

International Journal of Learning, Teaching and Educational Research
Vol. 18, No. 12, pp. 240-257, December 2019
<https://doi.org/10.26803/ijlter.18.12.14>

Analysis of Accountability Ratings for Elementary Schools Texas, USA: Considering the School Growth and Students' Demographics

Felipe Sepulveda*

Universidad Católica de la Santísima Concepción,
Departamento Fundamentos de la Pedagogía.
Centro de Investigación y Desarrollo CIEDE-UCSC

Christie Bledsoe, Marlene Zipperlen and Robert Rose

University of Mary-Hardin Baylor

Abstract. The purpose of this quantitative study was to compare the Texas school accountability approach used from 2004 to 2011 based on a status-based model to a model that incorporates academic growth and student demographic variables. Information from three large urban school districts in Texas was analysed considering data from 398 elementary schools and the reading and mathematics TAKS test scores of 24,065 fourth and fifth grade students. Results indicated that the average school growth did not present significant differences between institutions with different accountability rating defined by TEA in 2011. Statistical differences were found when the average school growth was analysed disaggregating the students' population by ethnicity, socioeconomic status, and English language proficiency. A systemic approach must prevail for designing an accountability system used to judge the quality of education delivered at an elementary school. This research was an analysis of how an accountability system could be useful to evaluate school effectiveness based on standard-based assessment results. However, the use of only one measurement to judge a school's efficacy regarding the quality education provided to students may be a reductionist and narrow approach.

Keywords: school accountability; standard-based assessments; academic growth; school ranking.

* Corresponding Author:

Mailing Address: Alonso de Ribera 2850, Concepción, Chile

Phone Number: (56) 9523 63019,

E-mail: fsepulveda@ucsc.cl

1. Introduction

In an increasingly competitive, knowledge-based, and globalized world, having a strong public educational system is a priority for developing future generations of college-ready students and skilled workers. Education was a priority reform under President George W. Bush, and the No Child Left Behind Act 2001 (NCLB) had support of the public as a fundamental step to develop a high quality, equitable public education in the United States. The new reform improved federal financial support allocated to education, but required the state legislators to define content standards for testing each student from third to eighth grade. A core principle contained in NCLB was to hold educators accountable for providing quality education that might be reflected in performance on standards-based assessments. The NCLB policy defined school achievement goals based on the Adequate Yearly Progress (AYP). This indicator addresses the percentage of students who scored at or above the proficient level and defines school achievement for the academic year. Therefore, school accountability is defined around the concept of achievement status rather than school effectiveness (Forte, 2010; Linn, 2008). Based on this assumption, a higher number of students meeting the standards are an indicator of a higher quality of education provided by the school. Thus, differences on students' performance on standardized tests illustrate the school's quality. However, an accountability model based on a status approach does not necessarily represent a school's effectiveness in terms of student learning and progress over time. Schools with students who score at or above the proficiency standard still achieve AYP even if the students did not show improvement from year to year. Consequently, concerns were raised whether accountability systems based on status or growth students' performance is more appropriate (Ladd & Lauren, 2010; Mittleman and Jennings, 2018). Holding schools accountable for growth rather than status may become a fairer option, however further research should be conducted to determine whether growth-based accountability has positive impact on building a more equitable accountability system (Lauen & Gaddis, 2016). Ready (2013) suggested schools with substantial proportions of students living in high poverty may have an advantage compared with students from low poverty groups in showing academic growth based on value-added accountability systems. However, many of the growth-based accountability models that use value-added measures may produce ratings that are highly unstable highly influenced by class composition (Mathis & Trujillo, 2016). Certainly, more research is needed on determining the best use of accountability based on academic growth.

The purpose of this quantitative study is to compare the Texas school accountability approach used during 2003 to 2011, based on a status model to a model that incorporates academic growth and student demographics. The central research question in this study is: What is the difference between school accountability rating in terms of TAKS tests scores growth and the students' demographic variables? The results of this research might guide the discussion about how a public education system, with a diverse population of students, should use end-of-the-year standards-based assessments scores to hold schools accountable about the quality of education provided. This research aims to

contribute to the body of evidence, regarding the debate about using accountability ratings based on high stakes assessments.

2. Theoretical Framework

2.1 Standards-based assessments under NCLB

The use of a standards-based test as a tool for measuring school performance is popular among policy makers and the general population, because it provides an idea about a school's educational performance. When NCLB was implemented, there was strong public support for using tests to track student progress from third to eighth grade (Hess, 2006). Also, parents, teachers, administrators, and authorities perceived standards-based assessments as a valuable instrument and tool for improvement (Ballard & Bates, 2008). It is a common belief that analysing summative standards-based test scores facilitates the work of teachers by providing clear objectives on what to teach (Ballard & Bates, 2008). An accountability system based on high-stakes testing has the inherent assumption that educators will try harder and will implement effective instructional strategies (Heilig & Darling-Hammond, 2008).

After more than 10 years of full implementation of the NCLB policy, there was an increasing concern that high-stakes tests have moulded the classroom instructional design (Finkel, 2010). With an accountability system based on a student's performance on standardized tests, teachers tend to focus their lessons on rote memorization of facts, preparing the students for multiple-choice questions (Finkel, 2010). Students expend increasingly large amounts of time preparing for the test and learning strategies for gaming the test, such as discarding the obviously wrong answers (Krieg, 2011; Schlechty, 2011; Zimmerman & Dibenedetto, 2008). Even experienced teachers might believe their instruction is not fulfilling the students' academic needs, allocate more time for language and mathematics instruction (Wills & Sandholtz, 2009), and modify interactions with the students (Barrier-Ferreira, 2008). Also, the emphasis on defining teaching effectiveness based on test results may prevent collaboration among teachers and promote practices that do not have the best interest of learners at heart (Brevetti, 2014). Teachers tend to focus instructional resources on academically average students (also called bubble kids), which might cause the "zero-sum," referring to the increased gains for bubble kids that coincide with decreases in gains for other students resulting in a sum of zero in relation to overall academic progress (Ho, 2008, p. 357).

