122nd ASEE Annual
Conference & Exposition
June 14 - 17, 2015
Seattle, WA

Seattle
Making Value for Society

Paper ID #11956

HLM modeling of pre/post-assessment results from a large-scale efficacy study of elementary engineering

Dr. Cathy P. Lachapelle, Museum of Science

Cathy Lachapelle leads the EiE team responsible for assessment and evaluation of our curricula. This includes the design and field-testing of assessment instruments and research on how children use EiE materials. Cathy is particularly interested in how collaborative interaction and scaffolded experiences with disciplinary practices help children learn science, math, and engineering. Her work on other STEM education research projects includes the national Women's Experiences in College Engineering (WECE) study. Cathy received her S.B. in cognitive science from the Massachusetts Institute of Technology and her Ph.D. in educational psychology from Stanford University.

Dr. Yoonkyung Oh, Pennsylvania State University

Yoonkyung Oh is a research associate in the College of Education at Pennsylvania State University. She received her Ph.D. in educational policy from University of Wisconsin-Madison. Her research focuses on investigating family, school, and community as contexts for children's education and development. She is interested in applying experimental, quasi-experimental, and longitudinal research methods to understand the effects of educational practices, policies, and interventions.

Mr. Muhammad Faiz Shams, Museum of Science

Muhammad Shams is a Research Associate who has been with the Museum of Science, Engineering is Elementary team for almost 2 years. He graduated from the University of Massachusetts Dartmouth with a Bachelors of Science in Mathematics. Prior to his time with EiE, Muhammad worked as an undergraduate researcher in the field of applied numerical analysis. Muhammad assists the team with quantitative analysis, paper writing, and database management.

Jonathan D Hertel, EiE/Museum of Science

Jonathan manages the Examining the Efficacy of Engineering is Elementary (E4) project (an NSF-funded study of the efficacy of the EiE curriculum), overseeing and organizing a research effort that involves 240 teachers in the different states. He also provides evaluation support for the Engineering Adventures and Engineering Everywhere projects. He holds an Ed.M. in learning and teaching from the Harvard Graduate School of Education. In 2013-2014, he was named a CADRE Fellow as part of the NSF Community for Advancing Discovery Research in Education (CADRE) project.

Dr. Christine M Cunningham, Museum of Science

Dr. Christine Cunningham is an educational researcher who works to make engineering and science more relevant, accessible, and understandable, especially for underserved and underrepresented populations. A vice president at the Museum of Science, Boston since 2003, she founded and directs Engineering is Elementary[TM], a groundbreaking project that integrates engineering concepts into elementary curriculum and teacher professional development. As of September 2014, EiE has served 6.2 million children nationwide and 71,000 educators. Cunningham has previously served as director of engineering education research at the Tufts University Center for Engineering Educational Outreach, where her work focused on integrating engineering with science, technology, and math in professional development for K-12 teachers. She also directed the Women's Experiences in College Engineering (WECE) project, the first national, longitudinal, large-scale study of the factors that support young women pursuing engineering degrees. Cunningham is a Fellow of the American Society for Engineering Education and was awarded the 2014 International Society for Design and Development in Education Prize. She holds B.A. and M.A. degrees in biology from Yale and a Ph.D. in Science Education from Cornell University.

HLM modeling of Pre/Post-Assessment Results from a Large-Scale Efficacy Study of Elementary Engineering (Evaluation)

Introduction

More than 10 years ago, we began development of an engineering curriculum for elementary school, Engineering is Elementary (EiE). Our goals were, and are, to (1) Increase children's knowledge about engineering technology, and their skills in solving engineering problems; (2) Support teachers to improve their ability to teach engineering and technology; and (3) to advocate for the inclusion of engineering in elementary schools[1].

Engineering is new in elementary school classrooms. Teachers and administrators want to know, before adopting new curricula, what evidence exists that it works. Does a curriculum help students achieve a better understanding of the problems and processes of engineering? Does it support them to become more skilled in engineering and science practices as laid out by the Next Generation Science Standards[2]? Does it contribute to students' understanding of science?

To answer these questions, we began the Exploring the Efficacy of Engineering is Elementary (E4) study of our curriculum to examine the effects of a set of critical curriculum design components on student learning of engineering and science concepts at the elementary level. In this paper, we present an intermediate-stage model of the science outcomes from one of four units examined during this study: the Environmental Engineering unit. In future work, we will examine engineering outcomes as well, and extend the analysis to the other three units examined in the study.

Theoretical Framework

Inquiry and Project-Based Learning

EiE curricular units place the teacher in the role of facilitator, while students have agency to make design decisions. This approach is consistent with inquiry learning, often used in science. Research comparing well-scaffolded inquiry-based methods to direct instruction has shown that students' cognitive gains are comparable both immediately and after a delay, while their attitudes towards science and their procedural skills are improved[3-9]. Geier et al.[10] found that an inquiry curriculum helped to narrow the achievement gap for underserved urban students while improving achievement and engagement.

EiE also meets the criteria for project-based learning (PBL): design projects are central, and focused on questions that engage students with key ideas in science and engineering; these projects and activities are open-ended, largely student-driven, and concerned with realistic problems. Engaging in PBL is challenging for both teachers and students, but when compared to

traditional learning methods, has been shown to lead to improved attitudes, self-efficacy, and learning gains on both traditional subject matter and problem-solving skills[11]. Underrepresented minorities specifically have also been found to benefit in these ways from PBL curricula and training for their teachers[12].

