

The Texas Projection Measure: Ignoring Complex Ecologies in a Changing World

Warren Roane

Humble Independent School District, Humble, Texas

**May 3, 2010**

**Presented at the Annual Meeting of the American Educational Research Association  
in Denver, Colorado, Friday, April 30 – Tuesday, May 4, 2010**

The Texas Projection Measure: Ignoring Complex Ecologies in a Changing World  
Warren Roane  
Humble Independent School District, Humble, Texas

### **Abstract**

The Texas Projection Measure (TPM) has grown out of the state's need to meet the requirements of No Child Left Behind (NCLB). An examination of the state's method of predicting 8<sup>th</sup> grade mathematics scores reveals that several factors have been ignored in the process of developing the model, including assumptions in its underlying statistical analysis as well as ease of use for its stakeholders. Although the TPM was based on value-added models, it has deviated from that foundation in significant ways. Alternatives are given to the TPM as currently used by the Texas Education Agency (TEA).

### **Background**

#### *The Texas Projection Measure (TPM)*

Texas' need to comply with No Child Left Behind (NCLB) has led to the Texas Projection Measure (TPM). In order for students to progress from certain levels of P-16 education to the next, they must pass the reading and mathematics components of the Texas Assessment of Knowledge and Skills (TAKS). The vertical scale of the TAKS ranges from 300 to 1000, with 700 as the passing score for 8<sup>th</sup> grade, which represents about thirty items correct out of forty-two (Texas Education Agency, 2009b). Retention in middle school is a cause of concern to all stakeholders, and Texas has created local Grade Placement Committees (GPC) to ameliorate some of these anxieties, where parents, teachers and administrators decide using a variety of measures if students have made sufficient progress to be promoted (Texas Education Agency, 2007a). In particular, 8<sup>th</sup> grade is considered one of the two Student Success Initiative (SSI) grades, in that accelerated instruction must take place with those students who do not meet the passing standard on the reading and mathematics TAKS test. At the same time, there are pressures to improve on the rating of Adequate Yearly Progress (AYP). The TPM is a statistical model that addresses both grade promotion and AYP concerns by estimating the number of students in 7<sup>th</sup> grade who do not meet the standard for the current year, but given time, will pass in 8<sup>th</sup> grade. These additional students are counted as passing the 7<sup>th</sup> grade standard for both AYP and the state accountability system.

In 2007, Texas conducted a pilot study to determine how to predict annual improvement in test scores. It considered two models, *Reaching the Standard*, which relied on vertical scales, and the *SAS EVAAS*, which was regression-based. Due to its inclusion of multiple variables, the state chose to continue to investigate the use of the second model (Texas Education Agency, 2007b). However, the decision was made to not include demographic variables in the final model, the TPM.

#### *Value added models*

Can value-added models disregard student and campus characteristics? Sanders and Wright (2008) noted that many value-added models include student and classroom level variables. To avoid this, Sanders advocated at least five tests of

The author would like to thank the reviewers of the AERA session, Division L - Educational Policy and Politics/Section 3: Curriculum, Testing, and Instructional Practice, for their comments, as well as his colleagues in the Humble Independent School District. The viewpoints herein are solely those of the author and do not necessarily reflect those of the school district or of its Board of Trustees.

achievement over five years so that the twenty-five values would eliminate the need for demographic data (Lockwood & McCaffrey, 2007). The Tennessee Value-Added Assessment System (TVAAS) substituted test performance for background variables (Ballou, Sanders & Wright, 2004). These background factors are already reflected in the pre-test score. However, Ballou et al. acknowledged that a study in Florida found student income and race to be statistically significant and teacher and school effects were sensitive to them. They objected to including SES into the model, because if disadvantaged students are systematically assigned to less effective schools, it would mask “genuine differences in school and teacher quality” (p. 39). Even in the TVAAS system, while considering a variation in the modeling, it was found that the percent of students on free or reduced lunches had a “substantively significant impact on the standardized teacher effect” for math (p.56). Correlations between scores in the same subject across grades were about .8, while same grade scores for different content areas were .6 to .7. Ballou et al. maintained that these high correlations can “serve as a substitute” for other student data (p.60). The authors admitted that for simple models (rather than TVAAS), it makes a “considerable difference” whether the model includes SES and demographics (p.60). Based on the literature cited by the Texas Education Agency, the Texas Projection Measure is a simplified form of the TVAAS. In addition, the TVAAS vertically linked the tests over time on the same development scale, equated across the years. Presumably, the TVAAS 7<sup>th</sup> grade test would have material from 4<sup>th</sup> to 8<sup>th</sup> grades so that growth could be measured. The 7<sup>th</sup> grade TAKS test, by comparison, is focused on 7<sup>th</sup> grade objectives of the Texas Essential Knowledge and Skills (TEKS), although its results are reported on a vertical scale from grade 3 to grade 8.

One complication cited by Rubin, Stuart, and Zanutto (2004) is the Stable Unit Treatment Assumption, where all individuals in the school are assumed to receive the same treatment. Another is missing data, and it is “likely to be students whose performance is worse than average” (p. 109). Ballou et al. (2004) also cited the problem of “unclaimed students”. Background covariates are important to consider when looking at control and treatment groups. If the groups are “very different,” the results will be unreliable. (Rubin, Stuart & Zanutto, 2004, p. 109). Furthermore, estimates are sensitive to the choice of the statistical model. In particular, they were hesitant to look at estimates of school effects, but rather the value of incentives for performance. Raudenbush (2004) also discouraged the use of value added models for accountability purposes. Instead, Rubin et al. and Raudenbush said value-added models should be descriptive of schools, not causal. McCaffrey et al. (2004) noted that their work was misinterpreted as advocating using value-added models for school accountability. Instead, they argued for developing databases with a “broad collection of measures of student and contextual characteristics” (p. 140).

### *Current TPM*

By 2009, Texas had decided to modify the model into a much simpler design, called the Texas Projection Measure, or TPM (Texas Education Agency, 2009a). It was considered “easy to implement,” having a short “turnaround time between test completion and projection calculation,” impacting “instructional planning as early as possible,” and although “difficult for stakeholders to understand,” the basic premise was “straightforward” (Texas Education Agency, 2009a, p. 8). The TPM predicts 8<sup>th</sup> grade

math scores by using linear regression with variables of individual 7<sup>th</sup> grade reading and math scores and the campus average score for 7<sup>th</sup> grade math. In order for the student to be considered as passing for the current year, she must earn a total of 670 points outright or 700 with the TPM, with the prediction that though failing this year (earning less than 670 on the vertical scale), she is on course to pass the next year. The TPM is loosely based on a state graduation prediction model used in Maryland. That study emphasized ordinary least squares regression using the 8<sup>th</sup> grade math score, reduced lunch status, attendance, English Language Learner (ELL) status, and GPA to predict 11<sup>th</sup> grade reading score on the Maryland high school examination. The models explained from fifty-eight to seventy percent of the variance in the reading score (Lissitz & Fan, 2006). The Maryland study looked at using both OLS regression and HLM. The TPM diverges from this theoretical basis, using only regression and avoiding examining differences in student achievement based on ethnicity, ELL status, or SES. In exchange for student-level variables, the TPM relies only on campus average scores.

### **Theoretical Framework**

Using the CRESST conceptual model, the theoretical framework behind the TPM will be analyzed to answer the question of why Texas chose the TPM model. The CRESST model includes validity, fairness, credibility, educational improvement, substantive research and development, utility, knowledge, and public engagement, as well as teaching and learning (Gipps, 1999). An integral part of the model, acceptance of an assessment, is determined both by cultural values and political necessities. The subset of the CRESST model sections of validity, fairness and credibility can be re-cast as a question of efficiency, equity and effectiveness (Bishop, 2006; Hanushek, 1988). This triad will be considered to see how well the TPM speaks to this concern. Of these three, perhaps equity is the most complex, because there are several issues to be addressed to ensure equity in assessment, including the impact of assessment on low-income students' lives, the impact of language on the transmission of knowledge, and the lack of theoretical undergirding of items during test development in general (Garcia & Pearson, 1994). Equity is one of the fundamental pillars of modern mathematics instruction and assessment (National Council of Teachers of Mathematics, 1989). In addition, mathematics achievement interacts with language and ethnicity in complex ways (Fillmore & Valetz, 1986; Secada, 1992; Tate & Rousseau, 2007). These social issues regarding math assessment impact mathematical knowledge and educational policy, and assessments are often in danger of oversimplification with its undesired results, particularly when schools are considered as behaving as though they were individuals (Rochex, 2006).

### **Method of Inquiry**

This paper will employ mixed methods to explore the appropriateness of both the statistical model underlying the TPM as well as its theoretical underpinnings, following the eight step process model of Johnson and Onwuegbuzie (2004). The stages in the inquiry include research questions, purpose, methodology, data collection, data analysis, data interpretation, legitimization, and conclusions.

## **Research questions**

The following are the research questions to be explored:

1. Why did Texas adopt the TPM?
2. What is the theoretical basis of the TPM ?
3. Does the TPM meet the test of efficiency, equity and effectiveness?
4. Is there evidence to support that the TPM predicts 8<sup>th</sup> grade math TAKS scores?
5. Is there an alternate statistical model to predict scores that arises from this foundation?

