Abstract Title Page

Title:
Improving gender, racial, and social equity in elementary science instruction and student achievement: The impact of a professional development program

Author(s):
Rheta E. Lanehart, M.S.P.H., University of South Florida; Kathryn M. Borman, Ph.D., University of South Florida; Theodore L. Boydston, Ph.D., University of South Florida; Bridget A. Cotner, M.A., University of South Florida; Reginald S. Lee, M.A., University of South Florida

Abstract Body

Background/context:
The mission of Teaching SMART, a science inquiry professional development program for elementary school teachers, is to encourage the performance and persistence of all students, particularly female and minority youth, in elementary science. Teaching SMART provides instruction, hands-on training, and long-term technical assistance and support for third through fifth grade teachers in a given school to increase their awareness and comfort level in using equitable, hands-on inquiry, and exploration-based approaches to teaching science, ultimately enhancing student achievement. The Teaching SMART intervention has been fully developed and implemented in school districts around the country.

Purpose / objective / research question / focus of study:
The study was designed to establish the efficacy of Teaching SMART (Teaching Science, Mathematics and Relevant Technologies); a science professional development program for teachers with students in grades 3 through 5 that specifically addressed the following research questions: 1) To what extent does participation in the Teaching SMART professional development program by elementary school teachers in grades 3-5 improve their practice of teaching science; and 2) To what extent does teachers' participation in the Teaching SMART professional development program improve students' outcomes in science? Focusing on the second research question, the purpose of this paper is to investigate the impact of the Teaching SMART program on gender, racial, and social equity as well as student and teacher attitudes toward science. In this paper, the Partnership for the Assessment of Standards-based Science (PASS) multiple choice assessment is evaluated as the outcome measure.

Setting:
There are 35 elementary schools ranging in size from 400 to 1100 students in a large Florida school district, characterized by low-SES, diverse minority groups, and linguistic minority groups with three schools having a non-white majority student population. The district is home to 18 Title I schools with a free/reduced lunch eligibility rate greater than 50%. The mobility rate among students is 44%. In fall 2005, there were approximately 62,768 students attending pre-kindergarten through 12th grades including 28,831 students at the elementary level in grades k-5. The racial composition of the elementary students is 78% White, non-Hispanic, 4% African American, 12% Hispanic, 6% other (including American Indian, Asian, and Multi-racial).
The school district's organizational priorities include three guiding principles: continuous progress, continuity of caring, and ensuring equity and excellence. The intent of continuous progress is to enhance students' academic, social/emotional, and physical performance with a curriculum that allows students to progress at their own rate. The "blurring" of ages across grades results in multiage or non-graded classrooms. About 60 percent of the teachers in the study taught classes with multiple combinations of grade levels, including three and four, four and five, and three through five. The goal of continuity of caring allows students to have individual teachers or teams of teachers across multiple years. Most of our schools' buildings (14 of 20 schools) were designed with four classrooms arranged with a common core area that included common restrooms, an open area, and storage or support staff rooms called pods. The four teachers in a pod work together to varying degrees as a team. Some schools follow the continuous progress plan while others allowed teams to move students among the four teachers.

In some teams we found examples of teachers teaching by grade levels for science where, for example, one teacher on the team taught all the third graders, and instances where teachers taught multi-grade science within their classrooms.

When the Teaching SMART program was implemented in the fall of 2005 the Florida state assessment, the FCAT, including fifth grade science was administered as a measure of accountability in each school. The district also began a district-wide professional development program based on Learning Focused Strategies (LFS), developed by Max Thompson and Julia Thompson. This program was implemented in groups of elementary schools each year for three years. Focused on general teaching strategies, LFS included such approaches such as essential questions, activating, teaching, and summarizing strategies. All schools within the district were required to participate in LFS professional development.

At the time of this research, there was growing pressure placed upon teachers to focus on reading principally due to the perception of district personnel that this area needed special attention. Some principals would not allow teachers to use any science reading activities in their required, uninterrupted 90 minute reading/language arts block. The district uses an integrated thematic program. Elementary intermediate science themes were clustered across three years by semester among science, social studies, and health, according to district strands and Florida Sunshine State Standards. The district curriculum map for science instruction included 60 days of science themes for year one, 80 days for year two, and 55 days for year three. The district encouraged integration of content in line with Teaching SMART guidelines so teachers could include applicable science lessons as part of their social studies or health themes. The district did not offer other science specific professional development programs during the implementation.