Efficacy Study

The intent of an efficacy study is to test whether an educational innovation, implemented under ideal conditions, has a causal effect on student outcomes[13]. Resources should be ample and fidelity carefully monitored, with teachers given feedback to help them to improve fidelity[14,15]. Efficacy should be tested, if possible, using random-assignment controlled tests (RCT's) because only RCT's and related designs are suitable for determining causality[13]. Researchers must implement randomization at the school or teacher level, not the student level, to prevent bias, and must recruit sufficient subjects to achieve the power to detect potentially small effect sizes[15,16].

In addition to randomization, it is important to establish the validity of independent and dependent variables[15,17], including potential moderators[18]. The treatments must be carefully differentiated from each other, and implementation monitored to ensure they are implemented with fidelity[14,19], because if they are or become too similar in implementation, the study risks Type III errors[15]. On the other hand, researchers must take care to ensure that both treatments address the learning objectives measured in the assessments--otherwise effect sizes may be inflated due to over-alignment of the intervention with outcome measures[16].

Methodology

Study Design

The E4 study employs a Random-assignment Controlled Test (RCT) design to test whether participation in EiE has a causal effect on student outcomes. We randomized our sample at the school level, to avoid cross-contamination of samples within schools. We use Hierarchical Linear Modeling (HLM)[20] to implement an ANCOVA model to measure the effect sizes of treatment and demographic variables. We choose to implement hierarchical models because our data is hierarchically organized, with students nested within classrooms within schools, and a multilevel analysis provides more accurate estimation of the effects of variables at different levels of hierarchy than other analytic methods. We use pre-assessment class means as covariates to reduce intercluster correlation, as this is a more efficient way to improve power than a matched pairs design[21].

In this paper, we examine the following question from our proposal:

1. What effect does the EiE curriculum have, when implemented under ideal conditions, on children's learning of engineering and science content and processes compared to lessons

that address similar learning objectives but do not include EiE's hypothesized "critical components" (see below)?

In designing the EiE curriculum, we adhered to key critical components that we believe are essential for optimal learning by all students. Critical components include that:

(a) engineering content is introduced in a context
(b) students learn about and use the engineering design process
(c) engineering challenges specify a challenge and constraints and permit many possible solutions
(d) children use math and science to design solutions
(e) children use failure constructively and design iteratively
(f) students work collaboratively.

These critical components are congruent with the definitions of both inquiry learning and project-based learning. We developed a rubric to measure the presence of critical components in curriculum materials; this rubric was employed by an independent researcher to independently verify that the critical components are strongly represented in the EiE curriculum units, while they are only weakly represented in the comparison curriculum units developed for the E4 study.

Instrument Development Process

Engineering instruments suitable for elementary school students are only beginning to be published in the literature. Therefore, we have invested in the development of our own instruments and instrument development process. This process begins with (1) the specification of learning objectives to be tested; (2) development of a large set of candidate questions for each learning objective; (3) analysis of questions to ensure they do not violate basic principles of item development as specified by Taylor and Smith[22]; (4) validity testing of revised items using a talk-aloud protocol for cognitive interviews with the target population of students; (5) revision and further validity testing as necessary; (6) reliability testing, IRT analysis, determining items to be dropped using PCA, finalizing of scales, and further revision as necessary. In-depth description of our process can be found in Lachapelle & Cunningham[23].

Materials Development Process

The E4 project is examining the implementation and effect of four out of 20 EiE units, as testing all 20 would require considerably more resources than available. We selected the four EiE units to represent a range of types of science disciplines—physical, life, and earth sciences—and a range of engineering fields (see Table 1). Each of the four units is designed to guide students to apply relevant science content in the process of completing their designs.

In addition, we have included a fifth pair of units addressing the civil engineering of structures to be implemented by half of participating teachers as an additional "dose" of engineering. The content of the additional unit is not assessed.

The five comparison *Engineering For Children* (E4C) curriculum units are intended to match engineering content and process learning objectives with the chosen EiE units but to differ from those units, inasmuch as possible, in critical components. These were chosen to represent a more "traditional" pedagogy, with more direct instruction. To represent a more "business as usual" approach to engineering, we started with lessons freely available on the web, and modified the lessons until the comparison curriculum could address the same learning objectives and engage children in engineering for the same amount of time. E4C units are weak in the critical components listed above:

(a) the challenge is presented with little or no context
(b) no engineering design process is formally presented, though from time to time steps are mentioned
(c) engineering challenges may not specify either a goal and/or constraints
(d) math and science are not incorporated into lesson design
(e) opportunities to analyze results and make improvements are not part of the lessons
(f) students sometimes work independently, often competitively, and sometimes in teams

To gauge adherence to (or in the case of the comparison curriculum, differentiation from) our critical components, we developed a rubric of adherence to the critical components. An outside reviewer judged all 10 units against the rubric. For cases of low adherence for the EiE curriculum, or high adherence for the E4C curriculum, we modified the curriculum materials.

