Investigation of Video-based Multidisciplinary Online Professional Development for Inservice High School Science Teachers

Susan M. Kowalski, Betty Stennett, Mark Bloom, and Karen Askinas

BSCS Science Learning

Susan M. Kowalski, Betty Stennett, Mark Bloom, and Karen Askinas, BSCS Science Learning.

This project was funded by a grant from the National Science Foundation, Award #1118643. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Susan Kowalski, BSCS Science Learning, Colorado Springs, CO 80918

Contact: skowalski@bscs.org

Abstract

We report findings from the development and testing of an online, multidisciplinary, video-based analysis-of-practice professional development (PD) model for inservice high school science teachers. The study used a cohort control, quasi-experimental design to investigate impacts of the PD on students, and a pre-post design to investigate effects on teachers. This study explores whether a multidisciplinary, online, video-based PD model can enhance science teaching and learning as narrowly focused, face-to-face models have done in the past. We find that translation of complex analysis-of practice PD models for a multidisciplinary audience in an online environment is not without challenges. Although we found significant changes in teacher knowledge and nearly significant changes in teacher practice, the effects did not lead to significant enhancement of student achievement, and effects on students varied strongly by teacher. We investigated teachers' online reflections and comments to better understand the affordances and challenges of the multidisciplinary, online model.

Teacher knowledge and instructional practice are influential in supporting student learning (Hill, Rowan, & Ball, 2005; Kersting, Givvin, Sotelo, & Stigler, 2010; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In math and science education, there have been important developments in using video-based, analysis-of-practice professional development (PD) to enhance teacher knowledge and practice (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Seidel, Blomberg, & Renkl, 2013; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011; Sherin & van Es, 2005) and enhance student learning (Roth, Garnier, Chen, Lemmens, Schwille, & Wickler, 2011; Taylor, Roth, Wilson, Stuhlsatz, & Tipton, 2017; Wilson, Stuhlsatz, Hvidsten, & Stennett, 2017). Given the critical role that teachers play in classrooms everywhere and the striking findings from video-based, analysis-of-practice PD models, it's important to consider how to make video-based, analysis-of-practice PD accessible to broad audiences.

There are two characteristics of existing video-based PD that limit its accessibility. First, most video-based, analysis-of-practice PD has a narrow content-focus (Borko et al., 2008; Roth et al., 2011; Taylor et al., 2017). For example, Borko and colleagues (2008) describe PD experiences that revolve around solving specific mathematics problems, first among a group of teachers and then as teachers guide students in solving the same problems. Likewise, Taylor and colleagues (2017) describe PD for teachers linked to approximately three weeks of instruction. Second, many video-based, analysis-of-practice PD models require extensive face-to-face time with PD leaders—Borko and colleagues describe a two-year program, and Taylor and colleagues describe 89.5 hours over one full year.

The combined effect of a narrow content focus and extensive face-to-face interactions is particularly problematic for high school science teacher PD. High school science teachers tend to occupy distinct disciplinary niches, focusing entirely on biology, chemistry, physics, or earth science. Across a single district, there may be only a handful of physics or chemistry teachers and possibly a single earth science teacher. The narrow disciplinary focus and extensive face-to-face hours of video-based, analysis-of-practice PD thus pose a limitation for all but the largest districts in the country: there simply are not enough high school teachers within a single discipline in a single district to warrant the expense of extensive face-to-face, discipline-specific, video-based PD.

Cognizant of potential accessibility limitations of video-based, analysis-of-practice PD, we developed a multidisciplinary, online, high school science PD course and rigorously studied it over two national field tests. The course is known as *Energy: A Multidisciplinary Approach for Teachers*, or EMAT. Our overarching research questions were (1) to what extent does facilitated, online, video-based, multidisciplinary, analysis-of-practice PD support teacher knowledge and practice and enhance student learning? and (2) what can we understand from teacher online reflections and comments about the affordances and limitations of the multidisciplinary online model? To answer our research questions and in consideration of calls to rigorously study PD models with both teachers and students (Dede, Jass Ketelhut, Whitehouse, Breit, & McCloskey, 2009), we collected teacher knowledge and practice data using a pre-post design as well as student achievement data using a cohort-control, quasi-experimental design.

Although this study was not conducted as a direct comparison of online and face-to-face PD, we contextualized results from our study considering those from comparable outcome measures for the narrowly focused, face-to-face model. The comparison is possible because we used similar outcome measures and instruments used by the narrowly focused, face-to-face model. Overall, the multidisciplinary, online model had substantially weaker positive effects than the narrowly focused, face-to-face model. Most importantly, the multidisciplinary, online model failed to significantly enhance student achievement. In this paper we describe in detail our findings related to both teachers and students to better understand the limitations of the multidisciplinary, online model and consider the potential implications for those with an interest in improving the reach and accessibility of face-to-face, video-based, analysis-of-practice PD models.

Literature Review and Conceptual Framework

Characteristics of effective PD are emerging from research over the past two decades (Ball & Cohen, 1999; Desimone, 2009; Desimone & Garet, 2015; Roth et al., 2017). We discuss in this literature review two components of effective PD central to our PD model: (1) active learning (Borko, 2004; Desimone & Garet, 2015; Garet et al., 2001) and (2) situated cognition (Putnam & Borko, 2000; Roth et al., 2017). We describe how video as a tool can support both active learning and situated cognition. We conclude our literature review by describing general principles associated with facilitating effective video-based PD and discuss current literature related to online, video-based PD.

Active Learning

Active learning of various flavors has been touted for decades as foundational to deep learning (Dewey, 1910; Inhelder & Piaget, 1964; Schwab, 1960; Vygotsky, 1960). More recently, researchers included active learning as a core feature of effective professional development (Borko, 2004; Desimone, 2009; Garet et al., 2001). Desimone (2009) described active learning related to pedagogy as including features such as observing teaching or being observed in conjunction with discussion and/or feedback, analyzing student work, and leading discussions. Likewise, Borko (2004) included opportunities to solve mathematics problems or conduct scientific experiments as hallmarks of active learning related to content knowledge.

We embrace active learning as a key ingredient to effectively support teacher learning. But given the plethora of definitions and theoretical frames associated with active learning, we describe how we instantiate the active part of active learning in our online, video-based, analysis-of-practice PD model. Within our framework, active learning begins when a learner articulates his or her current understanding about a key idea (related to either content or pedagogy). The learner then examines data associated with the key idea: data from a physical phenomenon for science concepts or data from classroom artifacts for pedagogical concepts. Depending on the goals of instruction, the learner either collects his or her own data or simply analyzes provided data. In either case, actively identifying patterns in data and interpreting what those patterns might mean in relation to the key idea is paramount.

