Using Log Data to Analyze Teacher Implementation of Framework-aligned Curriculum

Cynthia D'Angelo (SRI International), Savitha Moorthy (SRI International), Carrie Allen Bemis (CU Boulder), Carrie-Anne Sherwood (University of Michigan)

A New Focus for Science Instruction

The *Framework on K-12 Science Education* (hereafter referred to as the Framework) (National Research Council [NRC], 2012) and the Next Generation Science Standards (NGSS, 2013) call for ambitious change in the way students learn science. One of the most important of these changes is integrating content and scientific practices together in instruction. Teachers are asked to help their students use their scientific knowledge and understanding while engaging in the various scientific practices. There are three ways that the Framework aims to change science education: (1) students should continually build on and revise their science knowledge and abilities over time, (2) education should focus on a limited number of core ideas, and (3) knowledge and practices should be integrated throughout instruction (NRC, 2012). One of the most difficult of these three changes for teachers will be the last of these, finding ways to structure instruction to integrate scientific practices with the appropriate core ideas.

The Framework lists eight scientific practices that students should be able to participate in by the end of 12th grade. Some of these practices are ones that students and teachers typically engage in to some extent in current instruction, such as analyzing data and communicating information. Other practices are ones that students and teachers have not typically engaged in during instruction, such as argumentation and developing models. Helping teachers to shift their instruction to a place where they are comfortable engaging their students in these practices in a rigorous and frequent fashion will take time and will require appropriate professional development and resources to support them in this transition.

Some curricula exist that hold promise for helping teachers become familiar with some of these practices, although because the Framework and NGSS are still very new, these curricula were created to be aligned with past standards and previous paradigms for instruction. We can call these curricula "partially aligned" since they have some elements that will support teachers in implementing the Framework and NGSS ideas, but they were created with other overall intentions in mind. Project-Based Inquiry Science (PBIS) is one of these partially aligned curricula.

PBIS is focused on conceptual understanding and is organized around core science ideas (such as energy). It includes driving questions for each learning set and requires students to build on their understanding over time. Students also are scaffolded into creating and critiquing explanations while they are building this understanding, bringing in new evidence as they go through each unit. Creating explanations (one of the eight Framework practices) is a central instructional shift for teachers using PBIS. Particular activities also target other practices, such as modeling and data analysis, all integrated with the specific content of that lesson.

Due to the partially aligned nature of PBIS, we would expect teachers using PBIS to have more support and resources to help them change their instruction towards what the Framework and NGSS call for when compared to teachers using a more traditional curriculum.

This paper presents findings from the weekly teacher logs of the first year of a two-year study of the implementation of the PBIS curriculum in a large urban district in the U.S. Results from teachers' weekly log self-reports of the frequency of engaging their students in the scientific practices are discussed. The insights from these findings have broad implications as researchers and educators help the nation's teachers and students implement the new Framework and NGSS in the coming years.

Study Context

The work reported here is part of a larger study involving approximately 100 sixth grade science teachers and their students from 42 middle schools in a large urban school district. The district is highly diverse, with 42% African American, 32% Caucasian, 18% Hispanic, and 6% Asian American students. The goal of the study is to understand the impact of the curriculum on student achievement, classroom implementation, and teacher practice. Accordingly, schools were assigned randomly to the treatment (i.e., the curricular intervention) or comparison condition (i.e., the regular district-wide curriculum materials). Teachers in both conditions received professional development in the Framework and NGSS.

Multiple sources of data are being collected in this larger study. Data sources include student assessments, video classroom observations of curriculum implementation, and surveys and field notes from teacher professional development. We are also collecting weekly log data from the teachers (discussed in detail below), a baseline survey (information on teachers' degrees, teaching experience, etc.), teacher assigned student work (specific assignments, information about the assignments, and examples of student work representing a range of ability levels for certain focal units), Edmodo data (an online discussion forum for the teachers), and end of unit surveys (teacher perceptions of the unit, implementation timing, etc.). In this paper we report on results from the weekly logs from the first year of implementation and discuss the implications of these results for connecting weekly log data to the other data sources.

When looking to connect across multiple data sets it is important to recognize the affordances and limitations of each data set and the ways in which it is both possible and theoretically meaningful to connect or link across data sets. The weekly logs give us a unique view into teachers' everyday practice that other forms of data collection do not. However, it is important to remember that the weekly logs are teacher self-reports of their instructional practice and as such, are reflections of not just what is actually going on in their classrooms, but also their *perceptions* of what is going on in their classrooms.

Background

Implementation Logs

Theories of action underlying educational interventions typically assume that instruction—specifically the enactment of key instructional practices emphasized in the materials—is a critical component of efficacy. While investigating the efficacy of interventions, researchers need to (a) develop a picture of instruction; (b) examine how the intervention influences teaching practices; and (c) associate evidence of instruction with evidence of student performance. A major challenge, especially in the context of large-scale studies, therefore, is the "measurement of teaching" (Shavelson, Webb, & Burstein, 1986), namely developing instruments to reliably document and measure instruction during the period of implementation.

Instructional logs offer a valid, reliable, cost-effective way of gathering data regarding the enactment of academic content and instructional practices emphasized by curricular materials (Ball & Rowan, 2004; Rowan & Correnti, 2009). Also referred to as teacher logs or time diaries, instructional logs are self-administered instruments that ask teachers to report on instruction (content topics, instructional strategies, etc.) at specific time points during implementation; data from these periodic reports is aggregated to create a description of instructional practice over time.