Table 1: Intervention and Comparison Curricular Materials

Engineering Field	EiE® Units	Related Science
Package Engineering	Thinking Inside the Box: Designing Plant Packages	Plants: structures and functions; survival needs
Geotechnical Engineering	A Stick in the Mud: Evaluating a Landscape	Landforms and erosion
Environmental Engineering	A Slick Solution: Cleaning an Oil Spill	Ecosystems: reading and analyzing food webs.
Electrical Engineering	An Alarming Idea: Designing Alarm Circuits	Electricity: conductors and insulators; properties of electricity; forms of energy
Civil Engineering	To Get to the Other Side: Designing Bridges	N/A

Procedure

Materials: Intervention and Comparison Curriculum

Data analyzed for this paper is from the Environmental Engineering pair of units. The engineering learning objectives for these units include that students can demonstrate (1) an understanding of environmental engineering and the role of environmental engineers in cleaning up pollution; (2) an understanding of the nature of pollution and different ways that pollution can spread, particularly via rain and the action of water moving above and underground; and (3) that they can communicate their ideas through drawings, justify design decisions, and understand testing as an important step in engineering design. In addition, the EiE Environmental Engineering unit is designed to reinforce the following learning objectives that should have been previously addressed in science: that students can demonstrate (1) an ability to read food web and food chain diagrams; (2) an understanding of the major components of ecosystems (producers, consumers, decomposers, and the physical environment) and how they are interdependent; and (3) an understanding of how a change in one part of an ecosystem can be related to other changes.

The EiE Environmental Engineering unit consists of four lessons. In the first lesson, students are introduced to the design challenge and the engineering field through a fictional story about a girl from the Elwha tribe in Washington State, who finds an oil spill on the river, reports it to authorities, and learns about oil spill clean-up methods from an environmental engineer as she provides documentation of the prior state of the river and helps with the wildlife rescue efforts. In the second lesson students learn more about environmental engineering by exploring a fictional challenge to identify the source of pollution in a town by reading about recent changes in different areas and taking "pH" measurements. They discuss the effects of pollution on an ecosystem in the third lesson; they also test common household materials for their ability to contain, absorb, or remove oil. In the fourth lesson, they design and test a plan to clean up an oil spill in a tub of water, using what they've already learned.

For the matching E4C unit, we combined multiple lessons so as to match time spent on the lessons across treatment conditions. We revised lessons selected for inclusion in E4C to be age-appropriate and share a similar format across E4C lessons and units. We also designed a few lessons to address learning objectives for which we could not find engineering lessons available on the web. In lesson 1, students learn about pollution and how pollution moves through the environment by observing a model. In lesson 2, students read about environmental engineers and draw pictures of environmental engineers. In lesson 3, students design and build water filters to clean muddy, "polluted" water. In lesson 4, students explore a fictional challenge to identify the source of pollution in a town by reading about recent changes in different areas and taking "pH" measurements—this lesson is adapted from the EiE lesson, with changes to pedagogy and materials so that critical components are not as well represented. In lesson 5, students test different materials to use as booms to contain an oil spill. Finally, in lesson 6, students design

and test a plan to clean up an oil spill in a tub of water—this lesson, again, is adapted from the EiE version described above.

Instruments

A large number of assessments, surveys, observations, and video data were collected; those that inform this paper are the student pre-/post-assessments that probe students' knowledge of science concepts. The science assessment for the Environmental Engineering unit includes 14 questions addressing the three learning objectives listed above (interpretation of food webs, components of ecosystems, and the implications of changes in ecosystems). The COs instrument is available in Appendix A.

We recoded student responses to each question on the assessment as correct (1) or incorrect (0). We separately checked the initial pre- and post-assessment scores for internal reliability and factorability by means of reliability analysis and principal component analysis with direct oblimin rotation in SPSS v22.0. Where scales demonstrated sufficient internal reliability, we used principal component analysis to identify which questions could or should be dropped. Item 3 was removed to improve reliability. Details of the reliability analysis are presented in Table 2 below.

Table 2: COs Instrument Reliability

Scale	N of Students	# Items	Items	Reliability: Cronbach's α
Reading Food Webs	333	7	1, 2, 4-6, 7a, 7e	0.692
Analyzing Food Webs	333	6	7b - 7d, 8a - 8c	0.775

We completed further data cleaning after analysis of reliability of scores, but before beginning analysis. Any student who had missed more than twenty percent of the questions on either the pre- or post-assessment was excluded from the dataset. All remaining missing responses were replaced with a zero – a coding of "incorrect." We also excluded any students who were missing ethnic and gender demographic information.

Sample Selection and Randomization

To recruit teachers, we sent mail and email invitations in January of 2013 to principals of schools in Massachusetts, Maryland, and North Carolina, inviting them to put up flyers and spread the word with teachers who may be interested in participating in an engineering education research study. We promised teachers free professional development, materials, and a stipend for implementing an assigned engineering unit and collecting data from their students. We

encouraged teachers to apply as "school teams" as our goal was to have two teachers from each school participate. We collected applications and principals' letters of support online.

We considered applications only from generalist or science specialist teachers of grades 3, 4, or 5. Applications were rejected if the teacher had already taught the intervention curriculum or a similar curriculum, or if a colleague in the same school was teaching EiE. We accepted school teams of non-consecutive grades (all one grade, or grades 3 & 5) wherever possible, with some of those from consecutive-grade teams rejected, so that students would not participate in the study over two consecutive years. Ineligible teachers made up the majority of those not accepted to the study.

Of the 613 applications we received, we either accepted or placed on a wait list 359 teachers from 231 schools, with 131 of the schools from North Carolina, and the remainder split between Massachusetts (52 schools) and Maryland (48 schools). Ten teachers from 10 schools in North Carolina were placed on a waitlist. For each of the three states, we listed the accepted and waitlisted schools alphabetically in Microsoft Excel, with each school assigned a random number between 0 and 1 using the RAND function. The schools were then sorted by random number, and assigned alternately to intervention (EiE) and comparison (E4C) groups in order of their random assignment.