Having considered data, the learner reflects on whether the data match what he or she would have anticipated, based on initial ideas. Here, reflection is an opportunity to surface any contradictions between data learners are analyzing and initial conceptions. When learners in a group all analyze common data, they gain a reference frame from which to engage in discussions and interpretation about what the data mean. Depending on the complexity of the concept and the availability of an instructor or peers, the learner can either work with others to develop an explanation for the data or can participate in more traditional learning activities such as reading an explanatory passage.

Once learners have had the opportunity to consider an explanation, they reflect again on their evolving ideas, noting any changes that have occurred in their thinking. They then either apply their learning in a new context or extend their learning to incorporate related ideas. Our approach is aligned with the work of Bybee and colleagues (2006) and pays strong attention to the importance of providing learners the opportunity to track how their thinking has changed over time (Bransford, Brown, & Cocking, 2000). We applied this active learning framework to support the development of both pedagogical and science disciplinary understanding in the EMAT course.

Situated Cognition

Situated cognition theorists reject the notion that knowledge exists independent of social contexts and applications (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991; Putnam & Borko, 2000). Creating opportunities for teachers to learn scientific and pedagogical concepts in the context of how they might teach key science concepts to students should enhance teacher learning over what might occur in a traditional lecture course. According to Borko and colleagues (2008), PD does not need to occur in the classroom for learning to be situated. Instead, artifacts such as video, curricula, student work, and lesson plans can be studied in a PD setting.

If artifacts can be studied in a PD setting to support situated cognition, how closely should those artifacts match a teacher's instructional context? Should the artifacts include exact lessons or problems that teachers will use with students? Alternatively, is it sufficient if artifacts focus on key concepts of relevance to all teachers and showcase those key concepts in varied classroom contexts, but do not strictly situate the crosscutting concepts in the exact lessons that teachers will enact? Ball and Cohen (1999) indicate that practice-based PD requires three things: identification and instruction in instructional practices that are central to teaching and learning, the creation and use of PD materials that showcase the work of teaching, and then using the materials to support teacher learning. These characteristics of practice-based PD suggest there may be room for some flexibility in how closely matched supporting artifacts are to contexts in creating situated learning environments—but a question remains: how situated is 'situated enough' to enhance teacher learning and ultimately impact student achievement?

Video as a PD Tool to Support Active Learning and Situated Cognition

With appropriate scaffolding (van Es, Tunney, Goldsmith, & Seago, 2014), video can promote active learning. Desimone (2009) included observing other teachers and being observed, followed by discussion or feedback, as one form of active learning—and video readily affords the opportunity for observation, discussion, and feedback. Teachers can analyze and interpret video for evidence of student thinking and reasoning and consider teacher moves in response to that thinking and reasoning.

Video analysis in concert with facilitated discussion can also support situated cognition. With video analysis, cognition is situated in relevant contexts because teachers are seeing the application of content knowledge and pedagogical practice in the settings in which teachers themselves will be working. Although there appear to be differences in the type of learning afforded by analyzing one's own video (intimate context) compared with analyzing the video of others (more distal context), both contexts can promote learning (Seidel et al., 2011).

General Principles Associated with Effective Video-Based PD

Because video is a medium for content, not the content itself (Brophy, 2004), facilitators must formulate clear objectives for using video in PD (Blomberg, Renkl, Sherin, Borko, & Seidel, 2013; van Es, 2009) and carefully attend to facilitation of video analysis to meet those objectives (van Es et al., 2014). Van Es and colleagues (2014) identified four practices facilitators can use to enhance the productive use of video, including (1) sustaining an inquiry stance, (2) supporting group collaboration, (3) orienting the group to the video task, and (4) maintaining a focus on the video and subject matter.

Existing Literature Related to Online, Video-based PD

Most of the literature for video-based professional development relates to face-to-face PD (Borko et al., 2008; Roth et al., 2011; Taylor et al., 2017). There is no reason to assume that media or the modality of instruction should influence learning (Clark, 1983), and some studies have shown little or no difference in efficacy of PD by changing the modality of delivery (Fisher, Schumaker, Culbertson, & Deshler, 2010; Fishman et al., 2013). But with a highly complex model such as video-based, analysis-of-practice PD that requires careful facilitation and support for teachers, it is not obvious that an online modality would be as effective as face-to-face. There is little research related to the effects of online, video-based analysis-of-practice PD. Studies by Koc, Peker, and Osmanoglu (2009) and Nemirovsky and Galvis (2004) examined teacher discussions of video in online forums, but the focus of the research was on the nature of teacher discourse rather than on the effects of the PD on either teachers or their students. Our study goes beyond these earlier studies by describing the multidisciplinary, online, video-based, analysis-of-practice PD in detail; reporting results from two national field tests of the PD; and considering those results in the context of a narrowly focused, face-to-face model on which the online course was based.

Design of the EMAT Course

Course Structure: Overview

BSCS, in collaboration with Oregon Public Broadcasting, designed EMAT as a 10-week (120 hours) summer PD course for inservice high school science teachers. In our field tests, teachers received three graduate credits from Montana State University's National Teachers Enhancement Network for their successful completion of the PD. EMAT situates crosscutting energy concepts within alternative energy contexts and embeds video-based, analysis-of-practice PD throughout the course, intertwining content deepening with pedagogy.

EMAT incorporates active learning and situated cognition in both the content deepening and pedagogical components of the course. It uses classroom video extensively to support the pedagogical components and attends to the characteristics of effective video use (van Es et al., 2014) as described in the section, EMAT Course Structure: Pedagogy.

EMAT Course Structure: Energy Concepts

EMAT emphasizes three key energy concepts across six units:

1. Energy can be neither created nor destroyed.
2. Tracking energy and matter inputs and outputs within a system helps promote understandings about the system's potentials and limitations.
3. Energy transfer is never 100% efficient; some energy always leaves the system as heat.

These key ideas are associated with two Next Generation Science Standards (NGSS Lead States, 2013) crosscutting concepts, namely, *Energy and Matter: Flows, Cycles, and Conservation* and *Systems and System Models*.

The first unit, Coal, provides a foundation for teachers to learn about one of the most common energy sources in the United States and use it for comparison with other energy technologies. Teachers investigate electromagnetic induction in steam turbine generators and use a systems approach to consider both monetary and environmental costs of a coal-fired power plant. The remaining five units focus on nuclear energy, wind power, geothermal heat exchange, biofuels, and solar energy.