Population / Participants / Subjects:

This is an analysis of the 3rd grade cohort of students (n=1037) in treatment and control schools who participated in the study in grade 3 and continued with at least 3 data collection points to grade 5 (see Table 1). The 3rd grade cohort was a subset of the total 'intent to treat' sample (n=8429) that included all students in the study (Montori & Guyatt, 2001). The number of 3rd grade students in the study at Fall 2005 was 1438: 848 treatment students and 590 control students. Overall attrition bias for the cohort was 27.9% (treatment students = 598; control students= 439) and differential attrition between treatment (29.5%) and control groups (25.6%) was 3.9% indicating an "acceptable level of bias even under conservative assumptions" (What Works Clearinghouse [WWC], 2008). Only 19 out of the 20 randomized schools were included in the analysis because one treatment school had no third graders in Fall, 2005.

Intervention / Program / Practice:

Teaching SMART is a comprehensive, three-year teacher professional development program designed to produce systemic changes in the classroom and the school through improving science education at the elementary school level. The professional development program is intended to change teaching practice in accord with the National Research Council's National Science Education Standards. The program is designed to support the enhancement of teachers' confidence in teaching science through the development of their science content knowledge and pedagogical skills. In turn, students' attitudes toward and confidence in science are expected to improve. Teacher professional development consists of three annual trainings with classroom follow-up visits by district resource teachers who lead the training and are assigned a set of teachers and two half-day networking sessions. The program is delivered by district resource teachers who have been trained and monitored by Teaching SMART program staff. The

Teaching SMART program provides teachers with 103 lessons that were adapted to follow the Teaching SMART format. The lessons were reorganized by the district resource teachers to align with the district's three year thematic cluster program. There was an average of 45 Teaching SMART lessons recommended a year and 85 unduplicated recommended lessons across the three year cluster. At the conclusion of each level of teacher training, teachers signed a "contract" stipulating participation in Teaching SMART activities that included completing a minimum of two lessons per month. The resource teachers prepared science kits for each teacher and school. These kits contained the majority of the materials needed to conduct the recommended activities. The resource teachers also assisted teachers by providing perishable supplies. These kits were "refurbished" each year by the resource teachers.

Research Design:

A randomized controlled trial (RCT) is used in which schools (n=20) were randomly assigned to either a treatment (n=10) or control conditions (n=10). Teachers in treatment schools participated in the Teaching SMART professional development program while teachers in control schools participated in a business-as-usual condition.

Data Collection and Analysis:

Data were collected in elementary schools at four time points: Fall, 2005 (baseline), Spring, 2006, Spring, 2007, and Spring 2008. The items used in the teacher and student survey were adapted from the work of the Council of Chief State School Officers (Blank, Porter & Smithson, 2001). Scores from the instructional practice items were explored through factor analysis where six categories were created: empowering, hands-on, exploring, inquiry, conventional practices, and self-efficacy. In this analysis, we focus on the reported impact of professional development on self-efficacy as an indicator of attitude toward science. Three measures of science achievement are used to evaluate the impact of the Teaching SMART professional development program on student outcomes. Two measures were developed by the Partnership for the Assessment of Standards-based Science (PASS) and consist of a performance task and a multiple choice assessment. Students' responses to multiple-choice PASS items are computer scored while performance items are scored using analytical rubrics with high interrater reliability (Klein, Stecher, Shavelson, McCaffery, Ormseth, Bell, Comfort & Othman, 1998). The third measure is the state of Florida standardized test, the Florida Comprehension Assessment Test (FCAT). In this paper, the PASS multiple choice assessment is evaluated as the outcome measure. The PASS uses an approach to measuring science learning that varies from standardized testing approaches. To account for this variation, item response theory (IRT) analysis for binary data is used to scale the multiple choice achievement outcome using BILOG-MG3 software (Zimowski, Muraki, Mislevy & Bock, 1996). A multilevel repeated measures model is used to analyze students' scaled scores using HLM 6.06 software (Dedrick, et al., 2009; Raudenbush & Byrk, 2002; Singer & Willett, 2003).

Models

The unconditional model with scaled multiple choice score as the outcome and no predictor variables had a level-2 intraclass correlation coefficient (ICC) of 0.12 and a level-3 ICC of 0.02 (see Table 2) indicating that 12% of the variance in scaled multiple choice scores occurred between teachers and 2% of the variance occurred between schools (Hox, 2002). The level-2 ICC indicates a moderate degree of data dependency and clustering, but a higher ICC value is not unusual in repeated measures designs.