Under a strong standards-based accountability system, teachers are encouraged to develop instructional plans based on the academic standards being tested, discouraging creativity and reducing teaching motivation (Lobascher, 2011). The emphasis on defining education quality based on standardized testing and making teachers accountable for the test results, has led to an excessive focus on teaching to the test. Furthermore, the test results are not available in a timely manner, so the data is not useful because teachers' own experiences with the students quickly supersede the information provided by those test scores (Young & Kim, 2010). NCLB policy, centred on accountability through standardized testing, represents a tangible threat to the teacher's sense of

autonomy (Moloney, 2006). As a consequence, teachers might feel discouraged to extend their lessons above what is defined by a specific academic standard. The standard can limit instruction when the teacher centres lessons on mastering only those competencies measured on the exam to the exclusion of other topics (e.g., collaboration, research, project design) which may be educationally important but untested (Horn, 2003).

2.2 Defining school accountability

With educational policies like NCLB, lawmakers seek to make schools accountable for delivering education excellence for all students. The current educational policy on United States relies on testing as the main source of information regarding the quality of education delivered in schools. The classroom instruction should be focused on the important skills and content defined in the curriculum, which will be tested (Madaus & Russell, 2010). The NCLB policy has school achievement goals based on Academic Yearly Progress (AYP). This indicator reflects the percentage of students who scored at or above the proficient level for academic achievement. The federal goal is for all students to score on or above grade level according to the state standards by the end of 2014. Therefore, it is necessary to increase students' performance on standardized testing consistently in order to reach the NCLB imposed goal of 100% proficiency. If low performance based on the AYP goals occurs, school administrators and teachers must engage in an improvement process directed to increase students' academic achievement on state standards-based assessments. Based on this model, a higher number of students meeting the standards are an indicator of higher quality education being provided by the school. Consequently, school accountability is defined around the concept of achievement status rather than school effectiveness (Forte, 2010; Linn, 2008). The current accountability model does not necessarily represent if a school is effective in supporting student learning and progress over time. One of the main concerns of this accountability system is that schools with students scoring at or above the proficiency standard still achieve AYP even if the students are not showing improvement from year to year.

The school achievement model, defined by NCLB and implemented in most states, focuses on the performance of successive cohorts of students, rather than individual students. For example, improvement is measured by performance of Grade 4 students one year, but the scores are compared to the performance of a different group of students in Grade 4 the following year. This accountability approach, typically called "status model," involves using sets of annual achievement targets and compares them with the status of current students' achievement (Linn, 2008, p. 700). Presumably, the reason for using a status orientation as the reporting vehicle in NCLB is to provide no excuse for student failure, regardless of student's background or when the student entered a particular school (Marzano & Waters, 2009). Following this accountability model, the AYP goals are attained in conjunction with the performance of cohorts from year to year in terms of the proportion of students attaining proficiency on state standards-based tests, rather than examining whether a previously underperforming student is making progress toward attaining

proficiency (Choi, Seltzer, Herman, & Yamashiro, 2007). Consequently, school accountability mainly indicates the proportion of students in each grade level (from third to eighth grade) achieving the academic proficiency level (William, 2010). The use of the status accountability model raises the issue of what really is being measured: student achievement or educational effectiveness (Forte, 2010). This vision for measuring year-end performance levels rather than improvement of student achievement has occasioned debate. Defenders support this design stating that it is imperative if the focus is to close the achievement gap (Hess, 2006).

2.3 Status accountability model versus measuring academic growth

The discrete results on improving public schools academic performance, since NCLB was implemented, have raised discussions related to improvement of the school accountability model. Most critics focus on the technical deficiencies associated with the current status approach to define school academic proficiency. The lack of randomization of the current accountability systems has serious limitations with regard to the available data about instructional practices, teacher characteristics, and students' characteristics (Linn, 2006). Education policymakers are beginning to realize that inferences based on school accountability systems that use single-year comparisons of school performance aggregate data simply reflect accumulated inequities and enrolment characteristics, and not necessarily the school effectiveness in influencing student academic progress (Goldschmidt, Choi, Martinez, & Novak, 2010). A status oriented accountability approach is well intentioned, but it may be unfair as a method of determining the effectiveness of a district or school because populations with different demographics are not equal (Marzano & Waters, 2009). Those differences strongly relate to disparities in student achievement, but might have nothing to do with the quality of the education provided (Marzano & Waters, 2009). As a consequence, this accountability model is vulnerable to rewarding schools that serve students from economically advantaged backgrounds and penalize schools located in more disadvantaged context (Zvoch & Stevens, 2008). Furthermore, the status approach to determine school effectiveness toward meeting the AYP goals presents inconsistencies based on meeting assumptions upon which this accountability model is founded. Raudenbush (2004) concluded inferences about differences in school effectiveness are scientifically indefensible. NCLB contains language such as evidence-based decisions and scientifically based research, but ironically, the NCLB accountability system leads to causal inferences about school quality based on evidence, which does not satisfy the tenets of scientifically based research (Linn, 2006).

