Education and Science tedmem

Vol 41 (2016) No 185 1-17

Is STEM Academy Designation Synonymous with Higher Student Achievement?

Ayşe Tuğba Öner [1], Robert M. Capraro [2]

Abstract

STEM education has received greater attention with increasing need of technology and engineering knowledge; therefore to improve young adults' knowledge in STEM, schools have been designated as STEM academies all over the world, especially in the US. The authors examined and compared Texas STEM (T-STEM) academies and non T-STEM schools' achievement longitudinally-2009 through 2011 to determine whether STEM schools fulfill their promises. Propensity score matching and HLM was used to determine the T-STEM and non T-STEM schools with similar backgrounds and analyze the longitudinal mathematics and science achievement of both types of schools, respectively. The results showed that from year to year for both school types, there was a statistically significant difference between students' mathematics and science scores; however, there was not a statistically significant difference between T-STEM academies' and their counterparts' academic achievement over time.

Keywords

STEM academies' achievement
STEM education
T-STEM

Article Info

Received: 26.05.2014
Accepted: 27.04.2016
Online Published: 09.06.2016

DOI: 10.15390/EB.2016.3397

Introduction

There are many proposed interventions for improving student performance in science, technology, engineering, and mathematics (STEM) fields. In fact, many hope that the interventions are linked to post secondary matriculation into STEM fields. Some even subscribe to the notion that those K-12 interventions are linked to students' entering a STEM job when they complete their education. However, there is little evidence to support these strongly held believes. Through this study, we will answer one aspect, whether or not the investments that go into STEM schools accounts for measurable or practically important variance on student learning.

STEM schools were designed to improve students' mathematics and science achievement and promote students' interest to STEM fields and career. Therefore, one can expect that STEM schools should outperform their non-STEM counterparts. In the present study, we examined STEM schools and their counterparts to determine if STEM schools and their model designs account for any variance on state accountability tests. The results of this study could address: researchers about assessing how well STEM schools performs and examining outcomes; administrators about the trajectories of their schools; teachers about how well the STEM model was implemented; and policymakers about whether it was worth to make investments to designate schools or to improve the quality of schools as well.

---

[1] (Corresponding Author) Texas A&M University, College of Education and Human Development, Department of Teaching, Learning & Culture, USA, atugbaoner@gmail.com

[2] Texas A&M University, College of Education and Human Development, Department of Teaching, Learning & Culture, USA, rcapraro@tamu.edu

The Need for STEM Education

Science, technology, engineering, and mathematics are fundamentals of cultural advancement that affects attainment of a high standard of living and economic power (National Research Council, 2011). The United States (U.S.) needs precollege students who are STEM career inclined. In order to get successful precollege students, they need instruction in integrated STEM disciplines. Students, who are taught in integrated STEM disciplines, will be able to overcome 21st century difficulties such as energy conservation, environmental protection, and health. There is a need for more integrated strategies for STEM professionals in order to overcome difficulties of the 21st century (Bybee, 2010). The need for more people in STEM careers generated the need of STEM education.

STEM education commonly focuses on science and mathematics, but technology and engineering should not be forgotten, because those two components have tremendous influence over everyone's daily life (Bybee, 2010). A less typical, but growing understanding of STEM education, is the systematic acquisition of knowledge that is dependent on some level of knowledge of each S-T-E-M field while being an expert knowledge in one (Capraro, Capraro, & Morgan, 2013). Therefore, STEM education needs to improve students' understanding of how the world and things work and how they can use technology (Bybee, 2010). Bybee (2010) stated three essential features of "true" STEM education, which are: a) improving students' understanding of how things operate, b) increasing the use of technology, and c) integrating engineering principles into students' education. In comparison to the three categories, a four category model was suggested (Scott, 2009): a) integration of science and mathematics content with the implementation of technology, b) blending academic coursework with career-technical education, c) application of STEM courses' concepts into other disciplines, and d) 'well-rounded education with outstanding science and mathematics with technology integrated across the curriculum' (p. 15). Having Scott's (2009) comprehensive definition in the literature is not surprising, because it was found that the conceptualization of STEM differs by person (Breiner, Johnson, Harkness, & Koehler, 2012). Thus, the definition of STEM education is important for both the sender and receiver, in this case, students who are interested in STEM majors.

STEM education should include science, technology, engineering, and mathematics subjects situated in an applied context. Science and mathematics are commonly taught as stand alone subjects in elementary grades. While technology is sometimes used in elementary grades but rarely taught engineering is most likely omitted. Even though, some science and mathematics' concepts are not applicable for integration (Huntley, 1998; Lonning & DeFranco, 1997), in general STEM's integrated structure, acknowledges an important nexus for the four subject areas that teachers need to understand and be able to communicate. Therefore, how STEM education is disseminated is important. Because labeling old paradigms associated with science and mathematics instruction as STEM, and/or retaining an antiquated curriculum is insufficient for developing students' interest in STEM fields or in promoting post secondary STEM matriculation.

Students who enter the post-secondary STEM track are expected to enter STEM related fields. This is referred to as the STEM pipeline. Students, who study in STEM fields in high school then progress into STEM majors in college or trade school and finally enter a STEM field, and only these students have successfully navigated the STEM pipeline. The pipeline was expected to remain constant, however there is a leak in the pipeline (Blickenstaff, 2005; Lee, 2011; Subotnik, Tai, Rickoff, & Almarode, 2010; Xu, 2008). To prevent this leak, the emphasis on STEM education has increased.

STEM education has become widespread around the world. The studies in Korea showed that there was an increase on students' interest in STEM and STEAM (Science, Technology, Engineering, Art, and Mathematics) (Jeong & Kim, 2014; Jon & Chung, 2013). However, even though students' interest increased, they did not prefer to choose STEM careers (Jon & Chung, 2013), which shows that there was a leak in the STEM pipeline. The importance of and need for STEM education was emphasized in Australia (Manufacturers' Monthly, 2015; Panizzon, Corrigan, Forgasz, & Hopkins, 2015). There were studies focusing on teachers' professional partnership in STEM in Australia (Bissaker, 2014). The need for STEM education in Malaysia was also stated in studies (Osman & Saat, 2014). For instance, the STEM teachers training programs in Malaysia improved teachers' beliefs, attitudes, perceived efficacy, and knowledge positively (Shahali et al., 2015). In addition, Malaysian students' STEM efficacy was investigated and students showed positive attitude towards their STEM assignments and exams (Meng, Idris, & Eu, 2014). The need for more STEM education programs for gifted and talented students in India was also emphasized (Kurup, Chandra, & Binoy, 2015). Out of school STEM programs in Turkey also showed that these programs were effective to increase students' STEM knowledge and skills as well as their idea for future usage of what they have learned in these programs (Baran, Canbazoğlu Bilici, Mesutoğlu, & Ocak, 2016). In another study, the importance of robotics camp to improve students' engineering skills was mentioned (Ayar, 2015). Teachers, who attended integrated teacher education programs in Turkey, had positive attitudes towards integration of mathematics and science and improved their self-efficacy beliefs to the integration (Çorlu, 2012). Besides these countries, the most amount of studies were prepared in the U.S. because of the emphasis on STEM education for a long time in the U.S. and increasing number of designation of STEM schools over time.