## **Purpose and Scope**

The purpose of the paper is to serve as a critique of the TPM as a statistical model used to predict student outcomes on the TAKS. Multiple forces have served to shape both the TAKS test and the TPM. Some of these forces will be investigated, but the scope of this paper will be limited to the TPM as it relates to 8<sup>th</sup> grade math TAKS scores.

## **Methodology**

### *Qualitative*

Statistical reports and other supporting documents from the Texas Education Agency will be used to examine the TPM's theoretical foundation. Assumptions will be explored, and advantages and disadvantages will be weighted using the CRESST model as a framework. The first three research questions will be considered using this methodology.

### *Quantitative*

Both ordinary least squares (OLS) multiple regression and hierarchical linear modeling (HLM) will be used to investigate the TPM using school district TAKS data. Student score changes from 7<sup>th</sup> to 8<sup>th</sup> grade will be examined and compared to the TPM's prediction. The fourth and fifth research questions will be considered using this methodology. The TAKS test was developed by Pearson, and in a study by Pearson Educational Measurement, five growth models were considered. Of these, OLS and HLM were the most promising (Tong & O'Malley, 2006).

## **Data Sources**

TEA provides data for region, district, and campus level data files through its website (Texas Education Agency, 2009d). In addition, the 2008 and 2009 TAKS data files at the student level were used which were provided to a large suburban district by Pearson Education, the company which holds the state contract for the TAKS tests. For the State, 318,810 students were in 7<sup>th</sup> grade in 2008, and 317,831 students were in 8<sup>th</sup> grade in 2009 who took the mathematics TAKS test (Texas Education Agency, 2008b; Texas Education Agency, 2009c). Of these 7<sup>th</sup> graders in Texas, the demographics breakdown was: 49.4% female, 13.8% African American, 45.9 % Hispanic, 36.2 % White, 53.2% economically disadvantaged (free or reduced lunch participation), 8.0% Limited English Proficient (LEP), 5.9% in Special Education, 11.5% Gifted and Talented (GT), and 55.7% were enrolled at a Title I campus (Texas Education Agency, 2008b). For the district data, there were 2368 students in the sample who had scores for both 7<sup>th</sup> grade (2008) and 8<sup>th</sup> grade (2009). These students were: 51.8% female, 19.9% African

American, 23.2 % Hispanic, 53.6 % White, 26% economically disadvantaged, 4.2% LEP, 7.6% in Special Education, 9.7% GT, and 24.7% were enrolled at a Title I campus.

## **Data analysis**

### **Procedure**

#### *1. Qualitative*

According to Johnson and Onwuegbuzie (2004), there are seven stages of data analysis: data reduction, data display, data transformation, data correlation, data consolidation, data comparison, and data integration. The data reduction phase produced the 8<sup>th</sup> grade math TAKS score as the main TPM model to be studied. It was decided to narrow the study to this model due to several factors including content and grade level to be explored, the fact that it greatly impacts grade placement and retention, as well as the results impact a campus change (from middle school to high school). This step involved examining state, district, and local data with t-tests and correlations to see what variables might be good predictors of 8<sup>th</sup> grade math TAKS. Graphs of data, particularly histograms of the variables of interest (scores by gender, scores by ethnicity, 7<sup>th</sup> grade reading and math scores, meeting standards on 7<sup>th</sup> grade reading and math) were examined for outliers, normal distributions, etc. TPM graphical reports from the state were also analyzed to see if they met the CRESST criteria. Amrein-Beardsley, who raised several methodological concerns about value-added models, maintained that student background variables seem to affect measures of growth of student achievement (2008). Since research indicates that there may be an “achievement gap” in mathematics, some variables were recoded to indicate that tendency. In general, the review of literature indicates that students who are female, Hispanic or African American, economically disadvantaged, or performed poorly on prior mathematics tests are less likely to score well on mathematics tests (Baker, Goesling, & Lentendre; 2002, Gandara, 2010). If a student was a member of that group, they were recoded as “1” to indicate membership in the disadvantaged group to predict not meeting the passing standard on 8<sup>th</sup> grade math TAKS. Thus, there was directionality of all of the variables of interest. The data transformation stage allowed variables of interest to be converted into numeric codes to be represented in statistical models (female, Hispanic, African American, and “not passing 7<sup>th</sup> grade math” were recoded in a similar way as “not passing 8<sup>th</sup> grade math”).

#### *2. Quantitative*

In order to determine which variables might predict performance on 8<sup>th</sup> grade mathematics TAKS, district data from the Texas Education Agency (TEA) was used to get a large picture of state-wide trends, and then individual data from one district was analyzed. TEA’s model to predict an individual’s 8<sup>th</sup> grade scores (the Texas Projection Measure, or TPM) was considered and then alternative models were considered for suitability. Not only linear regression, but also nested analysis was used since “individuals drawn from the same classroom or school tend to share certain characteristics...observations based on these individuals are not fully dependent” (Osborne, 2000, p. 2).

#### a. Variables used

Several demographic variables were taken into consideration. In addition to coding as above for gender, ethnicity, and prior testing, economically disadvantaged status was based on free or reduced lunch participation. Enrollment at a Title I campus was also included. The vertical scale for reading and mathematics ranged from 300 to 1000 for both 7<sup>th</sup> grade and 8<sup>th</sup> grades. Another concern mentioned in the literature review is missing data. Of 2502 students who had values for 8<sup>th</sup> grade math, only 2368 had scores for 7<sup>th</sup> grade math. Missing data is one reason TPM is not generated. Only students who had scores for both 7<sup>th</sup> and 8<sup>th</sup> grade were included in the present study. The issue of missing data will be addressed later in the limitations section of this study.

#### b. Data correlation

Johnson and Onwuegbuzie (2004) stated that the data correlation step of data analysis entails the quantitative data being correlated with the “qualitized data,” or the qualitative data being correlated with the “quantitized data”. This is followed by the process of data consolidation of new or consolidated variables, or data sets. Data from both quantitative and qualitative sources are compared and then integrated to answer the research questions. Statistical tests were conducted at the state, district, and campus levels to investigate variables of interest. For example, at each level, ANOVA and t-tests were conducted to see if there were mean differences in 7<sup>th</sup> grade math scores based on demographic characteristics. Using campus data from Texas, the relationship of demographics to campus mean scores in mathematics and reading was explored using correlations in SPSS. These values are available from the Texas Education Agency (2008c).

For the 1125 school districts in Texas with middle school grades, the 8<sup>th</sup> grade TAKS math score was significantly correlated (Pearson correlation,  $\alpha < .01$ ) to 7<sup>th</sup> grade reading scores (.72), 7<sup>th</sup> grade math scores (.78), as well as the proportion of district students who are Hispanic (-.28), African American (-.17), or economically disadvantaged (-.50). These statistically significant variables were examined using partial correlations. While controlling for 7<sup>th</sup> grade math scores, predictors of 8<sup>th</sup> grade mathematics scores at  $\alpha < .01$  were: 7<sup>th</sup> grade reading scores (.23), as well as the proportion of district students who are Hispanic (-.07), African American (-.09), or economically disadvantaged (-.18). It was decided that ethnicity, 7<sup>th</sup> grade reading and math scores, and economically disadvantaged status should be included in the more focused district study.

There is a much stronger relationship between reading and math scores for both 7<sup>th</sup> and 8<sup>th</sup> grade. Besides the question of the correlations being statistically significant, the overlap might indicate that the mathematics test does not measure strictly mathematics, but how to read technical writing, how to test well, or some other construct. In fact, same grade tests are (slightly) more correlated to each other than tests of the same content area. This is in contrast to Ballou, et al. in which correlations between scores in the same subject across grades were about .8, while same grade scores for different content areas were .6 to .7 (Ballou, Sanders, & Wright, 2004). Ballou et al. maintained that these high correlations can “serve as a substitute” for other student data (p.60). For the TAKS test, it was found that correlations same grade scores were higher across content areas than same subject scores across years, the opposite of Ballou et al. (see

Table 1). To turn this on its head, do high correlations of tests by grade level indicate more influence of the same student factors than subject or content knowledge differences? Should test scores substitute for student demographic variables?

Table 1: Correlations of TAKS test scores

	7 <sup>th</sup> reading	7 <sup>th</sup> math	8 <sup>th</sup> math
7 <sup>th</sup> math	.823**	1	.778**
8 <sup>th</sup> reading	.782**	.689**	.828**
8 <sup>th</sup> math	.721**	.778**	1

\*\* Correlation is significant at the 0.01 level (2-tailed).

### District/Individual level results

Using a suburban district in Texas of 33,000 students, the author used 2008 and 2009 district data of 7<sup>th</sup> and 8<sup>th</sup> grade scores collected in the TEA dataset to explore the TPM model’s ability to predict individual future scores (i.e., 2008 data was used to estimate 2009 results). The TPM was based on 7<sup>th</sup> grade reading scores, 7<sup>th</sup> grade mathematics scores, and 7<sup>th</sup> grade campus mean mathematics score from the previous year. Additional factors based on the district analysis were included in ordinary linear regression and HLM to see if the TPM model could be improved—ethnicity, SES (free or reduced lunch status), gender, and Title I campus status. For the district data, there were 2368 students at 7 campuses in the sample for both 7<sup>th</sup> grade (2008) and 8<sup>th</sup> grade (2009). For this data set, the following demographics were found: 51.8% female, 19.9% African American, 23.2 % Hispanic, 53.6 % White, 26% economically disadvantaged, 4.2% LEP, 7.6% Special Education, 9.7% GT, and 24.7% enrolled at a Title I campus.