In a status-based accountability system, all students must demonstrate grade level proficiency regardless of the students' prior academic and social background. Therefore, current test-scores on standardized assessments are contaminated with factors other than school performance. Meyers (2000) found student, family, and community characteristics, prior to entering Grade 1, affected student achievement in terms of growth from first grade thorough the grade in which students are tested. By a set of related academic and social co-

variables acting in concert to define the student's academic performance, this example clearly illustrates how a student's performance on a standardized assessment might be determined.

Accountability models based on students' academic growth are an option to the current status approach. Researchers promoting the academic growth model to define school accountability concluded this approach might overcome some deficiencies inherent to the status accountability approach. Based on empirical information, it is possible to demonstrate the measure of academic growth has a weaker correlation to students' socioeconomic characteristics than status accountability measures (Braun, 2009). Centring the accountability model on the students' gains helps to eliminate some alternative explanations of differences in students' achievement based on the status model (Linn, 2006). Furthermore, emphasizing students' academic growth remedy disproportionate attention to bubble kids by locating all non-proficient students in the same bubble (Ho, 2008). Growth models also enable the recognition of teachers and schools dealing effectively with students at various levels of proficiency (Klein & Hoff, 2007). Considering the diverse student population within the public-school system and among districts, evidence shows a status accountability approach might foment an uneven accountability system.

The widespread interest in the growth approach to accountability led U.S. Department of Education officials to introduce the Growth Model Pilot Program (GMPP) in which educators in nine states implemented a growth model to determine AYP (Linn, 2008). For the first time, the GMPP allowed individual student growth trajectories to factor into the educational accountability framework mandated by NCLB (Briggs & Weeks, 2009). Results from the GMPP showed using the growth model approach reduced the number of schools with inadequate AYP. From this perspective, the number of schools below AYP decreased by 16% overall because of the GMPP (Hoffer et al., 2011). The growth accountability model is a method of identifying schools in which students are making progress even though they may not yet be reaching proficiency standards. The principles included in the GMPP are a consideration for the upcoming reauthorization of the NCLB (Ho, 2008).

2.4 Texas state public school accountability

The NCLB act is a federal mandate, but each state was required to define an accountability system. Texas legislators were among the pioneers in implementing a state-wide accountability system in United States. In 1993, Texas legislators enacted the Texas public school accountability system to rate school districts and evaluate campuses. The state assessment is a direct measure school performance. Based on school performance, officials at Texas Education Agency (TEA) can sanction schools not meeting performance targets and reward schools exceeding targets (Kress, Zechmann, & Schmitten, 2011). After 10 years of the first standards-based accountability system implementation, the TEA developed a new assessments and accountability plan. The Texas Assessment of Knowledge and Skills (TAKS) was used during 2003-2011 and assigned school a categorical performance rank (that ranged from Exemplary, Recognized,

Academically Acceptable, or Academically Unacceptable) based on assessment results, longitudinal completion rates, and dropout rates (TEA, 2013). Finally, in 2012 the end-of-the-year standards-based assessments (TAKS) was replaced by the State of Texas Assessment of Academic Readiness (STAAR), (House Bill 3 Transition Plan, 2010). This new assessment was designed considering a more rigorous implementation a the grade level curriculum (“TEA Brochures”, 2011). Also, the STAAR provided families with results that explained individual performance ranging from Advanced Academic Performance (Level I), Satisfactory Academic Performance (Level II), or Unsatisfactory Academic Performance (Level III). At the time the STAAR was implemented, the system to evaluate schools was still under discussion by the state legislators. TEA officials proposed developing an indicator of school performance that included four indexes: student achievement, student progress, closing the performance gap, and postsecondary readiness (Texas Education Agency, s. f.).

The general problem is school effectiveness is determined by the proportion of students meeting the standard on a standards-based assessment regardless of school demographic characteristics or academic growth from one year to another. In the current mandated assessment system, student achievement is compared with the proportion of students meeting the standards of the previous cohort of students (Anderman, Anderman, Yough, & Gimbert, 2010). The results of this accountability status model indicate the level of excellence or the improvement needed. However, there is widespread professional consensus that the current status model accountability approach does not provide a scientifically, defensible basis for concluding that one school is more effective than another (Linn, 2008). In Texas, the accountability system, based on a status model approach, considers students’ sub-populations in rating schools, which does not account for growth. Due to the achievement gap, schools with higher population of students at risk may receive lower ratings. The rating of a school may not reflect that the school made progress from one year to another. Otherwise, a school that barely showed academic progress may keep a better rating (e.g., Recognized or Exemplary) since a high proportion of the student population meets the test standard. The problem is the status model used in Texas to determine school accountability does not account for students’ academic progress from one year to another or the schools’ demographic characteristics.

3. Method

3.1 Research design

Publicly accessible data, regarding Texas Assessment of Knowledge and Skills (TAKS) test scores and student demographics of public schools in Texas during years 2009-2011, were analysed. The variables included in the study have nominal characteristics (e.g., students’ socioeconomic status), ordinal characteristics (e.g., school accountability rating), and ratio characteristics (e.g., TAKS vertical scale score). The analysis of data was based on quantitative techniques including inferential statistics. The researcher used a quasi-experimental research design to test the suggested null hypotheses. Since the

grouping patterns for the independent variables included in the study are not controlled, the quasi-experimental model was the most appropriate research design because random assignment of participants was not possible (Creswell, 2012). To analyse how the independent variables, associated with each school, related statistically with the dependent variables, the investigator used a factorial analysis of variance (ANOVA). Consequently, a factorial design was used to determine whether the effects of the independent variable are generalizable across all levels or whether the effects are specific to particular levels (Gay et al., 2009). Also, a factorial experimental design allows studying the independent and simultaneous effects of two or more independent treatment variables on an outcome (Creswell, 2012). The status-based model of school accountability ratings was compared to a growth model using standards-based assessment scores and student demographics.

