Effects of a teacher incentive program on 4th grade state assessment scores

Moises Alfaro
Texas A&M University – Kingsville

Don Jones
Texas A&M University - Kingsville

Glenda Holland
Texas A&M University – Kingsville

Marie-Anne Mundy
Texas A&M University – Kingsville

ABSTRACT

Recent national and state standards of accountability have focused on increasing student performance and achievement as well as teacher quality. Included in this challenge is the issue of teacher compensation and how it has evolved with the efforts of providing quality instruction in order to improve the performance of students. Texas has developed and implemented innovative pay systems that focus on student achievement through the process of improving teacher quality. Most recently, the 79th Legislature of Texas passed House Bill One which funded the largest investment in teacher incentives with two teacher incentive programs that would provide $320 million annually to eligible school districts and campuses. This study focused on the Texas Educator Excellence Grant (TEEG) program which provided over $100 million annually to Texas' most economically disadvantaged campuses that demonstrated high levels of student achievement as indicated on the state's academic accountability system.

Relationships between student academic performance at the 4th grade level and the total number of years of grant implementation within the 3 cycles of TEEG were analyzed, specifically, passing percentages on reading, math, writing and all tests taken for students meeting minimum standards. Overall results of this study showed statistically significant differences in passing percentages in the areas of reading, math, and all TAKS tests. No statistical difference was found in the area of writing TAKS. The year 2008 had the highest positive coefficient in the areas of reading and math, while 2007 had the highest positive coefficient in writing.

Keywords: teacher incentives, teacher compensation, school improvement, assessment, accountability

INTRODUCTION

Teacher compensation in the United States has historically evolved slowly. Protsik (1995) notes that "since the 1800's, there have only been three major changes in the method of teacher pay: an initial rural tradition of paying teachers' room and board, a move to a grade-based salary schedule, and finally the shift to today's single salary schedule" (p. 23). It is only fairly recently in that history that merit based pay was considered. Thus, the research literature on the effectiveness of performance-based pay is limited; nonetheless, it holds enough potential to support extensive policy experiments in combination with careful follow-up evaluations. In 1984, six states, including Texas, developed and implemented at least to some degree, career ladder programs but Texas had discontinued its program by 1994 never fully funding it (Cornett & Gaines, 2004).

Recently, Texas education policy efforts have once again more closely focused on improving teaching quality throughout the state. Since Texas often leads the nation in innovative education reforms, including both school and direct accountability programs and performance-based incentive pay policies, this renewed focus culminated in the creation of the nation's largest statewide performance incentive system. In May 2006, the 79th Legislature, passed House Bill One which included two-teacher incentive programs that, when fully funded at $320 million annually, created the largest investment in teacher incentive in the nation. One, which is highlighted in this study, is the Texas Educator Excellence Grant (TEEG) Program. TEEG provides approximately $100 million annually to the state's most economically disadvantaged campuses that demonstrate the highest levels of student achievement or improvement. According to the Texas Education Agency (1998), "TEEG is a non-competitive grant available to a targeted group of elementary, middle, high school, all grades, and alternative education campuses who fall in the top half of economically disadvantaged campuses and demonstrate the highest levels of student achievement or comparable improvement, as measured through the state accountability system. Campuses that receive these grant funds must use seventy-five percent of the funds to award incentives to classroom teachers. In determining which teachers receive awards, campuses must create an incentive program plan that relies on objective and quantifiable measures for two required criteria; impact on student achievement and collaboration and can include up to two optional criteria; teacher initiative, commitment, personalization, professionalism and campus involvement and teacher assignment to hard to staff or high turnover subject areas. Campuses may use the remaining twenty-five percent of funds on additional incentive, mentoring and induction, professional development, and other promising practices aimed at improving teacher quality and raising student achievement" (p. 98).

Overall findings on the effectiveness of the Texas Educator Excellence Grant (TEEG) programs seem to abate the traditional critique raised against performance incentive programs. Specifically, performance incentive programs appear to be having an encouraging impact on student performance, schools' organizational dynamics, teachers' perceptions of performance incentives, and teachers' instructional practice (National Center on Performance Incentives, 2008b).

The problem of the study specifically focused on performance incentive pay systems in Texas and the effects on student academic performance. Long a staple of the business environment, incentive or merit pay is a new concept in education. This has not been a popular idea in education as it breeds competition among teachers, grades, schools, and school districts.

However, it may be a worthwhile endeavor as tying teacher pay to performance clarifies teaching goals and attracts productive teachers (Lavy, 2007). Compounded with the opportunity to improve student performance, this provides incentive for the research that is currently lacking. The success and continuation of incentive programs will be tied to quantifiable results. This study emerged from this lack of research as it focused on student performance. It provided new insights into the Texas Educator Excellence Grant (TEEG) Program's Teacher Incentive Programs by utilizing a quantitative approach to understanding the phenomenon of performance based incentive programs.