Using ANOVA for 8<sup>th</sup> grade math scores, means were statistically different based on gender ( $F=8.7, \alpha <.003$ ), ethnicity ( $F=193.4, \alpha <.001$ ), and economic disadvantaged status ( $F=126.2, \alpha <.001$ ). The mean score for females was 10.5 points lower than males, while the mean score for White and Asian students was 47.5 points higher than their Hispanic and African American counterparts. The mean 8<sup>th</sup> grade score for students not on free or reduced lunch was 46 points higher for than participants in the program.

Using paired t-tests, there was a significant difference in an individual’s score from 7<sup>th</sup> to 8<sup>th</sup> grade in math with 36.9 increase on the vertical scale ( $t = -27.34, \alpha <.001$ ). Overall, the mean TAKS math score was 734 for 7<sup>th</sup> grade, and 773 for 8<sup>th</sup> grade, with standard deviation of 88 and 91 respectively for the district sample. The state expectation is for students to “grow” 30 points on a vertical scale from 7<sup>th</sup> to 8<sup>th</sup> grade, with “met standard” rising from 670 at 7<sup>th</sup> grade to 700 for 8<sup>th</sup> grade. For state-wide data, the mean TAKS math score was 725 for 7<sup>th</sup> grade and 757 for 8<sup>th</sup> grade.

Based on the previous analysis cited above with district 7<sup>th</sup> and 8<sup>th</sup> grade scores, variables included in the model were gender, ethnicity, and economic disadvantage (free or reduced lunch). These variables have also been associated with mathematics ability testing (Diversity in Mathematics Education Center for Learning and Teaching, 2007; Khisty, 2007; National Council of Teachers of Mathematics, 2001; Secada, 1992;



Thomas, 1997). These and other demographic variables were explored using partial correlations in SPSS. The variables that were correlated (at  $\alpha < .05$ ) to individual 8<sup>th</sup> grade TAKS mathematics scores were: gender (.042), ethnicity (-.213), economic disadvantage (-.192), 7<sup>th</sup> grade reading TAKS (.373), 7<sup>th</sup> grade math TAKS (.532), and Title I campus status (-.265). Partial correlations yielded similar results. When gender was controlled for, individual 8<sup>th</sup> grade math scores were statistically significantly correlated (at  $\alpha < .001$ ) to 7<sup>th</sup> grade math scores (.777). Nearly identical results happened when controlling for ethnicity, Title I status, and economic disadvantage. Controlling for 7<sup>th</sup> grade reading scores, (at  $\alpha < .001$ ) partial correlations to individual 8<sup>th</sup> grade mathematics scores were: gender (-.092), ethnicity (-.174), economic disadvantage (-.084), 7<sup>th</sup> grade math (.654), and Title I campus status (-.152).

However, the correlation dropped from .78 to .65 between 7<sup>th</sup> and 8<sup>th</sup> grade math while holding 7<sup>th</sup> grade reading scores constant. Controlling for both 7<sup>th</sup> reading and math scores, only two variables were correlated to 8<sup>th</sup> grade math scores at  $\alpha < .05$ : ethnicity (-.102) and Title I campus status (-.085), though both of these were significant at  $\alpha < .001$ . Since demographic variables did not alter 7<sup>th</sup> to 8<sup>th</sup> grade math score partial correlations, and inclusion of control for 7<sup>th</sup> grade reading scores shrank all demographic variable correlations to 8<sup>th</sup> grade math, it gave credence to Ballou et al. that other test scores could substitute for demographic variables.

### c. Additional Variables of interest

Statewide, 76% of students met standard on the 7<sup>th</sup> grade 2008 math test (Texas Education Agency, 2008b). The next year, 79% of students passed the 8<sup>th</sup> grade math test (Texas Education Agency, 2009c). However, 79% of 7<sup>th</sup> graders met standard for the 2009 math test (Texas Education Agency, 2009m). This tendency is true at the campus level as well. Of the 8385 middle school campuses, 76.7% had increases in mathematics scores in 7<sup>th</sup> grade from 2008 to 2009 and also those 7<sup>th</sup> graders in 2008 increased their mean score in 2009 as 8<sup>th</sup> graders. For 336 campuses (4% of the total) had only an increase on student scores by cohort, 7<sup>th</sup> (2008) to 8<sup>th</sup> grade (2009). 279 campuses (3.3%) had an increase in 7<sup>th</sup> grade scores from 2008 to 2009. Thus, almost 82% of campuses had students improve scores as they moved from 7<sup>th</sup> to 8<sup>th</sup> grade. In addition, 80% of the campuses improved 7<sup>th</sup> grade scores from 2008 to 2009. Only 16% of campuses had middle school math scores decline from 2007 to 2008 by cohort and by grade level. Since successive cohorts are getting higher 7<sup>th</sup> grade scores, it follows that they will have higher vertical scores in 8<sup>th</sup> grade, and the TPM will be of limited benefit. For 2009, the TPM increased the passing rate for 7<sup>th</sup> grade math from 79% to 83%.

The TPM assumes that all students have an equal chance at passing or failing. However, the relationship across 7<sup>th</sup> and 8<sup>th</sup> grade math scores is probably not linear, because only those near boundary of “700” have a chance of altering pass/not pass status. Unless drastic changes occur, those on low side have only a slight chance of passing while those on high side have almost no chance at failing. For 2009, 21% fail the 7<sup>th</sup> grade test, while only 12% fail the 8<sup>th</sup> grade test. “Not met mathematics standard in 7<sup>th</sup> grade” should also be a variable to consider. Using district data in Table 2, those who passed 7<sup>th</sup> have only a 5% chance of failing 8<sup>th</sup>, while those who failed 7<sup>th</sup> have a 50% chance of failing 8<sup>th</sup>. If passing 7<sup>th</sup> grade math TAKS was the only consideration, using conditional probability, there is 12% probability of error; 4% false positive (predicting to

pass in 8<sup>th</sup> when a student failed 7<sup>th</sup>) and 8% false negative (predicting to fail 8<sup>th</sup> when a student passed 7<sup>th</sup>). This analysis was conducted for 7<sup>th</sup> graders in April 2008 projected to April 2009 as 8<sup>th</sup> graders. Thus, using only 7<sup>th</sup> grade scores, this simplified model is 88% accurate.

Table 2: District TAKS Math Scores in Two Grades

	2009 Pass 8 <sup>th</sup>	2009 Fail 8 <sup>th</sup>
2008 Pass 7 <sup>th</sup>	80%	4%
2008 Fail 7 <sup>th</sup>	8%	8%

The TPM, in contrast, is 86% accurate for projecting 7<sup>th</sup> grade math scores to 8<sup>th</sup> grade math scores (Texas Education Agency, 2009l, p.16). This is comparable to the EVAAS model’s accuracy which was also considered by TEA.

d. District results compared to State results

The current formula to calculate 8<sup>th</sup> grade math scores, based on TPM and linear regression is:

$$139.24 + .1392 * (7^{\text{th}} \text{ read}) + .6851 * (7^{\text{th}} \text{ math}) + .0265 (\text{camp mean math})$$

(Texas Education Agency, 2009h, p.11).

For the district dataset used, linear regression produced a slightly different model. 61.7 percent of the variance was explained, and 7<sup>th</sup> grade reading TAKS, 7<sup>th</sup> grade math TAKS, and 7<sup>th</sup> grade math TAKS campus mean were all significant at  $\alpha < .001$ . The resulting formula for finding 8<sup>th</sup> grade vertical math score was:

$$46.27 + .084 * (7^{\text{th}} \text{ read}) + .677 * (7^{\text{th}} \text{ math}) + .215 (\text{camp mean math}).$$

Correlations were conducted between the 8<sup>th</sup> grade vertical score predicted assigned by TEA and the vertical score predicted by the formula above. For 8<sup>th</sup> grade vertical scores over 700, the correlation between these two methods is .897. However, overall, the correlation was .753. Using the district dataset, the TEA formula results in 139 students who passed, but weren’t predicted to pass (false negative 5.8%), and 112 students with a false positive (4.7%), or 10.6% total error. Using the adjusted formula, total error was slightly less (10.1%) with 4.3% false negative, and 5.8% false positive.

**Data interpretation**

In the following discussion, we return to the first three research questions:

1. Why did Texas adopt the TPM?
2. What is the theoretical basis of the TPM ?
3. Does the TPM meet the test of efficiency, equity and effectiveness?

*1. Qualitative*

The TPM predicts 8<sup>th</sup> grade math scores by using linear regression with variables of individual 7<sup>th</sup> grade reading and math scores and the campus average score for 7<sup>th</sup>

grade math. In order for the student to be considered as passing for the current year, she must earn a total of 700 points on the TPM, with the prediction that, though failing this year, she is on course to pass the next year. In exchange for student-level variables, the TPM relies only on campus average scores. It thus skirts the equity test. In terms of efficiency, the TPM serves both to move students from 8<sup>th</sup> to 9<sup>th</sup> grade, as well as to increase the state's rate of AYP, but its model requires high campus average scores, high reading scores, or a combination of the two to assist math scores. It effectively (with a 86% rate of accuracy) predicts future eight grade math scores using seventh grade data (Texas Education Agency, 2008d). The district dataset used above indicated the TPM was about 90% accurate.