3.2 Participants

The study data included information from schools of three urban independent school districts from the state of Texas. These public educational organizations serve approximately 440,000 students attending 628 elementary, middle, and high schools. A representative sample of the Texas student population attending the fourth and fifth grades was included in the study by analysing data from the three largest public school districts in the state. Altogether, the districts' demographics was represented by 447,253 students in which 64% were Hispanics, 22% African American, 10% White, and 4% other ethnicities. The districts' combined populations were considered 79% economically disadvantaged and 32% were classified as limited English proficiency.

The data analysis included information from 398 public elementary schools represented by 24,065 students. The students' demographic variables included in the study were ethnicity, English language proficiency levels, and socioeconomic status. Also, the sample included schools rated at different levels of accountability based on the criteria defined by the Texas Education Agency in 2011 (Table 1).

Table 1. Elementary School Rating Based on TAKS Test Results 2011 - Sample

School Rating	Number of Schools	Number of Students
Exemplary	78	5,740
Recognized	174	10,471
Acceptable	130	7,156
Unacceptable	16	698
Total	398	24,065

3.3 Procedures for data collection

The investigation included public data available through TEA's website. The Texas Assessment Knowledge and Skills (TAKS) results from 2003 to 2011 are

available through a public database (TAKS Aggregate Data System), and maintained by members of the Student Assessment Division of TEA. The data includes the TAKS results aggregated by region, district, and campus. A report was generated to gather information about fourth and fifth grade TAKS results for reading and mathematics from 2010 and 2011. The aggregate data represented the results at a campus level containing 2,064 variables for each campus including average scores by grade level and results detailed by student demographics (i.e., ethnicity, socioeconomic status, and English language proficiency). A second database containing the schools' accountability ratings from 2010-2011 periods was generated from TEA public records. Merging both documents resulted in a new database containing information about TAKS test scores, school accountability, and students' demographics at a campus level. Since the TAKS results are represented as a vertical scale score, it is possible to compare gains from one grade to another. The vertical scale is a measure to compare students' performance to the existing academic achievement standards at each grade level and additionally to evaluate students' progression from one grade to another (Colunga, 2010). The school academic growth considered the school vertical scale score of a cohort of students taking the reading TAKS test in fourth grade (2010), and fifth grade (2011). This procedure was repeated for mathematics TAKS results. A new variable created to represent the net school academic growth by analysing how the cohort of students in fourth grade (year 2010) progressed by fifth grade (year 2011) in reading and mathematics. Since the current rating is based on TAKS scores in all content areas, it is appropriate to combine the reading and mathematics scores for a school score. The new variable called TAKS test scores growth was calculated using the following formula:

$$\text{TAKS test score school growth}_{(2010-2011)} = \frac{(\text{Reading } 5^{\text{th}} - \text{Reading } 4^{\text{th}}) + (\text{Math } 5^{\text{th}} - \text{Math } 4^{\text{th}})}{2}$$

4. Results

4.1 Differences in school academic growth considering the accountability ranking and students' ethnicity

A 4x4 factorial ANOVA was conducted to evaluate differences between the school reading and mathematics TAKS tests scores growth (dependent variable) considering the school accountability rating (year 2011) and students' ethnic groups (independent variables). The data were analysed for normality and homoscedasticity using the Kolmogorov-Smirnov test ($p = .184$), and Levene's test ($p = .285$) to comply with the underlying assumptions for ANOVA tests. The mean schools growth, as a function of the school ranking is shown in the Figure 1, where Unacceptable schools showed the higher score growth ($M = 71.8$ points of growth, $SD = 29.9$), followed by Exemplary schools ($M = 62.6$ points of growth, $SD = 27.3$). The ANOVA analysis revealed no significant differences between the mean school growth of schools with different accountability ranking, $F(3, 738) = 2.047$, $p = 0.106$. Therefore, schools with different accountability rankings (Exemplary, Recognized, Acceptable or Unacceptable,

based on 2011 TEA accountability system) showed no differences between TAKS reading and mathematics tests score gains from one year to the next (fourth and fifth grades).