PURPOSE OF THE STUDY

The purpose of the study was twofold: (1) to determine the effectiveness of the Texas Educator Excellence Grant (TEEG) Program's Teacher Incentive Programs on TAKS student performance at selected low-performing schools, and (2) to contribute to the literature review by examining the effectiveness of performance-based incentive programs on student success. This was achieved by determining the effects on the overall Texas Assessment of Knowledge and Skills (TAKS) scores as well as the individual scores on the reading, math and writing exams based on years of involvement and when involved in the three cycles of TEEG at selected low performing schools in the south Texas region.

METHOD

This study incorporated a quantitative methodology analyzing the effects on student achievement on the state TAKS assessments at the elementary level based on the number of years of involvement and when the involvement took place within the 3 cycles of the TEEG incentive program. Hence, the independent variable was the number of years in TEEG performance incentives and the dependent variable was student performance on the reading, math, writing and all test measures of the TAKS assessments.

Eligibility for the TEEG program was based on the percentage of economically disadvantages students and the school's accountability rating and was determined on a yearly basis. TEEG funds were offered for three consecutive years beginning with Cycle 1 in the 2006-2007 school year and was based on data from the 2004-2005 school year. Cycle 2 was implemented the 2007-2008 school year and was based on data from the 2005-2006 school year. And Cycle 3 was implemented the 2008-2009 school year based on 2006-2007 school data. Figure 1.1 (Appendix) depicts the eligibility in and out transitions of schools for the 3 years of the TEEG.

To guide the study the following research questions were derived:

1. Is there a difference in the percentage of 4th grade students meeting minimum standards on the TAKS reading test based on the number of years of grant implementation during the 3 cycles of TEEG?
2. Is there a difference in the percentage of 4th grade students meeting minimum standards on the TAKS math test based on the number of years of grant implementation during the 3 cycles of TEEG?
3. Is there a difference in the percentage of 4th grade students meeting minimum standards on the TAKS writing test based on the number of years of grant implementation during the 3 cycles of TEEG?

4. Is there a difference in the percentage of 4th grade students meeting minimum standards on all TAKS tests taken based on the number of years of grant implementation during the 3 cycles of TEEG?

RESEARCH DESIGN

A one-way analysis of covariance (ANCOVA) was used to determine whether group means on the dependent variable were the same across levels of a factor while adjusting for differences by using a covariate (Green & Salkind, 2005). In this study, the independent variable was the number of years of participation based on three cycles in the TEEG incentive program. The dependent variable was student performance on the reading, math, writing, and all-tests-taken passing percentages on the TAKS, and the covariate was year 0 of the TEEG which is the year preceding Cycle 1 of the TEEG. The control group included campuses that qualified, but never participated, in the TEEG and the experimental groups were campuses that participated at least one year in the TEEG.

This study utilized data from schools that met TEEG qualification criteria and participated in Cycles 1-3 of the TEEG. The control group included schools that met eligibility but did not participate. School eligibility was determined annually and was based on two criteria; the first was that the school had to be in the top half of Texas public schools in terms of percentage of economically disadvantaged students and the second was that the school had to have earned an exemplary or recognized rating according to the Texas state accountability rating or performed within the top quartile of Comparable Improvement in math or reading.

Population and Sample

The population of the study consisted of all elementary schools in Regions 1, 2, and 3 that housed fourth grade and that met the two criteria for participation in the TEEG. The first criteria was that the school had to be in the top half of Texas public schools in terms of percentage of economically disadvantaged students and the second was that the school had to have earned an exemplary or recognized rating according to the Texas state accountability rating or performed within the top quartile of Comparable Improvement in math or reading. Since not all elementary schools that were eligible for the TEEG participated in its implementation, the study sample consisted of those elementary schools that were eligible and participated in Cycles 1-3 of the TEEG and those that were eligible but chose not to participate, which were used as the control.

Instrumentation

The Texas Education Agency's website was used to collect a variety of data utilized in this study. Academic Excellence Indicator System (AEIS) reports were utilized to obtain TAKS passing percentages for each elementary school. AEIS is a culmination of information on the performance of students in each school and district in Texas. For the purposes of this study, data from TEEG participating elementary schools were utilized to analyze 4th grade student performance. A list of eligible and participating campuses for the three cycles of the TEEG program were also obtained from TEEG Eligibility lists found also on the TEA website.

Reliability and Validity

Reliability is a crucial quality of any quantitative study and is an indicator of the consistency of the measurements (Texas Education Agency, 2006). The TAKS test reliability is based on internal consistency measures of which include the Kuder-Richardson formula 20 (KR-20) for assessments that involve dichotomous tests items. For TAKS mixed-model tests (mixture of dichotomous items with polytomous items) such as the writing test, Pearson Educational Measurement (PEM) under the direction of TEA, provided reliability indicators such as the stratified coefficient alpha with reliabilities of 0.871, 0.884, and 0.834 for reading, math , and writing, respectively (Texas Education Agency, 2006).

