Measuring Fidelity of Implementation in a Large-Scale Research Study (RTP)

Dr. Cathy P. Lachapelle, Museum of Science, Boston

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. Christine M. Cunningham, Museum of Science, Boston

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. She focuses on developing research-based, field-tested curricula. For sixteen years, she worked as a vice president at the Museum of Science where she was the Founding Director of Engineering is Elementary, a groundbreaking program that integrates engineering concepts into preschool, elementary, and middle school curriculum and teacher professional development. Her recent book, Engineering in Elementary STEM Education, describes what she has learned. 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. At Cornell University, where she began her career, she created environmental science curricula and professional development. Cunningham has received a number of awards; in 2017 her work was recognized with the prestigious Harold W. McGraw Jr. Prize in Education. Cunningham holds joint B.A. and M.A. degrees in biology from Yale University and a Ph.D. in Science Education from Cornell University.

Measuring Fidelity of Implementation Using Engineering Implementation Logs in a Large-Scale Research Study (RTP)

Dr. Cathy P. Lachapelle

Museum of Science, Boston

Dr. Christine M. Cunningham

Abstract

When evaluating the implementation of an engineering curriculum, it's important to be able to measure the fidelity with which the curriculum is implemented by teachers. In this paper, we describe our instruments for and approaches to measuring fidelity of implementation of an elementary school engineering curriculum, and give evidence for reliability and validity of use of these instruments for an efficacy study of the curriculum. The most important instruments are the engineering lesson implementation logs, for which teachers were prompted to indicate (1) which portions of each engineering lesson they completed; (2) the duration and date of each lesson; and (3) indications of how they taught each portion of the lesson, to measure whether teachers were using a pedagogy more in line with the research study's treatment or comparison group—a measure of program differentiation. In this paper, we provide both qualitative and quantitative evidence for the suitability of use of the engineering logs to measure implementation fidelity.

Introduction

Since the Next Generation Science Standards, or NGSS, introduced engineering practices as core content that students should be learning prior to college [1], schools and teachers have been looking for ways to incorporate engineering into an already-crowded curriculum. With more and more engineering curricula on offer for pre-college settings, there is a particular need for new methods and instruments to test the efficacy of these curricula, so that educators can have scientific evidence about which curricula contribute to student learning.

This paper reports on aspects of a research study, Examining the Efficacy of Engineering is Elementary (E4), to analyze whether an engineering curriculum for elementary school students, Engineering is Elementary (EiE), has a positive impact on student outcomes, including science and engineering outcomes, as well as attitudes. EiE is designed to introduce students to engineering fields and the engineering design process in the context of science lessons. Students learn about engineering by engaging in the engineering design process and applying science content to solve a given problem. EiE is comprised of 20 units, each of which is intended to take approximately 8-10 hours of instruction to implement. The pedagogical approach is project-based learning, rooted in sociocultural learning theory, which posits that students learn best through actively engaging in well-supported, developmentally appropriate approximations of valid engineering practices. Krajcik and Blumenfeld [2] say that "project-based learning allows students to learn by doing and applying ideas. Students engage in real-world activities that are similar to the activities that adult professionals engage in."

E4 is designed to compare EiE units to a comparison curriculum, using a randomized controlled trial (RCT) design whereby participant schools were randomly assigned to one of the treatment conditions. Teams of teachers of grades 3-5 were recruited from schools in three eastern states to participate, with the approval of their school or district administration. Schools were then randomly assigned to either the treatment or the comparison condition: in treatment schools, teachers were assigned to teach EiE, and in comparison schools, they were assigned to teach Engineering for Children (E4C), which was assembled from engineering units and lessons freely available online, for the purpose of E4.

The EiE and E4C curricula were deliberately designed to differ on a number of EiE's design features, or "critical components," listed in Table 1. These critical components characterize a sociocultural approach to teaching and learning, where teachers are supported by professional development (PD) and the printed teacher guide to engage students as active participants in age-appropriate, yet valid to the discipline, engineering activity. The E4C curriculum activities were chosen because they embody a more traditional, didactic approach to learning. An independent researcher ranked all 10 curriculum units (5 treatment, and 5 comparison, matched by topic) according to a rubric designed to measure the presence of critical components, and found that EiE units scored very high on the rubric, while E4C units scored low.

Table 1. Critical components of EiE as compared to E4C.

EiE	E4C
Engineering content is introduced in a narrative context, designed to appeal to children from diverse backgrounds.	Engineering content is introduced in traditional textbook style. No context is provided for the challenge.
Students use a specified engineering design process.	Students are not explicitly taught an engineering design process.
Engineering challenges identify a problem with constraints and specifications for the solution requiring trade-offs.	Constraints and specifications for successful solution of the engineering challenge are not given. Trade-offs are not required.
Students use science and math as they design solutions.	Science and math are not explicitly featured nor is their use supported.
Students analyze data and use failure constructively as they design iteratively.	Students are not supported to analyze data or reflect on failures. Designs are not improved.
Students' collaborative work is supported and includes negotiating with team members	Students may work together in teams but are not given support to do so.
Students are encouraged to be creative, brainstorm, and consider a multiplicity of ideas and possible solutions.	The design challenge is open-ended but development of multiple design ideas is not discussed or supported in the curriculum.

Teacher guide supports engaging prior knowledge, prompting reflection, and modeling engineering thinking and practices.	Teacher guide focuses on how to explain content to students and the specifics of running the activity.

The goal of E4 is to determine if EiE and its critical components lead to improved student learning of engineering and science as compared to E4C. Teachers were assigned to one of four unit topics—environmental, electrical, geotechnical, or package engineering—based on the science content they planned to teach in the coming 2 years of the study: ecosystems, electricity, landforms and erosion, or plant structures and functions. All teachers, regardless of treatment condition, were required to teach the science content related to the EiE unit, so that all students would have "opportunity to learn" the content to be tested; however, teachers could use their usual teaching materials and methods.

Teachers from both conditions received PD and all materials from E4 prior to the first year of implementation. They received additional PD prior to the second year of implementation. They collected pre- and post-assessments as well as engineering journals from their students, and they completed engineering and science logs after each required lesson.

In this research paper, we explain our methodology for measuring teachers' fidelity to curriculum materials and their pedagogical intent as they implement their assigned engineering curriculum. We analyze data from these instruments using structural equation modeling to create factor scores which will be used in later analysis to predict the mediating effect of fidelity of implementation on the relationship between treatment and student outcomes in science and engineering.

What is implementation fidelity?

Teachers have control over how they will implement a curriculum. Frequently, teachers will skip lessons because of time constraints, supplement with additional lessons for struggling students, or make adjustments to lesson plans to address the needs and interests of their students [3]. Teachers also make changes to curricula based on their habitual approach to teaching and their philosophy of teaching and learning [4]. They make changes to better match district pacing guides and local or regional content standards [5]. Any change that teachers make to the curriculum as it is designed to be implemented can be thought of as lessening fidelity of implementation (FOI).