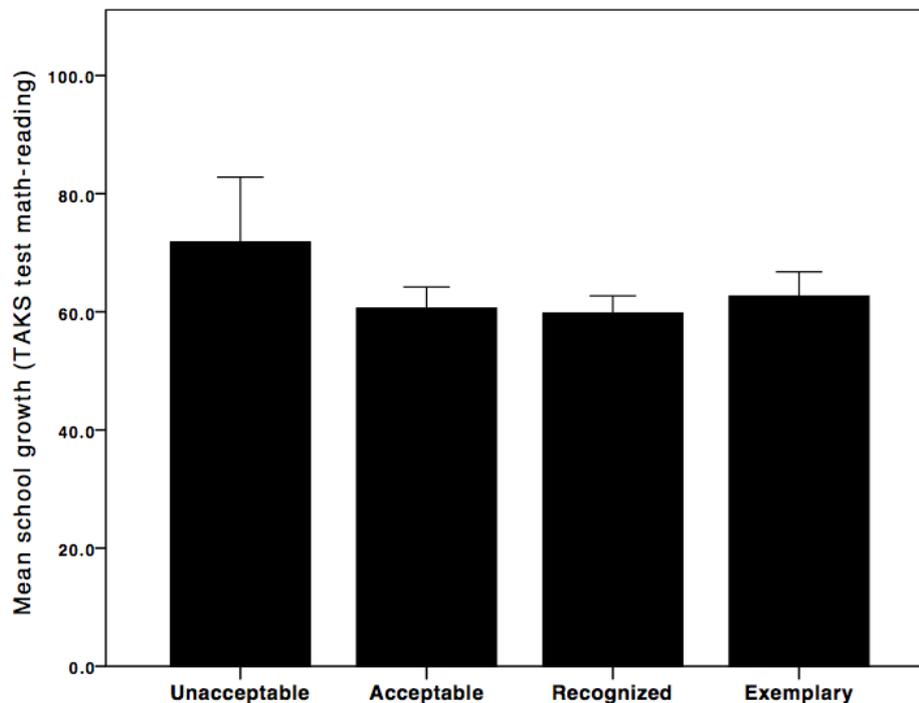


Figure 1. School growth average based on TAKS vertical score in reading and mathematics for school with different accountability rating (2011). The bars represent the mean and the error is represented by the 95% confidence interval.

The second factor included in the analysis was the differences among the students' ethnicity (Hispanic, Black, White, and Asian) in terms of the school growth. The Black students showed the highest growth compared with the other three ethnic groups included in the analysis ($M = 67.4$, $SD = 28.04$, Figure 2). The ANOVA analysis detected significant differences between the mean school growth of students from different ethnic groups, $F(3,738) = 3.446$, $p = 0.016$. Fisher's Least Significant Difference (LSD) post-hoc test was used to examine differences in the mean school growth from students from different ethnic groups. The LSD analysis indicated the mean school growth of Hispanic ($M = 59.4$) and Black ($M = 67.4$) students were significantly higher than the growth showed by Asian and White students, $p < .05$. The mean school growth between White and Asian students did not show significant differences, $p > .05$. The factorial ANOVA allows the evaluation of an interaction between the independent variables included in the analysis. This research data showed no statistically significant interaction between the school accountability ratings and students ethnic groups $F(8, 738) = 1.538$, $p = 0.140$.

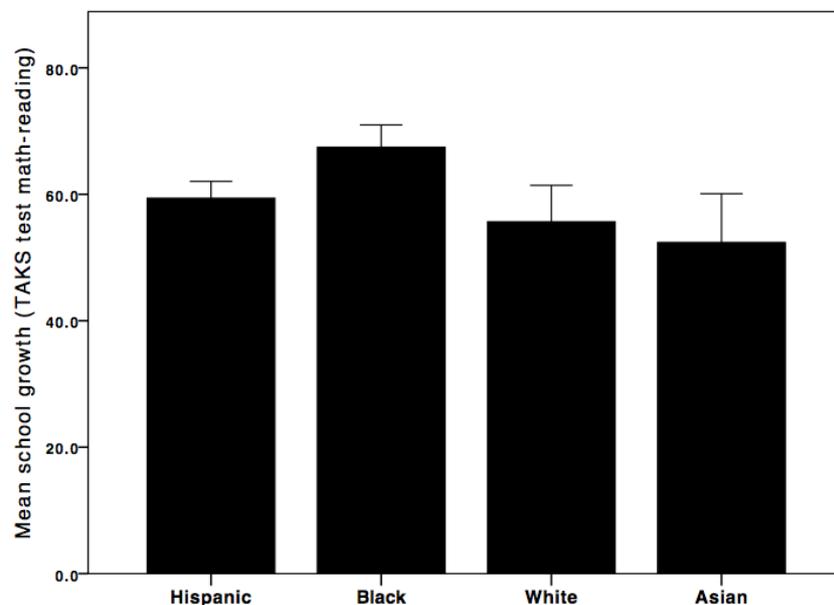


Figure 2. School growth average based on TAKS vertical score in reading and mathematics by student ethnicity. The bars represent the mean and the error is represented by the 95% confidence interval.

4.2 Differences in school academic growth considering students' socioeconomic status

A one-way ANOVA was conducted to compare the students' socioeconomic status as an independent variable and the reading and mathematics TAKS tests scores growth as dependent variable. Before the ANOVA test was applied, the data were tested for normality and homoscedasticity using the Kolmogorov-Smirnov test ($p = .200$), and Levene's test ($p = .007$). Since the three levels of the independent variable (free lunch, reduced lunch, or no free lunch) did not meet the homoscedasticity of variance requirements, the levels free lunch and reduced lunch were merged. Both variables represent a population of students with limited economic conditions compared with the population that does not qualify for this benefit. The rearranging of variables produced data that met the homoscedasticity assumption to correctly run the ANOVA analysis (Levene's test, $p = .483$). Descriptive statistics for the dependent variable, school growth average, as a function of the students' socioeconomic status are shown in Figure 3. There is a difference of 10.8 school growth points, where the group of students with free/reduce lunch presented a higher score growth ($M = 63.3$ points of growth, $SD = 29.9$) compared with the group that does not qualify for free lunch ($M = 52.5$ points of growth, $SD = 31.49$). The ANOVA analysis revealed significant differences between the mean school growth of schools of students with different socioeconomic status groups, $F(1,672) = 16.061$, $p < 0.001$. Based on these findings, it is possible to generalize that the schools' score growth averaged significantly higher in students with free/reduced lunch benefits compared with group of students that did not qualify for this assistance.



Figure 3. School growth average based on TAKS vertical score in reading and mathematics for students from different socioeconomic status. The bars represent the mean and the error is represented by the 95% confidence interval.