The Final Model (see Table 2 & Table 3) was specified as:

Y_{ij} (Scaled Multiple Choice Score) =

$$\gamma_{000} + \gamma_{001} * \text{FCAT} + \gamma_{010} * \text{Gifted} + \gamma_{020} * \text{Reading Baseline} + \gamma_{030} * \text{Math Baseline} + \gamma_{100} * \text{Time} + \gamma_{101} * \text{Group*Time} + \gamma_{102} * \text{Migrant * Time} + \gamma_{110} * \text{Teacher Science Emphasis* Time} + \gamma_{120} * \text{Non-White*Time} + \gamma_{200} * \text{Self-efficacy} + r_0 + r_1 * \text{Time} + r_2 * \text{Self-efficacy} + u_{00} + u_{10} * \text{Time} + u_{20} * \text{Self-efficacy}$$

where scaled multiple choice score Y_{ij} is the outcome for student *i* at timepoint *j*. Y_{ij} is predicted by the intercept, γ_{000}, the fixed effects of the independent variables , the random slopes of the intercept, time, and self-efficacy ($r_0 + r_1 * Time + r_2 * Self\text{-}efficacy$), and the variance of these slopes ($u_{00} + u_{10} * Time + u_{20} * Self\text{-}efficacy$).
Model 14 (See Table 2 & Table 3) was specified with the same parameters as the Final model with the exception of the inclusion of the linear growth predictor variable, % females at a school. Females are a subgroup of interest in the research and results for the linear growth in scaled multiple choice scores for schools with a high percentage of females are reported despite the lack of statistical significance.

Findings / Results:
Treatment schools' rate of change in scaled multiple choice scores was significantly higher (0.09, $p < 0.01$) per year resulting in positive growth for treatment schools ($slope_{tx} = 0.02$) and negative growth ($slope_{ctrl} = -0.07$) for control schools (see Figure 1, Table 2 & Table 3). Classrooms in schools with a high percentage of non-white students had a decrease in scaled multiple choice scores overtime (-0.18, $p < 0.01$). However, classrooms in treatment schools with a high percentage of non-white students had a rate of change in scaled multiple choice scores equivalent to control classrooms with a low percentage of non-white students, indicating a treatment effect for minority students receiving Teaching SMART (see Figure 2). Schools with a high percentage of migrant students had a large decrease in scaled multiple choice scores overtime (-0.72, $p < 0.01$). However, treatment schools with a high percentage of migrant students had a rate of change in scaled multiple choice scores greater than control schools with a high percentage of migrant indicating a treatment effect for low income migrant students in treatment schools receiving Teaching SMART (see Figure 3). Schools with a high percentage of females had an increase in scaled multiple choice scores (0.22, $p = 0.25$), but the increase was not significant. Treatment schools with a high percentage of females had the greatest rate of change in scaled multiple choice scores overtime indicating a treatment effect for females in treatment schools receiving Teaching SMART (see Figure 4). The level-1 predictor, self-efficacy increased school scores by 0.22 ($p < 0.01$) scaled points each year. Our results indicate that change in student self-efficacy was higher in treatment schools than control schools for both low and high self-efficacy scores. In addition, schools with low self-efficacy at baseline had lower scaled multiple choice scores overtime indicating a relationship between attitude and achievement ($r = 0.18$, $p < 0.001$).
A teacher's high emphasis on science activities (-0.09, $p < 0.01$) decreased a school's scaled score overtime. However, teachers in treatment schools with low emphasis on science activities at baseline experienced the greatest increase in classroom emphasis on science activities indicating a change in attitude towards science for this subgroup.