The U.S. takes a leadership role in science and technology around the world, but in order to maintain that leadership, young adults and students must be interested in STEM-related fields (Subotnik et al., 2010). The Congressional Research Service Report for Congress (Kuenzi, Matthews, & Mangan, 2006) stated that there was a concern that the U.S. was not preparing an adequate STEM workforce. There is an increasing concern about the insufficient number of student preparation in STEM fields in the U.S. In order to alleviate these concerns, a greater importance has been placed on STEM education. Therefore, schools were designated as STEM schools to address this need. There was an emphasis on designation of STEM schools in the U.S. and across fifteen states STEM education has been promoted through STEM designated high schools (Subotnik et al., 2010).

The Designation of Texas STEM Schools

The state of Texas is one of the states having highest number of STEM schools nowadays. Educate Texas, formally known as the Texas High School Project launched in 2004, began supporting STEM high schools and later the Texas Education Agency began a designation process. Designated schools were called Texas STEM (T-STEM) academies. The first schools were designated in the 2006-2007 academic year (SRI International, 2010). The purpose of the T-STEM initiative was to improve mathematics and science achievement across the state and raise the number of students who pursue STEM careers (SRI International, 2010).

The T-STEM academies have multiple purposes. The primary purpose of T-STEM academies was to increase achievement in STEM subjects. A secondary purpose is to nurture interest in STEM careers and foster college readiness (Pantic, 2007; Young et al., 2011). In addition, T-STEM academies were charged with developing students' 21st century skills (Young et al., 2011) such as "working in teams, using interdisciplinary approaches to problem-solving, applying technology, and communicating through multiple media' (Young et al., 2011, p. 15). Given all these expectation, one would expect increased academic achievement in mathematics and science from T-STEM academies as compared to non-T-STEM academies.

The two reports prepared by Young et al. (2011) and SRI International (2010) examined the academic achievement of T-STEM academies after their designation in 2006, and indicated that there were promising results for T-STEM academies. According to both reports, high school students in T-STEM academies had statistically significantly higher scores in mathematics and science from 2006 through 2009 in comparison to other schools. Every year, the number of T-STEM academies has been increasing in the state of Texas as well as nationally with the influx of public and private recourses. To better understand the value added in terms of academic achievement for STEM academies this study was conducted. The purpose was to provide a longitudinal examination of Texas STEM academies considering only whole school STEM Academies, excluding school within a school models and academies without at least two years of operation from 2009 to 2011 by comparing with their counterparts.

The Design of T-STEM Academies

T-STEM initiative was established with a specific design. This initiative included T-STEM Academy design blueprint benchmarks and rubrics that assessed how T-STEM academies performed on those benchmarks. The T-STEM initiative was intended to: 1) contribute to the existing efforts to improve students' mathematics and science achievement in Texas (Avery, Chambliss, Truiett, & Stotts, 2010), 2) enhance the number of students who wanted to study and have a career in STEM fields, 3) empower teachers through high quality professional development, and 4) promote school leadership (Educate Texas, 2013). T-STEM academies are expected to serve as display schools for model STEM teaching and learning (Avery et al., 2010; Educate Texas, 2013). The T-STEM design blueprint has seven benchmarks that have evolved over time, originally written in 2005, revised in 2008 and most current version was 2010 (Avery et al., 2010). These benchmarks were: a) mission-driven leadership, b) T-STEM culture, c) student outreach, recruitment and retention, d) teacher selection, development and retention, e) curriculum, instruction and assessment, f) strategic alliances, and g) academy advancement and sustainability. The T-STEM Design Blueprint was assessed with T-STEM Academies Design Blueprint Rubric. Academies were expected to make progress on each benchmark each year.

Each of the T-STEM Academy Design Blueprint benchmarks was comprised of various subcomponents that functioned to maintain STEM schools under an umbrella with seven edges. The first edge, *Mission- Driven Leadership*, included four subcomponents (Mission and Vision, Leadership and Governance, Program Review and Evaluation, and Leadership Development and Collaboration). Each of these subcategories were comprised of varying quantities of objectives. Mission and Vision (2 objectives) consisted of developing a shared mission and vision uniting every stakeholder and developing an Annual Actual Plan (AAP) to aid monitoring and evaluating T-STEM academies' mission and vision. Leadership and Governance (7 objectives) was comprised of establishing design teams, leadership teams, and an advisory board for the groundwork for school innovation, defining their roles before the employment of innovation, including stakeholders into AAP, and portraying a chart for mission-driven decision-making structure. Program Review and Evaluation (2 objectives) dealt with ensuring that mission-driven and data-driven performance occurs. Leadership Development and Collaboration (3 objectives) included collaborating with T-STEM Centers, T-STEM Coaches, and other T-STEM Academies to improve teaching and learning.

STEM Academy Culture and Design, benchmark two, was comprised of three subcategories (Personalization, Culture, and Post Secondary Success). Personalization (6 objectives) generally dealt with establishing small learning communities (SLC), time for collaboration, graduation planning, and celebrating students' successes. Culture (3 objectives) included developing a strong T-STEM identity, mutual respect, and professional learning communities for teachers. Post Secondary Success (6 objectives) generally consisted of, coursework aligned to college requirements, post-secondary exam preparation, college preparation and assistance to students and parents, develops higher educations partnerships, and provides a venue for earning college credit while in high school.