EMAT Course Structure: Pedagogy

The video-based analysis-of-practice components of EMAT are based heavily on the *Science Teachers Learning from Lesson Analysis* (STeLLA) PD model developed by Roth and colleagues (2011). STeLLA is a face-to-face PD designed for inservice elementary teachers. STeLLA uses video analysis to support teachers in learning key teaching strategies within the STeLLA Conceptual Framework (Table 1).

By applying the STeLLA PD model to EMAT, we created opportunities for teachers to examine video of students learning crosscutting energy concepts, analyze video for student thinking and coherence of science content storylines, collaborate with colleagues in synchronous discussions while analyzing video, learn and reflect on the STeLLA strategies, and consider ways to apply the STeLLA strategies to their own teaching. Due to course time limitations, we were unable to include all STeLLA strategies in EMAT. Based on consultation with Roth and based on her earlier findings (Roth et al., 2011), we selected a subset of 12 strategies for the course. We used both nonparticipant video and teachers' own video in the course.

Table 1. STeLLA conceptual framework adapted from Roth and colleagues, 2011. Items with an asterisk (*) were part of EMAT.

Strategies to Reveal, Support, and Challenge Student Thinking	Strategies to Create a Coherent Science Content Storyline
1. Ask questions to elicit student ideas and predictions*	A. Identify one main learning goal*
2. Ask questions to probe student ideas and predictions*	B. Set the purpose with a focus question or goal statement*
3. Ask questions to challenge student thinking*	C. Select activities that are matched to the learning goal*
4. Engage students in analyzing and interpreting data and observations*	D. Select content representations and models matched to the learning goal and engage students in their use*
5. Engage students in constructing explanations and arguments	E. Sequence key science ideas and activities appropriately
6. Engage students in using and applying new science ideas in a variety of ways and contexts	F. Make explicit links between science ideas and activities*
7. Engage students in making connections by synthesizing and summarizing key science ideas*	G. Link science ideas to other science ideas*
8. Engage students in communicating in scientific ways	H. Highlight key science ideas and focus question throughout
	I. Summarize key science ideas*

Group discussions of nonparticipant video. Group discussions centered around nonparticipant video directly related to each EMAT unit. For example, classroom video in the Coal unit showcased a teacher uncovering student thinking about energy transfers in electromagnetic induction. Groups of five to seven teachers participated in six 2-hour, synchronous, online discussions (one near the end of each unit). To accommodate teachers with slow internet speeds, discussions used only audio (no live video of participants).

Facilitation of online discussions attended to the elements of effective video-based PD identified by van Es and colleagues (2014). To support group collaboration, facilitators established norms of discussion, encouraged participation by all teachers, and shared a model video of teachers engaging in a video-based discussion. Norms of discussion emphasized using evidence, avoiding discussion of things that may simply be annoying (e.g., a teacher saying “umm” too much), and avoiding the trap of “this doesn’t look like my classroom.”

To orient the group to the video task, facilitators identified specific STeLLA strategies that should be the focus of the discussion (e.g., a single discussion might focus on elicit, probe, and challenge questions; see Figure 1). To sustain an inquiry stance and maintain a focus on the video and subject matter, facilitators asked teachers to analyze classroom video for evidence of student thinking, evidence that students are constructing a coherent story related to the key concepts in the lesson, and teacher actions to move student thinking forward and advance the science storyline. To that end, facilitators asked teachers to complete an analysis protocol prior to the discussion in which they make a claim about something they see in the video, support their claim with evidence from the transcript (including time stamps), and consider alternative claims for what they saw in the video. During discussions, the facilitator encouraged teachers to talk to each other, pose claims and alternative claims, and use evidence from the video to support their discussion. Throughout the discussion, facilitators highlighted and encouraged teacher comments that led to productive and substantive discussion of the video.

Personal analysis of own video. Near the end of the summer, teachers engaged in private analysis of their own video using STeLLA lesson analysis protocols, made suggestions for improving their own teaching, and received feedback on their video and their analyses from the instructor. We transcribed teachers' videos prior to analysis and teachers applied the STeLLA lesson analysis protocols (used in the small group discussions) to their own video.

EMAT Course Structure: Coal Unit Example

In this section we describe how the Coal unit espoused both active learning and situated cognition, where and how teachers made connections to the three key energy ideas, and connections between the unit and the STeLLA framework.

Active learning. Teachers began the unit reflecting on their current understanding of a coal-fired power plant, including what goes into the plant, the energy transformations that occur therein, and what comes out of the plant. They also reflected on their initial ideas of what constitutes high quality instruction.

Teachers engaged with several types of data in the Coal unit. They (1) used an interactive learning experience to see what happens to a lightbulb when they moved a magnet near a coil of wire connected to the bulb; (2) used an interactive to examine matter and energy inputs and outputs associated with a coal-fired power plant system (including mining, transport of coal, and the power plant itself); (3) considered the origin of coal by examining data on the elemental composition of rocks, plants, animals, and coal; (4) examined evidence for how the use of a cooling reservoir can improve the efficiency of a power plant; and (5) analyzed an interview with a student in which the interviewer uncovers the student's thinking about electromagnetic induction in a hand-crank flashlight. They noted any incomplete or nonscientific initial conceptions the student articulated, along with any accurate scientific ideas. Across all data, teachers considered patterns in the data and interpreted those patterns.

Teachers also reflected on how the data they analyzed aligned (or misaligned) with their initial thinking. As part of a facilitated, synchronous discussion teachers shared their analyses of data and began to develop consensus around scientific concepts and characterizations of the students' understanding of induction (using evidence from the video to make claims and counterclaims). To further support their emerging understanding of the science concepts and questioning strategies, teachers viewed an animation that describes in detail the processes of converting coal to electrical energy, watched an animation that explained Carnot's theoretical maximum for efficiency and its relevance for power plants and cooling reservoirs, and viewed a video of how to use elicit, probe, and challenge questions to uncover student thinking and begin to move student thinking forward.

Situated cognition. To situate learning in teachers' own contexts, we tied teachers' explorations of energy transfer, efficiency, and systems concepts in electrical induction to classroom video in which a high school teacher is using elicit, probe, and challenge questions with physics students exploring magnets and coils of wire. The context was most situated for physics and physical science teachers and least situated for biology, chemistry, and earth science teachers Across the course, teachers moved in and out of contexts that were of greatest relevance to their own instruction. In synchronous discussions, teachers considered not only the questioning strategies and science ideas but also challenges a student might be having with the ideas. Frequently, as teachers considered challenges a student may have with a concept, they surfaced their own incomplete understanding (Roth et al., 2011).

