Impact of Video-Based Analysis-of-Practice Professional Learning: Comparing Online and Face-to-face Models

Susan M. Kowalski[1]. Amy Belcastro[2], Connie Hvidsten[2], Guy Ollison[2], Karen Askinas[2], Renee De Vaul[2], and Gillian Roehrig[3]

[1]NWEA

[2]BSCS Science Learning

[3]STEM Education Center, University of Minnesota

Author Note

Susan M. Kowalski https://orcid.org/0000-0002-2012-6378

Susan M. Kowalski is now at NWEA, Department of Research and Development.

We have no known conflict of interest to disclose.

This project was funded by a grant from the National Science Foundation, Award #1813127.

Correspondence concerning this article should be addressed to Susan M. Kowalski. Email: susan.kowalski@nwea.org

Abstract

Video-based analysis of practice models have gained prominence in mathematics and science teacher education inservice professional learning. There is a growing body of evidence that these intensive professional learning (PL) models lead to positive impacts on teacher knowledge, classroom instructional practice, and student learning (Roth et al., 2018; Taylor et al., 2017), but they are expensive and difficult to sustain. An online version would have several benefits, allowing for greater reach to teachers and students across the country, but if online models were substantially less effective, then lower impacts would undercut the benefits of greater accessibility. We designed and studied a fully online version of the face-to-face *Science Teachers Learning from Lesson Analysis* (STeLLA) PL model (Roth, et al., 2011; Roth et al., 2018; Taylor et al., 2017). We conducted a quasi-experimental study comparing online STeLLA to face-to-face STeLLA. Although we found no significant difference in elementary student learning between the online and face-to-face versions ($p = .09$), the effect size raises questions. Exploratory analyses suggest that the impact of online STeLLA on students is greater than the impact of a similar number of hours of traditional, face-to-face content deepening PL, but less than the impact of the full face-to-face STeLLA program. Differences in student populations, with higher percentages of students from racial and ethnic groups underserved by schools in the online STeLLA program, along with testing of the online STeLLA model during the pandemic, complicates interpretation of the findings.

Impact of Video-Based Analysis-of-Practice Professional Learning: Comparing Online and Face-to-face Models

Background

Video-based professional learning (PL) models for math and science teachers have gained prominence over the past two decades (Borko, Koellner, Jacobs, & Seago, 2011; Sherin, 2003), with demonstrated impacts on teacher and student outcomes (Sun & van Es, 2015; Kersting, Givvin, Sotelo, & Stigler, 2010; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Seago, Jacobs, Heck, Nelson, & Malzahn, 2014). *Science Teachers Learning from Lesson Analysis* (STeLLA) is one such model. The face-to-face version of STeLLA, an analysis-of-practice PL model for elementary science teachers, has demonstrated positive impacts on teachers and students alike in a cluster randomized trial (Taylor et al., 2017; Roth et al., 2019). What Works Clearinghouse identified STeLLA as a program that demonstrates positive impact on students.

About STeLLA. STeLLA emphasizes two key aspects of teaching and learning: 1) attending to coherence of science ideas through a science content storyline; and 2) attending to student thinking to inform instruction. These two aspects of teaching and learning form the STeLLA Framework with a *Science Content Storyline Lens* and a *Student Thinking Lens*, guiding all PL experiences throughout the year.

STeLLA is a year-long PL experience with three phases. In the summer, teachers 1) learn about instructional strategies to support each of the lenses of the framework, 2) analyze exemplar classroom video of non-participating teachers to develop a common vision of the strategies, 3) study detailed model curriculum materials to examine how the strategies might be embedded in a science classroom, and 4) experience activities to enhance teachers' own science content knowledge. The summer phase includes 60 hours of learning experiences.

In the fall, teachers enact the model curriculum materials they studied in the summer, video-record themselves teaching one complete lesson from the unit, and participate in study group sessions to analyze video of themselves and their peers. Facilitators select short clips of instruction from the fully recorded lesson, create transcripts to support analysis, and lead the teachers in making evidence-based claims about the coherence of instruction and/or student thinking visible in the video. The fall session includes 15 hours of small group work to analyze participating teacher videos.

In the spring, facilitators provide scaffolding for teachers to develop their own lesson plans. Facilitators provide the key curriculum learning goals and activities for teachers, and teachers develop the STeLLA framing – sequencing the activities; noting when and where individual STeLLA strategies will be focal to the instruction; and creating detailed plans to ensure coherence of a science content storyline and attention to student thinking to inform instruction on a day-to-day basis. The spring session includes 15 hours of PL experiences to develop and analyze lesson plans.