Generally speaking, lower FOI is thought of as reducing the potential effectiveness of a curriculum. However, not all changes and adaptations made by teachers will necessarily reduce the quality of implementation – some adaptations increase effectiveness and improve the quality of curriculum, by adjusting the design of lessons to account for the needs, interests, and backgrounds of particular students [3], or to better match district and state learning objectives [5] Such changes are often said to align with "integrity of implementation" [6].

Why is it important to measure fidelity in an efficacy study?

Key to measuring the true effect of an intervention like EiE is measurement of FOI to improve and assure internal validity [7], [8]. In an RCT study, it's vital that the treatments are differentiated, to ensure that statistical models can discern differences between the intervention and comparison: when teachers change their instruction in ways that blur differences between the treatments, internal validity is lowered, and the power of statistical models is reduced [7]; [9]. Clear differentiation of the treatments, which we have addressed by specifying critical components and testing materials for each treatment against a rubric (as described above), is one way to work towards internal validity. FOI—the adherence of teachers to their prescribed curriculum—is another important variable to address.

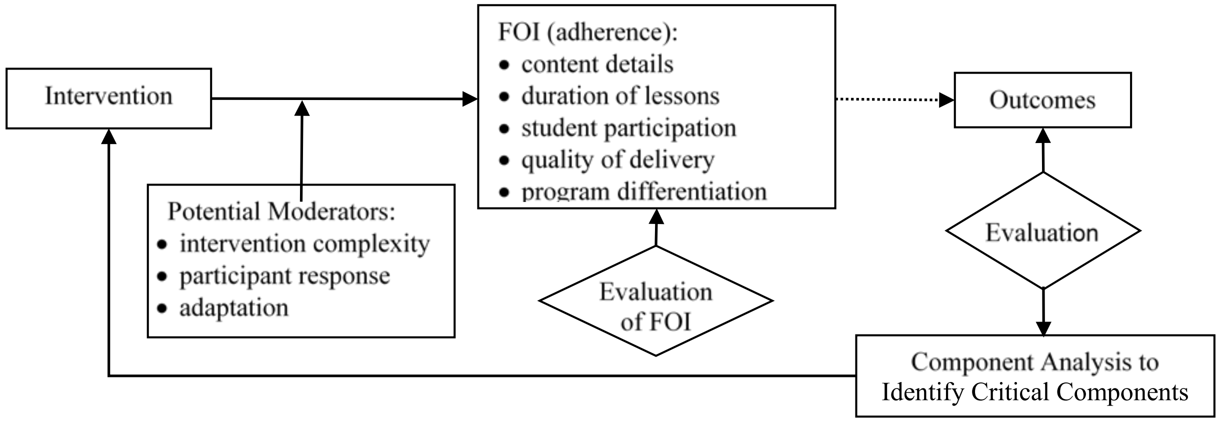
A research study testing the efficacy of an intervention will generally take steps to ensure that participants implement the intervention with high fidelity, to ensure that the test of differences between treatment and comparison is a valid one. Part of addressing fidelity is through coherent teacher training that includes planning time and specific support by trainers [10], which E4 implemented with separate PD workshops for EiE and E4C teachers, each addressing fidelity both indirectly, through the style of pedagogy modeled in the workshop, and directly, with a discussion of expectations about what fidelity would look like. However, we also expect that teachers will modify their assigned curriculum, for the reasons discussed above. To gather information about where and why teachers deviated from the print materials during E4, we used instruments designed for the purpose of measuring aspects of instruction that pertain to fidelity, as recommended in the literature (e.g. [7]; [8]).

Our conceptual framework for implementation fidelity

In seeking to understand FOI, we rely on a framework defined by Carroll et al. [11] in the field of social work, where much of the pioneering work has taken place in defining and measuring FOI of behavioral health interventions. Science education researchers have built upon this work to study the implementation of science curricula, e.g., [12], [13].

We have modified the framework from [11] with the work of O’Donnell [7], who conducted a research review of implementation fidelity studies in K-12 education in order to better define and measure FOI in the context of testing educational interventions (Figure 1). In this framework, the causal effect of an intervention on outcomes depends primarily upon FOI [11]. FOI includes specifics of what content was addressed, frequency and duration of lessons, and student participation—the “structure” of implementation [14]. The relationship between the intervention as intended and adherence may be modified by other factors, including the complexity of the intervention, facilitation strategies (teacher guide, PD workshops, incentives, etc.), and participant response (teacher and student attitudes and judgments of value). Carroll et al. [11] categorize quality of delivery by the teacher (also called an “intervention process” [13]), or “the way in which services are delivered” [14]—as a potential moderator; however we follow the majority of studies in considering this an element of FOI.

Figure 1. Conceptual Framework for Implementation Fidelity (adapted from [11])



Mowbray et al. [14] advocated that researchers identify and develop valid and reliable measures for "fidelity criteria" of an intervention. The identification of critical components is the first step in developing fidelity criteria. Critical components must include both structural components (specifying elements of adherence) and process components (specifying quality of delivery and program differentiation). Program differentiation—the implementation of critical components unique to the intervention—is particularly important because it affects whether evaluation of the outcomes will find an effect of the intervention beyond that of the comparison [7].

How can fidelity be measured?

Many research studies in education use classroom observations to characterize FOI. However, this method is expensive and time consuming, and can be infeasible for a large-scale research study. Instructional logs are another common way to track whether and how teachers have implemented a lesson. Research has shown that such logs are not only much more cost-effective than class observations, they can also be a valid and reliable way to measure aspects of instruction like completion and some features of pedagogy [15].

Methods

Development and content of fidelity measures

We developed two kinds of implementation logs for this study: engineering lesson logs, to track the implementation of the E4-assigned unit, and science lesson logs, which track the implementation of required science content. For this paper, we will focus on the engineering logs: in particular, the yes/no question portions of the engineering lesson logs.

The engineering lesson logs were tailored to each lesson of each unit, with the intention that teachers would complete each form soon after completing each lesson. We developed 68 engineering logs, eight or nine per unit, one for each part of a lesson intended to take an instructional hour, to track fidelity to the 8 units (4 E4C, 4 EiE) that were used and tested in E4. At the start of each engineering lesson log, teachers were asked for the date, time spent, and duration of the lesson. Prompts then asked, for each part of the lesson, whether the teacher had completed it, and if so, how. For the E4C lessons, this transfer was more straightforward than for EiE lessons, because EiE lesson instructions were more detailed, and had to be summarized. As an example, teachers of the environmental engineering units (both treatments) were prompted to indicate whether "I helped students understand pH," with the same standard sub-prompts, designed to ascertain the type of teaching (direct instruction or sociocultural engagement) based on the type of activity being prompted: "I explained to students" or "I prompted students to try to come up with the idea themselves." Finally, all teachers were prompted to indicate whether they did something differently than what was instructed in the teacher guide, or whether they made cuts or additions to lessons, and if so, to explain what they did. For a sample lesson log, see Appendix A.