Methods

Despite the rapid expansion of online PD, little is known about best practices for its design and implementation and most PD research is evaluative in nature (Dede et al., 2009). To address the overarching questions stated in the introduction, we posed the following specific research questions:

1) After participating in the EMAT course, do teachers demonstrate improved
   a) content knowledge about energy concepts,
   b) ability to analyze video for evidence of student thinking and coherent science content storylines, and
   c) teaching practice through appropriate use of key strategies as described in the STeLLA framework?
2) After participating in EMAT, do teachers help students attain higher posttest scores (pretest adjusted) than they did for their prior year's students, taught before teachers' participation in EMAT?
3) What can we understand from teacher online reflections and comments about the affordances and limitations of the multidisciplinary, online model?
4) How do the effects from EMAT compare to those from the related STeLLA model?

Our outcome measures for teachers and students parallel those studied in a cluster-randomized trial of the STeLLA model (Taylor et al., 2017), allowing us to better understand how the multidisciplinary, online, EMAT model influenced teaching and learning in relation to the narrowly focused, face-to-face, STeLLA model. In the discussion we consider reasons for differences in outcomes between EMAT and STeLLA and the implications of our research for developers who wish to expand the reach and accessibility of video-based, analysis-of-practice, PD models.

We used a pretest-posttest design to investigate teacher outcomes and a cohort-control, quasi-experimental design to investigate student outcomes. Teachers in the research project participated over two school years. Teachers' students the first year constituted the comparison group, and students in the second year constituted the treatment group. Teachers completed EMAT in the summer between the two school years. We conducted the entire quasi-experiment twice over two national field tests.

Participants

We recruited 35 teachers for the first field test and 39 teachers for the second field test by emailing teachers, district leaders, and state science coordinators in our database. Teachers from across the United States participated in the project. Because the study required that teachers participate over a two-year period, the project suffered from extensive attrition. Attrition was primarily caused by non-study-related factors: teachers were reassigned, moved away, left teaching, or suffered illnesses. Furthermore, we did not collect complete data from all teachers. After attrition we had 25 teachers participating in the second field test and complete data on only 18 of those teachers and their students.

Measures

Teacher content knowledge. To assess changes in teacher content knowledge, we developed six unit-specific content assessments. Before and after each unit, the teachers completed a pretest and a posttest consisting of 20–25 multiple-choice and open-ended response items. Each unit test included items directly related to each of the three crosscutting energy concepts as well as unit-specific items. For example, all unit tests included items assessing the idea that energy leaves a system as heat during energy transfers, but only the Solar unit test included items related to energy transfers related to the photovoltaic effect.

Teachers' ability to analyze videos. We measured teachers' ability to analyze videos through their written reflections as they watched video clips. The written clip reflections provided a proximal measure of teachers' pedagogical learning. Teachers completed a pretest prior to taking the course and a posttest at the end of the course. Video clips with accompanying transcripts were between five and nine minutes long. We used identical clips to those used in the face-to-face, elementary STeLLA PD study (Taylor et al., 2017) to facilitate cross-project comparisons.

Roth, Askinas, and Gardner (2013) developed a rubric to score teachers' written responses. Two coders jointly coded and discussed their scores on 20 written responses, then divided and independently scored the remaining responses, including an additional 20 overlapping responses to measure interrater agreement. The final interrater reliability statistics reveal that the coders remained well calibrated throughout coding (intraclass correlation, two-way mixed effects, absolute agreement = 0.898).

Teacher classroom practice. Teachers filmed themselves teaching a full class prior to their participation in EMAT and again following their participation. We transcribed all videos. Roth and Kowalski (2015) developed a video analysis coding protocol to score classroom sessions (approximately one hour in length) for the teachers' use of the STeLLA strategies. The scoring rubric emerged from the STeLLA conceptual framework (Figure 1) and the STeLLA strategy booklet designed to help teachers understand the STeLLA framework (BSCS, 2015). The coding protocol was extensive, requiring six to eight hours to code one hour of recorded classroom instruction. Roth and Kowalski used and refined the rubric to jointly score seven master videos that showcased a wide array of teaching practices. We used two master videos for training purposes and the remaining five for calibration of coders. We calibrated individual coders to the master codes. The intraclass correlation values ranged between 0.875 and 0.947 (mixed effects, absolute agreement).

Student content knowledge related to energy concepts. Participating high school students were enrolled in classes across biology, chemistry, physics, earth science, and environmental science and grade levels (9–12). We developed a multiple-choice assessment to examine students' understanding of the same three crosscutting energy concepts that formed the backbone of the EMAT course for teachers. We situated assessment items in varied science disciplinary contexts and provided explanation of potentially unfamiliar contexts. For example, items situated in the context of a coal-fired power plant included a picture of a power plant and a brief paragraph about how a coal-fired power plant works. The assessment included 35 items and had a Rasch person reliability of 0.77.

Teacher comments and reflections about the EMAT course. We archived a wealth of computer-based teacher discourse, reflections, and commentary from the EMAT course. To focus our analyses, we identified case study teachers based on the relative performance of treatment and comparison students, including teachers with strong positive, flat, and strong negative student treatment effects. We examined case study teacher survey comments, unit reflections, reflections on their own video, and lessons plans for the following fall.

Analyses

We used Rasch analyses to convert all outcome scores to person measures, anchoring the pretest for each measure to its posttest. For the three continuous teacher outcome variables, we used matched-pairs t-tests and linear regression to predict teachers' posttest scores from their pretest score, highest degree attained, gender, and years of science teaching experience. We lacked resources to fully code teacher videos from both field tests, so our analyses of teacher outcomes focus on data from the second field test.

For the student outcome, our primary independent variable was a dichotomous treatment variable. We used multilevel modeling to account for the nested nature of the data, with students at level 1, class at level 2 (with the treatment variable at level 2 indicating whether the class was a treatment class or a comparison class), and teacher at level 3. At level 1 we included students' pretest as a covariate along with student demographic variables. Level 2 included the treatment variable as the only predictor. Level 3 included the mean student pretest score for the given teacher, teacher gender, years of experience, highest degree, and random effects slope for the treatment variable, such that the coefficient of the treatment variable could vary by teacher. We then estimated the mean treatment effect across all teachers. We combined and analyzed student data across both field tests.