Conclusions:
We conclude that teacher participation in the Teaching SMART professional development program improved science achievement outcomes as seen by the significant treatment effect of 0.9 ($p < 0.01$). Although a treatment effect was seen in science achievement outcomes among minorities and low income subgroups, a wide gap in achievement still exists when compared to their counterpart subgroups. Treatment schools with a high percentage of females experienced the greatest improvement in scaled multiple choice scores indicating that the Teaching SMART professional development program can have a positive impact on gender equity in science education. Student attitudes toward science were significantly improved in the treatment schools. Teacher attitudes towards science were improved as seen by the change in emphasis on science activities among low science emphasis treatment teachers. Teacher attitude plays a major role in determining whether socially equitable education and "science for all" is practiced in the classroom (Howes, 2002, p.849). The small but statistically significant growth trend for the treatment schools is encouraging and suggests pursuing a scale-up of Teaching SMART professional development program in other school districts. This study indicates that initiating inquiry-based science instruction in elementary grades has an impact on science attitudes and achievement that may prevent subsequent negative patterns in science learning (Andre, et al. 1999). A criterion of scientific literacy for the National Standards for Science Education (National Research Council, 1996) includes the dual roles of inquiry and content in science education. Based on the findings in this study, a greater emphasis should be placed on incorporating strategies for enhancing the achievement and success of both minority and low SES students, especially the girls for whom Teaching SMART was developed in the first place. The findings for students, especially for the girls in this study as well as the findings for teachers are most encouraging.

Appendices

Appendix A. References

Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36*, 719-747.

Blank, R.K., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science: Results from survey of enacted curriculum project final report*. Washington, DC: CCSSO.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*, 69-102.

Howes, E. V. (2002). Learning to teach science for all in the elementary grades: What do preservice teachers bring? *Journal of Research in Science Teaching, 39*, 845-869.

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

Klein, S., Stecher, B., Shavelson, R., McCaffrey, D., Ormseth, T., Bell, R., Comfort, K., & Othman, A.R. (1996). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education, 11*, 121-137.

Montori, V. M., Guyatt, G. H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal, 165*, 1339-1341.

National Research Council. (1996). National science education standards. Washington, D. C.: National Academy Press.

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

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal analysis: Modeling change and event occurrences*. New York: Oxford University Press.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items.* Chicago: Scientific Software International, Inc.

Appendix B. Tables and Figures

Table 1. *School characteristics at baseline*

	Variable	Group	M	SD	Skewness	Kurtosis	Outliers
Level One	Scaled Multiple Choice Score	Control	-0.02	0.89	0.10	-0.40	0
(n=1037)		Treatment	-0.02	0.22	0.40	-1.16	0
	Student Self-Efficacy	Control	4.16	0.65	-1.30	2.70	0
		Treatment	4.16	0.16	-0.1	1.01	3
Level Two	Baseline Reading Score	Control	1316.10	244.48	-1.02	1.90	9
(n=231)		Treatment	1331.98	286.30	-1.82	4.10	10
	Baseline Mathematics Score	Control	1439.02	182.96	-1.34	5.73	6
		Treatment	1447.14	178.59	-.054	1.10	6
	Teacher Emphasis on Science	Control	2.30	0.55	-0.18	0.35	1
		Treatment	2.57	0.44	-0.49	0.91	4
	Teacher Equity	Control	3.60	0.40	-0.23	0.26	3
		Treatment	3.77	0.33	-0.51	-0.98	4
	Teacher Self-Efficacy	Control	3.80	0.41	0.04	0.57	3
		Treatment	3.66	0.43	-0.31	0.08	1
	Teacher Inquiry	Control	2.70	0.40	-0.05	1.34	3
		Treatment	2.77	0.41	-0.78	2.20	3
	Teacher Hands-on High (%)	Control	0.54				
		Treatment	0.63				
	Non_White (%)	Control	0.15				
		Treatment	0.22				
	Black (%)	Control	0.03				
		Treatment	0.05				
	Hispanic (%)	Control	0.09				
		Treatment	0.13				
	Free/Reduced Lunch (%)	Control	0.47				
		Treatment	0.48				
	Gifted (%)	Control	0.04				
		Treatment	0.01				
Level Three	Group (N)	Control	10				
(n=19)		Treatment	9				
	Females (%)	Control	0.50				
		Treatment	0.50				
	Migrant (%)	Control	0.08				
		Treatment	0.11				
	Science FCAT	Control	301.26	6.07	0.75	-0.54	0
		Treatment	298.15	9.14	-0.23	-1.30	0