The third benchmark was *Student Outreach, Recruitment, and Retention* including three subcomponents (Recruitment, Open Access, and Student Support and Retention). Recruitment (3 objectives) subcategory was about improving structures to increase minority students and families' participation, working with students in their early ages to increase their interest towards STEM, and improving recruitment plans for all stakeholders. Open Access (2 objectives) was consisted of giving access to all students for admission without any selection process and accepting economically disadvantaged and minority students mostly. Student Support and Retention (5 objectives) comprised of improving strategies to support students, preparing sessions and summer bridge programs to help students' get experience in a STEM environment, supporting students' school-sponsored activities, and helping parents' understanding about college readiness and expectation of STEM academies.

Teacher Selection, Development and Retention was another benchmark consisting of three subcategories (Highly Qualified Teachers, Teacher Support and Development, and Teacher Retention). Highly Qualified Teachers (5 objectives) subcategory was comprised of working with higher education institutes and faculties for STEM project-based learning and creating classrooms that build self-efficacy for minority students, improving teachers' job definitions to encourage them to use research-based activities for minority students, collaborating with stakeholders, T-STEM Centers, and T-STEM Coaches, and employing innovative program to choose qualified STEM teachers. Second subcomponent, Teacher Support and Development (6 objectives), was about developing a professional development plan (PD) to fulfill academies' needs and providing PD to counselors, teachers, staff and parents for students' success, maintaining professional learning communities (PLC) with job embedded activities, providing STEM coaches for students and teachers, and building expertise for developing and assessing STEM curriculum. Teacher Retention (5 objectives) informed about supporting creative instructional practices, arranging a common time for interdisciplinary collaboration, presenting PD to develop qualified STEM teachers to make sure the STEM pipeline is stable, and AAP's support on teachers' educational improvements with STEM related educational activities.

The fifth benchmark was *Curriculum, Instruction, and Assessment* including six subcomponents. The first subcomponent was Rigor (6 objectives). Rigor consisted of alignment between curriculum, instruction, and assessment; vertical and horizontal curriculum alignment with students taking 4-years of mathematics, science and STEM elective courses; identification of students' achievement gaps; students earn 12-30 college credit hours. The second subcomponent was STEM-focused Curriculum (6 objectives). This subcomponent consisted of supporting academies with innovative STEM programs and developing assessments to evaluate these programs, providing plans for students who are deficit in classes, and supporting student involvement in extracurricular STEM activities. Instructional Practices (6 objectives) was another subcomponent comprised of organizing instruction with pre-defined standards by problem and project-based learning approaches; ensuring teachers' use of interdisciplinary standards; and ensuring students have opportunities to express choice and voice in various contexts. STEM Education Integration (6 objectives) dealt with providing teaching strategies that foster critical thinking and problem solving skills, supporting contextual environments, integrating STEM literacy and new instructional tools, and encouraging students to learn collaboratively. The Literacy (4 objectives) subcomponent included improving students' academic and technical vocabulary about STEM, encouraging 21st century literate graduates in many disciplines, using both culturally and STEM relevant materials, and offering opportunities to students to show the content by language skills. The Assessment (5 objectives) subcomponent consisted of creating formative and summative assessments, diagnosing students' gaps, and using performance based assessments for STEM learning.

Strategic Alliances, the sixth benchmark, had four subcomponents (Parent and/or Family Participation, Business and School Community, Institutions of Higher Education, Communication with Alliance Members and Stakeholders). Parent and/or Family Participation (4 objectives) dealt with learning the needs of students, educating parents about academy expectations, and involving parents in student performance. Business and School Community (3 objectives) consisted of finding business and community partners, informing and engaging with these partners, and identifying partners to provide short-term STEM job experiences for students and teachers. Institutions of Higher Education (4 objectives) was the third subcomponent including creating a Memorandum of Understanding (MOU) for dual credit, building bridges with higher education institutions, and support for students to get college services. The last subcategory was Communication with Alliances Members and Stakeholders (2 objectives) comprised of informing stakeholders about academy success and about academy graduates entering STEM college majors and careers.

The last benchmark was *Advancement and Sustainability* including four subcategories (Strategic Planning, Continuous Improvement and Evaluation, Sustainability and Growth, and Program Advancement). Strategic Planning (5 objectives) dealt with creating 3-5 year plans including academy mission and vision, collaborating with centers, coaches, networks, etc., sharing analysis and results of plan with stakeholders, developing AAP and making sure that it is sustainable. Continuous Improvement and Evaluation (3 objectives) consisted of checking AAP and ensuring the compensation of expectations from academy, reviewing instructional plan to assure the growth of academy according to state accountability measures, and creating internal assessments to evaluate the academy's growth. The third subcomponent was Sustainability and Growth (4 objectives), dealing with assuring balanced budget and investment for PD of personnel, developing plan to sustain and get grants, and protecting the components of SLC. Program Advancement (2 objectives) was about collaborating with centers and universities to write grants, and working with centers and universities to show how academy's innovative teaching was successful.

The T-STEM Academy Design Blueprint was designed to improve students' mathematics and science achievement and promote students' interest to STEM fields and career. These benchmarks were designed to be more rigorous and more exacting than what non T-STEM schools were doing. Therefore, we sought to answer following research questions:

1) Is there any statistically significant difference between students' mathematics achievement (mathematics scores and mathematics meeting standard percentage) studying in STEM schools and non-STEM schools?
2) Is there any statistically significant difference between students' science achievement (science scores and science meeting standard percentage) studying in STEM schools and non-STEM schools?

Method

To examine the academic performance of students within schools, the aggregate student average was used for each school. Each school's average mathematics and science standardized test score (Texas Assessment of Knowledge and Skills- [TAKS]) and one accountability indicator (percent meeting the mathematics and science standard; ~2000/2800 scale score) were examined. The instrument was an adequate benchmark from year to year and served as a change indicator across all the schools in the study. This longitudinal study followed the same schools for three academic years, 2008-09, 2009-10, and 2010-11. The scores were obtained through an information request with the fee being paid to the Texas Education Agency by Aggie-STEM Center.