Findings

Descriptive Statistics

Table 1 summarizes the descriptive statistics for both teachers and students. Overall, 29% of the students were eligible for free or reduced-price lunch. Thirty-one percent were from racial or ethnic groups that are traditionally underserved in the sciences, and the median grade level was grade 9. The mean years of teaching experience was 10 years, and 40% of teachers had an advanced degree. The student populations for the treatment and comparison groups were very similar.

Table 1. *Descriptive statistics of the sample: Means and sample sizes by treatment group.*

Variables	Total sample (N = 2,462)	EMAT condition (N = 1,133)	Comparison condition (N = 1,329)
Students			
Mean pretest score person measure (SD)	-0.482 (0.762)	-0.460 (0.775)	-0.500 (0.751)
Mean posttest score person measure (SD)	-0.285 (0.901)	-0.248 (0.906)	-0.317 (0.895)
FRL (%)	29	28	29
Female (%)	48	46	49
Underrepresented minority (%)	32	32	33
Asian (%)	19	20	18
White (%)	50	49	52
African American (%)	9	10	09
Native Hawaiian or Pacific Islander (%)	6	6	7
American Indian or Alaska Native (%)	11	11	11
Hispanic or Latino/a (%)	18	18	19
Grade 9 (%)	35	35	35
Grade 10 (%)	21	21	22
Grade 11 (%)	28	28	25
Grade 12 (%)	16	16	18
English learner (%)	22	22	23
Teachers (first field test, N = 22; second field test, N = 25)			
Mean years of experience	10.07		
Percent with master's degree (none had doctoral degree)	40		

Teacher Outcomes

Teacher content knowledge. The pre to post change in overall teacher content knowledge was significant (matched-pairs t-test, $p < .001$) and had an effect size of $d = 1.71$. We calculated the effect sizes of the subscales for each key idea and found effect sizes of just under one standard deviation (0.94, 0.81, and 0.81) for concepts related to conservation, efficiency, and systems. Although almost all teachers in our study had at least a bachelor's degree (half of them in science), 40% had an advanced degree, and their average years of experience was 10, we still see evidence of teachers showing improved understanding of foundational energy concepts associated with their participation in EMAT.

Teacher ability to analyze videos. Teachers demonstrated an increased ability to analyze videos with a pre to post effect size of $d = 1.38$ ($p < .001$, matched-pairs t-test). Thus, teachers improved in their ability to identify and comment on key elements in a video related to revealing, supporting, and challenging student thinking and creating a coherent science content storyline.

Teacher classroom practice. Teacher practice scores also increased from pre to post with an effect size of $d = 0.57$, but the change was not statistically significant at the $\alpha = .05$ level ($p = .063$). The lower and upper limits for the confidence interval (CI) around the effect size for teacher practice includes a negative value (95% CI of effect size, [-.04, 1.17]). While most teachers showed increases from pre to post on their practice scores, some teachers' scores went down.

Table 2 summarizes the pretest-posttest results for teacher outcomes. The mean person measure of 1.13 logits on the content posttest corresponds with a raw score of approximately 110 (SD = 13.5) on the combined measure out of 167 possible. Likewise, the average post Rasch person measure for the ability to analyze video outcome (-0.92) corresponds with a mean post score of 25.3 (SD = 18.0) out of a possible 61 points. And the post practice measure (mean = -0.64) corresponds with a mean of 24.9 (SD = 26.8) out of a possible 75 points.

Table 2. Pretest-posttest changes for teacher outcomes.

Outcome	N	Pre mean (SD)	Post mean (SD)	Effect size (d)	95% CI of the effect size	
					Lower	Upper
CK: Overall (person measure)	28	0.20 (0.50)	1.10 (0.60)	1.71	1.39	2.03
CK: Conservation of energy (raw score)	20	6.8 (2.50)	9.0 (2.00)	0.94	0.63	1.25
CK: Efficiency (raw score)	17	15.1 (1.40)	18.4 (3.40)	0.81	0.51	1.11
CK: Systems (raw score)	22	21.4 (0.50)	25.0 (4.40)	0.81	0.46	1.16
VA (person measure)	23	-1.82 (0.64)	-0.92 (1.03)	1.38	0.85	1.92
Practice (person measure)	20	-1.06 (0.69)	-0.64 (0.81)	0.57	-0.04	1.17

Note: CI = confidence interval; CK = content knowledge; VA = video analysis task; effect sizes are standardized mean difference effect sizes.

The STeLLA face-to-face PD with its effect sizes for elementary teacher outcomes provides our most relevant effect size benchmarks (Hill, Bloom, Black, & Lipsey, 2008) for teachers' ability to analyze video and for teacher practice. The instruments and coding rubrics were identical for both EMAT and STeLLA. We conducted additional analyses (not reported elsewhere) on the STeLLA data and found STeLLA pre-post effect sizes for ability to analyze video and classroom practice were STeLLA $d_{VideoAnalysis} = 2.61$ (double that from EMAT) and

STeLLA $d_{practice}$ = 2.09 (approximately four times that from EMAT). Thus, the EMAT pre-post effect sizes for teacher outcomes were substantially smaller than those from the original STeLLA face-to-face PD for elementary teachers.

Predicting teacher outcomes. We used analysis of covariance (ANCOVA) to better understand factors associated with changes in teacher outcomes. Across all three measures, teachers' years of experience and highest degree were not significantly predictive of teacher outcomes at the $\alpha = .05$ level, with p-values ranging between .26 and .93, depending on the analysis. Teachers' pretest scores for both content knowledge and video analysis significantly predicted their posttest scores ($p < .001$ and $p = .010$), but teachers' pre practice scores did *not* significantly predict teachers post practice scores ($p = .906$).