An example of the use of instructional logs in large-scale implementation research is the Study of Instructional Improvement (SII)—a large-scale, multi-method longitudinal study of three large Comprehensive School Reform programs (CSR) in a sample of 90 high poverty urban elementary schools. SII researchers have pioneered the extensive use of instructional logs (e.g., Ball, Camburn, Correnti, Phelps, & Wallace, 1999), and have contributed a great deal of literature on instructional logs (see http://www.sii.soe.umich.edu/). Findings from the SII indicate that instructional log data can be (a) equivalent to classroom observation data gathered by highly trained observers, especially for coarse-grained, high frequency aspects of instruction (Camburn & Barnes, 2004); and (b) more reliable than annual surveys of teaching practice, which ask for retroactive reports of activities occurring over long periods of time (Camburn & Han, 2006). Moreover, instructional logs offer an inexpensive method for collecting data in large-scale contexts, as compared to in-person classroom observations or video-recordings of instruction.

Scholarship from the SII also offers guidelines regarding sampling—that is, the number of logs required for each teacher to draw reliable conclusions about the instructional practice in a classroom. Findings from the SII indicate that variation in instructional practice tends to be greater within classrooms rather than across classrooms or schools. In order to obtain a reliable picture of the instruction taking place in a given classroom and to discriminate among classrooms, therefore, requires a large number of logs—approximately 20—per teacher (Rowan & Correnti, 2009; Rowan, Jacob, & Correnti, 2009). Additionally, sampling a larger number of teachers per school helps discriminate across schools, and increases the reliability of school level parameter estimates.

At the same time, instructional logs also pose some data collection and analysis challenges (Rowan & Correnti, 2009). Instructional log data require a large number of logs per teacher and a high response rate from the sample as a whole, increasing the respondent burden considerably as compared to classroom observations or surveys; this can be particularly difficult in large-scale studies involving numerous schools and teachers. Additionally, the data that can be gathered reliably from instructional logs tend to be at a coarser grain, capturing the presence or frequency of instructional practices rather than assessments of instructional quality. Therefore, the conclusions that can be drawn from instructional log data are limited, unless combined with another method of data collection, such as classroom observations.

Scientific Practices

Current reforms in science education encourage the development of science content knowledge through the process of engaging in science practices. According to the Framework, "engaging in the practices of science helps students understand how scientific reasoning develops…and gives them an appreciation for the wide range of approaches that are used to investigate, model, and explain the world" (p. 42). Developing and using models to advance questions and explanations and to communicate ideas is a core practice undergirding science learning, and the framework maintains that students should be asked to “use diagrams, maps, and other abstract models as tools that enable them to elaborate on their own ideas and findings and present them to others” (p. 58). Constructing explanations is also a crucial core practice that is extremely important in both science learning and in the day-to-day use of practicing scientists. Students “should be encouraged to develop explanations of what they observe when conducting their own investigations and to evaluate their own and others’ explanations for consistency with the evidence” (p. 69).

The practice of developing and using models will be one of the more challenging ones for teachers to integrate into their current instructional practice because it asks teachers to do more than just use models to demonstrate ideas (which is the current typical classroom use). The practice of constructing explanations will also be difficult, especially if their students are not used to having to explain their science understanding to others. As part of this larger study of implementation of the PBIS curriculum, we have focused most of our effort on the take-up and integration of the two practices of modeling and explanation. The PBIS curriculum includes explicit modeling activities throughout the units and also, as mentioned earlier, has students continually create and build on their explanations of scientific phenomenon throughout the school year.

Methods

This study uses the log data collection that was part of the larger study of the efficacy of the PBIS curriculum. The study is a randomized control study using an entire school district with teachers assigned by school to one of the two research conditions. Approximately half of the teachers in the school district were randomly assigned by school to the treatment condition which involved using the new project-based curriculum in their classes. The other half of teachers in the comparison condition continued teaching as they normally would.

Online Logs were sent out to all teachers involved in the study each week during three sixth grade science units: an introductory unit at the beginning of the year, the physical science unit, and the earth science unit. For many teachers, these three units constituted most of their school year.

All teachers were sent an email towards the end of each week with a link to fill out their individual implementation log. The logs were to be completed online, similar to a typical online survey. Teachers were encouraged to complete the logs as soon as possible and most completed them within a week of receiving the link. If teachers did not complete them within a week, they were sent an automated reminder email and could fill it out later.

Within each log teachers were asked a series of questions about their teaching during that week of instruction. Each teacher was asked to select one of their classes as a “study class” and they answered the questions about their teaching in that class. Teachers were asked to identify the lessons/activities they taught that week, whether they made any modifications to those lessons (if they were from another source), which state standards they addressed that week, how often they asked students to engage in each of the eight Framework scientific practices (NRC, 2012), and what instructional successes and challenges they faced that week. Comparison and treatment teachers both received logs, however some questions on the logs were tailored specifically for one group or another (for instance, the treatment teachers received specific questions about the intervention curriculum while the comparison teachers received general questions about their curricular context).

The findings presented here are based on the questions about scientific practices. For each of the eight scientific practices (NRC, 2012), teachers were asked how often they engaged their students in these practices. For example, the online log asked teachers not how often they themselves constructed explanations in class, but to indicate how often they provided opportunities for their students to construct explanations. There were four response options for each of the eight scientific practices: *during every class, several times during the week, once or twice during the week,* or *not at all*.

Research Questions

Three research questions led the investigation of this set of weekly log data:

1. How often do teachers report engaging their students in the scientific practices?
2. Does use of a partially aligned curriculum increase classroom engagement in the scientific practices?
3. What factors are associated with increased classroom engagement in the scientific practices?

Findings

This paper presents findings relating to the teacher weekly log data set from the first year of the study. The results presented here are from the first year of data collection in this two-year study and from the Physical Science (Energy) unit. First, we describe the basic results from the completed logs. Then we explore visualizations of the variation in log results across and within teachers and look for ways to account for this variation, both for just the treatment teachers (for whom we have more data) and across conditions when appropriate.

Completion of weekly logs

There were 45 teachers in the comparison condition and 57 teachers in the treatment condition. There was a wide disparity in the completion rates and average number of completed logs between the two conditions. 15 teachers in the comparison condition did not complete any physical science logs while only 2 teachers in the treatment condition did not complete any physical science logs. For those teachers that did complete logs, the comparison teachers still completed substantially fewer logs than the treatment teachers.