Table 2. *Multilevel models predicting scaled multiple choice score*

Fixed Effects	Variable	Unconditional	SE	Unconditional Growth	SE	Unconditional Group	SE	Model 14	SE	Final Model	SE
Baseline											
Level-1											
γ_{00}	Intercept	-0.03	0.04	-0.02	0.04	-0.02	0.04	-0.05	0.03	-0.05	0.03
Level-2											
γ_{01}	Gifted							0.60***	0.11	0.60***	0.11
γ_{02}	Reading Score at Basline							0.0004**	0.0001	0.0004**	0.0001
γ_{03}	Math Score at Baseline							0.002***	0.0002	0.002***	0.0002
Level-3											
γ_{001}	Group					-0.01	0.09				
γ_{002}	FCAT Score							0.009**	0.003	0.009**	0.003
Linear Growth											
Level-1											
γ_{10}	Intercept (Time)			-0.06	0.02	-0.07*	0.02	-0.06**	0.02	-0.07***	0.02
γ_{20}	Intercept (Self-Efficacy)							0.22***	0.03	0.22***	0.03
Level-2											
γ_{11}	Teacher Science Emphasis							-0.08*	0.04	-0.09*	0.03
γ_{12}	Non-White Status							-0.15*	0.08	-0.18*	0.07
Level-3											
γ_{101}	Group					0.04	0.05	0.08*	0.03	0.09**	0.03
γ_{102}	% Female							0.22	0.18		
γ_{103}	Migrant Status							-0.66**	0.24	-0.72**	0.22
Fit Indices											
	AIC	8943.89		8934.64		8937.82		8582.89		8582.24	
	BIC	8957.66		8965.62		8975.69		8668.95		8664.86	
	Deviance	8935.89		8916.64		8915.82		8532.89		8534.24	
	Estimated Parameters	4		9		11		25		24	
	Level-2 ICC	0.12									
	Level-3 ICC	0.02									

*Note: *p<.05.**p<.01.***p<.001.*

Table 3: *Random effects for multilevel models*

	Variance Estimate	Unconditional	SD	Unconditional Growth	SD	Unconditional Group	SD	Model 14	SD	Final Model	SD
σ^2	Level-1 variance	0.65	0.81	0.64	0.80	0.64	0.80	0.61	0.78	0.61	0.78
τ_{00}	Intercept Level 2	0.09***	0.30	0.09***	0.31	0.09***	0.31	0.002	0.05	0.003	0.05
τ_{11}	Slope (Time)			0.009	0.10	0.009	0.10	0.002	0.04	0.002	0.04
τ_{01}	Covariance of intercept & slope			-0.003		-0.003		-0.002		-0.002	
τ_{20}	Slope (Self-Efficacy)							0.02**	0.12	0.02**	0.13
τ_{02}	Covariance of intercept & slope							-0.001		-0.001	
U_{00}	Intercept Level 3	0.02***	0.14	0.02***	0.13	0.02***	0.13	0.01***	0.08	0.01***	0.08
U_{10}	Slope (Time)			0.002	0.04	0.002	0.04	0.0004	0.01	0.0002	0.01
U_{01}	Covariance of intercept & slope			0.006		0.004		0.001		0.002	
U_{20}	Slope (Self-Efficacy)							0.003	0.06	0.003	0.06
U_{02}	Covariance of intercept & slope							-0.004		-0.004	

*Note: ***$p<.001$.*

Figure 1. *Average change for scaled multiple choice score in treatment and control schools*