Qualitative evidence of content validity of the engineering logs

To collect qualitative evidence of content validity for the purpose of measuring FOI, we designed two versions of each engineering log: one for teachers, the other for an observer. Before beginning data collection for E4, we asked several teachers of EiE (who we were working with on other projects) to teach a lesson and then complete the associated log, while a member of the research staff observed and completed the same log. We then asked teachers for feedback on the engineering logs, and revised them using the feedback and our own experiences.

Once the E4 was underway, we continued to use the observer logs in all 24 classrooms that were videotaped as part of the data collection for the larger study. We then compared these observer logs to the engineering logs completed by the teachers of these classrooms. Our analysis showed they were completed similarly.

Data collection and treatment of missing data

Teachers completed all engineering logs online. We collected engineering logs from 524 of 604 participating classes; teachers failed to return engineering logs for 80 classes. To prepare for all types of quantitative analysis for E4, we imputed missing data at the student and class levels using chained equations via IVEware [16]. To accomplish the imputation, we created a data set with all available data for the 14,015 students in the study, including outcome measures, student demographics, teacher demographics, school demographics, and class-level variables such as unit, year of the study, and data from the engineering and science logs, so that the maximum relevant data was available to impute all the different kinds of missing data, from student demographic information to engineering log data. We estimated 20 imputed datasets, which were used in the factor analyses described in this paper.

Coding of engineering logs and translation into data

The bulk of data collected from the engineering logs is in the form of yes/no responses to prompts for the teacher to read and answer; this is the data of interest for this research paper. This data was first organized according to the content of each prompt (Table 2). It was then converted to percentage values for each type of prompt, according to the unit and treatment, so that the sets of engineering logs from each of the eight units could be combined into the same dataset and compared in a standardized way.

The content areas include four main categories: "Comprehend" prompts, which relate to instructions to the teacher to help students understand some content in the unit; "Explain" prompts, which specifically instruct teachers to explain to students what would be happening during the lesson; "HowToDo" prompts instructing teachers to ensure that students know what to do during the activity; and "StudentWork" prompts, for parts of the lessons where students are to work individually or in groups. Three of these main categories included sub-prompts. "Comprehend" sub-prompts were designed to distinguish between explaining to students (didactic pedagogy) and "prompting students" for their ideas (sociocultural engagement). "HowToDo" sub-prompts distinguished between modeling activities for students and explaining what to do. "StudentWork" had two possible sets of sub-prompts, one set designed for situations where students were working on design activities or experiments, and another for where students were responding to questions or prompts given by their journals, or sometimes set by the teacher. In the case of "StudentWork" involving design or experimental activities, teachers were asked to designate whether students worked individually, in pairs or small groups, or together as a whole class. In the case where students responded to questions or prompts, the sub-prompts were further differentiated to distinguished whether students discussed their answers with others in their group or as a whole class discussion, and whether they wrote answers individually, for their group, or to capture the consensus of the class.

Table 2. Coding guide for types of prompt given in the engineering logs.

Type of Prompt	Coding Guide
Comprehend	Prompts the teacher to help students understand something, e.g. "I helped students understand what engineers do, specifically electrical engineers."
Sub-prompts for Comprehend prompts:	
Comp_Prompt	"I prompted students to try to come up with the idea themselves."
Comp_Explain	"I explained to students"
Explain	Prompts the teacher to explain to students, e.g. "I told students that they would use a model to explore how pollution moves."
HowToDo	Use whenever the teacher is to "help students understand how" to do something, e.g. "I helped students understand how they would test their water filters."
Sub-prompts for HowToDo prompts:	
How_Demonstrate	"I demonstrated by doing an example"
How_Describe	"I explained or described what to do"
StudentWork	Prompts the teacher to have students work independently or in groups. e.g. "I had students build their improved designs." e.g. "I had students answer the questions on journal p. 44."
Sub-prompts for StudentWork prompts (NOT having students respond to questions or prompts):	

SW_Individually	"Individually"
SW_Groups	"In pairs/small groups"
SW_WholeClass	"As a whole class"
Sub-prompts for StudentWork prompts (having students respond to questions or prompts):	
SW_Resp_DG	"Students discussed answers in groups."
SW_Resp_DC	"Students discussed answers as a class."
SW_Resp_WI	"Students wrote their own answers"
SW_Resp_WG	"Students wrote their group's answer"
SW_Resp_WC	"Students wrote a class consensus answer"

Exploratory factor analysis and scale development

The purpose of exploratory factor analysis (EFA) is to explore and describe relationships between interrelated items—in particular, to describe groupings of items corresponding to latent (unobserved) variables that are theoretically plausible. For this study, we conducted EFA because the engineering logs we designed for this study were new, and we wanted to ensure that an interpretable factor structure was possible [17]. We used the imputed datasets for all analysis of the engineering logs variables. We conducted all factor analytic methods with Mplus 8.1 [18].

We used multiple methods to estimate the number of factors we would extract before beginning analysis, because multiple competing factor structures should be tested and evaluated as part of the process of developing a theoretically plausible factor structure [17]. We used parallel analysis as a quantitative method to predict the number of factors, and we examined scree plots as a qualitative method, both using a script available online [19] for use with SPSS 25 [20].

We used the MLR estimator, an extension of Maximum Likelihood (ML) estimation that is robust to multivariate non-normality, and adjusts for missing data using Full Information Maximum Likelihood (FIML), to handle the non-normality of our data. The ratio of sample size (604) to expected factors (<10) is minimal for our needs; 5 factors would be better, increasing the ratio of sample to factors to 120:1, which should be sufficient as long as communalities are not very low [21]. We used the oblique Geomin rotation, which is the default within Mplus 8.1, as we predicted that resulting factors were likely to be correlated.

A concern in modeling the engineering log items was the association of certain prompts, either with parallel wording or as grouped sub-prompts (Table 1). Test items with similar wording can lead to correlated error, also called correlated uniqueness, which can, if left unspecified in modeling, lead to inflated estimates of covariation and the extraction of factors that do not have a basis in theory [22], [23]. To address this concern, we specified and tested the possible sets of correlated uniquenesses during EFA.

We used multiple goodness-of-fit measures to compare a variety of plausible candidate factor solutions. We also used fit statistics from each of the categories of absolute fit and comparative fit measures [24]. Absolute fit measures test the congruence of the covariance matrices for the model as compared to the baseline data. We examined three absolute fit measures: (1)

standardized root mean square residual (SRMR) should be <.08; (2) the root mean square error of approximation (RMSEA) should be <.05; and (3) the $\chi 2$ statistic should show a difference between the fitted model and the baseline model at $p<.05$, according to standard rules of thumb. The comparative fit index (CFI) should be >.95 for a model to be considered a good fit [24]. Most importantly, we evaluated all candidate models for interpretability.

We followed EFA with exploratory structural equation modeling (ESEM) within confirmatory factor analysis (CFA), a more general framework of analysis of factor structures, which unlike CFA allows for testing of latent variable structures while still allowing for cross-loading of items on factors / latent variables, a phenomenon which is common and theoretically reasonable in social sciences research [25]. The use of ESEM within CFA (EwC) allowed us to refine the models generated by EFA, further compare them, and choose a final model. We then output factor scores for the confirmed model. Mplus 8.1 uses the regression method [26] to predict factor scores for each individual participant.