Teachers were assigned to one of the four units according to what science they taught. If teachers taught more than one of the target science units (Plants, Ecosystems, Electricity, or Landforms) they were assigned to the less-popular of the science units.

We then contacted teachers with news of their acceptance, waitlist, or rejection status by email. Any teacher who did not reply with confirmation of his or her participation was replaced from the waitlist by a teacher from the appropriate treatment group.

Teachers were invited to volunteer to implement the Civil Engineering unit in addition to their assigned unit. As 252 teachers, most of the sample, expressed interest, we randomly assigned those interested so that half of each treatment sample includes the second "dose" of engineering.

We performed an analysis of the equivalence of the two samples (intervention and comparison) at the school and classroom levels at the start of the study to determine the effectiveness of randomization. In the case of categorical variables we performed a chi-squared analysis to determine significance between samples. We observed a number of different variables, all showing there was no significant difference between the two samples, indicating that they are indeed equivalent. Details of the analyses are presented in Tables 3 and 4 below.

Table 3: School Level Variables Analyzed for Equivalence

	Total (N/%)		Intervention (N/%)		Comparison (N/%)		Chi Squared p-value
School Setting							0.271
City	8	13.6%	6	19.3%	2	7.1%	
Rural	24	40.7%	14	45.2%	10	35.7%	
Suburb	23	39.0%	10	32.3%	13	46.4%	
Town	4	6.8%	1	3.2%	3	10.7%	
School Location							0.799
MA	10	16.9%	5	16.1%	5	17.9%	
MD	15	25.4%	9	29.0%	6	21.4%	
NC	34	57.6%	17	54.8%	17	60.7%	
Title I Eligibility							0.697
Ineligible	14	24.6%	8	26.7%	6	22.2%	
Eligible	43	75.4%	22	73.3%	21	77.8%	

Table 4: Classroom Level Variables Analyzed for Equivalence

	Total (N/%)		Intervention (N/%)		Comparison (N/%)		Chi Squared p-value
"Bridges" Second Dosage							0.133
No	60	52.6%	26	45.6%	34	59.6%	
Yes	54	47.4%	31	54.4%	23	40.4%	
Grade Level							0.588
3rd	17	14.9%	9	15.8%	8	14.0%	
4th	18	15.8%	7	12.3%	11	19.3%	
5th	79	69.3%	41	71.9%	38	66.7%	

Teacher Preparation

All teachers in the study participated in 3-day professional development workshops during the summer of 2013. Separate workshops for intervention and comparison curricula were held in Massachusetts, Maryland, and North Carolina. Members of the EiE project team conducted professional development with the assistance of E4 staff and state coordinators. After being introduced to the subject of engineering (with which many had not had significant contact), teachers engaged in hands-on training for their assigned engineering unit as well as a second unit in order to increase exposure to the curriculum. Throughout the workshop, professional development staff modeled curriculum-specific pedagogy for teachers by placing them in the

role of students while engaging in the activities. Staff also helped participants to reflect as teachers on the classroom management and logistical necessities of the units. In addition, the workshop focused heavily on the data collection requirements of the E4 study for which teachers were responsible. Continued progress in the study was contingent upon participation in these workshops.

Data Collection

Teachers' participation during the school year included three phases:

1. Collect pre-assessment and demographic data
2. Complete all related instruction (related science, the optional Civil Engineering unit, and the assigned Environmental Engineering unit)
3. Collect post-assessments and attitudes survey

Each teacher is implementing one of the eight "core" units for two consecutive years; half of teachers additionally implement the Civil Engineering unit. All teachers must implement the related science content, using a curriculum of their choice. After completing the pre-assessment, teachers must implement the related science content, as well as the bonus civil engineering unit if assigned to do so, before implementing their assigned unit. Once all required units are complete, they must have students do the post-assessments.

The presented data was collected during the first year of implementation. We expect the second year of implementation to show stronger effects as teachers learn improved implementation and data collection practices.

Sample Demographics

Initially, we began the study with 114 classrooms in the Environmental Engineering condition, taught by 99 teachers in 59 schools. This number had a certain attrition rate as the first year of the study progressed. Reasoning for drops included: teachers electing to teach fewer classes than originally intended, teacher drops from the study, and unusable student or classroom data during the data cleaning process. For analysis purposes we had 1658 students within 91 classrooms, taught by 68 teachers in 48 schools. Details are presented in Table 5 below.

Table 5: N of Schools, Teachers, Classrooms, and Students in Environmental Engineering

	Start of Study	Intervention	Comparison	End of Year One	Intervention	Comparison
Schools	59	31	28	48	26	22
Teachers	99	55	44	68	34	34
Classrooms	114	57	57	91	45	46
Students*	N/A	N/A	N/A	1658	816	842

*Number of Students at the Start of Study is N/A as teachers would not have been assigned classrooms at that point.

Student level demographics collected for this analysis included gender and racial ethnicity. Ethnicity was further collapsed into the Racial Representation variable, with White and Asian making up the "represented" ethnicities within the field of engineering and Black, Hispanic, and Other making up the "underrepresented" ethnicities. (See Table 6).

Table 6: Student Sample Demographics

Student Demographic		Proportion: Intervention	N: Intervention	Proportion: Comparison	N: Comparison
Gender (Male))		45.8%	374	49.6%	418
Gender (Female)		54.2%	442	50.4%	424
Race/Ethnicity	White	67.9%	554	62.9%	530
	Asian	5.6%	46	2.5%	21
	Black*	10.2%	83	19.5%	164
	Hispanic*	12.4%	101	10.1%	85
	Other*	3.9%	32	5.0%	42

*Denotes underrepresented races/ethnicities.