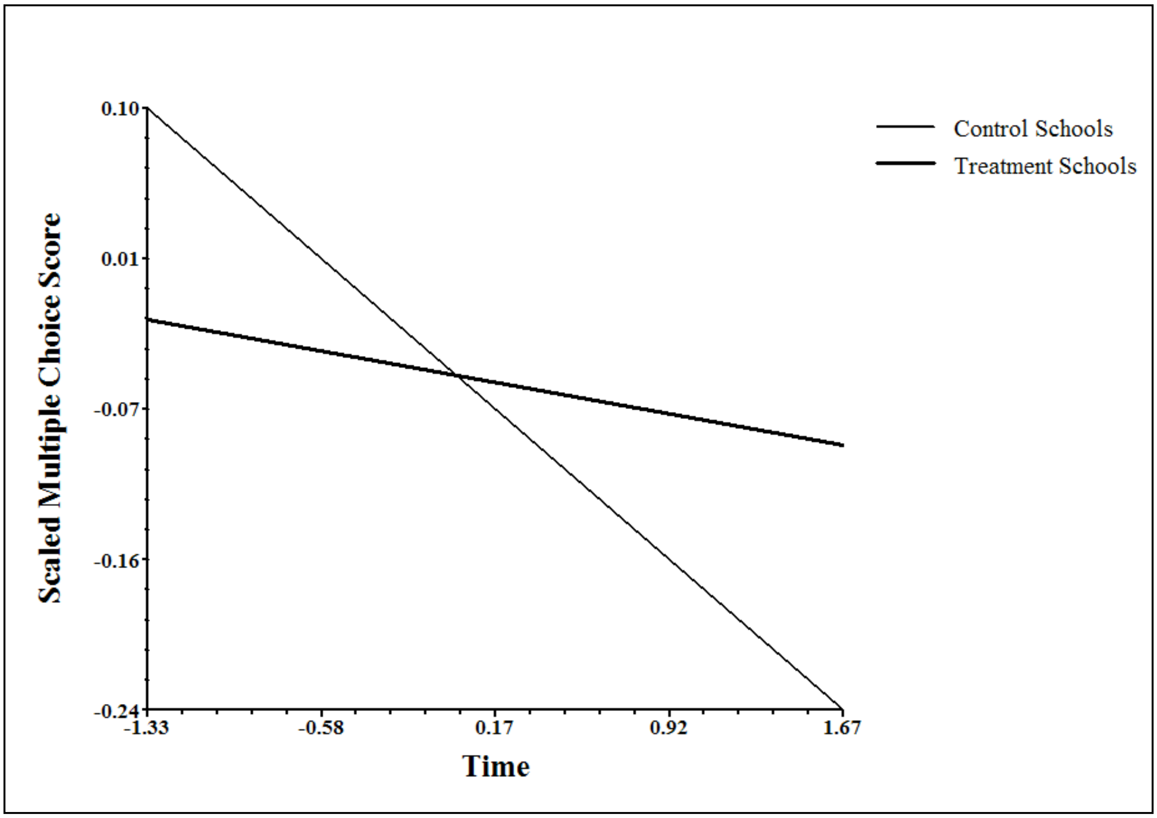

Figure 2. *Average change in scaled multiple choice score for treatment and control schools with low and high percentages of non-white students.*