Sample

The Texas High School Project (2011) (now Educate Texas) reported that there were 35 T-STEM academies designated before and on the 2008-2009 academic year with some of these being a school within a school model. In school within a school model, there are both STEM students and non-STEM students, where STEM students are a subgroup of school within a school model school. In school level data, it was not possible to separate STEM students from others in school within a school model. Therefore, because data from these models were not reported separately from the rest of the school, this model was eliminated from this study. Only high schools, for both T-STEM academies and their counterparts, that had 9th through 12th grades were included in the study. The number of designated T-STEM academies had consistently increased from 2006. In the present study, we included T-STEM academies that were designated on or before 2009, because it represents the most recent three-year time period for which longitudinal data were available. Therefore, T-STEM academies designated after 2009 were not included in the analyses because they would not have three complete years of school data and school-within-school academies were not included. This resulted in obtaining data from 10 T-STEM academies could be tracked for three years for the 2008-09 through 2010-11 school years. 9 out of 10 T-STEM academies were included in the final analysis. Table 1 represents the demographics of 9 T-STEM academies and the number of students by demographic variables and accountability ratings. The number of participants only from T-STEM schools was 2633. In addition to that, the number of participants from matched non-T-STEM schools increased the total sample size.

Table 1. The Number of Students in T-STEM Academies in Terms of Accountability Rating, Ethnicity, Gender, and Low-SES (L-SES)

ARA*	Ethnicity				Gender		L-SES	Total
	White	Hispanic	AA*	Other	Female	Male		
Acceptable	11	80	82	2	90	85	133	175
Recognized	2	95	1	0	45	53	89	98
Acceptable	47	722	451	72	576	716	1117	1292
Exemplary	22	89	5	2	64	54	70	118
Exemplary	5	349	87	7	246	202	392	448
Exemplary	3	111	0	0	53	61	102	114
Exemplary	104	86	6	0	85	111	104	196
Exemplary	44	11	24	0	30	49	41	79
Recognized	98	14	1	0	49	64	61	113
Total	336	1557	657	83	1238	1395	2109	2633

ARA*: Accountability Rating of the academy

AA*: African American

As other educational studies, non-probability sampling technique was used in this study. In fact, the sample of this study was purposefully selected; therefore, the sampling technique was purposive (Büyüköztürk, 2012) because the stand-alone T-STEM academies in Texas were included in the analysis. When sampling was random, the results will be more rigorous and accurate then non-random. Thus, in this study, propensity score matching strategy was used to select control group, which showed similar characteristics with treatment group, to mimic randomization. Therefore, by using matching technique, any problems resulting from sampling technique could be prevented.

In an educational study, protection of participants and confidentiality could be some ethnical problems. In this study, the data set was obtained from Texas Education Agency database and in this database, participants' personal information were kept confidential. In the data set, there was no identifying information about participants; therefore, using a dataset, which was open to the public, prevented this study from any possible ethnical problems. In addition, to use the data set, the permission from Institutional Review Board of the university was obtained.

Instrument

The Texas Assessment Knowledge and Skills test was the high-stakes test in the state of Texas until 2012. TAKS test was given to students from 3rd grade to exit level. Students took mathematics section in all grades in secondary level whereas students took science section only in 10th and 11th grades (Texas Education Agency [TEA], 2004, 2007). In 2012, the high-stakes test was changed to the State of Texas Assessment of Academic Readiness (STAAR) test. The STAAR test was an end of course exam; therefore, it was not available to run longitudinal analysis. As a result, in this study, only students, who were in 9th grade in 2009 and were followed over 3 years, scores were used. The reliability coefficients for the TAKS test for each academic subject were reported each year by TEA. According to the report, reliability coefficient for mathematics in 2008-2009 was 0.92, for mathematic and science in 2009-2010 were 0.91 and 0.90, in 2010-2011 were 0.90 and 0.89, respectively (Texas Education Agency [TEA] & Pearson, 2010a, 2010b, 2011). The validity agreement percentages were reported for 9th grades as 92, for 10th grade as 90, and for 11th grade as 85 (TEA & Pearson, 2010a, 2010b, 2011).

Data Analysis

The TAKS test subcomponents used for this study were mathematic and science, which were also dependent variables. Mathematics is tested each year of high school through the exit level in 11th grade and science is tested in 10th and 11th grades. To employ a longitudinal analysis, repeated measures are needed (van Belle, Fisher, Heagerty, & Lumley, 2004); therefore, at least two time points were needed. Thus, the 2008-09, 2009-10 and 2010-11mathematics; and, the 2009-10 and 2010-11 science tests were used to analyze the change in mathematics and science achievement. T-STEM academies that had only one-year of science or math data were removed from analyses, therefore one T-STEM academy and three non T-STEM schools had to be eliminated from the dataset before matching.

Propensity score matching was used to select a comparison non T- STEM academy group that was exactly matched for using hierarchical linear modeling as the data analytic technique.

Propensity-score matching. Propensity score matching was the first analysis that was employed in this study because it provided a useful approach for assessing treatment effects when using nonexperimental or observational data (cf. Guo & Fraser, 2010). It minimizes the effect of independent variables when estimating treatment effects, when creating a randomized control group is not possible (i.e., Austin, 2011; Wen, Leow, Hahs-Vaughn, Korfmacher, & Marcus, 2012). The variables used for the propensity-score matching were: (1) school accountability rating, (2) ethnicity, (3) gender, (4) socio economic status, and (5) school size. To specify the comparison sample for the study, nearest neighbor - many to one matching was used because of the benefit of this technique. Specifically the major benefit was maximizing the number of control schools to provide a better estimate of the true score for comparison. By many to one matching, the number of non T-STEM schools will be increased and this results in greater precision of the estimated treatment effect; thus, this method was used in the present study. As a result of the propensity score matching, 10 T-STEM academies were matched with 100 non T-STEM schools out of the 1063 possible non T-STEM schools. Prior to matching, 1 T-STEM and 3 and 4 non T-STEM schools were eliminated because of missing mathematics and/or science achievement scores, respectively.

Hierarchical linear model. When data are multilevel and the structure of data is hierarchical, observations usually are not completely independent (Hox, 1995). This characteristic leads to violation of the independence of observations assumption, and if conventional statistical tests are used, the non-realistic significant test results could be obtained because of the small estimates of standard errors (Hox, 1995). HLM prevents the violation of independence of observations assumption; thus it is one of the best analysis technique to analyze hierarchical and/or nested data like in this study (Hox, 2002; Raudenbush & Bryk, 2002). In addition HLM is a useful method to analyze longitudinal data (Snijders & Bosker, 1999). In this study, we used multilevel models to provide more robust results, because the schools were nested within districts. A two-level HLM was used to examine mathematics and science achievement over time by T-STEM academies and non T-STEM schools (HLM 7 software was used for all analyses [Scientific Software International, 2011]). The analyses were used to examine the effects of school designation on changes in mathematics and science achievement during high school.