4.3 Differences in school academic growth considering students' English language proficiency

A one-way ANOVA was conducted to compare the students' English language proficiency as the independent variable and the reading and mathematics TAKS tests scores growth as the dependent variable. The data were tested for normality and homoscedasticity using the Kolmogorov-Smirnov test ($p = .122$), and Levene's test ($p = .073$) complying with the underlying assumptions to proceed with the ANOVA. The school growth average, as a function of the students' socioeconomic status is shown in Figure 4. There is a difference of 13.6 school growth points, where the group of non-English language learners showed a higher score growth ($M = 62.1$ points of growth, $SD = 24.9$) compared with the group that was identified as English language learners ($M = 48.5$ points of growth, $SD = 36.63$). The ANOVA analysis revealed significant differences among the students' English proficiency (English language learners and non-English language learners) in terms of the school growth, $F(1,648) = 30.076$, $p < 0.001$. Based on these findings, the group of non-English language learners presented a school growth average significantly higher compared with students from the English language learners. Since the independent variable has two levels, in this case a post-hoc test was not justified.

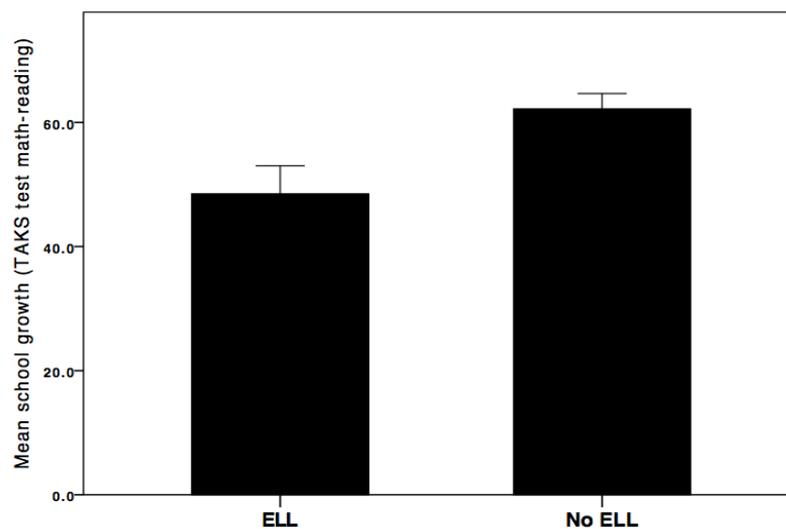


Figure 4. School growth average based on TAKS vertical score in reading and mathematics for students with different English language proficiency level. The bars represent the mean and the error is represented by the 95% confidence interval.

5. Discussion and Conclusion

An accountability system, based on a status approach, can measure the proportion of students attaining the defined passing standards considering the standardized assessment results. This was the accountability system used by The Texas Education Agency between 2003 and 2011. This accountability model did not include the students' academic growth to define the school accountability rating assigned to each school, except for the AYP goals defined by the NCLB policy. Based on this accountability system, a school with a higher proportion of students attaining the passing standard defined each year could receive a better accountability rating (i.e., Exemplary schools require 90% of students passing the end-of-the-year state assessments). Therefore just considering the percentage of students meeting the passing standards might not be the best representation of academic excellence. Consider that in 2011 the Texas Education Agency established the passing standard for the fourth and fifth grade reading and mathematics tests with 60% to 70% correct answers.

As indicated in this study, the school accountability rating may be a representation of the percentage of students meeting the passing standard, but does not necessarily reflect academic improvement from one grade to the next. No differences in the academic growth between schools with different accountability ratings were observed in this study. The causes associated with students attaining the passing test standard might not be explained only by the academic added value provided by that school to that child. Advantaged students, with families involved in their education, could have attended preschool, been exposed to diverse academic stimulus, and been exposed to varied printed material resulting in their success in school (Lareau, 2011; Neuman, 2013). Therefore, it is evident that subpopulations of students are unprepared to receive the base knowledge required to pass standardized end-of-the-year assessments. In general, it is expected that a student that passes the test one year is more likely to pass the test the next year, not requiring additional

academic support provided by the school. Thus, the student's academic success is significantly explained by the child's social background rather than the effectiveness of the school.

No statistically significant differences in the academic growth between schools with different accountability rating were observed, but these differences were present when considering the students demographic variables. In general, student populations associated with lower academic achievement presented higher academic improvement from one year to the next. This may be a response of the NCLB AYP requirements and may represent a positive tendency for educators to focus on struggling students to improve their academic performance. For example, when school and teacher resources are "zero-sum," that is, increases in gains for bubble kids coincide with decreases in gains for other students and sum to zero (Ho, 2008, p. 357). Therefore, the higher growth observed in African American, Hispanic, and low-income students (free/reduced lunch group) may represent a greater opportunity for improvement, since these populations have historically performed lower on standardized assessments. If this is the case, pressure on school administrators and teachers to meet state and federal accountability goals becomes a concern. Schools with limited resources and with a high proportion of students at-risk may face the greatest pressure to focus on bubble kids in a zero-sum manner, simply because there are no surpluses available for broader allocation. This scenario leaves behind students with high academic achievement that come to school prepared; little effort is required on behalf of the school to help them meet the test expectations.

English language learner students showed lower academic improvement compared with non-English language learners. These results may be attributed to the language acquisition process. This investigation considered data at the elementary school level, where it is expected that ELL students are in the process of developing the skills required to perform using their second language at an academic level (Abedi & Levine, 2013; Lakin & Young, 2013). Recent research results support these assertions where differences on performance on standardized assessment were reported considering English Language Learners children from school districts located at the Texas border with Mexico compared with non-border school districts (Tang, Wang & Min 2019; Vela, Jones, Mundy & Isaacson, 2017). The Hispanic population is growing faster than other minority student population in Texas. Appropriate support to ELL students may be a better implementation of bilingual and ESL programs that allow those students to achieve at the same academic level as any other student attending public schools. Efforts have been made to close the achievement gap of minority students' populations (i.e. Hispanic) by providing more option to access higher education, but the results show that Hispanic college enrolment significantly increased as a result these initiatives, but graduation did not (Tajalli & Ortiz, 2017).