Analysis and Results

Preliminary Analyses

As preliminary analyses, we conducted a series of independent sample t-tests to examine the group mean differences in each of pretest and posttest scores. We also conducted the analysis of variance (ANOVA) to examine the interaction of time and conditions, i.e., whether the change in test scores from pretest to posttest is significantly different between the intervention and the comparison groups.

The t-test results indicated that, for both outcomes (i.e., Reading Food Webs and Analyzing Food Webs), the pretest score was not significantly different between the groups, but the posttest score was significantly higher in the intervention group. For the Reading Food Webs scale, the mean pretest score was 4.95 (SD=1.63) for the intervention group and 4.80 (SD=1.65) for the comparison group. The mean posttest score was 6.03 (SD=1.19) and 5.75 (SD=1.43) for the

intervention and the comparison group, respectively. The pretest score mean was not significantly different between groups ($t=1.871$, $p=0.061$), but the posttest score mean was significantly higher for the intervention group than for the comparison group ($t=4.368$, $p<.001$). The results were similar for the Analyzing Food Webs scale. The mean pretest and posttest scores were 3.69 (SD=1.86) and 4.75 (SD=1.63), respectively, for the intervention group, and 3.55 (SD=1.89) and 4.44 (SD=1.76), respectively, for the comparison group. The group difference was not statistically significant in the pretest scores ($t=1.48$, $p=.138$), but significant in the posttest scores ($t=3.73$, $p<.001$).

The intervention group showed greater increases in test scores from pretest to posttest than did the comparison group, as graphically shown in Figures 1 and 2. However, the results of the treatment status-by-time interaction effect in ANOVA indicated that there were no significant interactions of time by treatment status (F=1.467, p=0.226 for Reading Food Webs; F=2.040, p=0.153 for Analyzing Food Webs).
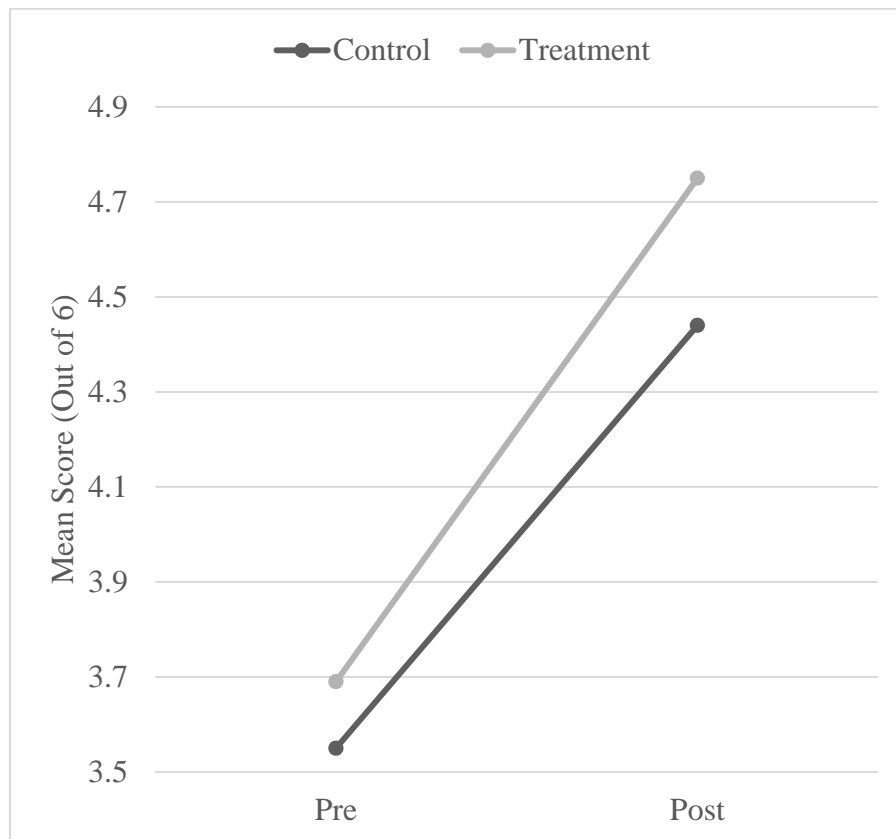
Figure 1: Reading Food Webs ANOVA
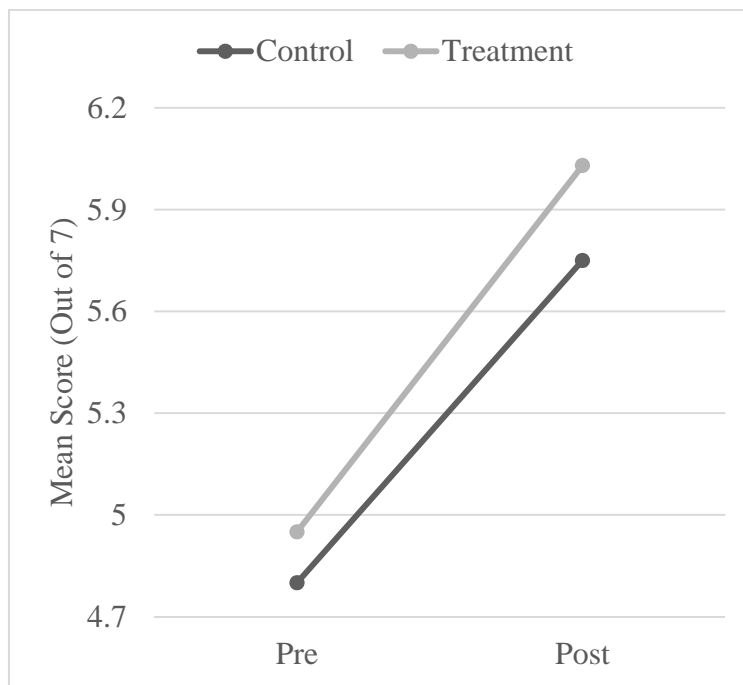
Figure 2: Analyzing Food Webs ANOVA