Table 1. Completion of weekly logs

	N	Energy logs	Earth Science logs
Comparison teachers	45	4.5 (56%)	3.8 (44%)
Treatment teachers	57	13.9 (89%)	5.3 (66%)

In Table 1 above, data on the number of logs completed and the completion rate by condition and unit are given. For the Physical Science/Energy unit, treatment teachers completed a much higher number of logs (an average of 13.80 versus only 4.11 for the comparison teachers) and also were much better about completing the logs (completing 89% of the logs they were sent, compared to the comparison teachers who only completed 56%). The treatment teachers spent a much longer amount of time on their Physical Science/Energy unit (from the PBIS curriculum) than the comparison teachers and much longer than their district pacing guide had projected (about 8 weeks).[1] For the Earth Science unit, which followed the Energy unit for most teachers, the number of completed logs (and therefore the amount of instructional time) was similar across conditions. However, the treatment teachers still completed a higher percentage of logs than the comparison teachers (66% versus 44%), although not as high as earlier in the school year.

Scientific Practices

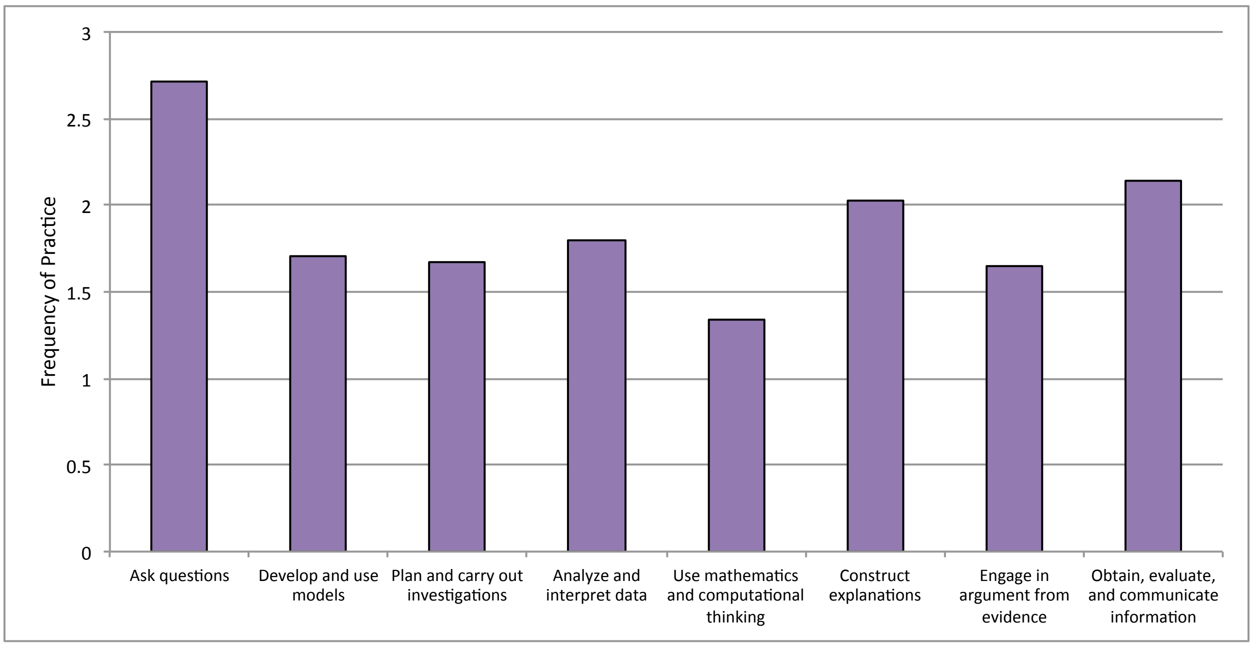

Figure 1. Teacher weighted averages for each of the eight scientific practices, showing how frequently they report they engage their students in the practice.

As shown in Figure 1 above, there was a difference in frequency of engagement in scientific practices for the eight different practices. A score of 1 on the figure above corresponds to *once or twice during the week*, 2 corresponds to *several times during the week* and 3 corresponds to *every class*. For many of the practices, teachers reported that on average they engaged their students in scientific practices a few times a week. There are two clear outlier practices, though. The first one, “ask questions” has an extremely high rate of engagement with nearly all teachers reporting having their students “ask questions” during every class. As described by the Framework, this practice works towards the goal of having students ask their own “well-formulated questions that can be investigated empirically” (NRC, 2012, p. 55). Another practice of note is the low frequency for the use of mathematics and computational thinking (this has the lowest of the scores above in Figure 1 and only 40% of logs reported this being a frequent practice).

---

[1] One possible explanation for this could be that teachers in the treatment condition were spending more time than expected because they were trying to get used to the new curriculum. It is also possible that the curriculum actually took much longer to implement than the district or publisher expected.

[2] There are some notable exceptions to this. Many of the teachers in Figure 3 have very little or no

Visualizing Variation

We now turn our attention to approaches to visualizing the weekly log data. Since we have data points for each week (of a completed log) then it makes sense that we might be able to gain a better understanding of the extent to which teachers are helping students engage with the scientific practices. One difficulty with doing this is now instead of a couple variables to compare and contrast or, as in our case, 103 variables to look across, we have more than an order of magnitude more data points to make sense of for each variable of interest.

This lends itself to a data visualization approach. As with any set of data, it is always good practice to be able to look at the data set before analyzing in order to ensure that the scope and range of it and the analytic options that are appropriate. Figure 2 below is an attempt at visualizing the data set of treatment teachers' responses about the practice of *developing and using models* for the physical science unit. This figure contains only treatment teacher data because we have more logs for these teachers and it is easier to visualize trends over time (and also due to space concerns).