#### *a. Assumptions*

TPM encompasses many grade levels and subject matters. For example, 5<sup>th</sup> grade science scores at elementary schools are used to predict 8<sup>th</sup> grade science scores at middle schools for the same students. Campus mean scores at the elementary schools, as well as 5<sup>th</sup> grade reading and math scores are part of the regression model. One can see that several assumptions must hold for the model to “project” a score three years into the future, from 5<sup>th</sup> grade to 8<sup>th</sup> grade: a school change makes no difference, feeder patterns for middle schools are not significant, teachers at a given school are constant, etc.

To return to the narrow focus of this paper, 8<sup>th</sup> grade math scores, there are several assumptions in the linear regression model. First, the distribution of scores is normal. Second, a large amount of variance is explained by the model. Third, school and individual variance should be included at the same level in the model. Fourth, calculations of coefficients based on a previous cohort of 7<sup>th</sup> graders continue to be valid. Fifth, the dichotomy of TPM (Yes or No) is appropriate for linear regression. Finally, the assumption is that outliers and missing data of 7<sup>th</sup> grade mathematics scores have no effect on the model.

#### *b. Advantages and Disadvantages*

Using the CRESST model described by Gipps (1999), the relative advantages and disadvantages of the Texas Projection Measure will be considered. The components under consideration are validity, fairness, credibility, educational improvement, substantive research and development, utility, knowledge, and public engagement, as well as teaching and learning.

#### *Validity*

Gipps echoed an expanded view of validity, including purposes and consequences of assessment. Validity can encompass basic statistics, but should also include politics and the decisions underlying the growth model selected. The basic projections may seem to be adequate, but may in fact obscure problems, or oversimplify them.

The Texas Education Agency conducted research to look at type I and type II errors, where TPM would predict a student to pass who would not pass the next year, or where TPM would indicate they would not, when in fact, they did (Texas Education Agency, 2009I). This reflects attention to one of the basic tenets of research (Wainer, 2010). In addition, the TPM is based on longitudinal data using a vertical scale, so that initial (prior year) and final scores (current scores) can be compared, although the 7<sup>th</sup>

grade and 8<sup>th</sup> grade math tests do not assess exactly the same material, nor is the similar material measured at the same depth each year. “Grade level assessments are not sensitive measures of growth”; multigrade assessments are needed to yield valid interpretations of student growth (Amrein-Beardsley, 2008, p.66). In addition to test scores, teacher evaluations or other qualitative data should be used to see if student growth can be attributed to teachers (Amrein-Beardsley, 2008).

Politicians often see accountability as the “least complex way to ensure quality learning” but laws and policies operate at a “high level of abstraction until socially processed” (Torres, 2004, p. 250). According to Torres, only after putting AYP and TPM into practice do issues of social justice, etc. come to light. AYP relies too heavily on politically-framed principles, rather than caring about organizations and individuals. He urged that students be judged according to “multiple sets of criteria”, rather than a single measurement (2004, p. 258). However, politicians, the public, and parents often prefer a single score as a clear indication of a student’s success.

Texas House Bill 3 indicates that a new data portal will be created so that district, campus, teacher, class, and students scores can be viewed (Texas Education Agency, 2010b). This seems to lend itself to analysis through nesting, rather than ordinary linear regression. Wainer (2010) cited several problems with value-added models. NCLB has fueled interest in such models, based on score trajectories, where attempts are made to estimate the contributions of schools to student learning. When NCLB measures AYP, it compares improvement in 7<sup>th</sup> grade from 2008 to 2009, rather than 7<sup>th</sup> graders in 2008 to the 8<sup>th</sup> graders in 2009. The “effectiveness of the school is confounded with intrinsic differences between the cohorts” (Weiner, p. 15). An additional difficulty of the TPM model is that Texas has recently mentioned that the TAKS test, on which the longitudinal data is based, will be changed in the future (Texas Education Agency, 2010a). Thus TPM will make projections from the TAKS exam to the new STAAR (State of Texas Assessment of Academic Readiness) test.

## Fairness

Justice and fairness are also part of the CRESST model, and form part of the substantive research and development focus of assessment. Is the result transparent? Is the test fair to all? Will all groups enjoy an equal footing entering into the assessment? Is fair the same as equal?

The TPM is transparent, with its coefficients that are published early so that anyone can use these numbers to calculate the TPM. In addition, there is an on-line calculator available (at <http://forwardfocus.pearson.com/tpmcalculator/> ). Furthermore, the literature review in the beginning of this article indicates that the TPM has reason to ignore individual variables, such as SES and gender. The partial correlations (above) in this study could lead one to that decision. However, the TPM lacks some of the features of value-added models that would allow it to ignore demographic variables. Using the initial value (7<sup>th</sup> grade math) and the final value (8<sup>th</sup> grade math) only to measure growth, TPM ignores the fact that certain groups have a lower mean initial score, and thus are required to grow more than their peers to meet standard. For example, thirty points of growth are required for all groups to show one year of growth in mathematics from grade 7 to grade 8, from 670 to 700 on the vertical scale. However, the mean scores for White students are 760 (grade 7) to 788 (grade 8): a

growth of twenty-eight. For African American students, the results are 694 (grade 7) to 728 (grade 8): a growth of thirty-four. Focused only on growth, African American students out-perform White students in middle school mathematics. However, despite this growth, African American students are still two years (sixty points) behind their White peers.

Equity and fairness are related to missing data (Wainer, 2010). Because prior year reading and mathematics scores enter into the calculation, a TPM is not generated if students have missing data, or change test versions. Some groups are more likely than others to suffer from these data restrictions on TPM. Special education students and English Language Learners are much more likely to change versions from one year to the next. Hispanic students are likely to have data entry errors, where the state system may not match records across years; e.g., DeLeon and De Leon, or Peña and Pena. The common use of Social Security numbers as a state identifier might also pose a problem. Also, attendance may be a problem for disadvantaged students. Texas maintains that “97.4% of students testing in mathematics had sufficient data in 2008 for making a projection” (Texas Education Agency, 2009k, p.16). For the district dataset, 52 students did not receive a TPM in 2009 due to version changes (2% of total). There were 125 students who did not have a TPM for all reasons (4.8% of total). However, of these 125 students, 110 were in Special Education. Using the district dataset, of 2587 eighth grade students in 2009, 289 students had missing 7<sup>th</sup> grade test scores. Students with missing data are slightly more likely to be male (57.8% compared to 48% of test takers), more likely to be African American (28.8% to 18.8%), more likely to be economically disadvantaged (41.2% to 24.2%), and more likely to receive special education services (30.1% to 5.3%).

Several of the statistical growth models take missing data into account, with varied degrees of success. HLM seems to perform best even when large portions of the dataset are missing (Tong & O’Malley, 2006). However, a simplistic model ignores problems of selection bias, with serious impact on estimates of the influence of schools on academic progress (Sanders & Wright, 2008).

### Credibility

Gipps (1999) stated that credibility interacts with trustworthiness and authenticity. Do the results seem dependable? The TPM is produced by Pearson, a leading company in education. It is also touted as very accurate, but there are exceptions (particularly as noted in the section above). “The percent of accurate projections typically exceeded 80% for students overall, and for all groups except the LEP [English Language Learners] and SPED [Special Education]” (Texas Education Agency, 2009k, p. 16). These groups have more missing or mismatched data and thus fewer TPMs generated. As noted in the data section of this paper, using 7<sup>th</sup> grade mathematics scores only to predict 8<sup>th</sup> grade math scores would be at least 85%, and would not have the disadvantage of mismatched or missing 7<sup>th</sup> grade reading data. The TPM must exceed this minimum standard of credibility.

There was a public relations backlash as TPM was introduced (Why Did DISD’s Ratings Go Sky High?, 2009). TEA was seen as manipulating scores, giving too much credit to campuses and districts in the rating system. The TPM helped higher achieving campuses (termed “Exemplary” or “Recognized”) more than underperforming campuses

("Academically Acceptable" or "Academically Unacceptable"). For state accountability, some 2,560 campuses used TPM to achieve a higher rating. but of these 358 used TPM to achieve Academically Acceptable, 1,088 used TPM to achieve Recognized and 1,114 used TPM to achieve Exemplary. Similarly, 331 of 1250 districts used TPM to achieve a higher rating, where 79 used it to achieve Academically Acceptable, 179 used it to achieve Recognized, and 73 used it to achieve Exemplary (Housson and Rinehart, 2009).

#### Educational improvement

As noted above, TPM allowed many campuses and districts to improve their state accountability ratings. Students who are "projected to pass" are counted as if they had met the standard (Texas Education Agency, 2009e). This phenomenon also occurred at the federal level for the ratings of Adequate Yearly Progress (AYP). The Texas Projection Measure (TPM) was used for 2009 AYP evaluations, and allowed an additional 10% (126) of districts to meet AYP that would have otherwise missed AYP; and 6% more (528) of campuses (Housson and Regalado, 2009). The State Commission maintained that the "use of the Texas Projection Measure will strengthen Texas' federal and state accountability systems and, in particular, will enhance the ability to close achievement gaps based on race, ethnicity, socio-economic and special program status." (Texas Education Agency, 2009f). However, TPM did not reflect educational change, but only a change in the way data was reported. The scores of individual students did not change; no improvement was made on the personal level. Only the campus and district ratings benefited.