Main Analyses

This study was a randomized controlled trial, with schools as the unit of random assignment and students as the unit of collection of outcome data. Because students were nested within teachers or classrooms, which in turn nested within schools, we employed hierarchical linear modeling [24] to take into account the nested structure of the data. By specifying a three-level hierarchical linear model, we were able to simultaneously estimate variation associated with student (level 1), teacher/classroom (level 2), and school (level 3) levels, and analyze the impact of our engineering curriculum intervention on student test scores.

We first ran the unconditional model with no covariates, and then included student background variables (i.e., gender, race, and pretest score), teacher or classroom characteristics (i.e., grade level and teaching bridge unit), and school level characteristics (i.e., school location in addition to the treatment status). We estimated the unconditional model to determine the intraclass correlation coefficient (ICC), i.e., the proportion of variance in posttest scores that occurred between schools. We added a set of student-, classroom-, and school-level variables in the final model because of the potential to reduce residual variance in the outcome variables and to increase statistical power and precision of the estimate of treatment effect [25]. The fully specified level-1 model can be written as:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(pretest) + \pi_{2jk}(male) + \pi_{3jk}(underrepresented) + e_{ijk}$$

This formula represents the posttest score for student i in classroom j in school k (Y_{ijk}) modeled as a function of the intercept for each classroom j in school (π_{0jk}), three student-level predictors, and a random student-level error (e_{ijk}). In Level 2, the coefficients from the level-1 model were further modeled as below.

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(Bridge) + \beta_{02k}(Grade4) + \beta_{03k}(Grade5) + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k}$$
$$\pi_{2jk} = \beta_{20k}$$
$$\pi_{3jk} = \beta_{30k}$$

We treated pretest score, gender and race as fixed effects and only the intercept as a random effect. The intercept from the level-1 model, that is, the mean test score of classroom j in school k (π_{0jk}), was modeled as a function of the intercept for each school k (β_{00k}), two dummy-coded classroom grade variables (with Grade 3 as a reference category), a dummy variable indicating whether the class taught a Civil Engineering (Bridge) unit, and a random classroom-level error (r_{0jk}).

At Level 3 of the model, we estimated the school-level effect of the curricular intervention on the mean posttest score in school (β_{00k}), controlling for two dummy variables indicating the location of school (i.e., Maryland and North Carolina with Massachusetts as a reference category). The level-3 equation is as below.

$$\beta_{00k} = r_{000} + r_{001}(Treatment) + r_{002}(MD) + r_{003}(NC) + u_{ook}$$

where r_{000} is the grand mean of the outcome and u_{ook} is a random school-level error. All the coefficients from the level-2 model except β_{00k} were treated as fixed effects.

The results of the analyses of unconditional models indicated that about 11% of the variance in the Reading Food Webs scores resided between schools and 7% of variance in the Analyzing Food Webs scores occurred between schools. The results of the analyses of final models are presented in Table 7. As shown in Table 2, the intervention-comparison group difference in the Reading Food Webs scale scores was close to significance at the 0.05 level controlling for pretest scores and other student-, classroom- and school-level covariates. The unstandardized coefficient of treatment status suggests that the EiE curriculum leads to about a 0.2 point increase in Reading Food Webs scores, on average. The result also indicates higher average scores of Analyzing Food Webs in intervention schools than comparison schools, above and beyond basic student-, classroom-, and school-level characteristics. The estimate of treatment effect suggests that the EiE curriculum intervention leads to about 0.22 point increase in Analyzing Food Webs

scores, on average. However, this difference was not statistically significant. The results also show that only two covariates were statistically significant predictors of the test scores: pretest scores and student race/ethnicity.

Table 7. Multilevel Analysis of Posttest Scores

	Reading Food Webs				Analyzing Food Webs			
	Estimate	SE	T	p-value	Estimate	SE	T	p-value
Intercept	5.75	0.20	28.47	<0.001	4.63	0.29	15.81	<0.001
Treatment	0.20	0.10	1.94	0.059	0.22	0.13	1.60	0.116
Pretest Score	0.35	0.03	13.76	<0.001	0.42	0.02	17.59	<0.001
Bridges	-0.04	0.10	-0.39	0.698	0.09	0.14	0.66	0.515
Male	0.07	0.06	1.36	0.173	0.05	0.07	0.70	0.484
Underrepresented	-0.16	0.07	-2.25	0.025	-0.37	0.09	-4.12	<0.001
Grade 4	-0.20	0.22	-0.94	0.354	-0.21	0.25	-0.82	0.42
Grade 5	0.03	0.17	0.18	0.859	-0.14	0.22	-0.62	0.542
School State MD	0.13	0.18	0.74	0.461	0.09	0.24	0.36	0.72
School State NC	0.07	0.14	0.51	0.616	-0.01	0.25	-0.05	0.962

Discussion, Implications, and Next Steps

For this initial foray into modeling treatment effects on student outcomes, we have found only marginal effects of the intervention on students' science scores for reading food webs—our first analysis does not establish that students learned the science concepts better with the intervention curriculum than the comparison curriculum. Pretest and student race/ethnicity were the strongest predictors of posttest scores, suggesting the need for strong science and engineering interventions for underrepresented minority students.