The level 1-model equations were a set of linear regression equations for school level data. These linear regression equations showed schools' longitudinal mathematics and science achievement on their corresponding test year (time), and school type (group). This model allowed for the analysis of growth in mathematics and science as well as the explication of the probable difference between school types. There were two outcome indicators: (1) school's average standardized test scores and (2) the percentage meeting standards, by two subject areas: (a) mathematics and (b) science, therefore, four different equations were used in model. At the same time, schools' changes in mathematics achivement could be quadratic. Therefore, examining the shape of the function helped to understand the relationship between time and mathematics achivement. Thus, to analyze the quadraric growth model, time square was added to the models as variable.

$$\text{Mathsscore}_{ij} = \pi_{00} + \pi_{10} * \text{time}_{ij} + \pi_{20} * time^2_{ij} + e_{ij} \quad (1)$$

$$\text{Mathmetstd}_{ij} = \pi_{00} + \pi_{10} * \text{time}_{ij} + \pi_{20} * time^2_{ij} + e_{ij} \quad (2)$$

$$\text{Sciencesscore}_{ij} = \pi_{00} + \pi_{10} * \text{time}_{ij} + e_{ij} \quad (3)$$

$$\text{Sciencemetstd}_{ij} = \pi_{00} + \pi_{10} * \text{time}_{ij} + e_{ij} \quad (4)$$

For both linear and quadratic models, Mathsscore_{ij} and $\text{Sciencesscore}_{ij}$ represented mathematics and science average standardized test scale scores, respectively; Mathmetstd_{ij} and $\text{Sciencemetstd}_{ij}$ represented mathematics and science meeting standard percentage, respectively; π_{00} represented the coefficient associated with the intercept; π_{10} represented the coefficient associated with linear slope; time_{ij} represented the observed time point; π_{20} represented the coefficient associated with quadratic slope; time^2_{ij} represented time squared; and e_{ij} represented the error score.

Limitations

There were three limitations in this study. One of the limitations of this study was the exclusion of school within a school T-STEM academies from two types of T-STEM academies in the state of Texas: stand alone and school within a school. The reason why this school type was excluded was explained in the sample section of this study. Because the study was longitudinal and students' academic achievement over three years was the focus, students, who had less than two data points, were omited from the analysis. This was another limitation of this study. Because the data was obtained from TEA, it was unpractical to make desired changes in the data that was the third limitation of the study. For instance, due to the third limitation, we lost the data in the school level mentioned in sample section.

Results

The first HLM anaylsis was used to examine the predictors (school type, time, time squared) of the schools' average mathematics scale scores. The analysis examining the relationship of school type and schools' average mathematics scale scores showed that schools' mathematics achievement was not statistically significantly different ($p= .451$) according to school type (see Table 2). Time had a statistically significant effect ($p< .001$) on average mathematics scale scores for both T-STEM academies and non T-STEM schools (see Table 2). The data consisted of three time points for mathematics achievement; therefore, the pattern of change in mathematics was not limited to being linear (see Figure 1). The analysis showed that there was a significant quadratic growth for mathematics achievement (see Table 2). Thus, only the quadratic results were interpreted. T-STEM academies performed the same as non T-STEM schools during 2008-09. The grouping variable of T-STEM designation was a covariate and it was not statistically significant (p=.451); therefore, the covariate school type had no effect on average mathematics scale scores (see Table 2). The variable school type had no effect on the linear slope (p=.583) nor on the quadratic growth of mathematics achievement outcome (p=.553). At the beginning of the longitudinal study (2008-2009 school year), there was a statistically significant variation on average mathematics scale scores, which showed each school's mathematics scale score could vary (see Table 2). The fixed and random effects' results are presented in Table 2.

Average mathematics scale scores from 2008-09 to 2009-10 decreased for both school types (see Figure 1). From 2009-10 to 2010-11, the mathematics scale scores increased for both school types. T-STEM academies and their counterparts had the same mathematics mean scale scores over three years (see Figure 1). Figure 1 demonstrates the growth of both school types over time and shows the quadratic change and it was obvious that both schools' growth curves were overlapped.

Table 2. Fixed and Random Effects of the Model with Level-1 Predictors for Math Scale Scores Outcome

Effects		Coefficient	*SE*		*df*	*p*-value
Fixed	For INTRCPT1, π_0					
	Intercept β_{00}	2212.74	9.12		104	<0.001
	Group β_{01}	23.7	31.3		104	0.451
	For TIME slope, π_1					
	Intercept β_{10}	-49.88	9.72		207	<0.001
	Group, β_{11}	-17.5	31.81		207	0.583
	For TIMESQ slope, π_2					
	Intercept, β_{20}	43.42	4.45		207	<0.001
	Group, β_{21}	9.07	15.28		207	0.553
Random		Variance	*SD*	χ^2	*df*	*p* value
	Intercept,	6788.14	82.4	1753.24	104	<0.001
	Level-1 effect,	1282.28	35.8			

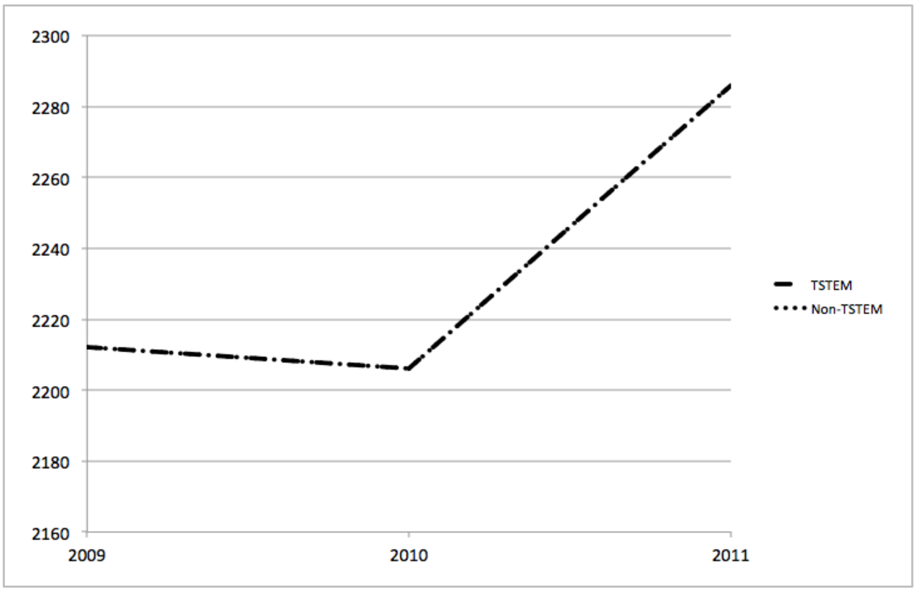

Figure 1. T-STEM Academies and Non T-STEM Schools' Mathematics Scale Scores Growth Over Time