#### Substantive research and development

The background section of this paper illustrates the large body of literature that exists for value-added and growth models. Texas has utilized research from several sources, including North Carolina (Texas Education Agency, 2009j), Maryland (Lissitz & Fan, 2006), Ohio, and Tennessee (Ballou, Sanders, & Wright 2004). However, it has chosen to simplify the model, moving away from the benefits of the more complex statistical models. It dropped demographic variables from the TPM without including multiple test scores per individual.

#### Utility

The Texas Projection Measure is relatively simple to calculate and understand. It is expressed as either Yes or No, and for 8<sup>th</sup> grade mathematics, only three variables are used in the calculation from linear regression. Its simplicity has some disadvantages; for example, a simple Yes or No is misleading, as it could be viewed as a guarantee that students will pass the next grade's high stakes examination. Plotting the distributions of 7<sup>th</sup> and 8<sup>th</sup> grade scores indicate that they are not normally distributed. In addition, the binary nature of the TPM result seems to indicate that a linear model is not the most appropriate one.

#### Knowledge

Gipps (1999) showed the connection of the consequences of assessments to knowledge. The assessment should enlighten, rather than obfuscate, and make sense for the learner and teacher. There are two situations where the TPM does not clarify, but

oversimplify, or in the worst case, confuse. In the case of modeling, TEA currently collects approximately 925 variables on students, of which about 350 would have values for most students in 8<sup>th</sup> grade. However, only three variables are used in the linear regression to predict 8<sup>th</sup> grade math scores: 7<sup>th</sup> grade reading, 7<sup>th</sup> grade mathematics, and campus mean mathematics score for grade 7. TEA has tried to keep the linear regression models uniform, so that only, at most, four variables are used in them (Texas Education Agency, 2009h). In the case of promotion, the situation is worse. Promotion to the next grade level is based on meeting standard, rather than TPM. Thus, TPM could predict a student to pass, but that student may still be held back a grade due to the Student Success Initiative (SSI), where a “student may advance to the next grade level only by passing these tests or by unanimous decision of his or her grade placement committee that the student is likely to perform at grade level after additional instruction” (Texas Education Agency, 2009g).

#### Public engagement

Besides claims that TEA has used TPM to inflate district and campus rating, TPM has confused educators and parents alike. Though the Yes or No result of the TPM is easy to understand, the linear regression on which it is based is not, particularly explaining negative coefficients to stakeholders. TEA has produced documents explaining the TPM to parents, in both English and Spanish, for grades 3 to 10 (Texas Education Agency, 2009i). However, because of the possibility the TPM may not be generated due to diverse reasons, parents may actually see a different score report than what is shown in the official explanation. For example, in 7<sup>th</sup> grade alone, there are many different reports possible—at least 12 possible charts configurations (due to student absence, LEP status, mismatch of Special education TAKS test versions, etc.) This problem is compounded when parents have children enrolled at different grade levels. In addition, there are three subject areas for 7<sup>th</sup> grade--reading, mathematics and writing. The reading and mathematics have a vertical scale (where 670 is passing), while the writing is on a horizontal scale (where 2100 is needed to meet standard). The use of two different scales occurs at 4th, 5th, 7th, and 8th grades.

#### Teaching and learning

The TPM can be used as a tool for intervention as it created a new group for educators to focus on: those students who were not projected to pass the next year, but did meet standard on current year TAKS. Prior to TPM, these students were invisible to interventions because they passed the previous year’s exam. Students who take the TAKS test and receive a TPM can be grouped into four categories for more focused interventions (Texas Turnaround Center, 2009).

#### Efficiency, equity and effectiveness

TPM is based on Ordinary Least Square regression. Although OLS assumes linearity in growth, it produces relatively stable school level results. It also has the disadvantage of regression towards the mean, but it is easier to use than other growth models (Tong & O’Malley, 2006). TPM also uses individual and campus scores. Due to the logic of nesting students within schools, a model that utilizes nesting effects, like Hierarchical Linear Modeling, might be more equitable in the sense that it better reflects

the nature of the data. HLM is superior to OLS with missing data, and can estimate school effects. However, it is “quite complex” and “the results can be hard to interpret” (Tong & O’Malley, 2006, p.20).

### Summary

Texas adopted the TPM for several reasons. It permitted additional campuses and districts to meet state and federal accountability requirements. The TPM simplified accountability into a simple Yes or No result. It allowed teachers to focus on a new student group that was heretofore undetected, the borderline students who had passed but would probably not pass the following year. Texas based the TPM on the value added and growth model studies in research literature and those adopted by other states, particularly the TVAAS. However, it greatly simplified the model, and in so doing, eliminated some of the safeguards and benefits the more complex models afforded, endangering credibility and accuracy. TPM does not meet the test of efficiency, equity and effectiveness. It has a slight advantage over using only prior year scores, but with the disadvantage of slighting certain groups of students who are less likely to receive a TPM. It is efficient in its binary form, but the way in which it is reported and explained is cumbersome.

### 2. *Quantitative*

The Texas Education Agency’s TPM regression model for 8<sup>th</sup> grade mathematics was explored (using 7<sup>th</sup> grade reading, 7<sup>th</sup> grade mathematics, and 7<sup>th</sup> grade campus mean mathematics score to predict 8<sup>th</sup> grade individual mathematics score). Linear regression was conducted with the addition of additional factors of economic disadvantage, gender, ethnicity, and attendance at a Title I school as these had been correlated to mathematics scores (see discussion above). These additional variables were coded following Raudenbush & Bryk, 2002; and Lee & Bryk, 1989. Economic disadvantage was assigned 1 if the student received free or reduced lunches, or -1 if she did not. Gender was coded 1 for female and -1 for male. Ethnicity was divided into two groups: 1 for African American or Hispanic students and -1 for White or Asian students. If a student attended a Title I school, the code was 1; else it was -1.

The same data was analyzed using HLM, both as a level 1 model, and as a level 2 model (student and school effects). Below is the level 2 model employed in the study. The Title I status is considered to affect the scores in general, while the school math mean affects the math scores in particular.

#### Level-1 Model

$$Y = B_0 + B_1*(\text{individual Grade 7 Reading}) + B_2*(\text{individual Grade 7 Math}) + R$$

#### Level-2 Model

$$B_0 = G_{00} + G_{01}*(\text{Title I status}) + U_0$$

$$B_1 = G_{10}$$

$$B_2 = G_{20} + G_{21}*(\text{School 7}^{\text{th}} \text{ Math mean})$$



Because the end use of the model was to predict passing the 8<sup>th</sup> grade math exam, analysis was conducted in both SPSS and HLM using binomial models as well. Thus, for part of the analysis, the independent variable was the 8<sup>th</sup> grade score, for others, it was obtaining a passing score on the 8<sup>th</sup> grade test (a 700 score was needed to pass). Once a prediction model was generated, the same data set was used to predict the accuracy of the model to the students' actual score on the 8<sup>th</sup> grade mathematics test. Tables 3 to 7 present the results of the analysis.

Table 3: Accuracy and Percent of Variance Explained by SPSS Models

Predicted Dependent Variable	SPSS Model	Levels	Variables at $\alpha < .05$	Percent Variance Explained	Accuracy
8th math score	1. Linear Regression	One	7 <sup>th</sup> math	61%	85%
8th math score	2. Linear Regression	One	7 <sup>th</sup> math, 7th reading Campus 7th mean	62%	90%
8th math score	3. Linear Regression	One	7 <sup>th</sup> math, 7th reading Campus 7th mean, ethnicity	62%	91%
Pass 8th grade math	4. Binary Logistic	One	7 <sup>th</sup> grade math	20%	87%
Pass 8th grade math	5. Binary Logistic	One	7 <sup>th</sup> grade math, 7 <sup>th</sup> reading	23%	89%
Pass 8th grade math	6. Binary Logistic	One	7 <sup>th</sup> grade math, 7th reading Title I status	23%	90%

Table 4: SPSS prediction of 8<sup>th</sup> math score, unstandardized coefficients (full model)

Independent variable	Coefficient	SE	t ratio
Intercept	88.0*	35.672	2.47
7 <sup>th</sup> grade reading score	.081***	.014	5.64
7 <sup>th</sup> grade math score	.663***	.017	39.10
Campus 7 <sup>th</sup> math mean score	.177***	.048	3.73
Gender	-1.536	1.106	-1.39
Ethnicity	-3.535**	1.298	-2.72

\*p<.05 \*\*p<.01 \*\*\*p<.001

Table 5: Accuracy and Percent of Variance Explained by HLM Models

Predicted Dependent Variable	HLM Model	Levels	Variables at $\alpha < .05$	Percent Variance Explained	Accuracy
8th math score	1. Linear Regression	One	7 <sup>th</sup> grade math, 7th reading Title I status, Ethnicity	63%	92%
8th math score	2. Linear Regression	Two	One: 7 <sup>th</sup> math, 7th reading Two: Title I Status, Campus mean	65%	87%

Table 6: HLM prediction of 8<sup>th</sup> math score, unstandardized coefficients (Level 1 model)