The face-to-face STeLLA model is effective, but costly, with a two-week summer institute and eight facilitated small-group sessions throughout the school year. The need for STeLLA facilitators to travel for the PL restricts the face-to-face model to large urban and suburban school districts that can afford the expense. It also precludes participation by some teachers who may find it difficult to attend the summer institute. However, face-to-face PL more easily allows teachers to establish high levels of trust essential to analyzing videos of their peers (van Es, 2012; Zhang, Lundeberg, Koehler, & Eberhardt, 2011).

From a policy perspective, an online STeLLA PL model would have several benefits. Teachers could enroll regardless of their district's location and can participate for 6 flexible hours per week over the entire summer (ten weeks), rather than during a more intensive two full weeks (30 hours per week). Furthermore, with no travel or food expenses, the total cost of the PL can be substantially reduced.

Although the potential benefits of an online STeLLA model are enticing, an unanswered question remains: is possible to achieve similar impacts on students with an online STeLLA model compared to the face-to-face model. If the impacts wane substantially in an online offering, then the cost savings and practicality of the online version would be less compelling.

Purpose

The purpose of this study was to design and test a fully online version of the STeLLA PL model with the overarching question, *Are the online and face-to-face STeLLA models similarly effective?* We wanted to determine if the online version is a viable option for expanding reach of the highly effective face-to-face STeLLA PL. We hypothesized that there will be little or no reduction in impact of STeLLA on student outcomes when shifting from the face-to-face to online models.

Setting

The present study of the online STeLLA model took place in the Canvas learning management system, through Zoom videoconferences, and in the classrooms of teachers across 18 states. The data collected for the online STeLLA PL model was compared with data leveraged from a cluster randomized trial (CRT) of the face-to-face model, collected in Colorado between 2011 and 2013.

Participants

Table 1 describes the participants. There were important differences across the groups. Students of teachers who participated in the online program included a higher percentage of Black and Hispanic/Latinx students; higher percentage of students eligible for free or reduced-price lunch; higher percentage of students learning science in an unfamiliar language; and included more schools in cities. Most notably, the teachers in the online version began their PL in the summer of 2020, just after schools were shut down. Thus, teachers in online STeLLA were implementing instruction during the most disrupted school years in our lifetimes.

Table 1. Participant characteristics

Characteristic	Face-to-Face STeLLA	Face-to-Face Content Deepening	Online STeLLA
Number of schools	26	11	20
Number of teachers (clusters)	41	18	26
Number of students	1134	399	472
Locations of teachers/students	Colorado	Colorado	18 states including Colorado
Student pretest mean (SD)	45.01(9.24)	40.79 (8.53)	43.16 (9.18)
Demographic data listed by school			
Percent Asian	4	4	3
Percent Black	7	4	14
Percent Hispanic/Latinx	27	28	31
Percent White	62	61	46
Percent City	27	36	50
Percent Suburban	46	27	30
Percent Town	12	18	0
Percent Rural	15	18	20
Percent of students receiving free or reduced-price lunch	42	37	54
Percent of students learni`ng science in an unfamiliar language context	13	13	20
Additional contextual information	Normal school years (between 2012 and 2014)	Normal school years (between 2012 and 2014)	COVID school years (between 2020 and 2022)

Intervention

The present study collected new data from teachers who participated in the online version of STeLLA and their students. For comparison, we used a subset of data collected from a cluster randomized trial (CRT) of two face-to-face models. The CRT compared face-to-face STeLLA wtih a face-to-face content-deepening PL of equivalent duration. It examined the relative impacts of the two face-to-face PL experiences on student outcomes and demonstrated that face-to-face STeLLA had significantly larger impacts on student science achievement than face-to-face content deepening PL (Taylor et al., 2017). We included original CRT data from teachers who taught plate tectonics, weathering and erosion, and the water cycle, to match the focus of the online STeLLA PL.

All three PL models required approximately 90 hours of PL across a calendar year. All had similar instructional goals for students. The online and face-to-face STeLLA PL models had similar designs in the summer, fall, and winter components. Table 2 provides more detail on the components of each intervention and how they were administered.

Research Design

We compared data collected from the online STeLLA PL program to data collected in the original face-to-face STeLLA PL cluster randomized trial. We used identical curriculum modules and instruments across the two versions to minimize bias. The assessment developer was external to the online project, and PL facilitators did not have access to the student assessment. Teachers were not randomly assigned in this research design. All teachers recruited for the 2020-2022 study experienced the online version of STeLLA, and teachers recruited for the original CRT were randomly assigned to either face-to-face STeLLA or face-to-face content deepening PL.