HLM was used to examine the influence of the predictors on the outcome variable of percent of students meeting the state's established standard in mathematics. Linear and quadratic model was used where appropriate.The results for the intercept showed that the relationship of school type and schools' mathematics meeting standard percentage was not statistically significantly different (p =.178) (see Table 3). The mathematics meeting standard percentage was obtained for three consecutive years, therefore the change of growth could be non-linear. Because of the availability of the three time points in the schools' mathematics meeting standard percentage data, quadratic change over time was examined as well as linear growth over time. The linear model showed that change over time was not statistically significant (p = .724) for schools' mathematics meeting standard percentage for both T-STEM academies and non T-STEM schools. However; the quadratic model showed that there was a statistically significant (p<.001) quadratic growth for schools' mathematics meeting standard percentage (see Table 3). The variable school type had no effect on the linear slope (p=.782) and quadratic growth (p=.919) of mathematics meeting standard percentage outcome. There was a statistically significant variation on average schools' mathematics meeting standard percentage at the beginning of the study (2008-2009). The statistically significant variation indicated that the schools' mathematics meeting standard percentage can vary according to each school (see Table 3). The fixed and random effects of the model with predictors are presented in Table 3. T-STEM academies performed the same as non T-STEM schools during 2008-09.

The schools' mathematics meeting standard percentage scores were increased over three years (2008-09, 2009-10, and 2010-2011) for T-STEM academies and non T-STEM schools. The difference between T-STEM academies and non T-STEM schools' change of mathematics meeting standard percentage over time remained the same; thus, their growth lines were overlapped.

Table 3. Fixed and Random Effects of the Model with Level-1 Predictors for Math Meeting Standard Percentage Outcome

Effects		Coefficient	*SE*		*df*	*p*-value
Fixed	For INTRCPT1, π_0					
	Intercept β_{00}	69.30	1.51		104	<0.001
	Group β_{01}	7.03	5.18		124	0.178
	For TIME slope, π_1					
	Intercept β_{10}	0.82	2.32		207	0.724
	Group, β_{11}	-2.21	7.96		207	0.782
	For TIMESQ slope, π_2					
	Intercept, β_{20}	4.89	1.11		207	<0.001
	Group, β_{21}	0.38	3.82		207	0.919
Random		Variance	*SD*	χ^2	*df*	*p* value
	Intercept,	141.42	11.90	651.88	104	<0.001
	Level-1 effect,	80.36	8.96			

The relationship of school type and schools' average science scale scores showed that the schools' science achievement was not statistically significantly (*p*=.384) different according to school type for intercept (see Table 4). Time had a statistically significant (*p*< .001) effect on average science scale scores for both T-STEM academies and non T-STEM schools (see Table 4). At start of collection of longitudinal data (2009-10 school year) for science achievement, schools had a statistically significant variation on average science scale scores (see Table 4) that showed each school's science scale score could vary. Table 4 shows the fixed and random effects of the model with predictors.

There was an increase for both school types in average science scale scores from 2009-10 to 2010-11. The growth of average science scale scores increased constantly for T-STEM academies and non T-STEM schools and the average science scale scores of both schools were same.

Table 4. Fixed and Random Effects of the Model with Level-1 Predictors for Science Scale Scores Outcome

Effects		Coefficient	*SE*		*df*	*p*-value
Fixed	For Intercept, π_0					
	Intercept, β_{00}	2197.63	6.97		104	<0.001
	Group, β_{01}	20.91	23.93		104	0.384
	For TIME slope, π_1					
	Intercept, β_{10}	77.24	3.60		103	<0.001
	Group, β_{11}	-5.35	12.30		103	0.664
Random		Variance	*SD*	χ^2	*df*	*p* value
	Intercept, r_0	4095.45	63.99	1467.13	104	<0.001
	level-1, e	623.05	24.96			

The HLM analysis for the intercept indicated a non statistically significant relationship between school type and schools' average meeting standard percentage in science (*p*=.390) (see Table 5). Time was a statistically significant predictor (*p*< .001) for students meeting standard percentage in science for both T-STEM academies and non T-STEM schools (see Table 5). In 2009-10 school year, schools had a statistically significant variation on the average science meeting standard percentage (see Table 5) that showed each school's science meeting standard percentage score could vary. Table 5 displays the fixed and random effects of the model with predictors.

T-STEM academies' and their counterparts' average science meeting standard percentage score did not differ, therefore the linear growth lines for both schools overlapped. There was an increase for both school types on average students meeting standard percentage in science from 2009-10 to 2010-11.

Table 5. Fixed and Random Effects of the Model with Level-1 Predictors for Science Meeting Standard Percentage Outcome

	Effects	Coefficient	*SE*		*df*	*p*-value
Fixed	For Intercept, π_0					
	Intercept, β_{00}	74.89	1.27		104	<0.001
	Group, β_{01}	3.76	4.36		104	0.390
	For TIME slope, π_1					
	Intercept, β_{10}	16.47	1.27		103	<0.001
	Group, β_{11}	-3.36	4.36		103	0.443
Random		Variance	*SD*	χ^2	*df*	*p* value
	Intercept, r_0	78.55	8.86	311.71	104	<0.001
	level-1, *e*	78.31	8.84			