Independent variable	Coefficient	SE	t ratio
Intercept	215.1***	14.2	15.15
7 <sup>th</sup> grade math score	.681***	.0001	13.42
7 <sup>th</sup> grade reading score	.075***	.02	5.19
Title I status	-5.92	3.48	-1.70
Ethnicity	-3.28*	1.44	-2.28

\*p<.05 \*\*p<.01 \*\*\*p<.001

Table 7: HLM prediction of 8<sup>th</sup> math score, unstandardized coefficients (Level 2 model)

Independent variable	Level	Coefficient	SE	t ratio
Intercept	One	775.2***	6.16	125.5
7 <sup>th</sup> grade math score	One	.68***	.022	31.5
7 <sup>th</sup> grade reading score	One	.076***	.021	3.7
Title I status	Two	-28.13**	6.30	-4.47
Campus math mean	Two	.001*	.0005	2.16

\*p<.05 \*\*p<.01 \*\*\*p<.001

## Discussion

For the analysis conducted using SPSS and HLM, compared to the models that were focused on 8<sup>th</sup> grade pass/fail status, the models using 8<sup>th</sup> grade mathematics scores account for a larger percentage of the variance, although the accuracy in predicting accurately 8<sup>th</sup> grade passing rates were similar for binary and linear models. Since the

linear regression models accounted for more variance, and are more easily used to calculate projections, these models were preferred. In addition, collapsing a wide range of scores into two categories (pass or fail) obfuscated interactions of demographic variables with math scores while accounting for little of the variance in the scores.

HLM indicated that only 14% of the variance was between schools. The second level HLM model was less accurate, but accounted for about the same amount of variance as the level one model. The variance improvement over the null model was at most 65%, and the accuracy of the HLM model to predict passing rates at 8<sup>th</sup> grade was about the same as the linear regression model. As expected, in the SPSS models, the 7<sup>th</sup> grade math score was the most important variable used to generate 8<sup>th</sup> grade math scores. However, in the HLM models, the intercept had a higher t-ratio. As noted above, the data at either extreme tended to disproportionately affect the results, so models that were more sensitive to outliers fared worse.

The TEA TPM model, model 2 in Table 3 (with variables including 7<sup>th</sup> grade reading score, 7<sup>th</sup> grade math score, and campus 7<sup>th</sup> grade mean math score) did account for a large portion of variance, and was reasonably effective, with  $r^2$  of more than .62 accounting for variation in future math scores. It accurately predicted 8<sup>th</sup> grade passing rates ninety percent of the time. An alternative (model 1) was tested resting on the same theoretical foundation as the TPM. Using linear regression with only the 7<sup>th</sup> grade math scores as input,  $r^2$  accounted for more than .61 of the 8<sup>th</sup> grade math score variation, and had an eighty-five percent accuracy rate. This model avoids the unfairness of including the campus where a student attends in the model, particularly when the HLM analysis indicates that only fourteen percent of the variance in 8<sup>th</sup> grade math scores is due to campus effects. Only math scores are needed to predict future math scores. Furthermore, the one variable model is simpler, and almost as effective and as accurate.

## Summary

Revisiting the research question about evidence to support that the TPM predicts 8<sup>th</sup> grade math TAKS scores showed that the TEA TPM model is as good as alternate statistical models considered in the study. It is about as accurate in its prediction and is relatively high in its explanation of variance. Its calculation is relatively straight forward as it is based on linear regression. However, there is an alternate statistical model (model 1) that uses just the 7<sup>th</sup> grade math score to predict the 8<sup>th</sup> grade math score, and is almost identical in accuracy and efficacy to the more elaborate TEA TPM model.

## Legitimization

According to Johnson and Onwuegbuzie, the stage of legitimization includes reflections on the trustworthiness of “both the qualitative and quantitative data and subsequent interpretations” (2004, 25). Furthermore, “[i]t is important to note that the legitimization process might include additional data collection, data analysis, and/or data interpretation until as many rival explanations as possible have been reduced or eliminated” (Johnson & Onwuegbuzie, 2004, p. 25). In this stage are the limitations of the study, the scholarly significance of this paper, and a review of the data integration phase.

### *Limitation of the study*

This study examined one portion of a Texas value-added model, the Texas Projection Measure, used for Adequate Yearly Progress for No Child Left Behind. A district sample of convenience of 2008 7<sup>th</sup> grade math scores projected to the 2009 8<sup>th</sup> grade was compared to the TPM model. The same conclusions may not be true for all content areas (mathematics, reading, writing, science, and social studies) and grades (3-10). To further investigate impact on certain student groups (ELL, Special Education, etc.) one should consider using propensity score methods (Graham, 2010). However, a preliminary study of 8<sup>th</sup> grade science indicates that the TPM model should be revisited for other content areas and grade levels. The TPM uses individual 5<sup>th</sup> grade reading, mathematics, science scores with campus mean science scores to project individual 8<sup>th</sup> grade science scores. The author's preliminary study indicates that campus, gender, and ethnicity are statistically significant ( $\alpha < .01$ ) predictors of meeting 8<sup>th</sup> grade science standards. In contrast to the linear regression model used in TPM, these variables are significant as nested data (gender within campus, ethnicity within campus).

Besides the limitations of linear regression for studies of this type noted above, it should be acknowledged that there are criticisms of multilevel modeling as well. Though HLM and other models take into account group covariates in individual-based estimates, group dynamic effects are assumed not to exist. At the very least, "thoughtful consideration of subjects' interactions must precede inferences based on estimates" of effects (Gitelman, 2005, p. 409). Because of interactions of school and classroom variables with individuals, "clustering students within groups generates design effects that considerably reduce the precision of impact estimates," so statistical power must be considered as well (Schochet, 2008, p. 62).

Other statistical models might be more appropriate given the final dichotomy represented by TPM (Yes or No). For example, Bahr (2010) used two-level hierarchical multinomial logistic regression to model variation in the probability in attainment of college skills and degrees. Tekwe et al (2004) cited several statistical models that could be used for value-added analysis, including Hierarchical Linear Models, Layered Mixed Effects Models, and Simple Fixed Effects Models.

The data indicate that a much larger sample will be needed to study interventions or treatment effects in middle school math. For the district dataset, the math mean score was 734 for 7<sup>th</sup> grade and 773 for 8<sup>th</sup> grade, with standard deviation of 88 and 91, respectively. The state expectation is for students to "grow" 30 points from 7<sup>th</sup> to 8<sup>th</sup> grade, from 670 to 700 on the vertical scale. This represents an impact of about .33 standard deviations for both 7<sup>th</sup> and 8<sup>th</sup> grade math scores. Following Schochet (2008), if a treatment were to study middle school math interventions that purported to give one year of growth, one would need at least 15 schools to study students within schools. To look at school and classroom-level clustering, one would need 51 schools. However, if the treatment added a more achievable ten week growth, the number of schools increase considerably to 133 and 534, respectively.

Treatment of missing data is also a concern. As noted above, five percent of 8<sup>th</sup> graders do not have a score for 7<sup>th</sup> grade math. This would mean that TPM is not generated for these students. A preliminary study conducted by the author indicates that students with missing 7<sup>th</sup> grade scores are dissimilar to those who have scores, on the measures of ethnicity, SES, gender, and special education status. Students with missing

data are slightly more likely to be male, more likely to be African American, more likely to be economically disadvantaged, and more likely to receive special education services. Sanders and Wright (2008) stated that there should be at least three prior scores for each student (in order to omit demographic data, among other reasons).

### *Scholarly significance*

The TPM impacts student promotion (SSI) and mandated programs of intervention. Although TPM is related to TAKS, the replacement of TAKS with the new STAAR system (State of Texas Assessments of Academic Readiness or STAAR.) will still require a growth measure (Texas Education Agency, 2010a). Texas has announced a new data warehouse that can or will be used to measure student, teacher, campus, and district scores on the TAKS system (Texas Education Agency, 2010b). This seems to imply a nested statistical model. More work should be done on the TPM before it is accepted as an appropriate value-added model, without meeting several of the important components of such a model. The criteria include: either demographic variables or sufficient numbers of exams that will eliminate their need, tests that measure multiple years of learning to allow for demonstration of growth, assessments not used for accountability, and consideration for missing data.

### *Data Integration of qualitative and quantitative analysis*

The simple alternate (model 1) to the TEA TPM is equitable, as it is based solely on one test score, and uses only math scores to predict future math scores. As shown in Table 3, it is as effective and efficient as the TEA TPM model. Nevertheless, the TPM and alternatives based on just a few test scores without demographic variables are on unsteady ground. The value-added model is often in a dilemma, between simplicity and statistical sophistication, leading to accuracy (Amrein-Beardsley, 2008). However, using a simple paired-means or simple regression is a “devil’s bargain” where simplicity is traded for reliability (Sanders & Wright, 2008, p. 7). There are additional unintended consequences that result from Texas’ choice of statistical model.

First, TPM depends heavily on campus variables. Holding reading scores at 740 (passing) and math at 638 (twenty-three questions correct where thirty correct is passing), only high achieving campuses have students that are counted towards TPM. A student at a campus with a ninety-five percent pass rate would receive a TPM score of 698, while a campus with a pass rate of ninety-eight percent would allow students to count as passing with a TPM of 700.