Results

Parallel analysis for the unimputed engineering log data (524 classes) is given in Table 3; scree plots are given in Figure 2. The parallel analysis shows that 9 factors are possible: the raw sample data value does not exceed the random data percentile until factor 9. The scree plot similarly shows the percentile and raw data lines crossing at factor 9, indicating the possibility of up to 9 factors. However, the scree plot also shows 2 major bends, one at 5 factors and another at 7 factors, so these are also possible factor structures.

Table 3. Parallel analysis for engineering log coded items: unimputed dataset eigenvalues.

Factor	Raw Sample Data	Random Data Means	Random Data Percentile
1	3.185	0.347	0.415
2	2.187	0.278	0.331
3	1.371	0.226	0.269
4	0.902	0.181	0.220
5	0.650	0.141	0.176
6	0.597	0.103	0.136
7	0.276	0.068	0.097
8	0.132	0.034	0.062
9	0.038	0.001	0.029

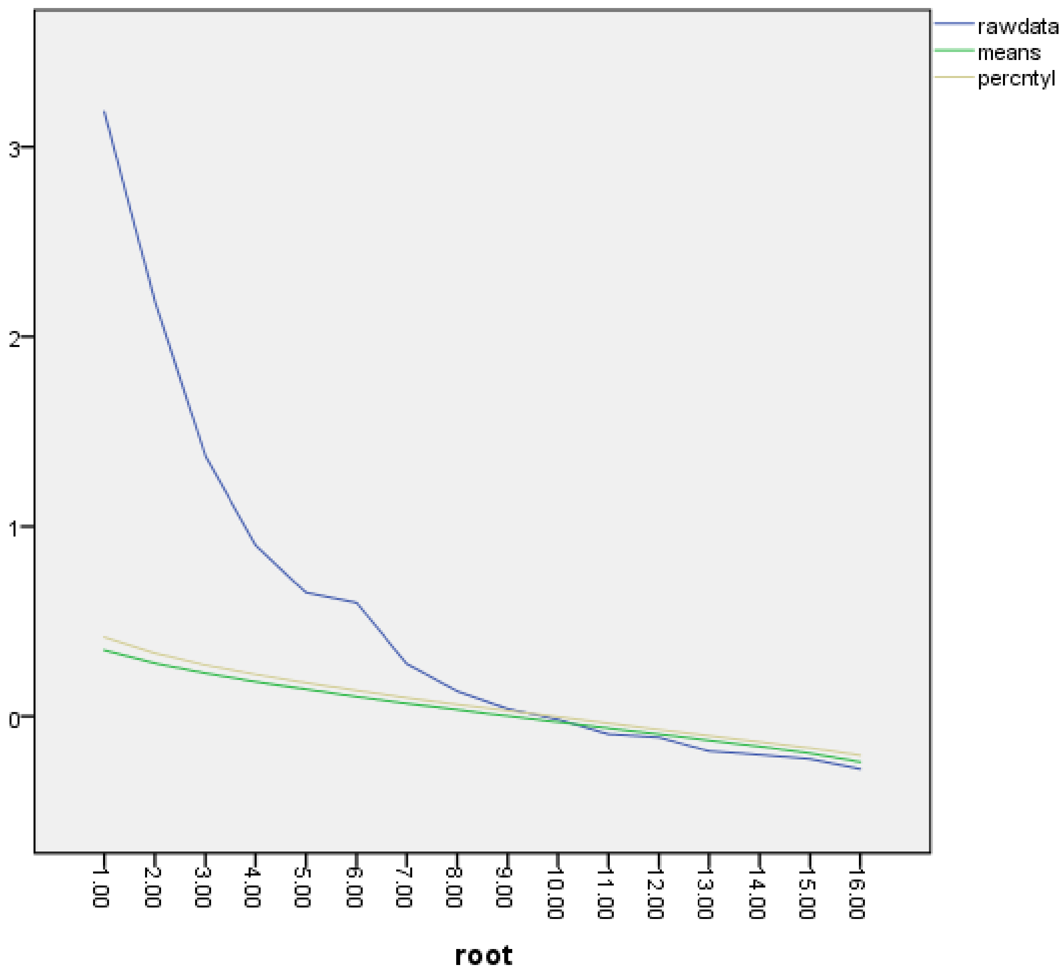
We ran parallel analysis and scree plots separately for each of the 20 imputed data files, and compiled the results (Table 4). We see, again, that scree plots show a first bend at factor 4 or 5 (most often 5) and a second bend at factors 6-8 (most often 7); parallel analysis indicates the crossing point (maximum possible factors) at 8-10, with 9 as mode and median. Therefore we chose to test models with between 4 and 9 factors, keeping in mind that 8-9 factors seems unlikely given the tiny eigenvalues, and would be difficult to estimate with our sample size.

Table 4. Parallel analysis for engineering log coded items: summary of 20 imputed data sets.

Datset #:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
# Factors	8	9	10	10	9	8	10	9	9	9	8	10	9	9	9	9	8	8	9	9
Scree bend 1	4	4	5	5	5	5	5	5	4	5	4	5	5	5	5	5	5	5	5	5
Scree bend 2	6	6	7	7	8	8	7	8	7	7	7	7	7	8	7	7	7	7	8	8

We tested three candidate organizations of correlated uniqueness (CU). The first (“CU sets”) was derived from the overall organization of prompts within the engineering logs into main prompts and sub-prompts. The second (“CU groups”) was based on the sub-prompt groupings alone, and the third (“CU wording”) was based solely upon wording (Table 5). All three candidate correlated uniqueness organizations were tested alone and in combination with the others.

Figure 2. Scree plot for engineering log coded items: unimputed dataset eigenvalues.



Running the EFA models with specified correlated uniquenesses resulted in a range of models (Table 6). On the one hand, the models with 4 or 5 factors had the most interpretable factor structures. These models were also the most likely to converge, and to have more than a handful of successful computations out of the 20 imputed models. On the other hand, models with more factors had better fit statistics. Models with more than 8 factors did not converge, probably because they were not identified (i.e. there was not enough information to calculate the models). We decided to stick with a 5-factor model for the next round of model testing, and to include the Groups CU's.

Table 5. Three candidate organizations of correlated uniqueness tested during EFA.

CU Sets:	*Main prompts correlated with sub-prompts*
1	Comprehend with Comp_Prompt and Comp_Explain
2	HowToDo with How_Demonstrate and How_Describe
3	StudentWork with SW_Individually, SW_Groups, and SW_WholeClass
4	StudentWork with SW_Resp_DG, SW_Resp_DC, SW_Resp_WI, SW_Resp_WG, & SW_Resp_WC
Groups:	*Groups of sub-prompts correlated with each other*
1	Comp_Prompt with Comp_Explain
2	How_Demonstrate with How_Describe
3	SW_Individually with SW_Groups with SW_WholeClass
4	SW_Resp_DG with SW_Resp_DC
5	SW_Resp_WI with SW_Resp_WG with SW_Resp_WC
Wording:	*Prompts with similar wording correlated*
1	SW_Resp_DG with SW_Resp_WG
2	SW_Resp_DC with SW_Resp_WC

Table 6. Mean fit indices (20 imputed datasets) for EFA models with correlated uniquenesses.