However, we have only begun our investigations and modeling for the E4 study. We continue to add variables to the HLM models that reduce variance and increase power. A vital next step is to add variables accounting for teacher fidelity, as this can have a major impact on intervention and comparison curriculum implementation, causing differences between the curricula to shrink in practice, and lessening the power of the study. We will also add interaction terms to our models to examine the moderating effects of demographic variables on the outcomes.

In addition, we will model the engineering student outcomes for the Environmental Engineering units, and we will extend our analyses to the other three pairs of engineering units that comprise the E4 study.

Acknowledgements

References

[1] C. M. Cunningham and K. Hester, "Engineering is Elementary: An engineering and technology curriculum for children," in *ASEE Annual Conference & Exposition*, Honolulu, HI, 2007, p. 17.

[2] NGSS Lead States, *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press, 2013.

[3] R. Anderson, "Reforming science teaching: What research says about inquiry," *J. Sci. Teach. Educ.*, vol. 13, no. 1, pp. 1–12, 2002.

[4] M. R. Blanchard, S. A. Southerland, J. W. Osborn, V. D. Sampson, L. A. Annetta, and E. M. Granger, "Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and traditional verification laboratory instruction," *Sci. Educ.*, vol. 94, no. 4, pp. 577–616, 2010.

[5] D. Kuhn, "Reasoning about multiple variables: Control of variables is not the only challenge," *Sci. Educ.*, vol. 91, no. 5, pp. 710–726, 2007.

[6] R. Lehrer and L. Schauble, *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York, NY: Teachers College Press, 2002.

[7] R. Lehrer and L. Schauble, "Cultivating model-based reasoning in science education," in *Cambridge handbook of the learning sciences*, R. K. Sawyer, Ed. Cambridge, UK: Cambridge University Press, 2006, pp. 371–388.

[8] M. Prince and R. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases," *J. Eng. Educ.*, vol. 95, pp. 123–138, 2006.

[9] W. A. Sandoval and A. Çam, "Elementary children's judgments of the epistemic status of sources of justification," *Sci. Educ.*, vol. 95, no. 3, pp. 383–408, 2011.

[10] R. GEiEr, P. C. Blumenfeld, R. W. Marx, J. S. Krajcik, B. Fishman, E. Soloway, and J. Clay-Chambers, "Standardized test outcomes for students engaged in inquiry based science curricula in the context of urban reform," *J. Res. Sci. Teach.*, vol. 45, no. 8, pp. 922–939, 2008.

[11] J. Thomas, "A review of research on project-based learning," Autodesk Foundation, 2000, 2000.

[12] D. E. Kanter and S. Konstantopoulos, "The impact of a project based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry based practices," *Sci. Educ.*, vol. 94, pp. 855–887, 2010.

[13] L. Towne, L. L. Wise, and T. M. Winters, *Advancing scientific research in education*. Washington, DC: National Academies Press, 2005.

[14] C. L. O'Donnell, "Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research," *Rev. Educ. Res.*, vol. 78, no. 1, pp. 33–84, 2008.

[15] S. W. Raudenbush, "Designing field trials of educational innovations," in *Scale-up in education: Issues in practice*, vol. 2, B. Schneider and S.-K. McDonald, Eds. Lanham, MD: Rowman & Littlefield Publishers, Inc., 2007, pp. 23–40.

[16] R. E. Slavin, "What works? Issues in synthesizing educational program evaluations," *Educ. Res.*, vol. 37, no. 1, pp. 5–14, 2008.

[17] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth, 2002.

[18] C. Carroll, M. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain, "A conceptual framework for implementation fidelity," *Implement. Sci.*, vol. 2, no. 1, pp. 40–40, 2007.

[19] G. D. Borman, "Experiments for educational evaluation and improvement," *Peabody J. Educ.*, vol. 77, no. 4, pp. 7–27, 2002.

[20] A. S. Bryk and S. W. Raudenbush, *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage, 1992.

[21] S. W. Raudenbush, A. Martinez, and J. Spybrook, "Strategies for improving precision in group-randomized experiments," *Educ. Eval. Policy Anal.*, vol. 29, no. 1, pp. 5–29, 2007.

[22] M. Taylor and S. Smith, "How Do You Know if They're Getting It? Writing Assessment Items that Reveal Student Understanding," *Sci. Scope*, vol. 32, no. 5, pp. 64, 60–64, 60, 2009.

[23] C. P. Lachapelle and C. M. Cunningham, "Assessing elementary students' understanding of engineering and technology concepts," in *ASEE Annual Conference and Exposition*, Louisville, KY, 2010.

[24] S. W. Raudenbush and A. S. Bryk, *Hierarchical linear models: Application and data analysis methods.* Thousand Oaks, CA: Sage Publications, Inc., 2002.

[25] H. S. Bloom, J. M. Bos, and S.-W. Lee, "Using Cluster Random Assignment to Measure Program Impacts Statistical Implications for the Evaluation of Education Programs," *Eval. Rev.*, vol. 23, no. 4, pp. 445–469, 1999.

