STEMscopes^{TM} Impact Study 2012-2013

Description of the Study

The purpose of this study was to determine if 5th grade students in classrooms utilizing the STEMscopes^{TM} curriculum were more likely to be proficient in science, as measured by the Texas state science assessment, than students in classrooms that did not use STEMscopes^{TM}.

STEMscopes^{TM} is a comprehensive, online K-12 science curriculum that is 100% aligned to the Texas science standards (the Texas Essential Knowledge and Skills) and that combines online content, activities, and teacher materials with hands-on experiments and explorations. The online component of STEMscopes^{TM} serves as both a support and a guide to teachers, as well as a platform through which students can interact with the material and get feedback on their progress.

STEMscopes^{TM} uses an inquiry-based approach to science, in which the teacher guides students towards the discovery of concepts and skills instead of using explicit direct instruction (Crawford, 2007). The specific way that STEMscopes^{TM} delivers inquiry-based instruction is by building on the Biological Science Curriculum Study's 5E inquiry model (Bybee et al., 2006). The 5E refers to five steps: engagement, exploration, explanation, elaboration, and evaluation. Engagement refers to how teachers activate students' prior knowledge about and interest in a new topic, building connections between what they know and what they are learning. Exploration is the step where students take part in activities and experiments that allow them to experience and learn new concepts and skills. Explanation requires students to explain those new concepts and skills learned during the explore phase. Elaboration challenges them to deepen their conceptual understanding through new, but related, experiences. Finally, in the evaluation phase, students' knowledge is assessed to inform teachers of their progress towards mastery.

The STEMscopes^{TM} pedagogical model adds two key steps: intervention and acceleration to make it the 5E+I/A model. Intervention means that STEMscopes^{TM} provides teachers with the tools both to identify where students are struggling and to provide them with additional opportunities to learn and practice those learning objectives. Acceleration refers to the activities that STEMscopes^{TM} provides for those students that have demonstrated mastery of a particular learning objective. For example, students can undertake a problem-based learning challenge, or connect science to art through a creative project. These two tools help teachers differentiate their instruction and address students' individual learning needs (Zuiker & Whitaker, 2014).

Design and Sample

The STEMscopes^{TM} research team conducted an impact study within elementary schools in one large, urban district in Texas. In the 2012-2013 school year, this district served approximately 100,000 elementary students in 160 elementary schools; these students were predominantly Hispanic (63%) and African American (25%), and a majority (80%) was considered economically disadvantaged. Of these 160 elementary schools, 51 had STEMscopes^{TM} school accounts. Using district data, we examined using the STEMscopes^{TM} curriculum was associated with higher science achievement. Specifically, we conducted two sets of analyses using 5th grade end-of-year STAAR™ science scores: one comparing STEMscopes™ schools to non-STEMscopes™ schools and one within STEMscopes™ schools. All analyses were conducted within a multi-level modeling framework to take into account that observations from the same teacher and school are not independent and to control for confounding factors at each level of analyses: student, teacher, and school (Raudenbush & Bryk, 2002).

In the first set of analyses, a case-matched comparison group was created from the elementary schools in the district that did not use STEMscopes™ through propensity-score matching, thus approximating an experimental design with randomly selected participants (Becker & Ichino, 2002; Rosenbaum & Ruben, 1983). For these analyses, propensity scores were created at the school-level using logistic regression based on the following variables: schools' 2011-2012 STAAR™ science passing rate, whether the school had a new principal, whether the school a magnet program, the number of 5th grade teachers and students, the number of new 5th grade teachers, years of experience of teachers in education, and the composition of the school based on students' sex, race and ethnicity, free or reduced lunch status, limited English proficiency (LEP), bilingual status (BIL), English as a second language (ESL), special education status, and gifted/talented status, as defined by state laws. Schools were selected into the comparison group based on their nearest neighbor match to each STEMscopes™ school. Descriptive statistics for each group are provided in Table 1.