#		# Models	Chi-Square Test			RMSEA	
Factors	CU	Successful	df	Mean χ^2	SD	Mean	SD
4	None, Word	0		*Failed to converge*			
4	Sets	9	50	684.1	87.1	0.145	0.010
4	Groups	5	53	152.3	19.7	0.055	0.005
4	G+W	19	51	187.1	71.3	0.065	0.016
4	All	4	39	100.9	13.4	0.051	0.006
5	None	1	101	502.5	0.0	0.122	0.000
5	S, W	0		*Failed to converge*			
5	Groups	13	41	143.6	34.2	0.063	0.012
5	G+W	6	39	115.6	16.5	0.057	0.006
5	All	9	27	110.2	42.9	0.069	0.019
6	N, S, W	0		*Failed to converge*			
6	Groups	1	91	62.8	0.0	**0.043**	**0.000**
6	G+W	3	28	57.2	12.0	**0.041**	**0.009**
6	All	7	16	56.2	17.8	0.063	0.015
7	N, S, W	0		*Failed to converge*			
7	Groups	2	20	33.9	4.8	**0.033**	**0.006**
7	G+W	11	26.70	26.7	9.4	**0.024**	**0.017**
7	All	3	43	5.2	3.9	**0.013**	**0.019**
8	N, S, W, A	0		*Failed to converge*			
8	Groups	6	11	15.9	4.6	**0.023**	**0.014**
8	G+W	2	57	32.4	20.1	**0.042**	**0.042**

In building EwC models, unlike in EFA models, items can be given a "starting value". We used the parameter estimates from EFA to set starting values for the items loading onto each factor. We fixed the variance of each factor at 1, to ensure the models had sufficient constraints to be identified. The final model with parameter estimates is given in Table 7. We found five factors: (1) Completion, representing the amount of each unit completed; (2) Didactic, representing use of a more traditional didactic method of instruction; (3) T_Teaching, representing use of a more socioculturally engaging method of instruction; (4) Classwork, a measure of how frequently teachers indicated they did student work and discussions as a whole class; and (5) Groupwork, a measure of how frequently teachers had students work and discuss in small groups or teams.

Table 8 gives the parameter estimates for the correlated uniquenesses. Though individual correlated uniquenesses are not statistically significant in size, most correlated uniquenesses within each group are significant, indicating the importance of specifying these relationships.

Table 7. EwC standardized parameter estimates/p-value for engineering logs factor structure.

Variable	Completion	Didactic	T_Teaching	Classwork	Groupwork
Comprehend	0.759/.000		0.398/.000		
Explain	0.626/.000				
HowToDo	0.849/.000				
StudentWork	0.806/.000	0.129/.040	-0.299/.000		
Comp_Prompt			0.439/.000		0.406/.000
Comp_Explain	0.216/.002	0.333/.000	-0.096/.208	0.210/.004	-0.206/.008
How_Demonstrate			0.372/.000	0.282/.000	0.132/.039
How_Describe	0.268/.000	0.306/.000			0.166/0.040
SW_Individually		0.491/.000			
SW_Groups		-0.509/.000			0.811/.000
SW_WholeClass	0.395/.000	0.496/.000		0.203/.008	-0.425/.000
SW_Resp_DG				-0.145/.145	0.648/.000
SW_Resp_DC		0.226/.000	0.234/.018	0.532/.000	
SW_Resp_WI	0.238/.001	0.491/.000	-0.118/.105	-0.148/.145	
SW_Resp_WG	-0.158/.004				0.449/0.000
SW_Resp_WC				0.608/.000	

Table 8. EwC standardized parameter estimates for correlated uniquenesses.

Correlated Uniqueness (CU)	Estimate	Standard Error	P-Value
Comp_Prompt with Comp_Explain	-0.733	0.038	0.000
How_Demonstrate with How_Describe	-0.585	0.045	0.000
SW_Individually with SW_Groups	-0.079	0.202	0.694
SW_Individually with SW_WholeClass	-0.608	0.095	0.000
SW_Groups with SW_WholeClass	0.043	0.194	0.823
SW_Resp_DG with SW_Resp_DC	-0.174	0.062	0.005
SW_Resp_WI with SW_Resp_WG	-0.431	0.059	0.000
SW_Resp_WI with SW_Resp_WC	-0.183	0.086	0.033
SW_Resp_WG with SW_Resp_WC	0.174	0.051	0.001

Fit indices for the final model are given in Table 9. All imputed models converged successfully. Fit statistics are adequate or good, with CFI > 0.95, RMSEA near to 0.05, and SRMR <0.08, meeting rules of thumb for a good model. Correlations between factors are given in Table 10. The only factors that correlate significantly are Groupwork and Completion.

Table 9. Mean fit indices for EwC model of factor structure of engineering logs.

#Factors	# Models Successful	χ^2 /SD	df	CFI /SD	RMSEA/ SD	RMSEA 95% CI	SRMR /SD
5	20	155.92 /25.707	64	.965 /.009	.048 /.007	.034-.062	.031 /.002

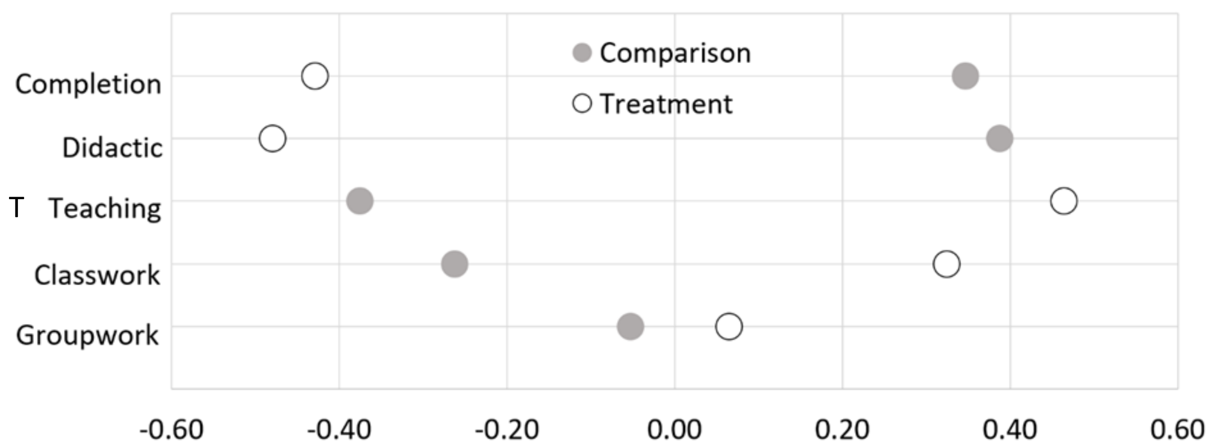
Table 10. Correlations between factors for engineering logs.