Student Outcomes

Across treatment and comparison groups, students showed improvement from pretest to posttest (Table 1). The mean post person measure of -0.285 corresponds with a raw score of 14.68 (SD = 5.68) out of a possible 35. The posttest average across the entire sample was 3.04 raw points higher than the pretest average. Data in Table 3 show that although the sample of treatment group of students outperformed the comparison group of students, the difference was not statistically significant ($p = .129$). We interpret the average treatment effect as follows: On average, the mean class posttest score for the treatment group was 0.081 logits (roughly 1.2 points on the raw scale) higher than for the comparison group, controlling for pretest and other demographic factors. Across the two treatment groups, we see that girls, English language learners, and students from racial/ethnic groups traditionally underrepresented in the sciences had lower achievement scores, but this is common across both groups and not particular to the EMAT group. Further, a teacher's gender, years of experience, and highest degree did not significantly predict student achievement. The effect size for the intervention was d = 0.13, the variance of effect size σ_d^2 = 0.20, SE_d = 0.452, and the lower and upper 95% confidence interval values for the effect size were [-0.733, 0.999]. The fact that the effect size had such a large variance and we see a wide 95% confidence interval for the effect size indicates that student effects from a teacher's participation in EMAT varied drastically from teacher to teacher. For comparison, the student effect from the cluster-randomized trial of the STeLLA PD model was d = 0.68, with 95% CI [0.60, 0.76] (Taylor et al., 2017). The STeLLA effect on student outcomes was consistently strong from teacher to teacher and was more than five times larger than the EMAT effect.

Table 3.

Test of Main Effect of Treatment on Student Achievement. Data combined across first and second field tests.

Variable	**Coefficient**	**Standard error**	**t-ratio**	**d.f.**	**p-value**
Level 3 (teacher)					
Intercept	-0.336	0.030	-11.024	57	< .001
YrsSci	-0.004	0.004	-0.896	57	.374
TchGen	0.032	0.055	0.586	57	.560
HighDeg	-0.030	0.052	-0.578	57	.565
MnPre	0.495	0.077	6.459	57	< .001
Level 2 (class)					
γ_{010} (avg. treatment effect across teachers)	0.081	0.053	1.537	61	0.129
Level 1 (student)					
Pre	0.582	0.029	20.316	2,448	< .001
Gender	-0.136	0.029	-4.669	2,448	< .001
FRL	-0.016	0.029	-0.540	2,448	.589
Grd10	-0.051	0.044	-1.156	2,448	.248
Grd11	-0.123	0.049	-2.491	2,448	.013
Grd12	-0.133	0.071	-1.868	2,448	.061
ELL01	-0.091	0.040	-2.300	2,448	.022
Race01	-0.115	0.037	-3.084	2,448	.003

From our main analysis (Table 3), the mean student pretest score aggregated at the teacher level (level 3) strongly predicts student achievement controlling for an individual's pretest score (at level 1) and it reflects something about the instructional context that impacts student achievement. Instructional context can include any factors that all students of a given teacher collectively experience—from teacher characteristics to classroom or school resources including curriculum.

Examining Variation in Student Outcomes: Quantitative analysis

Teacher outcomes and student outcomes. Because instructional context by teacher was influential in predicting student achievement, and because effects of EMAT varied extensively across teachers, we investigated the relationship between teacher outcomes and student outcomes. We used multilevel modeling to predict student posttest scores for treatment students, controlling for teachers' post EMAT scores. We included student pretest and demographic predictors at level 1 and teacher post outcome measures (content knowledge, ability to analyze video, and practice) along with mean student pretest scores at level 2. We found that variation in measured EMAT teacher outcomes does *not* statistically predict the variation in student outcomes, with p-values ranging between .631 and .780.

Treatment interactions with student demographics. We also investigated whether a teacher's participation in EMAT had a differential effect on students with different demographic characteristics. We found two significant interaction effects (treatment interacting with FRL status and with grade level, respectively). The results suggest that students who received free or reduced-price lunch and older students (11th and 12th grade) showed *stronger* treatment effects than students who are not eligible for FRL and 9^{th} and 10^{th} grade students in this sample. We hesitate to draw strong inferences from the significant interaction effects given the relatively small sample sizes and warnings that it is common to uncover spurious relationships in interaction analyses leading to overstated or misleading results (Wang, Lagakos, Ware, Hunter, & Drazen, 2007). However, it is notable that inclusion of interaction effects in the analytic model narrowed the 95% confidence interval [-0.306, 0.679] while only modestly altering the treatment effect estimate (d = 0.187). The interaction between the treatment and student demographics explains some of the variation in student outcomes.

Examining variation in student outcomes: Teacher comments

With quantitative models of teacher and student measures providing only limited explanations for the variable effects of EMAT by teacher, we identified five case study teachers for further analysis. Our case study teachers included two with large positive student effects (Teachers 1 and 2), one with flat/slightly negative student effect (Teacher 3), and two with large negative treatment effects for their students (Teachers 4 and 5). We studied the survey comments, personal video reflections, end of unit reflections, and lesson plans for implementing STeLLA strategies in instruction from the five case study teachers to help us better understand the variation in treatment effects. We examined teachers' comments in the contexts of active learning and situated cognition frames.

Active learning for case study teachers. Teachers 1 and 2 reflected positively (and explicitly) on several active learning elements of the course. For example, we frequently asked teachers to construct a concept map of their initial understanding of a key system at the start of a unit (e.g., a coal power-plant system) and then asked them to revise the map throughout the unit based on their coursework. Teacher 1 commented twice that drawing and revising the concept map were some of her favorite activities, stating, "It made me think about what I had learned." Teacher 2 made a similar comment, stating, "I always think I know the system diagram until I go through the process of writing it down and then correcting it. This is an excellent process for learning how all the parts of the process contribute to the whole.” Teachers 1 and 2 were also uniformly positive about opportunities to engage in lesson analysis. Both found the opportunity to analyze video to identify key strategies highly effective to supporting their own learning.

Teacher 3 made one positive comment about reflecting on earlier work. In other cases, he mentioned that he “didn’t learn anything” from activities designed to elicit his prior knowledge and “the concept map was the least effective part of the lesson.” In analyzing video, he made two comments that indicated that he had difficulty filling out the protocols designed to help him gather and analyze evidence from a video and was unclear about the purpose of the protocols. He indicated on seven occasions that the lesson analysis active learning opportunities were neither helpful nor effective in supporting his own learning.

Although Teacher 4 had a negative treatment effect on students, she indicated that she appreciated the opportunities to record her initial thinking and document how her thinking changed over time. She also identified concept maps as particularly effective elements of instruction. Her positive comments about some active learning components for science content go against the trend of the other four case study teachers (showing those with stronger student effects responding more positively to active learning of content, and those with weaker or negative effects responding negatively to some aspects of active learning). On the other hand, although she indicated that she enjoyed lesson analysis, she was often unsure of her analyses and would have appreciated greater feedback and discussion on her analysis of classroom videos.