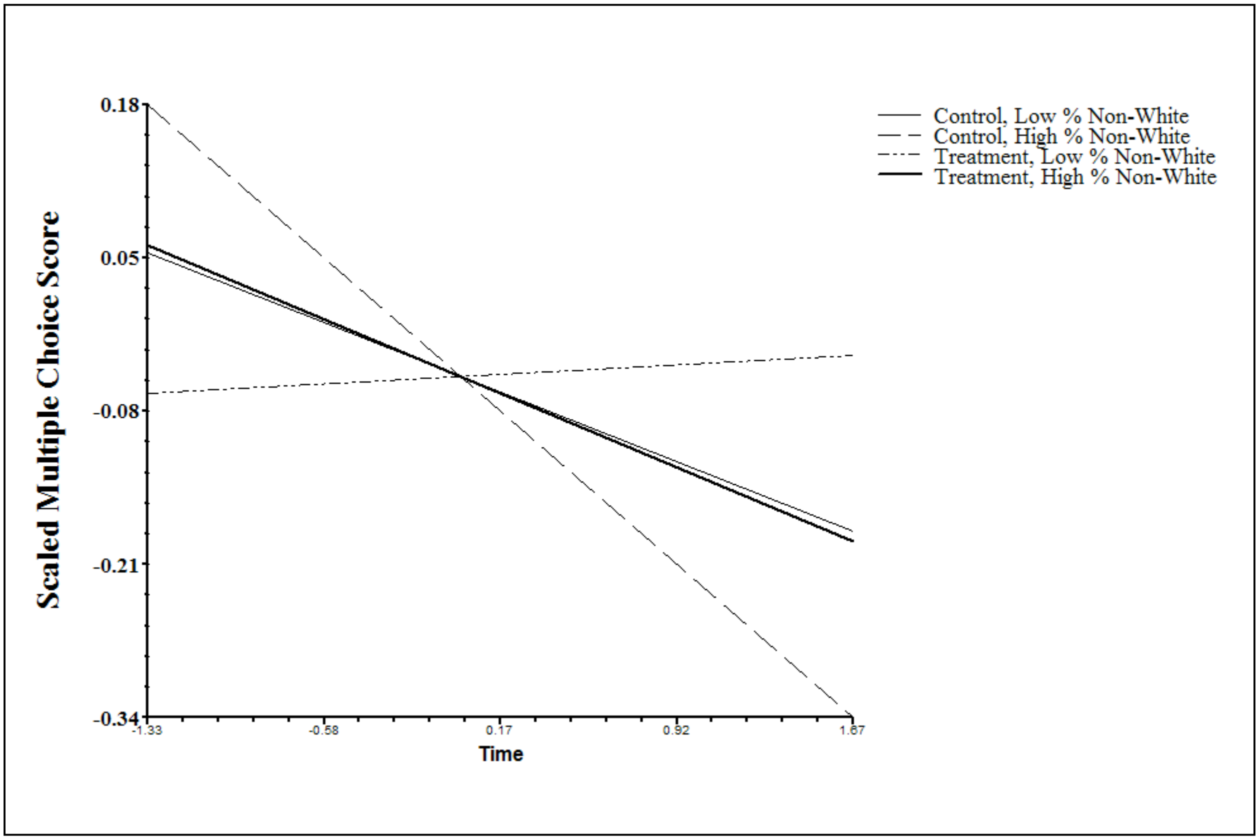

Figure 3. *Average change in scaled multiple choice score for treatment and control schools with a high percentage of migrant students.*

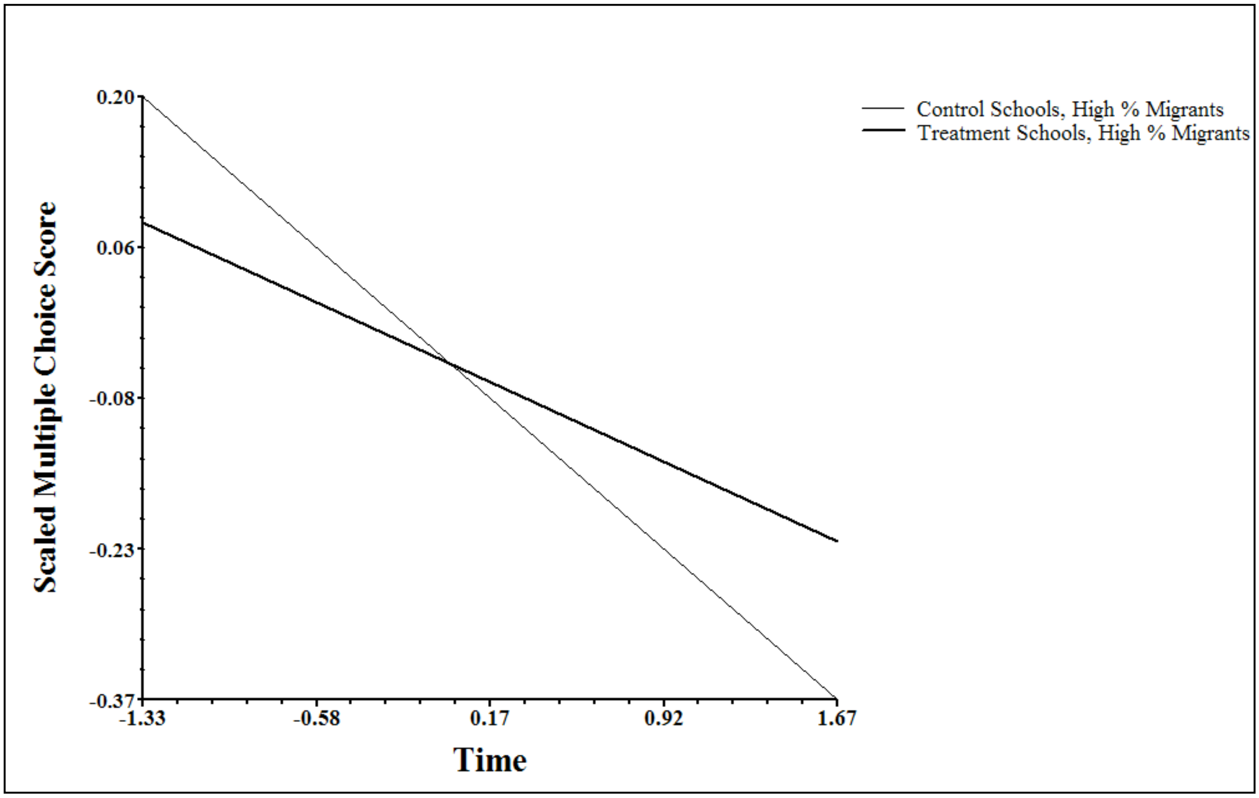

Figure 4. *Average change in scaled multiple choice score for treatment and control schools with low and high percentages of females.*