	Completion	Didactic	T-Teaching	Classwork	Groupwork
Completion	1.000				
Didactic	0.165	1.000			
T-Teaching	-0.050	-0.195	1.000		
Classwork	-0.130	0.098	0.136	1.000	
Groupwork	**0.488**	0.049	-0.100	0.142	1.000

Correlation in bold is significant at p<.000.

Factor scores derived from the engineering logs show differences between treatments, on all variables but Groupwork (Figure 3). Differences for the completion, didactic, and EiE-teaching scores approximated a standard deviation. The gray circles represent the comparison group, and the white circles the treatment group. As expected, E4C teachers are more likely to implement Didactic teaching, while EiE teachers are more likely to implement Treatment teaching; both treatment conditions spend about the same amount of time on groupwork, which is built in to both curricula. Less expected is that the comparison group completes more of their assigned units and lessons than the treatment group.

Figure 3. Mean scores on standardized fidelity variables by treatment (total mean=0; SD≅0.8).



Conclusions

The ability to derive fidelity information from instructional logs is important for the conduct of quantitative research testing the efficacy of curricular interventions. In this paper, we have described and analyzed a method to collect data from teachers about what parts of the engineering lessons within a unit they completed, and how they completed the lessons. This data was coded and combined into five factor scores for use in quantitative modeling. The factor scores represent (1) Completion: how much of each lesson was implemented; (2) Didactic: a representation of how much the teacher used direct instruction, such as explaining or describing to students, or replacing student independent work with a whole-class activity or individual activity; (3) Treatment-teaching, defined as involving more prompting of students to give their own ideas, demonstrating and modeling for students how they will complete work in their groups, and asking comprehension questions; (4) Classwork: spending more time doing parts of the lesson as a whole class, particularly with discussions and writing; and (5) Groupwork: spending more time with students working in groups or teams. These factor scores are useful in differentiating the treatment from the comparison group for the purposes of examining the mediating effects of fidelity of implementation on the relationship between treatment and student outcomes. Because we are able to differentiate between the EiE and the E4C using these variables, we expect to be able to use these variables to (1) account for teaching that deviated from the pedagogical style of the assigned curriculum; (2) account for differential rates of completion between the two treatment conditions; and (3) identify which variables explain the effect of EiE on student outcomes, so we can learn more about what aspects of EiE are important to student learning.

Other research studies examining the efficacy or effectiveness of curriculum interventions can learn from the methods we have experimented with here. With easy-to-answer yes/no questions, we have been able to quantify the amount of each lesson and unit that teachers have completed, as well as identify aspects of pedagogical approach and teacher use of whole class work versus group work. Four of the five factors show clear differentiation between the treatment conditions, as they were intended to do, which means they are promising variables for exploring aspects of instruction and fidelity that mediate (i.e. help to explain the effects of) the treatment.

With improved means to study the efficacy of curricula designed for pre-college settings, researchers and curricular designers will have the tools they need to collect and present evidence to educators and administrators to argue for incorporating engineering curricula with strong evidence of effectiveness on student outcomes such as engineering and science learning. Scaled-up implementation of engineering interventions is unlikely without such evidence.

Future Work

In our future work, we will analyze our data for between-group invariance, to determine whether the pattern of factors is the same for both the EiE and E4C groups.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1220305. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

[3] T. J. Power, J. Blom-Hoffman, A. T. Clarke, T. C. Riley-Tillman, C. Kelleher, and P. H. Manz, “Reconceptualizing intervention integrity: a partnership-based framework for linking research with practice,” *Psychology in the Schools*, vol. 42, no. 5, pp. 495–507, 2005.

[4] K. D. Squire, J. G. MaKinster, M. Barnett, A. L. Luehmann, and S. L. Barab, “Designed curriculum and local culture: Acknowledging the primacy of classroom culture,” Science education, vol. 87, no. 4, pp. 468–489, 2003.

[5] A. H. Debarger, W. R. Penuel, S. Moorthy, Y. Beauvineau, C. A. Kennedy, and C. K. Boscardin, “Investigating purposeful science curriculum adaptation as a strategy to improve teaching and learning,” *Science Education*, vol. 101, no. 1, pp. 66–98, 2016.

[1] NGSS Lead States, *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press, 2013.

[2] J. S. Krajcik and P. C. Blumenfeld, “Project-based learning,” in *The Cambridge handbook of the learning sciences*, R. K. Sawyer, Ed. New York, NY: Cambridge University Press, 2006, pp. 317–334.

[6] S. A. Barab and A. L. Luehmann, “Building sustainable science curriculum: Acknowledging and accommodating local adaptation,” *Science Education*, vol. 87, no. 4, pp. 454–467, 2003.

[7] C. L. O'Donnell, “Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research,” *Review of Educational Research*, vol.78, no. 1, pp. 33-84, 2008.

[8] M. Song and R. Herman, “Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I),” *Educational Evaluation & Policy Analysis*, vol. 32, no. 3, pp. 351-371, 2010.

[9] 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.

[10] W. R. Penuel, B. J. Fishman, R. Yamaguchi, and L. P. Gallagher, “What makes professional development effective? Strategies that foster curriculum implementation,” *American Educational Research Journal*, vol. 44, no. 4, pp. 921–958, 2007.

[11] C. Carroll, M. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain, “A conceptual framework for implementation fidelity,” *Implementation Science*, vol. 2, no. 1, pp. 40-48, 2007.

[12] P. Blumenfeld, J. Krajcik, T. Kempler, R. GEiEr, R. Kam, and S. Gallagher, “Opportunity to learn: Teacher instructional practices that account for variation in achievement in project-based science in urban middle schools,” Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, 2006.

[13] O. Lee, R. Penfield, and J. Maerten-Rivera, “Effects of fidelity of implementation on science achievement gains among English language learners,” *Journal of Research in Science Teaching*, vol. 46, no. 7, pp. 836-859, 2009.

[14] C. T. Mowbray, M. C. Holter, G. B. Teague, and D. Bybee, “Fidelity criteria: Development,

measurement, and validation,” *American Journal of Evaluation*, vol. 24, no. 3, pp. 315-315, 2003.

[15] B. Rowan and R. Correnti, “Studying reading instruction with teacher logs: Lessons from the study of instructional improvement,” *Educational Researcher*, vol. 38, no. 2, pp. 120–131, 2009.

[16]T. Raghunathan, P. W. Solenberger, P. A. Berglund, and van Hoewyk, J., *IVEware: Imputation and variance estimation software*. Ann Arbor, MI: University of Michigan, 2016.

[17] D. L. Bandalos and S. J. Finney, “Factor analysis: exploratory and confirmatory,” in *The reviewer’s guide to quantitative methods in the social sciences*, G. R. Hancock and R. O. Mueller, Eds. New York, NY: Routledge, 2010.

[18] L. K. Muthén and B. O. Muthén, *Mplus user’s guide*, 8th ed. Los Angeles, CA: Muthén & Muthén, 1998–2012.