Table 1. Descriptive Statistics for STEMscopes™ Schools and Comparison Schools

	STEMscopes Schools	Comparison Schools
Elmentary Schools	*n* = 51	*n* = 51
New Principal	14%	12%
Magnet Program	31%	35%
Average Number of 5th Grade Teachers	3.2	3.4
Average Number of 5th Grade Students	134.0	139.9
2011-2012 STAAR Science Passing Rate	63.7%	63.2%
5th Grade Teachers	*n* = 163	*n* =171
Average Teacher Experience	10.6 years	11.2 years
Average Teacher Experience in District	8.6 years	9.9 years
Average Number of Students	41.9	41.7
New Teachers	39%	36%
5th Grade Students	*n* = 7,056	*n* = 6,941
Female	49%	49%
White	6%	3%
African American	22%	22%
Hispanic	70%	72%
Asian	2%	2%
Multiracial	1%	<1%
Other Race/Ethnicity	<1%	<1%
Free/Reduced Lunch	46%	52%
Limited English Proficiency (LEP)	47%	52%
Bilingual (BIL)	28%	29%
English as a Second Language (ESL)	2%	3%
Special Education	11%	12%
Gifted/Talented	19%	18%
2012-2013 STAAR Science Raw Score	27.9	27.9

*All schools with STEMscopes accounts matched at school-level

Analyses

A three-level model was analyzed with students at Level 1, teachers at Level 2, and schools at Level 3. All covariates described were included in the analyses. A grouping variable indicating whether the school had a STEMscopes™ account was created (Comparison = 0, STEMscopes™ = 1) and entered as a school-level predictor of science scores. In addition, a variable indicating whether or not teachers had active S STEMscopes™ accounts was entered at the teacher-level as a predictor of science scores. Predictor variables were centered in order to make the coefficients more interpretable. Level 1 and 2 predictors were centered at the group mean in order to create coefficients that were independent from Level 2 and 3 variance, and Level 3 predictors were centered at the grand mean to create Level 3 coefficients that were relative to the average school (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002).

In addition to examining the overall impact of teacher access to STEMscopes™ and student STAAR™ science scores, we examined if access to STEMscopes™ benefitted specific subgroups of students more than similar students without access to STEMscopes™. We used multilevel modeling to examine cross-level interactions between student-level demographic variables and a teacher-level variable that indicated whether or not the teacher had a STEMscopes™ account. Separate models were conducted for each interaction between teacher-level account status and the following student demographics: gender (i.e., female), race/ethnicity (i.e., African American and Hispanic), economic disadvantage (i.e., free or reduced lunch status), LEP, BIL, ESL, Special Education, and Gifted/Talented status. Multilevel regression analyses were conducted to examine if subgroups of students had between STAAR™ science scores in STEMscopes™ classrooms compared to non- STEMscopes™ classrooms, and multilevel logistic regression were conducted to examine if subgroups of students were more likely to pass or score advanced on the STAAR™ science exam when in STEMscopes™ classrooms.

The second set of analyses examined the impact of curriculum *use* on students' science scores. For this analysis, our sample consisted only of schools with STEMscopes™ accounts and of teachers with active accounts. This sample was smaller than the sample of STEMscopes™ schools used in the first analysis (see Table 2 below) because we filtered by active teacher accounts; there were fewer schools (37), teachers (74), and fewer students (4,174).

Table 2: Descriptive Statistics for STEMscopes™ Elementary Schools

	STEMscopes Schools
Elementary Schools	*n* = 37
New Principal	7%
Magnet Program	31%
Average Number of 5th Grade Teachers	3.2
Average Percent of Teachers with Active Accounts	87%
Average Number of 5th Grade Students	139.0
Average Years with Curriculum	2.3
2011-2012 STAAR Science Passing Rate	60.8%
5th Grade Teachers	*n* = 74
Average Teacher Experience	8.8 years
Average Teacher Experience in District	8.4 years
Average Number of Students	57.3
New Teachers	37%
5th Grade Students	*n* = 4,174
Female	48%
White	8%
African American	20%
Hispanic	70%
Asian	2%
Multiracial	1%
Other Race/Ethnicity	<1%
Free/Reduced Lunch	47%
Limited Engish Proficiency (LEP)	46%
Bilingual (BIL)	28%
English as a Second Language (ESL)	2%
Special Education	10%
Gifted/Talented	21%
2012-2013 STAAR Science Raw Score	27.9