Discussion

The goal was to determine if T-STEM academies student achievement over time earned higher scores when compared to their counterparts. Given the amount of time, effort, and the academies blueprint, one might expect that they would show greater student learning. However, the results showed that T-STEM academies and non T-STEM schools' academic achievement growth over time were the same. The results of this study were not parallel with those of other researchers (e.g. SRI International, 2010; Young et al., 2011). SRI International (2010) and Young et al. (2011) evaluated T-STEM academies' and their counterparts' mathematics, reading and science achievement. Their studies differed in several ways from the current study. Young et al.'s (2011) sample was STEM standalone and school-within-school models. The problem with school-within-school models is that achievement scores were reported as whole school scores, thus a school within a school's achievement could highly represent the non-STEM part of the school's achievement. Further complicating the results is that school-within-school models are likely to be the smaller component. Even if, the data were accurate to only the STEM students in the school-within-school model, that data would have to be obtained from someone at that school who had knowledge of which students were in STEM, so they could visually scan test homeroom rosters to select the scores. This is a laborious and time-consuming process. Because in Texas, the test is not administered within program or teacher; therefore, a STEM student may be tested in mathematics by a teacher or proctor who is neither in STEM nor a teacher of mathematics. In the present study, the sample was intentionally drawn as standalone STEM schools to avoid any confounds that results from using a confederate at the school to cull the sample data from the whole school data. This is one of the strengths of this study. By using data obtained through the test-administrating agency, it was more likely to be impartial and accurate. Thus including school-within-school T-STEM academies would result in untenable results. In the SRI International (2010) study focused on early college high schools and T-STEM academies; therefore, their results were influenced by the performance of early college high school students. The confound of early college high schools and T-STEM academies treats early college high schools on par with T-STEM academies when in reality the academies blueprint is not used with early college high schools. Therefore, the obtained results are indicative of the performance of the supervising agency but not attributable to either one of the two programs.

The matching strategy for all three studies was similar but exactly the same. All the studies used propensity score matching many to 1 matching. However, SRI International (2010) and Young et al. (2011) used 8 matching variables that included grade span, campus rating, TAKS mathematics passing rates, TAKS reading passing rates, urbanicity, enrollment, Title 1 status, and percentage African-American and Hispanic students. While the added variables might provide a better match using background variables that have little no theoretical foundation should not be "tossed in" for the sake of using more variables. They focused on mathematics, science, reading, and social studies student outcomes for each grade separately in 2008-09 within considering the longitudinal effects. The strength of the present study is that we compared the overall growth of T-STEM academies and non T-STEM schools from 2008-09 through and including 2010-11 to provide a more accurate understanding of schools' performance. This strategy ensured that we used the same students across time and not a snapshot across grade levels in the same year, again resulting in a more robust finding, which is also another strength of this study.

The crucial question is why do T-STEM academies not show greater student growth over time than non T-STEM schools? There might be several reasons and further research could shed light on these. For instance, Peterson et al. (2013) indicated that one possible reason could be insufficient professional development (PD). In addition, Öner et al. (2014) found that there was not a statistically significant difference between T-STEM academies located in different education service centers and supposed to receive different PDs according to their needs. As a result, Öner et al. (2014) emphasized whether T-STEM academies receive high quality PDs that can also address their needs. Due to a lack of PD experience, there could be ineffective implementation such as unsuccessful project-based learning instruction. Therefore, the further investigation of differences after implementation of more qualified PD would help us to understand the effectiveness of the program and whether PD is one of the reasons. Demographics could be another reason. T-STEM academies were nonselective schools therefore percentages of students comprising the population for each study could be different (Avery et al., 2010). In addition, because other studies used stand alone and school within a school model T-STEM academies in their sample, it is inevitable to obtain difference between the demographics of samples.

There was a decrease in mathematics achievement from 9th to 10th grade for both school types and this could be due to the complexity and difficulty of the subject itself. However, if T-STEM schools are charged with preparing our future STEM professionals it is clear that they do not have a magic recipe nor the magic bullet to meet the U.S. needs for highly trained STEM professionals. It is not that there might not be promise in developing STEM academies, the criteria should be clearly developed and oversight should be aligned with an agency charged with ensuring success. Money should follow success while those floundering for three or more years should be returned to a non-designated status. Only when oversight is not based on politics but on metrics will STEM schools achieve their promise. Another likely solution is more STEM trained and certificated teachers in elementary, middle, and high school that clearly understand the higher education expectations and job markets for STEM professionals. These teachers would be able to facilitate a more STEM focused curriculum, work collaboratively across disciplines, and develop clear visions and missions for achieving higher achievement in STEM fields.

As indicated before, if Turkish pre-service teachers were in integrated teacher education programs, they were much successful and able to understand the integrated STEM education (Çorlu, 2012). Therefore, increasing number of qualified PDs, developing more partnerships, and promoting teachers who are able to understand and implement STEM education would avail improving STEM education. In this way, more qualified STEM teachers would serve in STEM schools and it would be possible to reach desired goals about STEM in Turkey as well as the U.S.

References

Austin, P. C. (2011). A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behavioral Research, 46*(1), 119-151.

Avery, S., Chambliss, D., Truiett, R., & Stotts, J. L. (2010). Texas science, technology, engineering, and mathematics academies design blueprint, rubric and glossary. Report of Texas High School Project T-STEM Initiative. Retrieved from http://www.edtx.org/uploads/general/pdf-downloads/misc-PDFs/2011_TSTEMDesignBlueprint.pdf

Ayar, M. C. (2015). First-hand experience with engineering design and career interest in engineering: An informal STEM education case study. *Educational Sciences: Theory & Practice, 15*(6), 1655-1675. doi:10.12738/estp.2015.6.0134

Baran, E., Canbazoglu Bilici, S., Mesutoglu, C., & Ocak, C. (2016). Moving STEM beyond schools: Students' perceptions about an out-of-school STEM education program. *International Journal of Education in Mathematics, Science and Technology, 4*(1), 9-19. doi:10.18404/ijemst.71338

Bissaker, K. (2014). Transforming STEM education in an innovative Australian school: The role of teachers' and academics' professional partnerships. *Theory Into Practice, 53*, 55-63. doi:10.1080/00405841.2014.862124

Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17,* 369-386.

Breiner, J. M., Johnson, C. C., Harkness, S. S., & Koehler, C. M. (2012). What Is STEM? A discussion about conceptions of stem in education and partnerships. *School Science and Mathematics, 112*, 3-11.

Büyüköztürk, Ş. (2012, Mayıs). *Örnekleme yöntemleri*. Retrieved from http://w3.balikesir.edu.tr/~msackes/wp/wp-content/uploads/2012/03/BAY-Final-Konulari.pdf

Bybee, R. W. (2010). What is STEM education? *Science, 329*, 996.