[19] B. P. O’Connor, “SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test,” *Behavior Research Methods, Instrumentation, and Computers*, vol. 32, no. 3, pp. 396–402, 2000.

[20] IBM Corporation, *IBM SPSS Statistics for Windows*. Armonk, NY: IBM Corporation, 2017.

[21] R. C. MacCallum, K. F. Widaman, S. Zhang, and S. Hong, “Sample size in factor analysis,” *Psychological methods*, vol. 4, no. 1, pp. 84–99, 1999.

[22] T. A. Brown, *Confirmatory factor analysis for applied research*, 1st ed. New York, NY: The Guilford Press, 2006.

[23] H. W. Marsh, “Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?,” *Journal of personality and social psychology*, vol. 70, no. 4, p. 810, 1996.

[24] E. K. Kelloway, *Using Mplus for structural equation modeling: A researcher’s guide*, 2nd ed. Thousand Oaks, CA: SAGE Publications, Inc., 2015.

[25] H. Marsh, A. Morin, P. Parker, and G. Kaur, “Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis.,” *Annual review of clinical psychology*, vol. 10, p. 85, 2014.

[26] C. DiStefano, M. Zhu, and D. Mindrila, “Understanding and using factor scores: Considerations for the applied researcher,” *Practical Assessment, Research & Evaluation*, vol. 14, no. 20, pp. 1–11, 2009.

Appendix A: Sample Engineering Lesson Log

Teacher Log

A Slick Solution: Cleaning an Oil Spill **Lesson 2:** *An Enviro-Mystery*

The purpose of this teacher log is to gather information about your use of the *Engineering is Elementary* unit *A Slick Solution: Cleaning an Oil Spill* for **Lesson 2** (Teachers Guide pages 53-72). Parts of teachers' plans and expectations for lessons often occur as anticipated, but lessons sometimes change during instruction for a variety of reasons. Please answer the following questions about what took place during Lesson 2 in your classroom.

☐ The class did not do Lesson 2: *An Enviro-Mystery*. (Do **NOT** complete the rest of this log.)	
1. About how many minutes did you spend reading *Lesson 2* in the Teacher Guide in preparation?	________minutes
2. About how many minutes did you spend preparing materials for *Lesson 2*?	________minutes
About how many class periods were spent on *Lesson 2*? *(Circle one.)*	0.5 1 1.5 2 2.5 3 3.5 4 or more
3. Who taught *Lesson 2*?	☐ I taught it. ☐ My colleague __________________________taught it.

Introduction

(Select one.)

☐ I introduced the lesson using at least a portion of the steps described in the Teacher's Guide.
☐ I introduced the lesson, but did not use any of the steps described in the Teacher's Guide. (**COMPLETE** Item 5, then **SKIP** to Item 8.)

4. Approximately how many minutes total were spent on the *Introduction* portion of Lesson 2?	_________minutes

If no time was spent on the *Introduction* portion of Lesson 2, **SKIP** to the *Activity* section beginning with Item 9.

5. **<u>Introduction:</u>** <u>Review Environmental Engineering</u> Which of the following were done during the *Introduction* portion of this lesson? *(Select all that apply.)*	
☐ The class did not review environmental engineering as part of the *Introduction* portion of the lesson.	(**SKIP** to Item 7.)
☐ 6a. I reminded students what they had learned by asking them to respond to questions: ☐ **What sorts of projects did Thomas, the environmental engineer in the storybook, work on?** ☐ **Do you think environmental engineers might work on other projects as well? What types of projects?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 6b. I explained that students would be acting as environmental engineers to help a town better understand problems observed in a local ecosystem.	
☐ 6c. I did something else to review environmental engineering. *If yes, please describe below in response to Item 8.*	

6. **<u>Introduction:</u>** <u>The Greentown Challenge</u> Which of the following were done during the *Introduction* portion of this lesson? *(Select all that apply.)*	
☐ The class did not discuss the Greentown challenge as part of the *Introduction* portion of the lesson.	(**SKIP** to Item 7.)
☐ 7a. I read *Letter from the Mayor* {2-7} aloud to the class.	
☐ 7b. I checked comprehension by asking students to respond to questions: ☐ **What are some of the problems Greentown is experiencing?** ☐ **Why is this surprising?** ☐ **Do you think that the plants and frogs in Greentown are likely to be the only organisms affected? Why?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 7c. I asked students to underline the key points in the *Letter from the Mayor* {2-7}.	
☐ 7d. I guided students to discuss Greentown's ecosystem by asking: ☐ **What living things do you think are in Greentown's ecosystem?** ☐ **What kinds of non-living things are also part of the ecosystem?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class

☐ 7e. Returning to the *Letter from the Mayor* {2-7}, I asked students: ☐ **Do you have any ideas about what might be causing problems in Greentown?** ☐ **What resources do we have to help figure out what might be causing the problems in Greentown?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 7f. I guided students to begin thinking about the connections between the parts of the environment by posting the Guiding Question: ☐ **How do environmental engineers use their knowledge of soil and water to investigate environmental problems?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 7g. I did something else to help students understand the Greentown challenge. *If yes, please describe below in response to Item 8.*	

7. If you did something else during or supplemental to the *Introduction* portion of Lesson 2 not listed in Items 6-7, please describe what you did:

Activity

(Select one.)

☐ I implemented the activity portion of the lesson using at least a portion of the steps described in the Teacher's Guide.

☐ I implemented the activity portion of the lesson, but did not use any of the steps described in the Teacher's Guide. (**COMPLETE** Item 9, then **SKIP** to Item 13.)

9. Approximately how many minutes total were spent on the *Activity* portion of Lesson 2?	________ minutes

If no time was spent on the *Activity* portion of Lesson 2, **SKIP** to the *Reflection* section beginning with Item 14.

10. **<u>Activity:</u>** <u>Pollution and pH</u>

Which of the following were done during the *Activity* portion of this lesson? *(Select all that apply.)*

☐ The class did not discuss pollution and pH during the *Activity* portion of the lesson.	(**SKIP** to Item 11.)