We measured use of the curriculum through the teacher user data from the STEMscopes™ website. We included their activity between August 1, 2012 and June 6, 2013. We measured use in four different ways to capture amount of and type of use:

- Time span in days during which the teacher used the STEMscopes™ account
- The total number of page visits to the main 5E+I/A steps
- IIC: An index measuring a teacher's visits to Engage, Explore, Explain relative to visits to the seven principal steps
- DIFF: An index measuring a teacher's use of the differentiation steps (i.e., Intervention and Acceleration) relative to the average number of visits to the seven principal steps

A three-level model was analyzed in a similar way as described above. For this analysis, the user analytics variables were entered as teacher-level predictors of students' science achievement, controlling for all school-, teacher- and student-level covariates. Separate models were analyzed for each user analytics variable because the analytics variables were highly correlated.

Results of the Study

Results for the analyses comparing STEMscopes™ and non- STEMscopes™ schools can be found in Table 3. According to the model, 11% of the variance in student science scores was accounted for by school-level differences, 5% by teacher-level differences, and 84% by student-level differences. We found that teachers who had active STEMscopes™ accounts also had students with higher science scores, controlling for demographic covariates, previous science achievement, and whether or not the school had a STEMscopes™ account. Specifically, teachers with an active STEMscopes™ account had an average student science score 108.26 points higher than teachers without active STEMscopes™ accounts (including teachers in non-STEMscopes™ schools and teachers in STEMscopes™ schools who did not have an active account). The effect size, calculated based on Tymms' (2004) recommendation for determining effect sizes in multilevel models, was 0.24. Though this size is considered 'small' in many contexts (Cohen, 1988), it is worth noting that it is comparable with benchmark effect sizes calculated for similar curricular interventions in 4th-5th grade (Hill, Bloom, Black, & Lipsey, 2008).

Table 3. Results Comparing STEMscopes™ and non- STEMscopes™ Schools

	STAAR Science Scores		
	B	*SE*	*p*-value
School-Level Predictors			
STEMscopes Account	-25.30	37.60	0.503
Number 5th Grade Teachers	-12.96	13.53	0.340
Magnet Program	120.95**	39.57	0.003
New Principal	-46.77	56.06	0.406
Teacher-Level Predictors			
Teacher has STEMscopes Account	108.26*	47.57	0.024
Number 5th Grade Students	1.42***	0.42	< .001
New to Campus	-8.31	27.91	0.766
Years Teaching Experience	-0.95	1.51	0.531
Student-Level Predictors			
Female	-88.29***	6.59	< .001
African American	-262.79***	20.37	< .001
Hispanic	-173.07***	19.40	< .001
Asian	37.79	28.93	0.191
Multiracial	-194.66***	48.39	< .001
Other Race/Ethnicity	-53.29	75.01	0.477
Free or Reduced Lunch	14.02*	7.10	0.048
Limited English Proficiency (LEP)	-32.17**	10.28	0.002
Bilingual (BIL)	-198.81***	10.23	< .001
English as a Second Language (ESL)	-324.66***	26.18	< .001
Special Education	-457.40***	11.55	< .001
Gifted/Talented	420.70***	9.09	< .001

Note. Science scores represent student scale scores (Range = 1,000-6,000). For ethnicity,White is the reference group.
*** $p < .001$, ** $p < .01$, * $p < .05$

Results examining if access to STEMscopes™ benefitted specific subgroups of students showed that two groups of students received additional benefit from being in STEMscopes™ classrooms. Two significant interactions were found in the multilevel logistic analyses. First, female students were less likely to pass the STAAR™ science exam compared to male students ($B = -0.28$, $SE = 0.17$, $p < .001$); however, the cross-level interaction between teacher-level account status and female was significant and positive ($B = 0.45$, $SE = 1.57$, $p = .008$). In other words, female students in STEMscopes™ classrooms were 1.56 times more likely to pass the STAAR™ science exam compared to female students in non- STEMscopes™ classrooms. Second, students with limited English proficiency (LEP status) were less likely to score advanced on the STAAR™ science exam ($B = -0.47$, $SE = 0.11$, $p < .001$); however, the cross-level interaction was significant ($B = 1.18$, $SE = 3.25$, $p = .023$). Therefore, LEP students in STEMscopes™ classrooms were 3.25 times more likely to score advanced in the STAAR™ science exam compared to LEP students in non- STEMscopes™ classrooms.