Capraro, R. M., Capraro, M. M., & Morgan, J. (Eds.). (2013). *Project-based learning: An integrated science, technology, engineering, and mathematics (STEM) approach* (2nd ed.). Rotterdam, The Netherlands: Sense.

Çorlu, M. S. (2012). *A pathway to STEM education: Investigating pre-service mathematics and science teachers at Turkish universities in terms of their understanding of mathematics used in science* (Doctoral dissertation). Texas A&M University, College Station, TX. (UMI No. 3524216)

Educate Texas. (2013). T-STEM academy design blueprint. Retrieved from http://www.edtx.org/uploads/general/pdf-downloads/misc-PDFs/EDTX_TSTEM_Academyblueprint.pdf

Guo, S., & Fraser, M.W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: SAGE.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam, The Netherlands: TT-Publikaties.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Huntley, M. A. (1998). Design and implementation of framework for defining integrated mathematics and science education. *School Science and Mathematics, 98*(6), 320-327.

Jeong, S., & Kim, H. (2014). The effect of a climate change monitoring program on students' knowledge and perceptions of STEAM education in Korea. *Eurasia Journal of Mathematics, Science & Technology Education*, *11*(6), 1321-1338.

Jon, J. E., & Chung, H. I. (2013). *Consultant report securing Australia's future STEM: Country comparisons-STEM Report Republic of Korea.* Retrieved from http://www.acola.org.au/PDF/SAF02Consultants/Consultant%20Report%20-%20Korea.pdf

Kuenzi, J., Matthews, M., & Mangan, B. (2006). Science, technology, engineering, and mathematics (STEM) education issues and legislative options. Report of Congressional Research Service. Washington, DC: Congressional Research Service. Retrieved from http://www.unm.edu/~cstp/articles/RL33434.pdf

Kurup, A., Chandra, A., & Binoy, V. V. (2015). 'Little minds dreaming big science': Are we really promoting 'children gifted in STEM' in India?. *Current Science, 108*(5), 779-781.

Lee, R. E. (2011). *Navigating the science, technology, engineering, and mathematics pipeline: How social capital impacts the educational attainment of college-bound female students* (Unpublished doctoral dissertation). University of Southern California, CA.

Lonning, R. A., & DeFranco, T. C. (1997). Integration of science and mathematics: A theoretical model. *School Science and Mathematics, 97*(4), 212-215.

Manufacturers' Monthly. (2015, May). News and analysis: STEM skills urgently needed. Manufacturers' Monthly.

Meng, C. C., Idris, N., & Eu, L. K. (2014). Secondary students' perceptions of assessments in science, technology, engineering, and mathematics (STEM). *Eurasia Journal of Mathematics, Science & Technology Education, 10*(3), 219-227.

National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics.* Committee on Highly Successful Science Programs for K-12 Science Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Osman, K., & Saat, R. M. (2014). Editorial. Science technology, engineering and mathematics (STEM) education in Malaysia. *Eurasia Journal of Mathematics, Science & Technology Education, 10*(3), 153-154.

Öner, A. T., Navruz, B., Bicer, A., Peterson, C. A., Capraro, R. M., & Capraro, M. M. (2014). T-STEM academies' academic performance examination by education service centers: A longitudinal study. *Turkish Journal of Education, 3*(4), 40-51.

Panizzon, D., Corrigan, D., Forgasz H., & Hopkins, S. (2015). Impending STEM shortages in Australia: Beware the 'smoke and mirrors'. *Procedia-Social and Behavioral Sciences, 167*, 70-74.

Pantic, Z. (2007). STEM sell. *New England Journal of Higher Education, 22*(1), 25-26.

Peterson, C. A, Oner, A. T., Nite, S., Bicer, A., Capraro, R. M., Capraro, M. M., … Sahin, A. (2013, September). *Aggie-STEM: Improving student outcomes with professional development that incorporates STEM-based project-based learning and professional learning communities.* Paper presented at the regional meeting of the Mathematics and Science Partnerships Program, Washington D.C.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Scott, C. E. (2009). *A comparative case study of the characteristics of science, technology, engineering, and mathematics (STEM) focused high schools* (Unpublished doctoral dissertation). George Mason University, Fairfax, VA.

Shahali, E. H. M., Halim, L., Rasul, S., Osman, K., Ikhsan, Z., & Rahim, F. (2015). Bitara-STEM training of trainers' programme: Impact on trainers' knowledge, beliefs, attitudes and efficacy towards integrated stem teaching. *Journal of Baltic Science Education, 14*(1), 85-95.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

SRI International. (2010). *Evaluation of the Texas high school project: Second comprehensive annual report.* Austin, TX: Author. Retrieved from http://www.earlycolleges.org/Downloads/SRI_THSPE_second_Nov2010.pdf

Subotnik, R. F., Tai, R. H., Rickoff, R., & Almarode, J. (2010). Specialized public high schools of science, mathematics, and technology and the stem pipeline: What do we know now and what will we know in 5 years?. *Roeper Review, 32*, 7-16.

Texas Education Agency [TEA]. (2007). *Texas assessment knowledge and skills information booklet: Mathematics exit level*. Retrieved from: http://tea.texas.gov

Texas Education Agency [TEA]. (2004). *Texas assessment knowledge and skills information booklet: Science grade 11 exit level*. Retrieved from: http://tea.texas.gov

Texas Education Agency [TEA], & Pearson. (2011). *Technical digest for the academic year 2010-2011*. Retrieved from http://tea.texas.gov

Texas Education Agency [TEA], & Pearson. (2010a). *Technical digest for the academic year 2008-2009*. Retrieved from http://tea.texas.gov

Texas Education Agency [TEA], & Pearson. (2010b). *Technical digest for the academic year 2009-2010*. Retrieved from http://tea.texas.gov

van Belle, G., Fisher, L. D., Heagerty, P. J., & Lumley, T. (2004). *Biostatistics: A methodology fort he health sciences* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Wen, X., Leow, C., Hahs-Vaughn, D. L., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's school performance in kindergarten. *Early Childhood Research Quarterly, 27*(4), 684-694.

Xu, Y. J. (2008). Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. *Research in Higher Education, 49*, 607-624.

Young W. M., House, A., Wang, H., Singleton, C., SRI International, & Klopfenstein, K. (2011). *Inclusive STEM schools: Early promise in Texas and unanswered questions*. Paper presented at the National Research Council Workshop on Successful STEM Education in K-12 Schools. Retrieved from http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072639.pdf