Appendix A: COs Instrument

1. The diagram below shows a food chain. Which of the following is a producer?

Grass → Grasshoppers → Birds → Snakes

● grass   ⑧ grasshoppers   ⓒ birds   ⑩ snakes
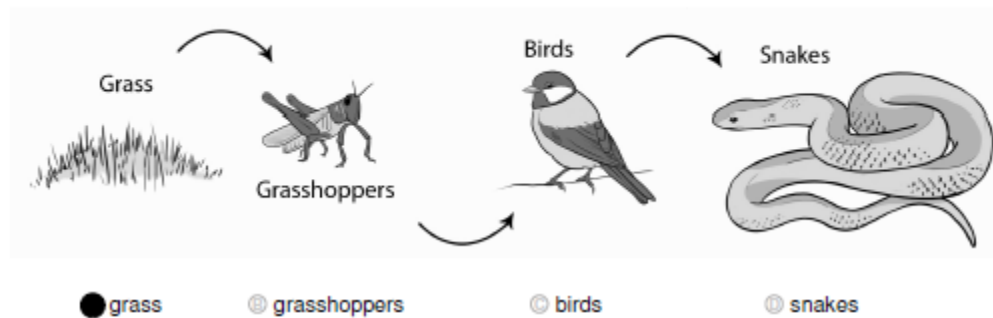
2. Which of the following events involves a consumer and a producer in a food chain?

● a deer eats a leaf

⑧ a snake eats a rat

ⓒ a cat eats a mouse

⑩ a hawk eats a mouse

The diagram below shows part of a food web for a meadow ecosystem.
Use it to answer questions 3 - 8.
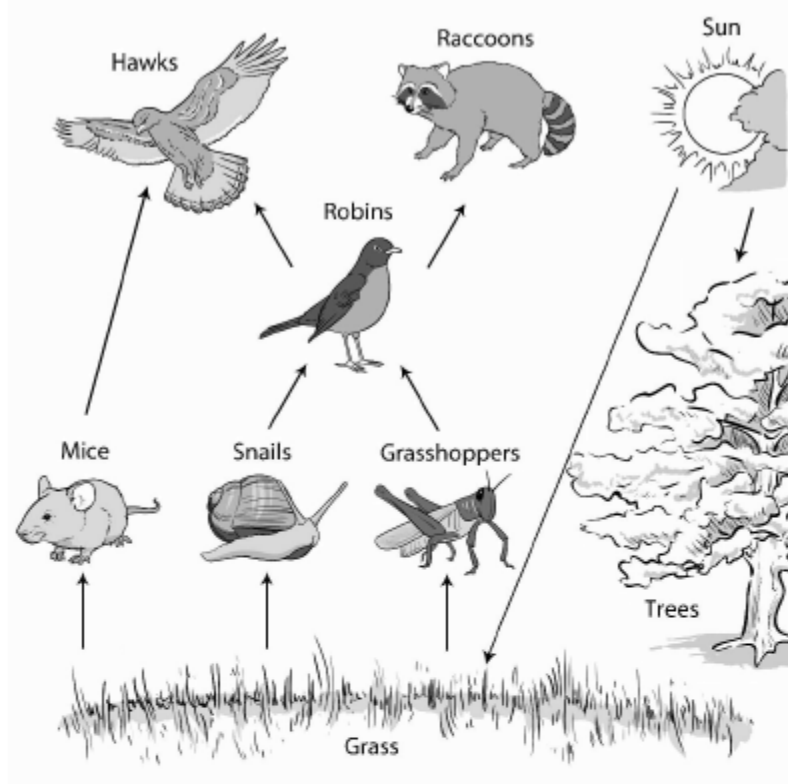
3. **What do the arrows in the food web diagram show?**

   Ⓐ what each animal is eaten by

   Ⓑ what each animal will eat this week

   Ⓒ what is most important in the ecosystem

   ● where each thing in the ecosystem gets its energy

4. **What does the arrow from the grasshopper to the robin in the food web diagram mean?**

   ● that robins get energy from grasshoppers

   Ⓑ that grasshoppers get energy from robins

   Ⓒ that grasshoppers and robins live next to one another

   Ⓓ that robins are more important than grasshoppers in the ecosystem

Use the diagram on the previous page to answer questions 5 - 8.

5. Which of the following lists all of the consumers in this food web diagram?

   Ⓐ grass and trees

   Ⓑ hawks, robins, and raccoons

   Ⓒ mice, snails, and grasshoppers

   ⬤ mice, snails, grasshoppers, hawks, robins, and raccoons

6. Which of the following is <u>NOT</u> a food chain found in this food web diagram?

   Ⓐ grass-mice-hawks

   ⬤ grass-mice-raccoons

   Ⓒ snails-robins-raccoons

   Ⓓ grass-grasshoppers-robins-hawks

7. If the grass in the meadow stopped growing, what would probably happen in the meadow ecosystem?

a. There would be...	⬤ fewer	○ the same number of	○ more	...snails.
b. There would be...	⬤ fewer	○ the same number of	○ more	...robins.
c. There would be...	⬤ fewer	○ the same number of	○ more	...hawks.
d. There would be...	⬤ fewer	○ the same number of	○ more	...raccoons.
e. There would be...	⬤ fewer	○ the same number of	○ more	...grasshoppers.

8. If a disease caused most of the robins to die, what would probably happen in the meadow ecosystem?

a. There would be...	○ fewer	○ the same number of	⬤ more	...snails.
b. There would be...	⬤ fewer	○ the same number of	○ more	...raccoons.
c. There would be...	○ fewer	○ the same number of	⬤ more	...grasshoppers.