Finally, results for the analyses within STEMscopes™ schools can be found in Table 4. The percent of variance at each level was identical to the model with the matched sample: 84% of variance was accounted for by school-level differences, 5% by teacher-level differences, and 11% by student-level differences. We found that teachers who used the differentiation steps more often in addition to the 5 principal steps had students with higher scores than teachers who used the differentiation steps less often. Specifically, teachers who used the Intervention and Acceleration steps in addition to their 5E lesson had an average student science score 935.83 points higher. The effect size for this result was 0.36.

Table 4: Results Comparing Levels of Use

	STAAR Science Scores		
	B	*SE*	*p*-value
School-Level Predictors			
Number 5th Grade Teachers	-9.46	23.02	0.684
Magnet Program	142.18*	67.83	0.044
New Principal	83.69	103.51	0.425
Teacher-Level Predictors			
Teacher Use of Differentiation	935.83**	273.77	0.002
Number 5th grade students	0.81	0.95	0.398
New to campus	-47.07	54.51	0.394
Years teaching experience	4.41	2.86	0.133
Student-Level Predictors			
Female	-77.98***	12.10	<0.001
African American	-294.81***	32.98	<0.001
Hispanic	-215.50***	29.86	<0.001
Asian	29.66	47.30	0.531
Multiracial	-140.56*	71.52	0.049
Other Race/Ethnicity	67.50	130.63	0.605
Free or Reduced Lunch	-15.73	13.39	0.240
Limited Engish Proficiency (LEP)	-11.58	19.10	0.544
Bilingual (BIL)	-184.70***	18.73	<0.001
English as a Second Language (ESL)	-416.55***	59.10	<0.001
Special Education	-425.84***	21.61	<0.001
Gifted/Talented	450.87***	16.53	<0.001

Note. Science scores represent student scaled scores (Range = 1,000-6,000). For ethnicity,White is the reference group.
*** $p < .001$, ** $p < .01$, * $p < .05$

Conclusion

Overall, the results of this study provide evidence for the impact of STEMscopes™ on student science achievement in elementary schools in a large, urban district. Teacher who used STEMscopes™ had students with significantly higher STAAR™ scores than teachers who did not use STEMscopes™. This finding is supported by previous research that has found a cumulative effect of inquiry-based science curricula, particularly in urban school districts serving students from low-income (Geier et al., 2008; Lee, Buxton, Lewis, & LeRoy, 2006).

In addition, when in STEMscopes™ classrooms, female students were more likely to receive a passing score on the science exam compared to female students in non- STEMscopes™ classrooms, and students with limited English proficiency were more likely to receive an advanced score. STEMscopes™ is meant to be engaging and accessible to all students, and provides all of its student material in English and Spanish. For these reasons, being in an inquiry-based classroom with lots of hands-on learning may mitigate the negative influence that sex and language proficiency usually have on student science achievement.

When examining how teacher use of STEMscopes™ was related to student scores, we found that teachers who utilized the differentiation steps more frequently in addition the 5E steps had students with higher STAAR™ scores. This finding also is supported by previous research that shows that when teachers provide students with multiple and varied opportunities to learn an objective or concept through differentiation tend have more mastery of content and perform better on state assessments (Subban, 2006).

In the 2012-2013 school year, STEMscopes™ was in its third year of implementation. Given that the curriculum is relatively new, the ability to show significant associations with student science achievement using rigorous statistical controls is very promising. The research team will continue to evaluate the efficacy of STEMscopes™ as well as to examine specifically how STEMscopes™ leads to higher science scores.

References

Becker, S., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal, 2,* 358-377.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 125-230.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-138.

Geier, R., Blumenfield, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching, 45*(8), 922-939.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. *Journal of Research in Science Teaching*, *43*(7), 607-636.

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

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal, 7*(7), 935-947.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London, England: National Foundation for Educational Research.