Teacher 5’s unit reflections tended to focus on science ideas (not activities in the course). Further, he responded to only a small fraction of surveys. However, his survey comments were uniformly negative, identifying active learning components (such as concept maps, revising concept maps, or analyzing data as part of an interactive) as “somewhat ineffective” or “not at all effective” to supporting his learning, with particularly negative comments about analyzing video.

Situated cognition for case study teachers. Teachers 1 and 2 made strongly positive connections between EMAT and their instructional contexts. Teacher 1 made nine separate comments indicating that a given activity would be useful for her students or indicating that the activity made her reconsider her practice. Teacher 2 indicated that analyzing her own lessons was particularly revealing, stating, “Wow. Very good. Very difficult from a self-esteem point of view.” In both cases, the teachers were clearly situating their learning in their own contexts.

Teacher 3 made a single positive comment about an activity that he thought might be relevant for his classes. He either did not see connections between the course and his daily instruction or he simply didn’t comment. Notably, Teacher 3 rated analysis of his own video (an element that should have been highly situated for all teachers) as “not at all effective” to supporting his learning.

Teacher 4 made eight comments connecting her participation in the course to her teaching. Of those, four negative comments indicated a difficulty of connecting what she was learning to her classroom. For example, she identified completing some of the reflection activities as challenging because she had difficulty remembering her school year lessons during the summer. In other cases, she indicated a greater need for examples of how to use strategies in her classroom.

Teacher 5’s comments primarily reflected his perception that he considered himself already highly skilled at implementing STeLLA strategies and highly knowledgeable about the science content.

For secondary students to perform well on the assessment, the students needed the opportunity to link ideas associated with daily instruction (e.g., energy flows into and out of an ecosystem, moving from one organism to the next with a reduction of energy in each trophic level) to larger crosscutting energy concepts (energy transfer is never 100% efficient; some energy always leaves a system as heat). Teachers needed to help students make those connections between science ideas. Teachers 2 and 5 expressed difficulty in helping students make these connections. In reflecting on analyzing her own video for connections between ideas, Teacher 2 noted, "Linking ideas to ideas was the least effective part of the lesson." Likewise, in one of Teacher 5's unit reflections, he shared a comment that revealed his challenge with connecting his daily instruction to crosscutting energy concepts:

> *When I think about the subject of energy, I have never felt like it fit into the storyline of the courses I have taught or taken in Chemistry. It's a strange situation because it is the basis for everything or almost everything, but it doesn't seem to flow in the storyline …. I'm not sure what the answer is, but it does bother me that energy in many ways seems like a disconnected unit.*

We note the challenge as important to the overall efficacy of EMAT and other PD models that focus on crosscutting concepts.

Finally, teachers' three-day lesson plans reveal the extent to which teachers were able to apply their learning to their own contexts. Across four of five case study teachers, lesson plans had serious flaws. Teachers 2, 3, 4, and 5 had difficulty selecting an appropriate main learning goal for at least one day of their three-day plans. Instead of identifying a single meaningful concept for a lesson, they identified facts (such as parts of a leaf) that should be memorized (Teachers 2, 3, and 4) or identified as many as six separate key ideas (Teachers 3 and 4) for a single lesson. Teacher 2 revealed a misunderstanding of what content representations are and how they might be used to support student learning. Her lesson plans further emphasized that she did not understand how to link science ideas to one another. Teacher 5 fundamentally misunderstood the difference between articulating a main learning goal for planning purposes (the intent of the STeLLA strategy) and writing a main learning goal on the board for students. His lessons emphasized procedures for students to carry out with little indication that students would make meaning from the activities.

Discussion

The present study is one of the first to empirically examine a model for bringing video-based analysis-of-practice PD to a wider audience by expanding focus (multidisciplinary science) and accessibility (online rather than face-to-face). It is also one of a small but growing number of studies of PD that examined how PD ultimately impacted student outcomes using quasi-experimental design. It investigated changes in teacher knowledge and practice and detailed comments from teachers' experiences to better understand effects on students, thus following research and data collection patterns demanded by the field (Dede et al., 2009).

Online PD is not a panacea, is not always low cost, and is appropriate only in specific contexts, particularly those in which teachers are spread across a wide geographic area (Fishman et al., 2013). Translating face-to-face, video-based, analysis-of-practice PD models that require extensive facilitation to an online environment makes sense for high school science teachers in medium and small districts, specifically because it would allow districts to leverage resources and bring together many teachers from a wide area. Upfront development costs are high, but facilitation costs can be lower for online PD due to reduction in travel expenses.

But our results demonstrate that translating narrowly focused, video-based, analysis-of-practice PD into a multidisciplinary, online environment is challenging. Our study design does not allow us to isolate challenges resulting from the multidisciplinary nature of the course from challenges arising from its online modality, but we consider each in turn.

Challenges associated with multidisciplinary focus

Roth and colleagues (2017) specifically mention using exemplar lesson plans to situate learning for teachers. STeLLA used model lessons extensively in the PD. Desimone and Garet (2015) emphasize tying PD to specific lessons. Taylor, Getty, Kowalski, Wilson, Carlson, and Van Scotter (2015) found positive effects for high school students for a multidisciplinary science PD program, but the intervention included comprehensive, two-year, multidisciplinary curriculum materials alongside curriculum-based PD.

EMAT teachers did not have common science content backgrounds and were approaching key energy concepts from different disciplinary perspectives. It was not practical to develop discipline-specific model lessons for EMAT. EMAT teachers did not analyze or use model lessons and did not jointly plan lessons with one another (although they did plan their own three-day lesson arcs. It may be that EMAT experiences were not sufficiently situated in teachers' daily practice to strongly support teacher learning and impact student achievement. Comments from case study teachers lend credence to this possibility. Exemplar lesson plans may help situate teacher learning in teachers' own disciplinary contexts and help teachers make connections among ideas, both for themselves and for students.

Challenges associated with online modality

Our study suggests that video-based, analysis-of-practice PD models are not easily translated into an online space. Analysis protocols and tools designed to maintain a focus on the video and subject matter (van Es et al., 2014) and support active learning may require modification or additional scaffolding for an online audience. In EMAT, teachers studied STeLLA strategies independently, watched videos on how to complete lesson analysis protocols, engaged in initial practice with protocols independently, and had intermittent small group discussions using the protocols. Case study teacher comments reveal that more guidance on using the protocols may be necessary. Further study is needed to identify how to best support teachers in learning to analyze video absent the give-and-take question and answer structure that face-to-face PD affords.