The results obtained in the study support advancing on developing accountability models that include measuring academic growth. Defining school

achievement based on the progress of each student might be a better estimator of educational effectiveness. This accountability approach allows teachers to be recognized for students learning and not just for students' performance at a fixed point in time (Buzick & Laitusis, 2010). Also, measuring the academic growth of each student prevents making unfair comparisons between schools performance with completely different demographic compositions. Using such systems can identify schools where teachers foster student learning but students may not achieve minimum proficiency levels (Ready, 2013). A growth model requires teachers to work with all the students, because every student has potential to progress in relation with their previous scores. Since, every student's score counts, a growth model might prevent the common practice of directing the educational resources on bubble kids, who are most likely to succeed (Booher-Jennings, 2005). This tracking model is inclusive, making each student accountable for the school performance. Furthermore, a growth model can be used to improve sensitivity and fairness of a summative assessment, by adjusting post-test decisions about mastery based on pre-existing levels of performance (Zimmerman & Dibenedetto, 2008).

The Texas Education Agency have considered the experiences of TASK assessment based on a status accountability system. Hence, since 2012 implemented a new standardized students assessment system called the State of Texas Assessments of Academic Readiness (STAAR) program. The current accountability model includes a closing the achievement gap component weighting the STAAR test academic achievement by 30% and academic growth by 50% (The Texas Education Agency, 2019). This method to determine the school ranking provides a fairer way to evaluate schools progress on standardized assessment considering the student's population characteristics.

As a concluding remark regarding designing an accountability system to determine schools' effectiveness considering the end-of-the-year results of standards-based assessments, is a complex mission. There is an unresolved question at the heart of discussions of accountability whether targets for student achievement should be expressed in terms of proficiency or progress. Proponents of an achievement growth models argue that considering just end of the year score holds schools accountable for factors beyond their control, whereas the latter allows a focus on value added, which is ostensibly fairer to schools and teachers (Ballou & Springer, 2017). Furthermore, the large diversity of students attending public schools in the United States, must be a considered. A systemic approach must prevail when defining an accountability system used to judge the quality of education delivered at an elementary school. Also, educators must provide straightforward information to parents regarding the academic progress of individual students in terms of standardized measurements. The way that school performance data is presented to the community may affect the public support and parental involvement in the local school (Jacobsen, Snyder, & Saultz, 2014). Relying on defining the school effectiveness based on average scores may mislead the learning community. Raising the bar of the United States public education system, characterized by a diverse society and uncertain context, is a process not achievable using simple,

cause-effect relationships and standardized solutions. This research was an analysis of how an accountability system may be useful to evaluate the school effectiveness based on standards-based assessment results. However, the use of only one measurement to judge a school's efficacy regarding the quality education provided to students may be a reductionist and narrow approach.

References

- Abedi, J., & Levine, H. G. (2013). Fairness in assessment of English learners. *Leadership, 42*(3), 26-38.
- Anderman, E. M., Anderman, L. H., Yough, M. S., & Gimbert, B. G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist, 45*(2), 123-137. doi:10.1080/00461521003703045
- Ballard, K., & Bates, A. (2008). Making a connection between student achievement, teacher accountability, and quality classroom instruction. *Qualitative Report, 13*(4), 560-580.
- Ballou, D., & Springer, M. G. (2017). Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains. *Education Finance and Policy, 12*(1), 77-106. doi:10.1162/EDFP
- Barrier-Ferreira, J. (2008). Producing commodities or educating children? Nurturing the personal growth of students in the face of standardized testing. *Clearing House, 81*(3), 138-140. doi:10.3200/tchs.81.3.138-140
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268. doi:10.3102/00028312042002231
- Braun, H. (2009). Discussion: With choices come consequences. *Educational Measurement: Issues and Practice, 28*(4), 52-55. doi:10.1111/j.1745-3992.2009.00162.x
- Brevetti, M. (2014). Re-evaluating narrow accountability in American schools: The need for collaborative effort in improving teaching performances. *The Delta Kappa Gamma Bulletin, 81*(1), 32-35.
- Briggs, D. C., & Weeks, J. P. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice, 28*(4), 3-14. doi:10.1111/j.1745-3992.2009.00158.x
- Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendations for research. *Educational Researcher, 39*(7), 537-544. doi:10.3102/0013189x10383560
- Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007). Children left behind in AYP and non-AYP schools: Using student progress and the distribution of student gains to validate AYP. *Educational Measurement: Issues and Practice, 26*(3), 21-32. doi:10.1111/j.1745-3992.2007.00098.x
- Colunga, M. (2010, June 9). TAKS Vertical Scale. Retrieved January 22, 2014, from http://www.tea.state.tx.us/index3.aspx?id=3818&menu_id3=793
- Creswell, J. W. (2012). *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. (4th ed.). Boston, MA: Pearson.
- Finkel, E. (2010). Gearing up for the new assessment: The next generation of standardized testing will focus on critical thinking skills. *District Administration, 46*(7), 78-82.
- Forte, E. (2010). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist, 45*(2), 76-88. doi:10.1080/00461521003704738