Second, TPM is overly dependent on reading scores for math prediction. Students who achieved a score of 675 in math (30 questions correct), but were low in reading (546), were estimated to not pass in math the following year, even though this would require just duplicating the same feat (30 questions correct). Even at campuses with a ninety-eight percent pass rate, students would only receive a TPM of 699.

Third, the TPM is overly sensitive to time or test-retest effects. For all re-takers on the 8<sup>th</sup> grade math test in one district (n = 223), the mean score was 653 on the first administration, but a mean of 657 on the second administration. This would translate to two more points on the TPM, just from one administration to another, in a matter of six weeks. For students who made 638 in 7<sup>th</sup> grade or better, but failed the first administration of the 8<sup>th</sup> grade test (n = 171), the test-retest results for 8<sup>th</sup> grade produced

a change in the mean score at the .05 significance level. The first time the group had a mean score of 630, but 625 six weeks later. This change in score is not atypical of mathematical examinations; in fact, it would even be unlikely to have identical scores on parallel tests on two consecutive days (Hattie, Jaeger, & Bond, 1999). As the TAKS math score is subject to fluctuations due to date of test administration, the TPM model loses its effectiveness.

Fourth, several assumptions underlying TPM's linear regression model listed above have been shown to be untrue. The distribution of math scores is not normal, which would need to be addressed before using linear regression. Successive cohorts have higher math scores, yet the TPM is calculated based on the previous cohort. Outliers affect the accuracy of the projections. Finally, missing data is ignored by the TPM despite its harm to certain student groups.

However, besides its predictive value for individuals, the TPM is used to create higher 7<sup>th</sup> grade scores, at the campus and district levels. Students who pass the 7<sup>th</sup> grade math test outright, and those who failed 7<sup>th</sup> but are predicted to pass 8<sup>th</sup> grade, all count as meeting standard for 7<sup>th</sup> grade math. TPM thus gives about a 5% increase to campus math scores. This translates to slightly higher campus and district accountability ratings. But, as noted above, it is the campuses and districts which are "acceptable" and become "recognized" that disproportionately enjoy the benefit of TPM. It is no wonder that critics have seen TPM as an accountability Matthew effect, where good districts become great, but struggling districts receive no benefit.

As an alternative to the current TPM, one could use high growth on TAKS (three years of progress in two years, or 97 points on the vertical scale from grades 5 to 7) to count as meeting standard on grade 7, since that student would be likely to pass at grade 8. For the dataset, about twenty percent of students achieve this kind of growth. A campus could earn credit for students who either met standard on 7<sup>th</sup> grade math TAKS, or for those who completed three years of growth in two years. Students that had this level of growth passed 8<sup>th</sup> grade TAKS 92.5% of the time, while those who passed 7<sup>th</sup> grade TAKS math (regardless of growth) met standard for 8<sup>th</sup> grade 87% of the time. By combining these two, the pass plus projected to pass rate, resulted in an accuracy of 87% of predicting 8<sup>th</sup> grade TAKS scores, and boosted campus ratings by 1.9%, compared to just counting students who actually passed 7<sup>th</sup> grade math TAKS. The Texas Education Agency's TPM increased ratings of the dataset (over the number of those passing 7<sup>th</sup> grade math) by 6.7% with an accuracy rate of 90% (for 8<sup>th</sup> grade projections). With this growth measure also added, the enhanced TPM would boost ratings of the dataset by an additional 1.3% with an accuracy rate of 89.5%. Compared to this simple high growth measure, the increase given by the TPM (in either the simple or enhanced case) seems excessive, especially in light of how it has tended to improve the ratings of already "recognized" campuses and districts.

If the current TPM system is kept, it has the advantage of being easy to understand, as it is based on Yes or No. It is transparent because it relies on a linear model with the coefficients that are published early so that anyone can use these numbers to calculate the TPM. Regardless of the statistical model adopted, the inclusion of the on-line calculator (now available) should make the calculation accessible to all.

## Conclusion

This paper considered the Texas Projection Measure as employed by the Texas Education Agency using the CRESST model with its elements of validity, fairness, credibility, educational improvement, substantive research and development, utility, knowledge, public engagement, as well as teaching and learning. The stages in the inquiry included research questions, purpose, methodology, data collection, data analysis, data interpretation, and legitimization (Johnson & Onwuegbuzie, 2004) to consider the following research questions:

Why did Texas adopt the TPM?

What is the theoretical basis of the TPM ?

Does the TPM meet the test of efficiency, equity and effectiveness?

Is there evidence to support that the TPM predicts 8<sup>th</sup> grade math TAKS scores?

Is there an alternate statistical model to predict scores that arises from this foundation?

This paper underscores the impact of how political forces shaped assessment policy and calls for reconsideration by the state of its decision. Texas has created the TPM based only on linear regression of a few variables and its need to increase grade promotion and achievement as measured by AYP. The state has ignored the complex ecology of middle school mathematics achievement by eliminating demographic information without incorporating the safeguards of value-added models, yet maintains that it is meeting the needs of a changing world. The TPM could return to its roots and use multiple measures across many years as a true value-added model. A more thoughtful process should be used by first creating a theoretical framework, such as the one promoted by CRESST, and then examining the data to create the statistical model. In addition, there are several questions to explore and take into consideration during the trial phase of this framework. Hattie, Jaeger, and Bond (1999) mentioned several assessment issues to address including conceptual models of measurement, test and item development, test administration, test use, and test evaluation. Finally, if the complex ecology must be ignored, an alternate, simpler model could be used to predict 8<sup>th</sup> grade TAKS math scores—7<sup>th</sup> grade TAKS math scores. If a growth measure is needed to increase campus ratings, it can easily be incorporated into this alternative model, and three years of growth serves as a reasonable measure. The simpler model meets the test of efficiency, equity and effectiveness. The current system in place—the Texas Education Agency’s Texas Projection Measure—fails this three-pronged test.

## References

- Amrein-Beardsley, A. (2008). Methodological concerns about the education Value-Added Assessment System. *Educational Researcher*, Vol. 37 (2), 65-75.
- Bahr, P. (2010). Revisiting the efficacy of postsecondary remediation: The moderating effects of depth/breadth of deficiency. *The Review of Higher Education*, 33 (2), 177-205.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29 (1), 37-65. Retrieved 1/29/2010, from <http://jeb.sagepub.com>
- Baker, D., Goesling, B., & Letendre, G. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review*, (46), 3, 291-312.
- Bishop, P. (2006). Effectiveness, efficiency and equity. *Australian Journal of Public Administration*, 65 (2), 108-109.
- Diversity in Mathematics Education Center for Learning and Teaching (2007). Culture, race, power, and mathematics education. In Lester, F. (ed.) *Second handbook of research on mathematics teaching and learning* (pp. 405-433). Charlotte, North Carolina: Information Age Publishing.
- Fillmore, L., & Valadez, C. (1986). Teaching bilingual learners. In M. Wittrock (Ed.). *Handbook of research on teaching*, 3rd edition (pp. 648-685). New York: Macmillan.
- Gandara, P. (2010). The Latino education crisis. *Educational Leadership*, 67 (5), 24-30.
- Garcia, G., & Pearson, P. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education*, 20 (pp. 337-392). Washington, D.C.: American Educational Research Association.
- Gipps, C. (1999). Socio-cultural aspects of assessment. In A. Iran-Nejad and P. Pearson (Eds.), *Review of Research in Education*, 24 (pp. 355-392). Washington, D.C.: American Educational Research Association.
- Gitelman, A. (2005). Estimating causal effects from multilevel group-allocation data. *Journal of Educational and Behavioral Statistics*, 30 (4), 397-412.
- Graham, S. (2010). Using propensity scores to reduce selection bias in mathematics education research. *Journal for Research in Mathematics Education*, 41 (2), 147-168.



Hanushek, E. (1989). Expenditures, efficiency, and equity in education: The federal government's role. *The American Economic Review*, 79 (2), 46-51.

Hattie, J., Jaeger, R., & Bond, L. (1999). Persistent methodological questions in educational testing. In A. Iran-Nejad and P. Pearson (Eds.), *Review of Research in Education*, 24 (pp. 393-446). Washington, D.C.: American Educational Research Association.

Housson, S., & Regalado, E. (2009). *Federal accountability/ AYP update*. Presented at Texas Assessment Conference, December 7, 2009 in Austin, Texas.

Housson, S., & Rinehart, N. (2009). *State accountability system update*. Presented at Texas Assessment Conference, December 7, 2009 in Austin, Texas.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed method research: A research paradigm whose time has come. *Educational Researcher*, 33 (7), 14-26.

Lee, V. , & Bryk, A. (1989). A multilevel model of the social distribution of educational achievement. *Sociology of Education*, 62, 172-192.

Lockwood, J., & McCaffrey, D. (2007). Controlling for individual heterogeneity in longitudinal models with applications to student achievement. *Electronic Journal of Statistics*, 1, 223-252.

Lissitz, R., & Fan, W. (2006). The projection of performance on the Maryland high school graduation exam: Magnitude, modeling, and reliability of results. A paper presented at the National Council on Measurement in Education, San Francisco, CA.

Khisty, L. (1997). Making mathematics accessible to Latino students: Rethinking instructional practice. In *Multicultural and gender equity in the mathematics classroom: The gift of diversity—1997 Yearbook*. Reston, Virginia: National Council of Teachers of Mathematics.

McCaffrey, D., Lockwood, J., Koretz, D., Louis, T., & Hamilton, L. (2004). Let's see more empirical studies on value-added modeling of teacher effects: A reply to Raudenbush, Rubin, Stuart and Zanutto, and Reckase. *Journal of Educational and Behavioral Statistics*, 29 (1), 139-143. Retrieved 1/29/2010, from <http://jeb.sagepub.com>

National Council of Teachers of Mathematics (1989). *Principles and standards for school mathematics*. Reston Virginia: NCTM. Retrieved July 13, 2009, [http://www.nctm.org/uploadedFiles/Math\\_Standards/12752\\_exec\\_pssm.pdf](http://www.nctm.org/uploadedFiles/Math_Standards/12752_exec_pssm.pdf)

National Council of Teachers of Mathematics (2001). *Teaching and learning mathematics in poor communities*. Reston, Virginia: NCTM.

Osborne, J. (2000). The advantages of hierarchical linear modeling: ERIC/AE Digest. ERIC Clearinghouse on Assessment and Evaluation, ED447198.

Raudenbush, S., & Bryk, A. (2002). Hierarchical Linear Models: Applications and data analysis methods. 2<sup>nd</sup> edition. Thousand Oaks, California: Sage.

Rochex, J. (2006). Social, methodological, and theoretical issues regarding assessment: Lessons from a secondary analysis of PISA 2000 literary tests. In J. Green and A. Luke (Eds.), *Review of Research in Education*, 30 (pp. 163-212). Washington, D.C.: American Educational Research Association.

Rubin, D., Stuart, E., & Zanutto, E. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 29 (1), 103-116. Retrieved 1/29/2010, from <http://jeb.sagepub.com>

Sanders, W., & Wright, S. (2008). A response to Amrein-Beardsley (2008) "Methodological Concerns about the Education Value-Added Assessment System". Retrieved 2/9/2010, from [http://www.sas.com/govedu/edu/services/Sanders\\_Wright\\_response\\_to\\_Amrein-Beardsley\\_4\\_14\\_2008.pdf](http://www.sas.com/govedu/edu/services/Sanders_Wright_response_to_Amrein-Beardsley_4_14_2008.pdf)

Schochet, P. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*, 33 (1), 62-87.

Secada, W. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623-660). New York: MacMillian.

Tate, W., & Rousseau, C. (2007). Engineering change in mathematics education: Research, policy and practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1209-1246). Charlotte, North Carolina: Information Age Publishing.

Tekwe, C., Carter, R., Ma, C., Algina, J., Lucas, M., Roth J., Ariet, M., Fisher, T., & Resnick, M. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29 (1), 11-36. Retrieved 1/29/2010, from <http://jeb.sagepub.com>

Texas Education Agency (2007a). Grade placement committee manual. Austin, Texas: TEA. Retrieved July 13, 2009, from <http://ritter.tea.state.tx.us/student.assessment/resources/ssi/gpcmanual07.pdf>

Texas Education Agency (2007b). Measuring annual improvement in student achievement: Pilot study results. Retrieved July 13, 2009, at [http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical\\_Reports/MeasuringAnnualImprovementInStudentAchievement.doc](http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical_Reports/MeasuringAnnualImprovementInStudentAchievement.doc)

Texas Education Agency (2008a). Technical digest 2007-2008. Retrieved July 13, 2009, from [http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/2008/chapter\\_16.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/2008/chapter_16.pdf)

Texas Education Agency. (2008b). Texas Assessment of Knowledge and Skills scale score frequencies, TAKS mathematics test, all students, grade 7, April 2008. Retrieved February 9, 2010, from [http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2008/gr7\\_08.html](http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2008/gr7_08.html)

Texas Education Agency. (2008c). Texas Assessment of Knowledge and Skills scale score frequencies, TAKS mathematics test, all students, grade 8, April 2008. Retrieved February 9, 2010, from [http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2008/gr8\\_08.html](http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2008/gr8_08.html)

Texas Education Agency (2008d). Appendix E: 2008 Classification Accuracy Tables. Retrieved February 28, 2010, from [http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/2008/appendix\\_e.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/2008/appendix_e.pdf)

Texas Education Agency (2009a). Procedures for developing the Texas Projection Measure equations. Retrieved July 13, 2009, from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/042009\\_tpm\\_development.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/042009_tpm_development.pdf).

Texas Education Agency (2009b). Texas Assessment of Knowledge and Skills-Raw score conversion table: Mathematics, Spring 2009 administration. Retrieved July 13, 2009, from [http://ritter.tea.state.tx.us/student.assessment/scoring/convtables/2009/taks/spring09\\_g07\\_math.pdf](http://ritter.tea.state.tx.us/student.assessment/scoring/convtables/2009/taks/spring09_g07_math.pdf).

Texas Education Agency. (2009c). Texas Assessment of Knowledge and Skills scale score frequencies, TAKS mathematics test, all students, grade 8, April 2009. Retrieved February 9, 2010, from <http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2009/g08.pdf>

Texas Education Agency. (2009d). TAKS region, district, and campus level data files. Retrieved February 9, 2010, from <http://ritter.tea.state.tx.us/student.assessment/reporting/taksagg/dnload.html>

Texas Education Agency. (2009e). 2009 Accountability Manual. Retrieved February 9, 2010, from <http://ritter.tea.state.tx.us/perfreport/account/2009/manual/index.html>

Texas Education Agency. (2009f). Response to U.S. Department of Education concerning Texas Projection Measure. Retrieved February 9, 2010 from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/010709\\_ResponseToUSDEConcerningTPM.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/010709_ResponseToUSDEConcerningTPM.pdf)

Texas Education Agency. (2009g). Student Success Initiative. Retrieved February 9, 2010 from [http://www.tea.state.tx.us/index3.aspx?id=3230&menu\\_id3=793](http://www.tea.state.tx.us/index3.aspx?id=3230&menu_id3=793).

Texas Education Agency. (2009h). Calculating projections with the Texas Projection Measure. Retrieved February 9, 2010 from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/042009\\_calculate\\_student\\_growth\\_with\\_tpm.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/042009_calculate_student_growth_with_tpm.pdf)

Texas Education Agency. (2009i). Explanation of TPM results. Retrieved February 9, 2010, from [http://www.tea.state.tx.us/index3.aspx?id=6410&menu\\_id=793](http://www.tea.state.tx.us/index3.aspx?id=6410&menu_id=793)

Texas Education Agency. (2009j). Procedures for developing the Texas Projection Measure equations. Retrieved February 9, 2010, from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/042009\\_tpm\\_development.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/042009_tpm_development.pdf)

Texas Education Agency. (2009k). Texas Projection Measure: Frequently asked questions. Retrieved February 9, 2010, from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/030509\\_TPM\\_FAQ.doc#Q9](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/030509_TPM_FAQ.doc#Q9)

Texas Education Agency. (2009l). Texas Education Agency growth model pilot application for Adequate Yearly Progress determinations under the No Child Left Behind Act. Retrieved February 9, 2010, from [http://ritter.tea.state.tx.us/student.assessment/resources/growth\\_proposal/011209\\_USDE\\_Growth\\_Proposal\\_Texas.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/growth_proposal/011209_USDE_Growth_Proposal_Texas.pdf)

Texas Education Agency. (2009m). Texas Assessment of Knowledge and Skills scale score frequencies, TAKS mathematics test, all students, grade 7, April 2009. Retrieved February 9, 2010, from <http://ritter.tea.state.tx.us/student.assessment/reporting/freq/2009/g07.pdf>

Texas Education Agency. (2010a). STAAR to replace TAKS. Retrieved February 9, 2010, from <http://www.tea.state.tx.us/index4.aspx?id=7874>

Texas Education Agency (2010b). Texas Student Data System and Stakeholder Feedback. Retrieved February 9, 2010, from <http://ritter.tea.state.tx.us/taa/comm020110.html>

Texas Turnaround Center (2009). Considering data from the Texas Projection Measure to inform systems of instruction, intervention, and improvement. Retrieved February 9, 2010, from [www.txt turnaround.org/docs/webinar/TPM\\_Considerations.pdf](http://www.txt turnaround.org/docs/webinar/TPM_Considerations.pdf)

Thomas, J. (1997). Teaching mathematics in a multicultural classroom. In *Multicultural and gender equity in the mathematics classroom: The gift of diversity—1997 Yearbook*. Reston, Virginia: National Council of Teachers of Mathematics.

Tong, Y. & O'Malley, K. (2006). An empirical investigation of growth models. Pearson Educational Measurement. Retrieved 2/28/2010, from [http://www.pearsonedmeasurement.com/downloads/research/Growth%20Models\\_rr0601.pdf](http://www.pearsonedmeasurement.com/downloads/research/Growth%20Models_rr0601.pdf)

Torres, M. (2004). Best interests of students left behind? Exploring the ethical and legal dimensions of the United States federal involvement in public schools. *Journal of Educational Administration*, 42 (2), 249-269.

Wainer, H. (2010). 14 conversations about three things. *Journal of Educational and Behavioral Statistics*, 35 (1), 5-25.

Why Did DISD's Ratings Go Sky High? (May 27, 2009). Dallas Morning News. Retrieved 3/1/10 from <http://dallasisdblog.dallasnews.com/archives/2009/05/why-did-disds-ratings-go-sky-h.html>