		1	2	3	4	5	6	7	8	9	10	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
SCH_ID	TCH_ID	Sept	Sept1	Sept1	Sept2	Oct1	Oct8	Oct1	Oct2	Oct2	Nov5	Nov1	Nov2	Dec3	Dec1	Dec1	Jan7	Jan1	Jan2	Jan2	Feb4	Feb1	Feb1	Feb2	Mar4	Mar1	Mar1	Mar2	Apr8
10	171						2	4	3	4	4	1	3	3	3	4	3	3	1	3									
10	172																												
10	173		1		4	4	4	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1						
11	102						2	3	2			4	3	3	2	3	3	3	3	3	3	3	3	3					
11	225				3	3	3	3					3	3	3	2	2	3	3	3				4	2				
11	226						2	1	3	2	1	2	2	2	2	2	3	3	2	1	1	3	2	2	2				
13	103		1	2	2	2			2	2	1	2	2	2															
13	104				2	2	2	2	2	2	2	2	2	2															
13	162		2	3	3		3	4	3	3	4	3	4	3		4	2												
13	250							3	3	4	3	3	3	3	1	3	3	3	4										
13	251					3	2	2	3	3	3	1	2	2	3	4													
14	105			1	3	4	3	3	4	4	2	4	3	4	2	2	4	2	4	2	2	2	4		3				
14	177		1	2	3	3		2	4	2				2	3	4	3		2	2	3	2							
17	106						2	1										3	1	2	2	2							
17	264																1		4	1	1	1							
18	107				3	3	4	3	3	2	3	3	4	4	4	4	2	3	1										
20	109				2	1	1	1	2	2	1	1		1	1														
20	229			2	2	2	2	2	3		3	2	2	3	3														
20	252																												
21	184			1	2	2	1	1	1	3	1	1	2	3		1	3	3	3		2								
21	230					1	1	1	1	1	1	1	4	4	1		1	2	1	1	1								
21	253				2	3	1	1	2	1	3	1	2	4	2	3	3	4	4	2	2								
23	110							2	2	2	2		1	2	2		2	2	2	2	2	2	2						
23	112					2	3	1	2	2	1		3	2	3	3	3	3	3	3	3	3							
23	231							1	2			1				1	1	1			2		2						
23	232			2	2		1	2	2	2	2		2	2	2	2	3	4	3	2	2	2	2	2					
26	113										4		3				4	4	4	4	3								
26	114								2	2	3	3	3	2	4	1	1	3	1		2	4							
28	115			1	2	3	3	2	3	3	2	4	3	3	3	3	3	3	3	1	3								
28	116			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1							
31	118					1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4						
31	119									2	2								2			3	2	4					
31	233							2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3						
31	262								2	2		1	1	3	3		3	2	2			1	2	2	3				
35	120			2	2	3	3	3	3	2	2	2	2	1	3	3	1	3	3	3	1	1							
35	189							3	2	4	3	3	3	3	3	3	1	3	3	3	3	3							
35	254				2	2			2	2	2	3	2	2	1	2	2	3	3	2		2							
36	235							2	2	2	2	2	2	2	1	2	2	3	2	2	1	1							
36	236										1		4	4	4	4	4						1	3					
36	263																		1			1	1						
41	123									3	1	3				2	2	4	3	3	3	3	1	2	1	3	1	1	
41	190			2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	1	1				4	3	4
41	237								4	4	4	4	1	4	4	4	3	3	4	4	4	3	4		4	4			4
43	238					2	3	3	3	3	4	3	4	3	3	3	3	2	3	2	3	4	4						
43	241					2	2	2	1	3			1		2	4	2		2	3	3	3	3	3					
43	242					2	3	3	2	2													3						
43	255					2	2	2	2	2																			
44	126			1	3	2	2	3	3	3	2	3	2	2	3	3	1	2	2	2	3								
44	128					3	3	3	3	3	3	3	3	3	2	3	4	3	4	4									
44	243		2	2	2	2	1	1	2	2	1	1	1	2	2	2	1	1	2	2	1	1							
44	244		3	3			3	3		3	4	1	2	2	3	3	1	3	1	3									
47	129								3	3	3	3	4	4	3	3	4	4	3		4	3	4	2	2		3	3	3
47	245								3	2	1	1	1	1	1	1	1	1	2	1	1								
47	246			4	1	1									1	1	1	1		1	1	1	1	1	1		1	1	
48	247				1	2	4	3									2	3		3			2	2					
49	133									1	1	1	1	1	2	2	2	3	3	2	1	2	2	2	4	3	3	4	
50	200								3	3	3	4	4	3		4	3	3	2	3	3	3	3	1	2	3	3		

Figure 2. Modeling score heat map for treatment teachers in the physical science unit. Each row represents one teacher and each column represents one week of data collection. A darker color corresponds to more modeling that week.

In the visualization in Figure 2 cells are colored to correspond with response options. For example, the darkest colored cell corresponds to the option *every day* (modeling each day that week) while the very light shading corresponds to *not at all* (no modeling that week). Blank cells represent when a teacher did not complete a log (sometimes because they were not supposed to fill out one – mostly on the left and right sides of the figure – and sometimes because they forgot

– mostly the blanks in the middle). Each row represents a teacher and each column is one week of data collection.

This view of the teacher data gives us a better sense of the variation within each teacher. This is where visualization can make a big difference. We are able to quickly see what the spread in responses is by teacher and how some teachers who may look similar with just their average score are not as similar at this week-to-week view. For example, teachers 49_133 and 35_254 both have an average modeling score of 2.1, so from our earlier views we might conclude they are engaging their students in modeling in fairly similar ways or at least in similar quantities. However, this visualization shows us that actually teacher 49_133 (towards the bottom of Figure 3) spent a lot of time on modeling in the beginning of the physical science unit, but then less and less as the unit continued, while teacher 35_254 fairly consistently incorporated some modeling throughout the unit. Another pair of teachers, 21_253 and 20_229, also had very similar average modeling scores (2.3) and very different patterns of modeling implementation. Teacher 21_253 had a wide of response options and spent a good amount of time on modeling towards the end of the unit while teacher 20_229 only responded with 2s and 3s throughout the unit.