- Gay, L. R., Mills, G. E., & Airasian, P. W. (2008). *Educational research: Competencies for analysis and applications* (9th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Goldschmidt, P., Choi, K., Martinez, F., & Novak, J. (2010). Using growth models to monitor school performance: Comparing the effect of the metric and the assessment. *School Effectiveness and School Improvement*, 21(3), 337-357. doi:10.1080/09243453.2010.496597
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75-110. doi:10.3102/0162373708317689
- Hess, F. M. (2006). Accountability without angst?: Public opinion and no child left behind. *Harvard Educational Review*, 76(4), 587-610. doi:10.17763/haer.76.4.p245w24572592m62
- Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, 37(6), 351-360. doi:10.3102/0013189x08323842
- Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). *Final report on the evaluation of the growth model pilot project*. Jessup, MD: U.S. Department of Education.
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory into Practice*, 42(1), 30-41. doi:10.1353/tip.2003.0009
- Jacobsen, R., Snyder, J., & Saultz, A. (2014). Informing or shaping public opinion? The influence of school accountability data format on public perceptions of school quality. *American Journal of Education*, 121(1), 1-27. doi:10.1086/678136
- Klein, A., & Hoff, D. (2007). Impact is slight for early states using "growth." *Education Week*, 27(16), 24-25.
- Kress, S., Zechmann, S., & Schmitt, M. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Law School Journal on Legislation*, 48(1), 185-234.
- Krieg, J. M. (2011). Which students are left behind? The racial impacts of the no child left behind act. *Economics of Education Review*, 30(4), 654-664. doi:10.1016/j.econedurev.2011.02.004
- Ladd, H. F., & Lauen, D. L. (2010). Status versus Growth : The Distributional Effects of School Accountability Policies, 29(3), 426-450. <https://doi.org/10.1002/pam>
- Lakin, J. M., & Young, J. W. (2013). Evaluating growth for ELL students: Implications for accountability policies. *Educational Measurement: Issues & Practice*, 32(3), 11-26. doi:10.1111/emip.12012
- Lauen, D. L., & Gaddis, S. M. (2016). Accountability Pressure, Academic Standards, and Educational Triage. *Educational Evaluation and Policy Analysis*, 38(1), 127-147.
- Lareau, A. (2011). *Unequal childhoods: Class, race, and family life*. Berkeley, CA: University of California Press.
- Linn, R. L. (2006). Validity of inferences from test-based educational accountability systems. *Journal of Personnel Evaluation in Education*, 19(1), 5-15. doi:10.1007/s11092-007-9027-6
- Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40(6), 699-711. doi:10.1080/00220270802105729
- Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy education in Australia? *Literacy Learning: The Middle Years*, 19(2), 9-19.
- Madaus, G., & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education*, 190(1/2), 21-30. doi:10.1177/0022057410190001-205

- Mathis, W. J., & Trujillo, T. M. (2016). *Lessons from NCLB for the Every Student Succeeds Act*. Boulder, CO.
- Marzano, R. J., Waters, T. (2009). *District leadership that works: Striking the right balance*. Alexandria, VA: ASCD.
- Meyer, R. H. (2000). Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. *Issue Brief*, 3(3). Madison, WI: University of Wisconsin, National Institute for Science Education.
- Mittleman, J., & Jennings, J. L. (2018). Accountability, Achievement, and Inequality in American Public Schools: A Review of the Literature. <https://doi.org/10.1007/978-3-319-76694-2>
- Moloney, K. (2006). Teaching to the test. *International Journal of Learning*, 13(6), 19-25.
- Neuman, S. B. (2013). The American dream: Slipping away? *Educational Leadership*, 70(8), 18-22.
- Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-130. doi:10.3102/10769986029001121
- Ready, D. D. (2013). Associations between student achievement and student learning implications for value-added school accountability models. *Educational Policy*, 27(1), 92-120. doi:10.1177/0895904811429289
- Schlechty, P. C. (2011). The threat of accountablism. *Educational Leadership*, 69(1), 80.
- Tajalli, H., & Ortiz, M. (2017). An examination of Hispanic college enrolment and graduation: Has the Texas Closing the Gaps plan been successful? *Journal of Latinos and Education*, 17(4), 330-343. doi:10.1080/15348431.2017.1348301
- Tang, S., Wang, Z., & Min, Y. (2019). Texas-Mexico Border vs. Non-Border School Districts' Growth Trajectory of High-Stakes Reading Performance: A Multi-Level Approach. *Education Science*, 38(9), 1-15. doi:10.3390/educsci9010038
- Texas Education Agency. (2011, October 20). *Brochures*. Retrieved from http://www.tea.state.tx.us/index2.aspx?id=2147485434&menu_id=692&menu_id2=796&cid=2147483661
- Texas Education Agency. (2013). *Accountability rating system*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/2013/faq.html>
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107-122. doi:10.1080/00461521003703060
- Wills, J. S., & Sandholtz, J. H. (2009). Constrained professionalism: Dilemmas of teaching in the face of test-based accountability. *Teachers College Record*, 111(4), 1065-1114.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives*, 18, 19. doi:10.14507/epaa.v18n19.2010
- Vela, A., Jones, D., Mundy, M., & Isaacson, C. (2017). Determining the effectiveness of bilingual programs on third grade state exam scores. *Research in Higher Education Journal*, 33, 1-15.
- Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools*, 45(3), 206-216. doi:10.1002/pits.20291
- Zvoch, K., & Stevens, J. J. (2008). Measuring and evaluating school performance: An investigation of status and growth-based achievement indicators. *Evaluation Review*, 32(6), 569-595. doi:10.1177/0193841x08320398