Additional Challenges

Extensive research has emerged about the importance of both time span and contact hours for PD (Desimone, 2009; Desimone & Garet, 2015; Roth et al., 2017; Taylor et al., 2017). Allowing PD to span an academic year pays valuable dividends because teachers can try new strategies with students and reflect on those strategies in group discussions. To facilitate two full design and test iterations within the timeline from the funding agency, we decided to compress the EMAT course into a single summer. There are indications from teacher comments that teachers and students would have been better served with a year-long PD, even if the PD had the same 120 contact hours as the original course. Researchers designing studies of various PD models to fit within funders' time constraints should carefully consider how tradeoffs may hamper effects on teachers and students.

We have identified three other issues that may have reduced the ability of EMAT to strongly impact teaching and learning. First, teachers did not analyze their own videos in group discussions. Although EMAT teachers recorded videos of themselves prior to participation in the summer course, the videos invariably did *not* incorporate use of key STeLLA strategies. In our early attempts to have teachers discuss these videos in small groups, we found that teachers were embarrassed by their own videos and were reluctant to critique one another. Even when one teacher boldly and publicly critiqued his own video in a discussion, his peers did not engage. As a result, we made the decision early on to have teachers analyze nonparticipant video in small-group discussions. Seidel and colleagues (2011) found that analysis of nonparticipant video can promote attention to critical incidents but analysis of one's own video can be more motivating. We question whether our results would have differed has group discussions focused on participants' own videos.

A second issue that may have bearing on our findings is that secondary science teachers are likely different from elementary teachers in important ways. Secondary teachers are likely much more confident in their science content knowledge than elementary teachers given the number with science majors. We have not completed studies demonstrating that the STeLLA model is effective with high school teachers and their students. Evidence from a face-to-face STeLLA PD model with a secondary audience would add clarity to the issue.

A third possible issue relates to the fact that we did not use video conferencing and thus the facilitator could not see or respond to teachers' facial expressions during the synchronous discussions. Video conferencing may have afforded stronger rapport among the facilitator and participants, may have uncovered and rectified problems with analysis protocols sooner. We offer these additional considerations because they are worthy of further study.

Conclusion

This study provides a detailed description of a multidisciplinary, online, video-based, analysis-of-practice science teacher professional development model. The model applies and extends successful, narrowly focused, face-to-face, video-based PD elements to reach a broader audience. The study provides insight into some of the challenges associated with translating video-based PD for a multidisciplinary and online audience. We did not find statistically significant impacts on students arising from the EMAT PD, but there is some indication of positive results with teachers. Inclusion of exemplar lessons, careful consideration of scaffolding to help teachers use video analysis protocols, and the timing of PD to extend over a school year may result in a more effective model that is more accessible to high school science teachers in small- and medium-sized districts.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond and G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3–31). San Francisco, CA: Jossey-Bass.

Blomberg, G., Renkl, A., Sherin, M. G., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for educational research online*, *5*(1), 90.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, *33*(8), 3-15.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and teacher education*, *24*(2), 417-436.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience and school*. Washington, DC: Commission on Behavioral and Social Sciences and Education, National Research Council.

Brophy, J. (2004). *Motivating students to learn* (2nd ed.). Boston, MA: McGraw Hill.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

BSCS. (2015). *STeLLA Strategies* [Unpublished material]. Colorado Springs, CO: Author.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, Co: BSCS, 5, 88-98.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445-459.

Dede, C., Jass Ketelhut, D., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of teacher education*, *60*(1), 8-19.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher*, *38*(3), 181-199.

Desimone, L. M., & Garet, M. S. (2015). Best practices in teacher's professional development in the United States. *Psychology, Society & Education, 7*(3), 252-263.

Dewey, J. (1910). *How we think*. Boston, MA: DC Heath.

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a computerized professional development program on teacher and student outcomes. *Journal of Teacher Education*, *61*(4), 302-312.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of teacher education*, *64*(5), 426-438.

Garet, M., Porter, A., Desimone, L. Birman, B., & Yoon, K. (2001). What makes professional development effective? Analysis of a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945.

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

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Inhelder, B., & Piaget, J. (1964). *The growth of logic in the child*. New York, NY: Harper & Row.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*(1-2), 172-181.

Koc, Y., Peker, D., & Osmanoglu, A. (2009). Supporting teacher professional development through online video case study discussions: An assemblage of preservice and inservice teachers and the case teacher. *Teaching and Teacher Education*, *25*(8), 1158-1168.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Nemirovsky, R., & Galvis, A. (2004). Facilitating grounded online interactions in video-case-based teacher professional development. *Journal of Science Education and Technology*, *13*(1), 67-79.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational researcher*, *29*(1), 4-15.

Roth, K. J., Askinas, K., & Gardner, A. (2013). *STeLLA II PCK measure coding manual*. Colorado Springs, CO: Biological Science Curriculum Study (BSCS).

Roth, K. J., Bintz, J., Wickler, N. I., Hvidsten, C., Taylor, J., Beardsley, P. M., Caine, A., & Wilson, C. D. (2017). Design principles for effective video-based professional development. *International Journal of STEM Education*, *4*(1), 31.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*(2), 117-148.

Roth, K. J., & Kowalski, S.M. (2015). *STeLLA II Lesson video coding manual*. Colorado Springs, CO: Biological Sciences Curriculum Study (BSCS).

Schwab, J. J. (1960). What do scientists do? *Systems Research and Behavioral Science*, *5*(1), 1-27.

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, *34*, 56-65.

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and teacher education*, *27*(2), 259-267.

Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, *13*(3), 475-491.

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. *American Educational Research Journal*, *52*(5), 984-1017.

Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A., & Tipton, E. (2017). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, *10*(2), 241-271.

van Es, E. A. (2009). Participants' roles in the context of a video club. *Journal of the Learning Sciences*, *18*(1), 100-137.

van Es, E. A., Tunney, J., Goldsmith, L. T., & Seago, N. (2014). A framework for the facilitation of teachers' analysis of video. *Journal of Teacher Education, 65*(4), 340-356.

Vygotsky, L. S. (1960). *Selected psychological works*. Moscow: Akad. Pedagogika Nauk RSFSR.

Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., & Drazen, J. M. (2007). Statistics in medicine—reporting of subgroup analyses in clinical trials. *New England Journal of Medicine*, *357*(21), 2189-2194.

Wilson, C. D., Stuhlsatz, M.A.M., Hvidsten, C., & Stennett, B. (2017, April). *Examining the impact of lesson-analysis based teacher education across methods courses, student teaching, and induction.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Antonio, TX.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*.