We found that the average is a good measure of overall take-up of modeling practice, but not a good window into what is happening in each classroom on a smaller time scale. There turns out to be a lot of variation across the weeks for each teacher. Additionally, even if the frequencies of responses are similar, *when* these different responses is only clear here with the visualization.

This new view then also allows us to add in more information about the teachers and their instructional settings to help us better understand why these different patterns appear. For instance, we could expand upon the information about the learning sets to look at not just which weeks teachers were doing modeling or not doing modeling but also which learning sets they were teaching that week.

Accounting for Variation

Although this visualization allows us to see the amount of variation across all teachers and all across all responses, it does not easily allow us to see or account for variation within teachers. For most teachers, there is a good amount of variation in their reported modeling score across the weeks of data collection.[2]

Variation across teachers could be due to many factors, including how comfortable and/or familiar teachers are with the practices and whether or not their prior teaching included this type of instruction where engaging students in scientific practices was commonplace. When trying to account for this variation within teachers, however, different factors come into play. The most obvious one is the curriculum. As mentioned above, teachers also reported on which of the learning sets they taught each week. A basic look at the percentages did not reveal any information about the variation within teachers. But, looking at the data set with more detail, we can then match up modeling scores with information about learning sets to create statistical models.

---

[2] There are some notable exceptions to this. Many of the teachers in Figure 3 have very little or no variation in their scores. One possible explanation for this is that these teachers just reported doing lots of modeling to quickly finish their logs. Many of these teachers also reported consistently high scores on the other seven practices. There is more discussion of this later on.

Modeling

A hierarchical linear model was created to account for the variation in the treatment teachers' modeling scores in the physical science unit. The model, with teachers added in as a grouping variable revealed that 40.0% of the variation among modeling scores was found at the teacher level. Multiple models were then run, adding in each of the first four learning sets (those that teachers spent the most time on). Adding in any of the learning sets to the model accounted for an additional 4% of the variance to be accounted for. There was no difference among the four learning sets and a model including all four did not reveal additional information. Figure 3 below shows the relationship between mean teacher modeling score and their residuals associated with the learning set variable (learning set 1 in this case). There is little deviation from the line, showing the small advantage of adding in learning set information to the statistical model.

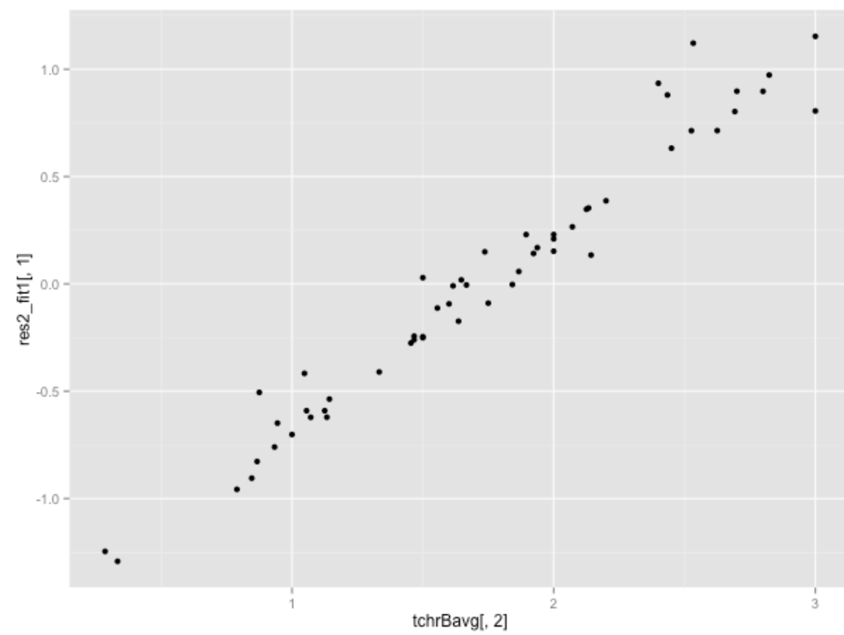

Figure 3. Residuals for learning set 1 plotted against each teacher's mean modeling score.

Explanation

A similar hierarchical linear model was created to help account for the variation in the explanation score for treatment teachers in the physical science unit. As above, with the practice of modeling, the model with only teachers as a grouping variable revealed that 43.6% of the variation in explanation scores was accounted for at the teacher level. Again, as with modeling above, adding in any of the learning sets to the model accounted for an additional 4% of the variance to be accounted for. There was no difference among the four learning sets and a model including all four did not reveal additional information.

Comparing Across Conditions

Next we look at three figures that explore, for three of the eight practices, the average (mean) score for each teacher in the study along with the standard deviation of their scores for that particular practice during the physical science unit. The standard deviation of the scores is important for two reasons: (1) we expect a certain amount of variation in the amount of the practice that the students do over time, as each lesson or learning set is focused on different types of knowledge-in-use (in other words, we don't expect a teacher who is well aligned with the Framework to teach every practice every day in her class) and (2) looking at the standard deviation alongside the mean can give us a sense for some teachers of how accurate their self-report data is. Looking at these two numbers and plotting them for all of the teachers in the study can give us an overall sense of how often teachers were engaging students in the different practices and how much this changed over time.

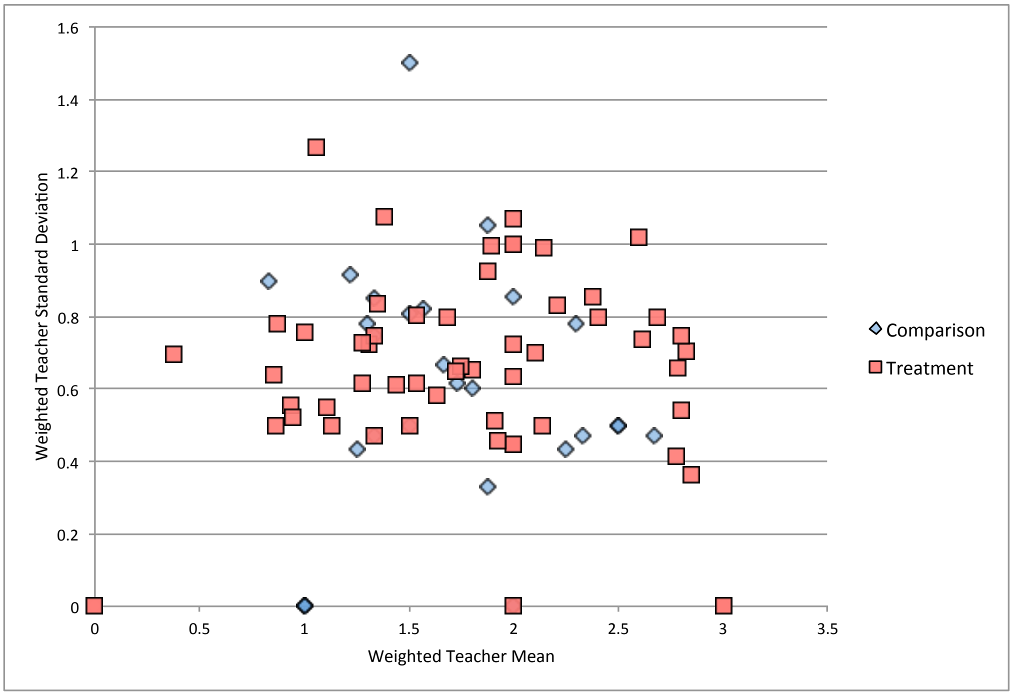

Figure 4. For the practice of Analyzing Data, a plot of teacher averages versus teacher standard deviations.

Figure 4 shows the means and standard deviations for the teachers, colored by research condition, for the practice of Analyzing Data. This is one of the practices that we expect teachers in the comparison and treatment conditions to implement at fairly comparable levels. We can see from the figure that the two groups of teachers overlap a lot and that there are few outliers among them. There do seem to be more treatment teachers with higher scores than comparison teachers, appearing in a clump on the upper right side of the figure.

Important to note is that there a few teachers, in both conditions, that have points that lie along the x-axis (a standard deviation of zero), meaning that for all of their logs, they always reported the same amount of that practice. Because we expect to see a fair amount of variation from week to week, these numbers are a little bit suspect, but especially so for the teachers that report doing the practice every day (a value of 3 in Figure 4) every single week.

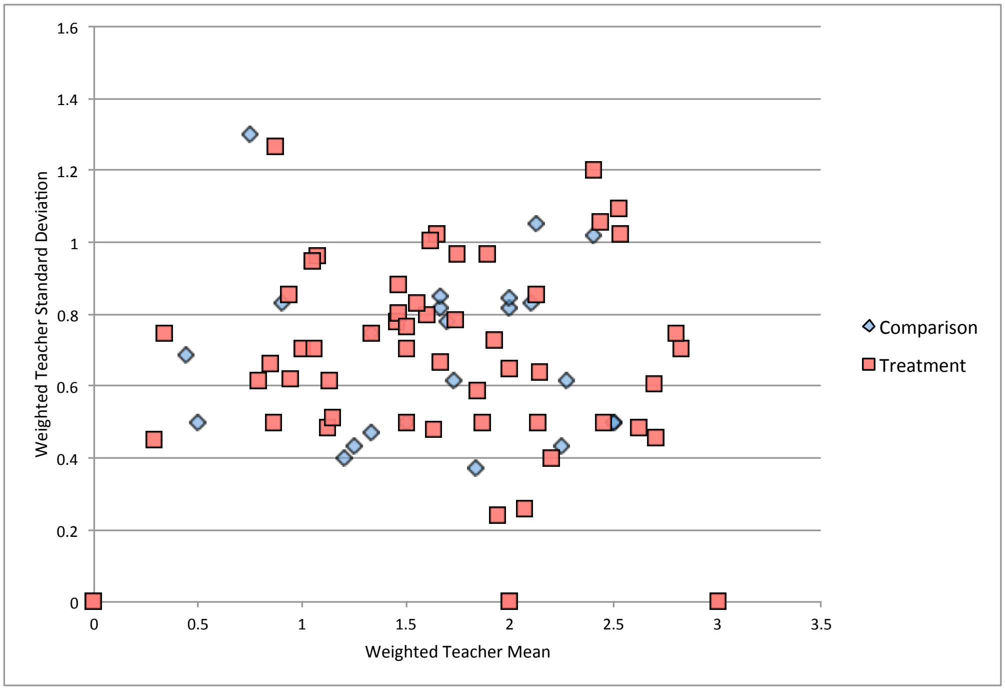

Figure 5. For the practice of Developing and Using Models, a plot of teacher averages versus teacher standard deviations.

Figure 5 shows the means and standard deviations for the teachers, colored by research condition, for the practice of Developing and Using Models. Although the PBIS curriculum does not focus on this practice frequently, it is a major emphasis in a few lessons in the physical science unit. Additionally, helping teachers understand and implement more modeling and model-based reasoning in their instruction was one emphasis in the professional development that the treatment teachers received related to the PBIS curriculum. We would expect some difference among teachers in the two groups if the curriculum-specific PD had an impact on their teaching. Both groups did have some professional development on the new Framework before the beginning of the school year. As we can see in the figure, the two groups mostly overlap so there is not good evidence for a difference between the two groups as a whole. There are a handful of treatment teachers that have high scores, so it seems possible that a teacher-level variable is influencing the amount of modeling that teachers implement in their classrooms.

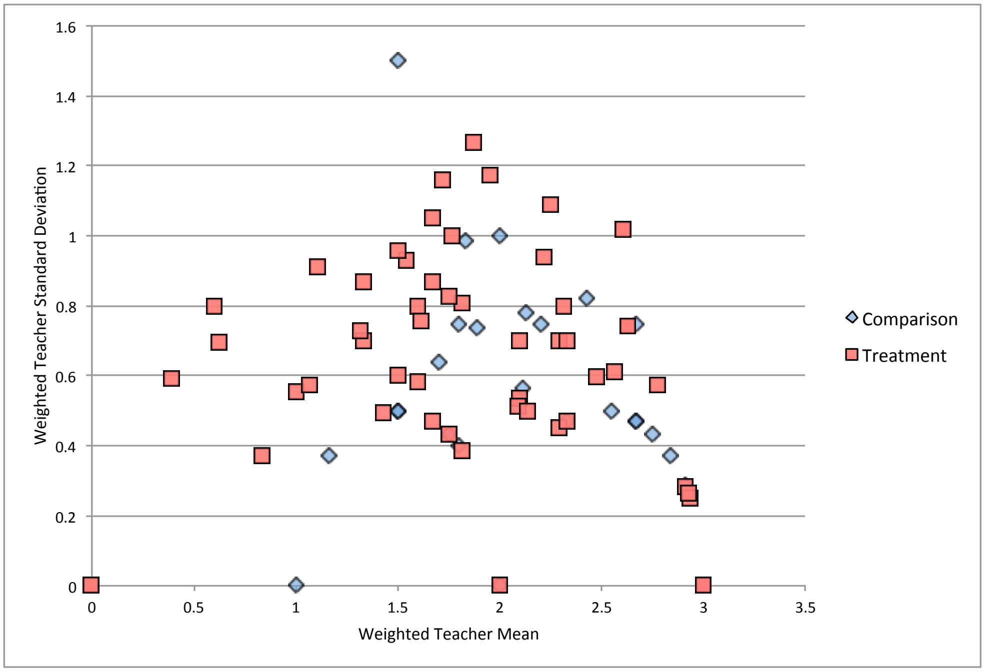

Figure 6. For the practice of Constructing Explanations, a plot of teacher averages versus teacher standard deviations.

Figure 6 shows the means and standard deviations for the teachers, colored by research condition, for the practice of Constructing Explanations. We can see that there are more treatment teachers on the low end of the means, meaning that there were a good number of treatment teachers that were reporting having their students construct explanations on average about once or twice a week. Most of the comparison teachers reported doing this on average almost every day. There is also another clear pattern here, where a group of teachers from both conditions (but more from treatment) have a very high mean score (very close to 3) and very low standard deviation, meaning that they almost always reporting engaging their students in the practice of constructing explanations every day of class. Most of the teachers in this group also reported similarly high and unvaried amounts of the other seven practices.

Discussion

There are two main findings that are particularly relevant for discussion. The first involves the amount of time that teachers spent teaching their physical science units. The second involves the differences in the amount of reported engagement in various scientific practices, especially constructing explanations.

As evidenced by their completion of the weekly logs, we saw that treatment teachers spent an average of 13.80 weeks on the physical science unit while comparison teachers spent an average of only 4.11 weeks. The comparison number is likely slightly lower than it actually was, due to fewer log completions by that group. However, both groups of teachers were using the same district pacing guide that asked teachers to complete the unit within eight weeks (this number was negotiated with the curriculum publisher as well). Very few treatment teachers were able to stay on this schedule which underlines the difficulty in implementing (a) a brand new curriculum and (b) a curriculum that focuses a lot more on student understanding of core concepts and integrating some scientific practices throughout content learning. While PBIS is only partially aligned to the Framework goals, it seems clear that teachers (and districts) who want to implement curriculum that fully or partially addresses the goal of the Framework to teacher fewer core ideas in more depth while integrating practices and cross-cutting concepts will need to allow more time to do this. It will take time for teachers to make the instructional shift that the Framework and NGSS ask teachers to make. We hypothesize that the second year of data collection (currently ongoing) will show that the treatment teachers are able to reduce their overall time of implementation by some amount.

When looking at the differences (or lack of differences) between comparison and treatment teachers on the eight scientific practices, there are at least three possible explanations for why we might or might not see differences between the two groups: (1) different teachers have different understandings of what that practice should look like in the classroom (which may or may not be affected by supports in the curriculum), (2) teachers have not changed their instruction much in order to align with the Framework and therefore show similar patterns of implementing the practices across conditions, and (3) different teachers are actually implementing the practices to differing degrees due to curriculum supports or professional development preparation.

Constructing explanations is the main practice in which we see difference across conditions. One possible explanation for this difference is that teachers in the treatment condition, who are explicitly supported via the unit materials to engage their students in meaningful activities around constructing and critiquing explanations, have a different understanding of the practice and therefore report it only when they are really engaging their students in constructing scientific explanations (and not just whenever they ask students to give or explain an answer). It seems likely that this more nuanced understanding of what the practice looks like has the unintended effect of resulting in a lower amount of that practice being reported (relative to comparison teachers).

It is of course possible that comparison teachers are engaging their students more frequently in the types of explanation-based activities that the Framework calls for, but this is less likely, due to what we know about typical classrooms (which the comparison ones are, by design). It seems much more likely that the comparison teachers have a more broad and less nuanced view of what *constructing explanations* looks like in the classroom, as defined by the Framework. A possible next step to uncovering how likely this interpretation of the data is, would be to conduct follow-up surveys and/or interviews with all of the teachers and ask them to articulate or describe what is going on in their classroom when their students are engaging in that practice. This would help us better understand if this difference in explanation self-report is due to differing views on what *constructing explanations* looks like or due to actual differences in implementation. If the comparison teachers were receiving additional professional development and resources to help implement a Claims-Evidence-Reasoning (CER) approach to explanations, this could also explain some of the difference. Checking on this possibility is a next step for our team.

In general, our findings suggest that many teachers may not yet fully understand the vision of the scientific practices as laid out in the Framework. This is most evident with the *asking questions* practice. In a typical classroom, students are not usually given opportunities to formulate their own questions. Nearly all teachers reported engaging their students in this practice in nearly every class. From our classroom observations we have seen that only on rare occasions are students asking each other questions about scientific phenomena and the conclusions of their investigations. These questions rarely, if ever, lead to new investigations and hypotheses. This result may indicate that teachers interpreted the survey as asking about any kind of generic questions that either they or their students ask. Due to this possibility, the wording of the questions was altered for the second year of data collection in order to reinforce the intended question about the Framework practice. We emphasized that the students, not the teachers, should be the ones that are engaging in the practice and also reminded teachers about the Framework definitions of the eight scientific practices.

Limitations

There is a clear limitation in relying solely on self-reported information in the form of weekly implementation logs. As mentioned above, the weekly implementation logs are one part of a larger study of implementation. Video observations of specific focal lessons of interest that are built around specific scientific practices have been collected and are being analyzed. Also, teacher assignments and student work that accompany these focal lessons have also been collected. Together, these three pieces of help triangulate information about the teachers and students in the study. This will help us, to a certain extent, and at least for a subset of the teachers in the study, better understand the magnitude of the limitations on the self-report log data.

Conclusion

This approach to analyzing and visualizing teacher implementation of new instructional practices and curriculum, can help us better understand the variation both within and across teachers in a large study. It will help us also understand how well teachers are able to understand what is asked for by the new Framework and NGSS and whether or not they are embracing the idea of knowledge-in-use.

Collecting this kind of observational data at this large of a scale would be prohibitive for many reasons (including cost and time), however online weekly implementation logs are able to give us an approximate snapshot into what is happening in our study classrooms. Teachers are also given an opportunity to reflect on their practice, which many of them have reported informally as being useful for them as professionals.

The visualization in Figure 2 also allows us to look at larger patterns that unfold, characterize variation in teacher responses, and eventually incorporate other instructional factors to help discern meaning behind the patterns. This approach, combined with other data collection techniques as part of a careful study, can give us new insights and perspectives on looking at teacher practice and student learning.

The use of weekly implementation logs and corresponding visualization and analysis tools allow us a unique insight into teacher practice and student learning and engagement. With the rollout of new science standards, this work begins to address how to analyze and understand teachers' roles in helping students learn with the new emphasis on integrating core content ideas, cross-cutting concepts, and scientific practices.

Collecting weekly log data of teachers' implementation is a relatively easy way to gather information about what is happening in teachers' classrooms while being minimally intrusive into those classrooms. The log data is one piece in this larger study, helping to shed light on teacher practice and implementation of a new curriculum over a long period of time (most of the school year) and also connecting various threads of data together. The analysis of the log data, both by itself and within the context of the rest of the study, can serve as a model for other large implementation and efficacy studies similar to this. Learning more about how teachers respond to these types of instruments and what kinds of inferences we can make from this data will be helpful for understanding teacher practice and implementation more broadly.

References

Ball, D. L., Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). New tools for research on instruction: A web-based teacher log. Seattle, WA: University of Washington: Working paper, Center for Teaching Policy.

Ball, D. L., & Rowan, B. (2004). Introduction: Measuring instruction. *The Elementary School Journal, 105*(1), 3-10.

Camburn, E., & Barnes, C. A. (2004). Assessing the validity of a language arts instruction log through triangulation. *The Elementary School Journal, 105*(1), 49-73.

Camburn, E., & Han, S. W. (2006). Factors affecting the validity of teachers, reports of instructional practice on annual surveys. Madison, WI: Wisconsin Center for Education Working paper of the Consortium for Policy Research in Education. Research.

Crawford, B., & Cullin, M. (2004). Supporting preservice science teachers' conceptions of modeling in science. *International Journal of Science Education, 26*, 1379 - 1401.

Davis, E., Nelson, M., Hug, B., Kenyon, L., Cotterman, M., & Teo, T. W. (2010). *Preservice teachers and scientific modeling: Synthesizing results of a multi-year, multi-site project*. Paper presented at the Annual Meeting of the Association for Science Teacher Education, Sacramento, CA.

Harrison, A. (2001). How do teachers and textbook writers model scientific ideas for students? *Research in Science Education, 31*, 401 - 435.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 371 - 388). Cambridge University Press: Cambridge, MA.

National Research Council. (2012). A framework for K-12 science education: Practices, cross-cutting concepts, and core ideas. Washington, DC: The National Academies Press.

Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian & P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery* (pp. 5 - 22). New York, NY: Kluwer Academic/ Plenum Publishers.

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. Educational Researcher, 38(2), 120-131.

Rowan, B., Jacob, R., & Correnti, R. (2009). Using instructional logs to identify quality in educational settings. New Directions in Youth Development: Research, Practice, Theory, 121(Spring), 13 - 31.

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. Handbook of research on teaching, 3, 50-91.

Schwarz, C. (2009). Developing preservice elementary teachers' knowledge and practices through modeling-centered scientific inquiry. *Science Education, 93*(4), 720 - 744.

Windschitl, M., Thompson, J., & Braten, M. (2008a). Beyond the scientific method: Model-based inquiry as the new paradigm of preference for school science investigations *Science Education, 92* (6), 941 - 967.
Windschitl, M., Thompson, J., & Braten, M. (2008b). How novice science teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction, 26*(3), 310-378.

Acknowledgments
This work was supported by a grant from the National Science Foundation.