In regards to the TAKS, assessment validity is based on the TEKS. Aligning the TAKS with the TEKS was done through several committees of Texas educators to identify TEKS student expectations that needed to be assessed and to agree on test objectives, guidelines, test items. A process of item development and review was developed to provide an adequate number of opportunities for educators to offer suggestions for the improvement of the TAKS and to provide a system of checks and balances (Texas Education Agency, 2006).

Data Analysis

The data obtained were entered into SPSS to run the appropriate statistical procedures. Pretest data were taken from 2006 AEIS report, which was the year preceding Cycle 1 of TEEG and posttest data were taken from 2008 AEIS reports, which is during Cycle 3 of TEEG. An ANCOVA was performed to determine if there were any differences between the passing percentages of reading, math, writing, and all-TAKS-tests-taken in 4th grade (dependent variable), based on the number of years of TEEG and which years of the three cycles the schools participated (independent variable/factor). The factor, or independent variable, divided the schools into 8 levels, depending on the number of years of implementation based on whether or not the schools participated in each of the three cycles.

RESULTS

The study involved 225 elementary campuses that housed 4th grade and met eligibility criteria for the TEEG in Regions 1, 2, and 3. Students in the 4th grade take 3 TAKS tests including reading, math, and writing. AEIS reports minimum passing percentages for each assessment, including percentages of students passing all 3 tests, labeled All Tests Taken. Students must achieve a score of 2100 to meet minimum standards. Descriptive statistics for campuses that participated in none of the cycles (control group), and the experimental groups, (Cycle 1, Cycle 2, Cycle 3, Cycles 1 and 2, Cycles 1 and 3, Cycles 2 and 3, and Cycles 1, 2, and 3) are seen in Table 1 (Appendix). Schools that were eligible but never participated in the TEEG represented 23 elementary schools that served as the control group that received no treatment, in this case, no implementation of the TEEG. Schools that participated in Cycle 1 only of TEEG had the most participation within the three cycles with 52 schools implementing TEEG while schools that participated in both cycles 2 and 3 had the least participation with only 11 schools implementing TEEG. See Table 1 (Appendix).

In the All TAKS Tests category, schools that participated in cycles 1 and 3 had the highest mean percentage of students meeting minimum expectations on all three TAKS with

80.15, while school that participated in cycle 2 only had the lowest mean percentage with 70.87. In the reading category, schools participating in cycles 1 and 3 had the highest mean percentage of students meeting minimum expectations with 86.31 and schools in the control group had the lowest percentage with 77.74. In math, schools participating in cycle 3 only had the highest mean percentage of students meeting minimum expectations with 93.94 while schools that participated in cycle 2 only had the lowest mean percentage with 82.03. Lastly, in the writing category, schools participating in all three cycles of the TEEG had the highest mean percentage of students meeting minimum expectations with 94.70 while the control group schools had the lowest percentage with 89.48 meeting minimum expectations.

Adjusted Mean Analysis

ANCOVAs conducted indicated no significant difference in the adjusted means as presented in Table 2 (Appendix). Schools that were eligible, but never participated in TEEG (control group), had the lowest adjusted means in each of the treatment areas: All TAKS Test Taken (M=71.87), Reading (M=77.88), Math (M=83.53), and Writing (M=89.830). Within the All TAKS Tests Taken and the reading ANCOVAs, cycle 2 & 3 schools had the largest adjusted mean, M=80.910 and M=87.40, respectively. In the math ANCOVA, cycle 1 & 3 schools had the largest adjusted mean with M=90.73. And in the writing ANCOVA, cycle 1 & 2 schools had the largest adjusted mean with M=93.92.

Difference in Reading Scores based on Number of Years of Incentives

Before the ANCOVA for the reading TAKS analysis, a homogeneity-of-slopes assumption was conducted to assess the interaction between the covariate, 2006 reading TAKS scores, and the factor, number of cycles participated in TEEG, as the dependent variable is predicted. Results indicate that the interaction was not significant, $F(7,209)=1.08$, $MSE=79.01$, $p=.375$ ($p>.05$), partial $\eta^2=.035$; therefore, the ANCOVA was conducted assuming homogeneity of slopes.

The reading TAKS ANCOVA was not significant, $F(7,216)=2.04$, $MSE=79.23$, $p=.052$ ($p>.05$). The strength of the relationship between the number of cycles participated in by each school and the dependent variable, TAKS reading posttest scores, was weak, as seen with the partial η^2 in which the number of cycles participated factor accounted for only 6.2% of the variance of the dependent variable, TAKS reading posttest scores, holding constant the TAKS reading pretest scores.

Since the significance level was nearly at the 0.5 level of significance and because there are several factors contributing to student TAKS scores, a MANOVA was conducted as an extension to the data analysis to identify the effect of the different cycles of TEEG participation on the dependent variable (reading TAKS tests). First, a multivariate test for homogeneity of dispersion matrices, or Box Test, was used to evaluate the null hypothesis that the covariance of the dependent variables is equal across groups. Results indicated that the homogeneity of dispersion matrices is significant, $F(70, 21320) = 1.42$, $p = 0.012$, meaning the homogeneity hypothesis was rejected. Results of the MANOVA indicated that overall, the Wilks's Λ of 0.76 was significant, $F(28, 773) = 2.19$, $p < 0.05$, meaning the null hypothesis that there is no statistically significant difference in the percentage of 4^{th} grade students meeting minimum standards on the reading TAKS test based on the number of years of grant implementation during the 3 cycles of TEEG was rejected. The multivariate η^2=0.067 indicated that 6.7% of multivariate variance of the dependent variables is associated with the factor.