☐ 10a. I posted the *Map of Greentown* and pointed out the location of Greentown Pond and Greentown Garden	
☐ 10b. To get students thinking about pollution in the environment, I had them respond to questions: ☐ **What do you think pollution is?** ☐ **Which parts of an environment do you think can be polluted?** ☐ **Do you have any ideas about how an environment might be polluted?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 10c. I explained that one of the tests environmental engineers might use to measure pollution in an area is a pH test. I asked students to respond to the following question: ☐ **Have you heard of the term pH before? Do you have any ideas about what it means?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 10d. I displayed the *pH scale* {2-8} and pointed out to students that pure water has a pH of 7, while acidic things like lemon juice have a pH of 2 or 3.	
☐ 10e. I demonstrated to students how to use pH strips by dipping one pH strip into the neutral vial, one into the acidic vial and one into the basic vial.	
☐ 10f. I had students to respond to: ☐ **What is the pH of each sample?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 10g. I had students think about why the pH of an environment might be important by having students to respond to: ☐ Do you have any ideas about how very acidic or very basic pH might affect the environment?	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 10h. Referring to the pH scale {2-8} I had students to respond to: ☐ According to the chart, in what range of pH can most frogs survive?	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 10i. I did something else to discuss pollution and pH. *If yes, please describe below in response to Item 13.*	

11. <u>**Activity:** Greentown Environmental Data: Collecting and Analyzing pH Data</u> Which of the following were done during the *Activity* portion of this lesson? *(Select all that apply.)*	
☐ The class did not collect or analyze pH data during the *Activity* portion of the lesson.	(**SKIP** to Item 12.)
☐ 11a. I posted the *Greentown Environmental Data: A Comparison {2-9}* and reviewed the data as a class	

☐ 11b. To get students thinking about the data on *Greentown Environmental Data: A Comparison* {2-9}, I had them respond to questions: ☐ **Do you notice any trends in the data from three years ago?** ☐ **What is the lowest pH?** ☐ **What is the highest pH?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 11c. I drew a box around the pH range (6.0-7.5) on Transparency of *pH Scale* {2-8}	
☐ 11d. I referred to the map of Greentown and explained that as environmental engineers, students can do pH testing in any of these areas of Greentown. I asked students to respond to the following questions: ☐ **Are there any areas of Greentown that you think might be related to the problems?** ☐ **How could we learn more about these areas?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 11e. I let groups choose a site to test for pollution with their pH strips.	
☐ 11f. I assigned groups a site to test for pollution with their pH strips.	
☐ 11g. Before testing, I had each group look up baseline data for their site on *Greentown Environmental Data: A Comparison* {2-9}.	
☐ 11i. Students tested samples, and recorded their findings on *Greentown Data: pH Testing {2-10}, journal page {x}.*	☐ Individually ☐ In pairs or small groups ☐ As a whole class
☐ 11j. Students reported their findings to the class and I asked of each site: ☐ **What was the site you tested?** ☐ **Did you have a soil sample or a water sample?** ☐ **What do you know about this site from three years ago?** ☐ **What was the pH of your sample today?** ☐ **Do you think there might be pollution at this site? Why or why not?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 11k. The class's pH data was recorded on *Greentown Environmental Data: A Comparison {2-9}.*	
☐ 11l. The class recorded the pH data on small pieces of paper and attached them directly to the *Map of Greentown* so that the map could serve as a data chart for class findings.	
☐ 11m. I guided students to think about the data and connections between areas of the environment by having students respond to the questions: ☐ **Based on your data, do you have any thoughts on what might be the source of the problem in Greentown Farm and Greentown Garden?** ☐ **Do you think the water and land in the environment are connected to one another? How?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class

☐ 11n. I did something else to collect or analyze pH data. *If yes, please describe below in response to Item 13.*	
12. Activity: Land and Water Connections in the Environment Which of the following were done during the Activity portion of this lesson? (Select all that apply.)	
☐ The class did not discuss land and water connections in the environment during the *Activity* portion of the lesson.	(**SKIP** to Item 13.)
☐ 12a. I placed 8-10 drops of food coloring in a pan of sand and explained to students that this represented a source of pollution.	
☐ 12b. I asked students to respond to the following question: ☐ **What do you predict will happen if I use the watering can to model rain falling on the land?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 12c. I used the watering can to pour water over the polluted spot, and asked students to respond to the following questions: ☐ **What do you notice?** ☐ **How did the water move? On top of the sand? Was it absorbed?** ☐ **Does this demonstration make you think differently about the pollution in Greentown? How so?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 12d. I guided students to think about next steps by asking them to respond to the following question: ☐ **Now that we have identified several sites in Greentown where there might be problems, what do you think we should do next**?	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 12e. I explained to students that environmental engineers would likely do more tests and further investigations to pinpoint the problems at each site.	
☐ 12f. I asked one student from each site/group to approach the map, open a door, describe what they see, and read the accompanying notes aloud to the class.	
☐ 12g. For each site I asked students to respond to the following questions: ☐ **What new information did we learn?** ☐ **Did this information help to explain our pH findings? How?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 12h. I did something else to reinforce land and water connections in the environment. *If yes, please describe below in response to Item 13*	

13. If you did something during or supplemental to the *Activity* portion of Lesson 2 not listed in Items 10-12, please describe what you did:

Reflection	
(Select one.) ☐ I implemented a reflection on the lesson using at least a portion of the steps described in the Teacher's Guide. ☐ I implemented a reflection on the lesson, but did not use any of the steps described in the Teacher's Guide. (**COMPLETE** Item 14, then **SKIP** to Item 16.)	
14. Approximately how many minutes total were spent on the *Reflection* portion of Lesson 2?	_______ minutes
If no time was spent on the *Reflection* portion of Lesson 2, **SKIP** to Item 17.	

15. **Reflection:** Pollution and How it Spreads Which of the following were done during the *Reflection* portion of this lesson? *(Select all that apply.)*	
☐ The class did not discuss pollution and how it spreads during the *Reflection* portion of the lesson.	(**SKIP** to Item 16)
☐ 15a. I guided students to think about the connections between each site by asking them to respond to these questions: ☐ **After reviewing our pH data and seeing the demonstration, do you have any new thoughts on how water and land in an environment might be connected?** ☐ **If there is pollution in one area of an environment, do you think it is likely to stay in that one area?** ☐ **Do you think the information our class has gathered can be used to explain the problems in Greentown? How?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 15b. I guided students to reflect on what they found from their pH testing and opening the doors on the Greentown map by asking them to respond to the following questions: ☐ **Are you surprised by the sources of some of the pollution in Greentown? Why?** ☐ **Do you have any suggestions to make to Mayor Higgins to help the situation?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 15c. I had students record the information they would share with the mayor and citizens of Greentown on *Greentown Presentation* {2-12}, journal page {x}.	

☐ 15d. I had students respond to the Guiding Question: ☐ **How do environmental engineers use their knowledge of soil and water to investigate environmental problems?**	☐ Individually (written) ☐ In pairs or small groups ☐ As a whole class
☐ 15e. I did something else to help students to reflect on what they learned pollution and how it spreads. *If yes, please describe below in response to Item 16.*	

16. If you did something else during or supplemental to the *Reflection* portion of Lesson 2 not listed in Item 15, please describe what you did:

16. <u>Extension and Reinforcement</u>

Which of the following were done as *Extension and Reinforcement* for the lesson? *(Select all that apply.)*

☐ The class did not complete any extension and reinforcement.	(Do **NOT** complete the rest of the log.)
☐ 17a. I made a three dimensional Greentown Map.	
☐ 17b. I gave students more pH strips and encouraged them to test liquids they find around their house.	
☐ 17c. I tried planting three of the same types of plants in three different pots, each with different soil pH.	
☐ 17d. I encouraged students to find out which species in their local ecosystem are considered indicator species.	
☐ 17e. I did something else to extend and reinforce what students learned in this lesson. *If yes, please describe below in response to question 18.*	

13. Please briefly describe the extension and reinforcement activities students completed: