Olivera-Ortiz

Introduction

The quality of education and teachers' effectiveness impact students' lives beyond their classroom performance and achievement as measured by standardized tests (Chetty, Friedman, & Rockoff, 2012). Regardless of the ample research documenting the impact an effective teacher has on the children's education and future, there is a prevalent failure to recognize teacher effectiveness, to effectively document poor teacher performance and to provide teachers with the support needed to improve their pedagogical practices (Weisberg, Sexton, Mulhern, & Keeling, 2009). Quite the opposite has taken over our education system since the passing of the *No Child Left Behind Act*. Teachers are regularly blamed for students' performance on state tests while little attention is placed on the identification of effective practices, the impact highly effective teachers could have in the learning, and the identification of marginal teachers who need support to improve their practices.

A Principal's Predicament

Aware of the importance of the quality of education and teacher effectiveness, I spent countless hours as a principal in the classrooms looking for effective practices and identifying those who needed additional support. As I visited classrooms, I often found myself in classrooms where I would not hesitate to place my own children given the high quality of instruction taking place. Students were engaged in learning at the appropriate cognitive level; they were challenged by activities that were directly aligned to the state's curriculum while providing students with relevant learning experiences. These observations provided evidence that the level of instruction taking place in many classrooms has significantly improved since I began visiting classrooms 13 years ago. Yet as a principal, I found myself consoling many of those same teachers when the state test scores were released in the spring. I often wondered how to help good teachers understand that they had added value to their students' education, when they felt they had failed their students. They, as the public, measure their effectiveness based on the state tests and not on the students' successes they had witnessed throughout the school year. Teachers are constantly blamed for failing to prepare students for standardized tests while they effectively implement research-based, instructional practices and have evidence that students are learning and mastering the state's standards through authentic learning experiences. The apparent disconnect between good teaching and the students' performance on the states tests, and the importance placed on these scores, is troublesome.

Researchers have been shown that tests do not measure the outcome of the educational system as policymakers assume tests could (Williams, 2010). As an instructional leader, I am amazed at the number of educational leaders that continue to fail to identify the effective teachers who serve



their students and persistently yield to the pressure of producing high student achievement scores, even when the practices used to produce the high scores may not always address what is right for every child. The pressure of knowing that the public will solely measure the teachers and schools' success based on the students' performance on the state's tests has resulted in "decontextualized test preparation" (Firestone, Schorr, & Monfils, 2004) with no consideration to authentic learning of the subject matter (Firestone et al., 2004; McCollum, 2011). Given the constant pressure principals and teachers have to produce high scores, I found it compelling as a school leader to take a closer look at the State of Texas Assessment of Academic Readiness (STAAR) results of teachers across a North Texas school district and the relationship the scores had to the teachers' effectiveness as identified by principals through the teacher evaluations. Were teachers who are deemed good teachers by their administrators producing higher student scores on the STAAR tests?

An Effort to Find Some Answers

In the fall of 2013, I conducted a correlational study at a North Texas Independent School District in an effort to determine the empirical relationship between teacher effectiveness, as identified by principals through teacher evaluations, and the impact teachers had on student achievement. A few of the criteria within each of the two analyzed Professional Development and Appraisal System (PDAS) domains had a positive impact in explaining the difference in students' STAAR scores; however, the impact was consistently low and was rarely supported by similar criteria within the domains. Furthermore, when the overall STAAR scores and raw teacher evaluation data were analyzed, the disparity between teacher evaluations and student performance on STAAR was startling. The PDAS raw data showed that 96% of reading teachers and up to 99% of all mathematics teachers were rated as proficient or exceeding expectations on their 2013 PDAS evaluations.

In contrast, according to the Texas Academic Performance Report (Texas Education Agency, 2013) 70% of the third through fifth grade students in the school district met the reading standards while 68% met the mathematics standards. It is important to point out that the 2013 STAAR reading passing standard ranged from 50 to 57% and the 2013 mathematics STAAR passing standard ranged from 54 to 60% (Texas Education Agency, 2014), anywhere from ten to 20 percentage points below what is popularly considered passing, 70%. Furthermore, only 28% of the district's third through fifth graders met the recommended final standard in reading, while 41% of Texas third through fifth grade students met the same standard. The recommended final reading passing standards, to be implemented as the passing standard in 2021-2022, range from 75 to 78% as passing. Similarly, 19% of the district's third through fifth grade students met the recommended final passing level in mathematics, set between 78 to 83% (Texas Education Agency, 2014) while 34% of the State's students met the same standard.

The analysis of the raw data indicate that the PDAS ratings of teachers in the North Texas school district were consistently high while the performance of the students was below the state's average. Correspondingly, in Pate's 2010 study, Texas teachers who were rated as highly effective on PDAS were associated with low student scores on the Texas assessments (Pate, 2010). The conclusions drawn from this study also compare to the findings of previous studies



from across the United States where teachers working at low performing schools received high ratings on their performance evaluations (Weisberg et al., 2009).

A Leader's Perspective

The inconsistencies between ratings of teacher effectiveness and student learning as identified by the state assessment system are troubling. If one believes that effective teaching practices should be reflected in tests scores that purport to measure student achievement, the findings concerning the impact of the instructional practices measured by PDAS and students STAAR scores lead to a wide range of potential conclusions and a host of concomitant questions. Some of these questions focus on issues related to teacher evaluation as a tool for identifying effective teaching practices while others relate to issues surrounding the use of standardized tests as the measure of student learning.

If one assumes that campus administrators used PDAS as designed; it could be argued that the instrument is measuring instructional practices that have little to no impact on student achievement as measured by high-stakes testing. Conceivably, the criteria valued and rated through PDAS promote good teaching with a focus on student engagement in student-centered, self-directed challenging learning, but fail to value instructional strategies that facilitate the students' transfer of knowledge and skills to standardized tests. The majority of the teachers might indeed have met the PDAS criteria to be rated as proficient or exceeding expectations, but failed to provide students with the necessary skills to demonstrate mastery of the acquired knowledge and skills on the tests like STAAR. On the other hand, the disconnect between the instructional practices documented through PDAS and the students' achievement could possibly be a product of the high-stake testing shortfalls. Standardized tests are unable to measure some of the objectives of learning (Koretz, 2008).

I would also argue that if the teachers were proficient or highly successful in the implementation of effective instructional strategies, the disparity between instructional practices and student achievement could be the result of factors beyond the educators' control, which could negatively impact their students' achievement on the STAAR tests. The study was conducted in a school district where 81 percent of all students are economically disadvantaged, 40 percent are English language learners and 72 percent are Hispanic. The challenges students from these historically underperforming groups face could have had an impact on the students' STAAR performance. Research has shown that Hispanic students enter kindergarten with lower reading and mathematics skills than White students with wider gaps upon entering kindergarten when the students come from economically disadvantaged homes (Reardon & Galindo, 2009). Therefore, one could make a case that PDAS does not take into account the different research-based practices that have been shown to be effective when working with the challenging population served at the school district where the study was conducted. Furthermore, I would question the effectiveness of the tests like STAAR in measuring these historically underperforming students' achievement.

Another claim, is that standardized tests, such as STAAR, are not accurate measures of what students are in fact learning in classrooms where teachers implement research-based instructional



practices. “All data do not measure the most important things, and some of it may be trivial, misleading, or unimportant” (English, 2010, p. 113). If the previous argument is accurate, campus administrators must help teachers find a balance between research-based instruction that effectively meets the needs of all students, while preparing for standardized tests that might not measure the most valuable educational experiences. Since the authorization of the *No Child Left Behind Act of 2001*, “whatever could not be measured did not count” (Ravitch, 2010, p. 21). School success and effectiveness are measured by the performance of students on standardized tests; testing has become a major concern in schooling (Ravitch, 2010). In order to be considered successful in an era of accountability and high-stakes testing, campus administrators and teachers have the responsibility to find a balance in engaging students in learning that is challenging, student-centered and self-directed while preparing students to transfer the knowledge and skills to the tests. Teachers must master the balance between the “art and the science of teaching” (Marzano, 2007) as they continue to operate in an education system that defines a year’s success on the students’ performance on the day of the state’s test.

Implications for School Leaders

Although many would argue that high stakes tests are necessary to hold schools and teachers accountable, researchers have shown that the quality of education has weakened after the implementation of higher standards, high-stakes testing and sanctions linked to state testing (Firestone et al., 2004; Gordon & Reese as cited in Haney, 2000). The expectations and the continuous pressure to deliver high test scores as an indication of effectiveness seemed to have changed instructional practices. The increasing pressure teachers feel to improve student achievement as measured by high-stakes testing has resulted in a growing trend in classrooms around the nation; teachers feel compelled to teach to the test. Teaching to the test implies that teachers are altering instructional practices with the sole purpose of helping students do well on the tests, with no consideration to authentic learning of the subject matter (Firestone et al., 2004; McCollum, 2011). Teaching to the test often results in the abandonment of high-yield instructional strategies to focus on the narrow scope of the curriculum that is tested (Marchant, 2004). Firestone et al. (2004) have labeled this practice as *decontextualized test preparation*; the planning and teaching of lessons that are loosely related to the curriculum and mainly focused on the test (Firestone et al., 2004). Researchers have advised about the danger of using high-stakes testing as a measure of school effectiveness. According to Rappelle as cited in William (2010):

True accountability in education should not be facilely linked to mechanical examination results, for there is a very distinct danger that the pedagogical methods employed to attain those results will themselves be mechanical and the education of children will be so much worse. (p. 108)

If one assumes that the majority of the teacher evaluations included in the study were accurate representations of what was taking place in the observed classrooms, one could argue that the teachers in this school district were not using “decontextualized test preparation” but rather were indeed implementing research-based, effective instructional strategies. Then, the issue school leaders must address is the lack of alignment between effective instructional practices and test readiness. But is that the goal of our school system? Dr. Walter Stroup testified at a House



education committee meeting and said that the Texas STAAR tests measure test-taking skills rather than mastery of the curriculum. In an interview released by The Dallas Morning News (2014), Stroup refers to the tests as insensitive to instruction. So what should school leaders and teachers do? Should they focus on preparing students for a test that might not measure what they have learned rather how well they can take a test? Should they prepare students for their future learning to increase their success in advanced courses? Do teachers continue to prepare students for the one day in the year when they will have to shine so their teachers are considered effective or should they focus on what educators and educational researchers know is for right for children? If school leaders are hired to advocate for their students, to protect our children's rights to an excellent education, how can we continue to pressure our teachers to prepare students for tests that might not measure what the students have learned? It is heartbreaking to visit a school in late August and find students working on reading passages and multiple-choice questions, in decontextualized test preparation, because the school principal and the teachers received the message clearly; they must increase student scores on the April exams. But who dares to blame these educators when their jobs might be on the line? I dare say that preparing students for a test was not the reason this principal and her teachers went into education but they find themselves in an impossibly difficult position.

Conclusion

After spending countless hours analyzing about the results of this study and reflecting on the state of our education system, the questions regarding what has happened to our education system are more rather than fewer. I embarked in this journey hoping to shed some light on the relationship between the classroom practices, as evaluated by Texas teacher evaluations, and students' performance on the reading and mathematics STAAR tests. My original intent was to identify any instructional practices considered effective under PDAS that are positively impacting achievement of students on the reading and mathematics STAAR tests. However, the students' scores lead to questions regarding the gap between learning taking place in the classrooms and students' achievement on the state's assessments. Furthermore, one must wonder what are the states tests truly measuring. The results show that the issues school administrators face are much deeper than instructional practices and test preparation. The main question remains, can we engage our students in authentic learning while preparing them for standardized tests? Can good teachers do what they know is right for children while working in an educational system that seems to value tests scores over authentic learning? The unanswered questions are many but our students have no time to wait until we find all the answers; our children are our counting on our leadership to protect and provide them with a strong education that prepares them for their future, not just a test. Regardless of how many questions remain unanswered and all the new questions that come to mind, school leaders must find a way to value and promote what's truly important in education and avoid getting bugged down by what's valued by outsiders which might not be what truly matters for our children's future. Educational leaders must protect their children's instructional time; they must ensure that our students are engaged in learning that's meaningful, relevant and will prepare them for the real future, not just an upcoming test.



References

- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2012). Great teaching. *Education Next*, 12 (3) Retrieved from <http://search.proquest.com/docview/1237826679?accountid=13158>
- English, F. W. (2010). *Deciding what to teach & test: Developing, aligning, and leading the curriculum*. Thousand Oaks, California: Corwin.
- Firestone, W., Schorr, R., & Monfils, L. (Eds.). (2004). *The ambiguity of teaching to the test: Standards, assessment, and educational reform*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved from <http://epaa.asu.edu/epaa/v8n41/index.html>
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, Massachusetts: Harvard University Press.
- Marchant, G. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science*. 104 (2), 2-7.
- Marzano, R. (2007). *The art and science of teaching*. Alexandria, VA: ASCD.
- McCollum, D. (2011). The deficits of standardized tests: Countering the culture of easy numbers. *Assessment Update*, 23(2), 3-5.
- Pate, G. (2010). *The Texas professional development and appraisal system: Links to student achievement* (Doctoral dissertation, Capella University). Retrieved from <http://search.proquest.com/docview/839826028?accountid=13158>. (839826028).
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grade. *American educational research journal*. 46 (3), 853-891. Retrieved from <http://www.jstor.org/stable/40284864>
- Texas Education Agency. (2013e). *2012-2013 Texas academic performance report*. Austin, TX: Retrieved from <http://ritter.tea.state.tx.us/perfreport/tapr/2013/srch.html?srch=D>
- Texas Education Agency. (2014). *STAAR® raw score conversion tables for 2012-2013*. Austin, TX: Retrieved from <http://www.tea.state.tx.us/index4.aspx?id=2147510637>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect. *The Education Digest*, 75(2), 31-35. Retrieved from <http://search.proquest.com/docview/218196265?accountid=13158>
- Weiss, J. (2014). How the Texas testing bubble popped: Part 3. Testing system shaken to its core. *Dallas Morning News*. Retrieved from http://res.dallasnews.com/interactives/2014_March/standardized_tests/part3/
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107-122.