As a follow-up test to the significant MANOVA, a discriminant analysis was conducted to determine whether there are differences among groups on the population means for percentages of 4th grade students meeting minimum standards on the reading TAKS tests taken in 2006, 2007, 2008, and 2009. The number of possible discriminant functions for the analysis is determined by the smaller of two values: the number of groups (N_g) -1 or the number of quantitative variables. For this study, the number of groups is 8 and the number of quantitative variables is 4; therefore, 4 is the smaller value and the number of discriminant functions since N_g-1 = 7. Preliminary statistics show that there were significant differences in the means on 3 of the 4 predictors. Reading TAKS tests for 2007, 2008, and 2009 showed p values of 0.007, 0.029, and 0.013, respectively. Reading TAKS for 2006 showed a nonsignificant p value of 0.055, which was expected because it was the year prior to TEEG. There were significant differences in the covariance matrices among the 4 groups (p value of 0.006 for the Box's M test).

Results of the descriptive discriminant function analysis, as presented in Table 3 (Appendix), reveal an overall Wilks's lambda showing significance, $\Lambda = 0.75$, χ^2 (24, N=225) = 55.43, $p < 0.05$, indicating that overall the groups differentiated among the 8 cycle implementation scenarios for TEEG implementation. The residual Wilks's lambda, $\Lambda = 0.86$, χ^2 (15, N=225) = 28.58, $p = 0.018$, that tested functions 2 through 4 was also significant, indicating significant differences among the different cycle implementation scenarios after partialling out the effects of the first discriminant function. Consequently, the first two discriminant functions were interpreted.

Table 4 (Appendix) shows strength-of-relationship statistics and revealed an Eigenvalue of 0.147 for the first discriminant function and a canonical correlation of 0.358. By squaring the canonical correlation ($0.358^2 = 0.128$), an eta square results that would have been obtained if conducting a one-way ANOVA on the first discriminant function. Therefore, 11.8% of the variability of scores for the first discriminant function is accounted for by differences among the 8 different cycle implementation scenarios for TEEG implementation. For the second discriminant function, strength-of-relationship statistics revealed an Eigenvalue of 0.90 and a canonical correlation of 0.287. By squaring the canonical correlation ($0.287^2 = 0.082$), an eta square results that would have been obtained if conducting a one-way ANOVA on the first discriminant function. Therefore, 8.2% of the variability of scores for the first discriminant function is accounted for by differences among the eight different cycle implementation scenarios for TEEG implementation.

In Table 5 (Appendix) , the within-groups correlations between the predictors and the discriminant functions as well as the standardized weights are presented. Based on these coefficients, the 2006 reading TAKS tests passing percentages demonstrate the strongest relationship with the first discriminant function according to the standardized coefficients but the 2009 reading TAKS demonstrated the strongest relationship according to the structure coefficient. With the second discriminant function, the 2008 reading TAKS passing percentages demonstrated the strongest relationship according to the structure coefficient, but the 2009 percentages showed the strongest relationship according to the standardized coefficients.

Values for group centroids, which are the mean values on the discriminant function show that cycle 1 and 2 participants (M = 0.5679) had the highest mean score on Function 1, followed by cycle 2 participants (M = 0.489), and cycle 1 participants (M = 0.104). On Function 2, cycles

1 and 3 participants (*M* = 0.550) had the highest mean value followed by cycles 2 and 3 participants (*M* = 0.327).

Difference in Math Scores based on Number of Years of Incentives

Before the ANCOVA for the math TAKS analysis, a homogeneity-of-slopes assumption was conducted to assess the interaction between the covariate, 2006 math TAKS scores, and the factor, number of cycles participated in TEEG, as the dependent variable is predicted. Results indicate that the interaction was not significant, $F(7,209)=1.17$, $MSE=82.22$, $p=.32$ ($p>.05$), partial $\eta^2=.038$; therefore, the ANCOVA was conducted assuming homogeneity of slopes. The math TAKS ANCOVA was not significant, $F(7,216)=1.95$, $MSE=82.69$, $p=.064$ ($p>.05$).

Since the significance level was nearly at 0.5 and because there are several factors contributing to student TAKS scores, a MANOVA was conducted to identify the effect of the different cycles of TEEG participation on the dependent variables (TAKS tests). First, a multivariate test for homogeneity of dispersion matrices, or Box Test, is conducted to evaluate the null hypothesis that the covariance of the dependent variables is equal across groups. Results indicated that the homogeneity of dispersion matrices was significant, $F(70, 21320) = 1.46$, $p = 0.008$, meaning the homogeneity hypothesis was rejected and there were differences in the matrices. Results of the MANOVA indicate that overall, the Wilks's Λ of 0.74 was significant, $F(28, 773) = 2.37$, $p < 0.05$. The multivariate $\eta^2=0.072$ indicated that 7.2% of multivariate variance of the dependent variables was associated with the factor.

As a follow-up test to the significant MANOVA, a discriminant analysis was conducted to determine whether there were differences among groups on the population means for percentages of 4th grade students meeting minimum standards on the math TAKS tests taken in 2006, 2007, 2008, and 2009. Preliminary statistics showed that there were significant differences in the means of all 4 groups. Math TAKS tests for 2006, 2007, 2008, and 2009 showed *p* values of 0.016, 0.000, 0.07, and 0.013, respectively. There were significant differences in the covariance matrices among the 4 groups (*p* value of 0.006 for the Box's *M* test).

Results of the descriptive discriminant function analysis, as presented in Table 6 (Appendix), revealed an overall Wilks's lambda showing significance, $\Lambda = 0.74$, $\chi^2 (24, N=225) = 60.06$, $p < 0.05$, indicating that overall the predictors differentiated among the 8 cycle implementation scenarios for TEEG implementation. The residual Wilks's lambda, $\Lambda = 0.86$, $\chi^2 (15, N=225) = 30.03$, $p = 0.012$, that tested functions 2 through 4 was also significant, indicating significant differences among the different cycle implementation scenarios after partialing out the effects of the first discriminant function. Consequently, the first two discriminant functions were interpreted.

As presented in Table 7 (Appendix), strength-of-relationship statistics revealed an Eigenvalue of 0.166 for the first discriminant function and a canonical correlation of 0.377. By squaring the canonical correlation ($0.377^2 = 0.142$), an eta square results that would have been obtained if conducting a one-way ANOVA on the first discriminant function. Therefore, 14.2% of the variability of scores for the first discriminant function is accounted for by differences among the 8 different cycle implementation scenarios for TEEG implementation. For the second discriminant function, strength-of-relationship statistics revealed an Eigenvalue of 0.0.098 and a canonical correlation of 0.299. By squaring the canonical correlation ($0.299^2 = 0.089$), an eta square results that would have been obtained if conducting a one-way ANOVA on the first discriminant function. Therefore, 8.9% of the variability of scores for the first discriminant function is accounted for by differences among the 8 different cycle implementation scenarios for TEEG implementation.

In Table 8, the within-groups correlations between the predictors and the discriminant functions as well as the standardized weights are presented. Based on these coefficients, the 2007 math TAKS tests passing percentages demonstrate the strongest relationship with the first discriminant function followed by 2009 and 2008 percentages showing weaker relationships. For function 2, the 2008 math TAKS tests passing percentages demonstrated the strongest relationship.

Values for group centroids, which are the mean values on the discriminant function show that cycle 1 and 2 participants ($M = 0.5679$) had the highest mean score on Function 1, followed by cycle 2 participants ($M = 0.489$), and cycle 1 participants ($M = 0.104$). On Function 2, cycles 1 and 3 participants ($M = 0.550$) had the highest mean value followed by cycles 2 and 3 participants ($M = 0.327$).

Difference in Writing Scores based on Number of Years of Incentives

Before the ANCOVA for the writing TAKS analysis, a homogeneity-of-slopes assumption was conducted to assess the interaction between the covariate, 2006 writing TAKS scores, and the factor, number of cycles participated in TEEG, as the dependent variable is predicted. Results indicate that the interaction was not significant, $F(7,209)=1.76$, $MSE=37.79$, $p=.097 (p>.05)$, partial $\eta^2=.056$; therefore, the ANCOVA was conducted assuming homogeneity of slopes.

The writing TAKS ANCOVA was not significant, $F(7,216)=2.01$, $MSE=38.69$, $p=.055$ ($p>.05$). A MANOVA was conducted to identify the effect of different cycles of TEEG participation on the dependent variables (TAKS tests). First, a multivariate test for homogeneity of dispersion matrices, or Box Test, was conducted to evaluate the null hypothesis that the covariance of the dependent variables is equal across groups. Results indicated that homogeneity of dispersion matrices was nonsignificant, $F(70, 21320) = 1.33$, $p = 0.03$. Results of the MANOVA indicated that overall, the Wilks's Λ of 0.86 was significant, $F(28, 773) = 1.15$, $p = 0.27$, meaning failure to reject the null hypothesis that there was no statistically significant difference in the percentage of 4th grade students meeting minimum standards on the writing TAKS test based on the number of years of grant implementation during the 3 cycles of TEEG. The multivariate $\eta^2=0.036$ indicated that 3.6% of multivariate variance of the dependent variables was associated with the factor.

As a follow-up test to the significant MANOVA, a discriminant analysis was conducted to determine whether there were differences among groups on the population means for percentages of 4th grade students meeting minimum standards on the writing TAKS tests taken in 2006, 2007, 2008, and 2009. Preliminary statistics showed no significant differences in the means of all 4 groups. Writing TAKS tests for 2006, 2007, 2008, and 2009 showed p values of 0.215, 0.176, 0.120, and 0.08, respectively. There were significant differences in the covariance matrices among the 4 groups (p value of 0.021 for the Box's M test). Results of the descriptive discriminant function analysis reveal no significance in any of the Wilks's lambda tests across all functions. Consequently, no discriminant functions were interpreted.

Difference in All TAKS Tests Taken Based on Number of Years of Incentives

Before the ANCOVA for the all TAKS Tests Taken analysis, a homogeneity-of-slopes assumption was conducted to assess the interaction between the covariate, 2006 All TAKS Tests Taken scores, and the factor, number of cycles participated in TEEG, as the dependent variable is predicted. Results indicate that the interaction was not significant, $F(7,209)=1.61$, $MSE=121.84$, $p=.135$ ($p>.05$), partial $\eta^2=.051$; therefore, the ANCOVA was conducted assuming homogeneity of slopes. The All TAKS Tests Taken ANCOVA was not significant, $F(7,216)=1.56$, $MSE=124.23$, $p=.148$ ($p>.05$).

A MANOVA was conducted to identify the effect of the different cycles of TEEG participation on the dependent variables (TAKS tests). First, a multivariate test for homogeneity of dispersion matrices, or Box Test, was conducted to evaluate the null hypothesis that the covariance of the dependent variables was equal across groups. Results indicated that the homogeneity of dispersion matrices was nonsignificant, $F(70, 21320) = 1.06$, $p = 0.35$. Results of the MANOVA indicated that overall, the Wilks's Λ of 0.80 is significant, $F(28, 773) = 1.86$, $p = 0.005$, meaning the null hypothesis that there was no statistically significant difference in the percentage of 4th grade students meeting minimum standards on all TAKS tests taken based on the number of years of grant implementation during the 3 cycles of TEEG was rejected. The multivariate $\eta^2=0.057$ indicated that 5.7% of multivariate variance of the dependent variables was associated with the factor.

As a follow-up test to the significant MANOVA, a discriminant analysis was conducted to determine whether there were differences among groups on the population means for percentages of 4th grade students meeting minimum standards on all TAKS tests taken in 2006, 2007, 2008, and 2009. Preliminary statistics showed significant differences in the means on three of four groups. All TAKS Tests 2007, 2008, and 2009 showed *p* values of 0.004, 0.014, and 0.021, respectively. All TAKS Tests for 2006 showed a nonsignificant *p* value of 0.154, which is expected because it was the year prior to TEEG. There were no significant differences in the covariance matrices among the 4 groups (*p* value of 0.225 for the Box's *M* test).

Results of the descriptive discriminant function analysis are presented in Table 9 (Appendix) and reveal only an overall Wilks's lambda showing significance, $\Lambda = 0.79$, $\chi^2(24, N=225) = 46.70$, $p = 0.004$, indicating that overall the predictors differentiated among the 8 cycle implementation scenarios for TEEG implementation. Residual Wilks's lambda tests were nonsignificant, indicating no significant differences among the different cycle implementation scenarios after partialling out the effects of the first discriminant function. Consequently, only the first discriminant function was interpreted.

Table 10 showed strength-of-relationship statistics and revealed an Eigenvalue of 0.134 for the first discriminant function and a canonical correlation of 0.344. By squaring the canonical correlation ($0.344^2 = 0.118$), an eta square results that would have been obtained if conducting a one-way ANOVA on the first discriminant function. Therefore, 11.8% of the variability of scores for the first discriminant function was accounted for by differences among the eight different cycle implementation scenarios for TEEG implementation.

In Table 11 (Appendix), the within-groups correlations between the predictors and the discriminant functions as well as the standardized weights are presented. Based on these coefficients, the 2007All TAKS tests passing percentages demonstrated the strongest relationship with the first discriminant function followed by 2008 and 2009 percentages showing weaker relationships.

Values for group centroids, which are the mean values on the discriminant function show that cycle 2 and 3 participants ($M = 0.566$) had the highest mean score on Function 1, followed by cycle 1 and 3 participants (M = 0.476), cycle 3 participants (M = 0.276), no years participation (M = 0.231), cycle 1 participants (M = -0.182), cycles 1 and 2 participants (M = -0.422), and cycle 2 participants (M = -0.473) having the lowest mean scores.

CONCLUSIONS

Findings support the argument in favor of incentive pay for educators. There were significant differences in the percentage of 4th grade students meeting minimum standards on the reading, math, and all TAKS tests taken categories based on the number of years of teacher incentive (grant years). In the three years with significant differences in the percentages, 2008 had the strongest positive coefficients indicating the highest mean scores for reading and math, while 2007 had the highest for the all tests taken category. Cycle 1 and 3 schools and Cycle 1 and 2 schools had the highest mean scores for reading and math respectively, while the Cycle 2 and 3 schools showed highest mean scores for the all tests taken category. Results indicate that schools that provided teacher incentives for two continuous years had higher scores. Schools that participated in only one cycle or in non-consecutive cycles were less effective in increasing and maintaining high levels of student achievement and consequently did not meet criteria for eligibility at some point.

RECOMMENDATIONS

States search for ways to provide competitive salaries for educators as well as increase teacher quality, maintain student achievement, and retain high quality educators. State with budget allocations that allow for teacher incentive programs may be more effective at increasing student achievement. Several key points that should be considered in future teacher incentive programs are ensuring that adequate financial resources are available to sustain these types of programs, ensuring the input of teachers in the development of incentive plans to increase acceptance and ownership, rewarding teachers for increased student achievement as well as professional development and collaboration, and continuously evaluating for improvement.

Other recommendations are for campus eligibility of incentive grants to include student academic growth versus raw passing percentages on state tests. With the current Texas Projection Measure as an example, educators can utilize student growth in academic areas to assess the quality of teaching. With added value systems, criteria can be set to analyze student academic growth and value added systems of accountability.

Additionally, professional growth should be considered as part of incentive plans. Professional development can lead to increased student performance and contribute to the quality of teaching in our classrooms. Lastly, legislatures should offer incentive programs to all districts regardless of socio-economic status to be equitable to all teachers and students.

REFERENCES

Cornett, L., Gaines, G., & Southern Regional Education Board, A. (2004). Resolve and Resources to Get a Qualified Teacher in Every Classroom. Challenge to Lead Series. *Southern Regional Education Board*.

Green, S. B. & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data (4th ed.).* Upper Saddle River, NJ: Prentice Hall.

Lavy, V. (2007). Using Performance-Based Pay to Improve the Quality of Teachers. *Future of Children*, *17*(1), 87-109.

National Center on Performance Incentives (2008b). *Texas Educator Excellence Grant (TEEG) Program: Year Two Evaluation Report.* Retrieved March 16, 2010 from http://ritter.tea.state.tx.us/opge/progeval/TeacherIncentive/TEEG_120108.pdf

Protsik, J., & Consortium for Policy Research in Education, M. (1995). *History of Teacher Pay and Incentive Reforms*.

Texas Education Agency. (1998). Incentive grant programs, *A Report to the 76th Texas Legislature.* Austin, TX, 6-16.

Texas Education Agency. (2006). Technical Digest. Retrieved May12, 2010, from http://www.tea.state.tx.us/index3.aspx?id=4351&menu_id=793

Figure 1.1 Eligibility in and out Transitions of Schools for the 3 Years of the TEEG.

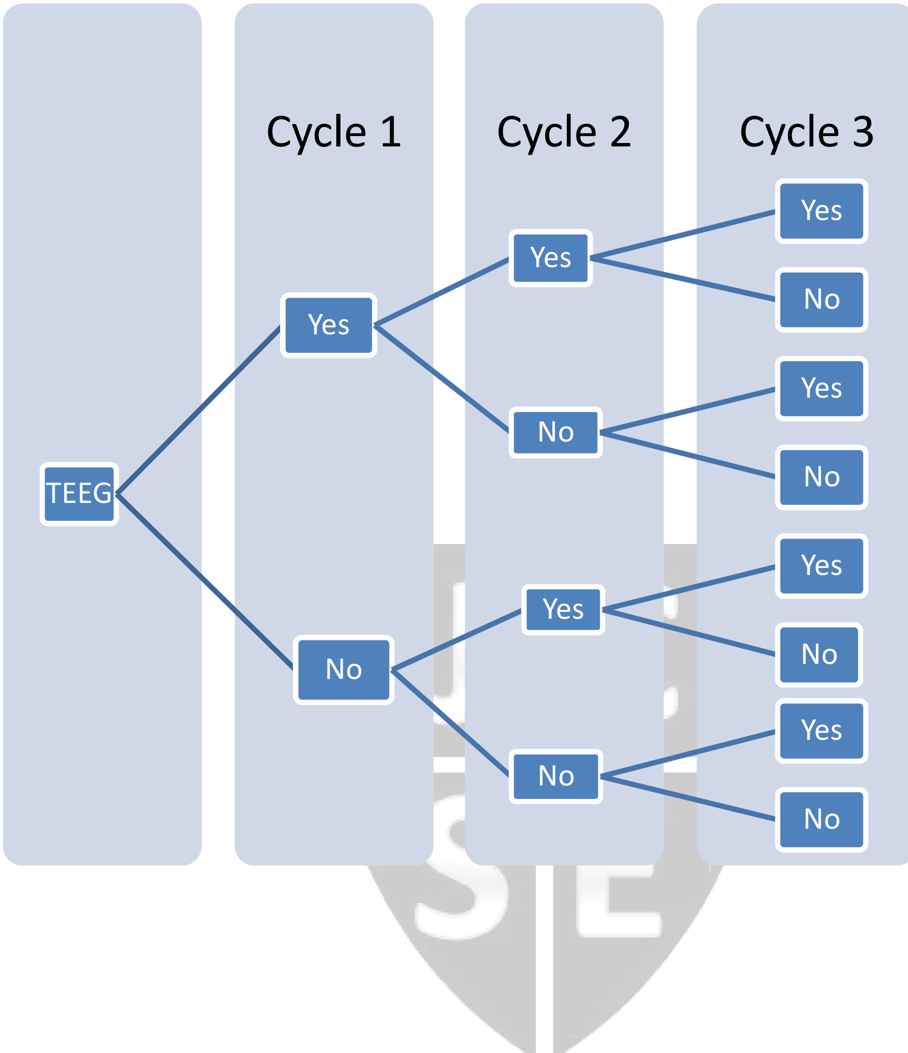

Table 1
Descriptive Statistics for Eligible Elementary Campuses Participating in TEEG

		All Tests		Reading		Math		Writing	
Cycles Participated	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
None (Control)	23	71.74	13.18	77.74	11.51	84.30	10.89	89.48	8.43
Cycle 1	52	76.83	13.23	84.13	9.69	86.81	10.62	92.54	6.74
Cycle 2	31	70.87	12.15	81.23	7.87	82.03	11.48	92.35	5.12
Cycle 3	36	72.28	12.18	80.50	9.46	93.94	11.45	89.83	8.26
Cycles 1 and 2	23	78.26	14.29	85.26	11.86	87.70	11.79	93.78	6.69
Cycles 1 and 3	26	80.15	9.58	86.31	7.95	91.50	6.65	94.23	5.49
Cycles 2 and 3	11	80.00	12.17	86.18	8.09	90.82	7.64	92.82	6.42
Cycles 1, 2, and 3	23	77.96	12.03	83.43	9.09	89.57	10.02	94.70	5.37

Table 2
Adjusted Mean Scores

Factor	All TAKS Tests Taken	Reading TAKS	Math TAKS	Writing TAKS
None (Control)	71.87	77.88	83.53	89.83
Cycle 1	76.23	84.11	86.63	92.40
Cycle 2	72.41	81.74	83.76	93.09
Cycle 3	74.28	81.91	86.35	90.09
Cycles 1 & 2	76.98	84.02	85.44	93.92
Cycles 1 & 3	79.01	85.23	90.73	93.27
Cycles 2 & 3	80.91	87.40	90.20	92.01
Cycles 1, 2, & 3	76.07	82.33	88.06	94.63

Table 3
Wilks's Lambda for Reading TAKS Variables

Test of Function (s)	Wilks's Lambda	Chi-Square	df	Significance
1 through 4	0.753	55.428	24	0.000
2 through 4	0.864	28.583	15	0.018

Table 4
Eigenvalues for Reading TAKS Variables

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.147	49.3	49.3	0.358
2	0.090	30.2	79.5	0.287

Table 5
Standardized Coefficients and Correlations of Predictor Variables with Discriminant Functions

	Correlation coefficients with discriminant functions		Standardized coefficients for discriminant functions	
Predictors	Function 1	Function 2	Function 1	Function 2
Reading TAKS Tests 2007	-0.552	0.716	-1.153	0.243
Reading TAKS Tests 2008	-0.011	0.876	-0.163	0.452
Reading TAKS Tests 2006	0.265	0.591	0.749	0.072
Reading TAKS Tests 2009	0.266	0.853	0.616	0.454

Table 6
Wilks's Lambda for Math TAKS Variables

Test of Function (s)	Wilks's Lambda	Chi-Square	df	Significance
1 through 4	0.735	60.059	24	0.000
2 through 4	0.858	30.026	15	0.012

Table 7
Eigenvalues for Math TAKS Variables

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.166	51.0	51.0	0.377
2	0.098	30.2	81.1	0.299

Table 8
Standardized Coefficients and Correlations of Variables with Two Discriminant Functions

	Correlation coefficients with discriminant functions		Standardized coefficients for discriminant functions	
Predictors	Function 1	Function 2	Function 1	Function 2
Math TAKS Tests 2006	0.172	0.719	-0.519	0.551
Math TAKS Tests 2008	0.305	0.863	-0.422	0.797
Math TAKS Tests 2009	0.480	0.677	0.328	0.181
Math TAKS Tests 2007	0.853	0.414	1.244	-0.498

Table 9
Wilks's Lambda for All TAKS Tests Taken Variables

Test of Function (s)	Wilks's Lambda	Chi-Square	df	Significance
1 through 4	0.788	46.699	24	0.004

Table 10
Eigenvalues for All TAKS Taken Variables

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.134	53.4	53.4	0.344

Table 11
Standardized Coefficients and Correlations of Variables with One Discriminant Function

	Correlation coefficients with discriminant functions	Standardized coefficients for discriminant functions
Predictors	Function 1	Function 1
All TAKS Tests 2006	-0.050	-0.720
All TAKS Tests 2009	0.189	-0.369
All TAKS Tests 2008	0.374	0.231
All TAKS Tests 2007	0.762	1.243