Data Collection and Analysis

Student assessments included multiple choice pre and posttests related to the water cycle (person reliability = 0.82; person separation = 2.10). We analyzed the impacts of the modality of STeLLA PL on student learning using a two-level model, with students nested within teachers. Student pretest data served as a level 1 covariate, mean pretest score by teacher was a level 2 covariate, and treatment condition was included as a level 2 explanatory variable. Figure 1 shows the analytic models. We estimated the effect size for student impacts (δ_τ, Figure 2). Our focal analysis (Model 1) compared face-to-face to online STeLLA. Our exploratory models compared STeLLA Online to the face-to-face content deepening PL (Model 2); face-to-face to online STeLLA while accounting for demographic characteristics (Model 3), and compared STeLLA Online to the face-to-face content deepening PL while accounting for demographic characteristics (Model 4). Models 1 and 2 are identical to those used in the original CRT (Taylor et al., 2017).

Figure 1. Analytic models

Models 1 and 2

Level-1 Model (student)

$$Post_{ij} = \beta_{0j} + \beta_{1j}*(Pre_{ij}) + r_{ij}$$

Level-2 Model (teacher)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Treat_j) + \gamma_{02}*(MeanPre_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Figure 1 (continued). Analytic models.

Models 3 and 4

Level-1 Model (student)

$Post_{ijk} = \pi_{0jk} + \pi_{1jk} *(Pre_{ijk}) + e_{ijk}$

Level-2 Model (teacher)

$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(Treat_{jk}) + \beta_{02k}*(MeanPre_{jk}) + r_{0jk}$

$\pi_{1jk} = \beta_{10k}$

Level-3 Model (school)

$\beta_{00k} = \gamma_{000} + \gamma_{001}*(PctUnrep_{k}) + \gamma_{002}*(PctFRL_{k}) + u_{00k}$

$\beta_{01k} = \gamma_{010}$

$\beta_{02k} = \gamma_{020}$

$\beta_{10k} = \gamma_{100}$

Figure 2. Effect size calculation

$$\delta_{\tau} = \frac{\gamma_{10}}{\sqrt{\tau + \sigma^2}}$$

Table 2. Intervention components

Component	Face-to-Face Content Deepening	Face-to-Face STeLLA	Online STeLLA
Total hours	88.5	88.5 hours	~88.5 hours (~46.5 hrs asynch + 52 hrs synch)
Summer Institute	2 weeks (60.5 hours) Full-day science content sessions with university science faculty	2 weeks (60.5 hours) learn strategies and watch stock video of lessons teachers will enact in fall	10 weeks (asynch + synch) (~40 hrs + 20 hrs = 60 hours) learn strategies and watch stock video of lessons teachers will enact in fall
Curriculum	Teachers' own curriculum targeting plate tectonics, weathering/erosion, and the water cycle	Earth's Changing Surface (Fall) Water Cycle (Spring)	Earth's Changing Surface (Fall) Water Cycle (Spring)
Number of student lessons	Variable dependent on teacher	7 lessons (Fall) 6 lessons (Spring)	7 lessons (Fall) 6 lessons (Spring)
Fall Sessions	Five sessions across Fall and Winter sessions (6 hours each = 30 hours) Content deepening sessions with university faculty	4 sessions (3.5 hours each = 14 hours) analyze own video and that of peers	6 synchronous sessions (2 hours each = 12 hours) analyze own video and that of peers
Winter Sessions		Homework plus 4 sessions (3.5 hours each = 14 hours) develop lesson plans for Spring teaching based on a given set of learning goals and key learning activities	5 weeks (asynch + synch) (~6.5 hrs + 10 hrs = 16.5 hours) develop lesson plans for Spring teaching based on a given set of learning goals and key learning activities

Results

Focal analysis. Table 3 provides descriptive statistics and Table 4 provides model-based results.

Table 3. Descriptive statistics

Condition	Number of individuals	Number of teachers	Pretest mean	Unadjusted posttest mean	Adjusted posttest mean	Unadjusted posttest standard deviation
F2F Content deepening	399	18	40.78	53.81	57.17	9.70
F2F STeLLA	1134	41	45.01	62.34	61.96	13.05
Online STeLLA	472	26	43.16	57.78	59.40	12.64

We emphasize that statistical significance tests are often misleading and provide limited information. Of greater interest are effect sizes and 95% confidence intervals around the effect sizes. Our focal analysis comparing face-to-face and online STeLLA found no significant difference in student achievement ($p = .098$) with a negative estimate of the treatment coefficient (-2.56) and an effect size, $\delta_\tau = -0.233$. The 95% confidence interval around the effect size is large [-0.57, 0.05]. Students of teachers who experienced face-to-face STeLLA were not significantly different at posttest from students of teachers who experienced online STeLLA, but the negative effect size and large confidence interval suggests further analysis is needed. In other words, “not significantly different” is not the same as “not different” and the negative effect size suggests that exploratory analyses are necessary.

Exploratory analyses. We conducted exploratory analyses comparing online STeLLA to face-to-face content deepening PL (the original comparison group in the CRT). Our goal was to examine how students of teachers experiencing the online STeLLA model fared relative to the original comparison group. Using the same analytic model, we found no significant difference in student achievement between online STeLLA and face-to-face content deepening PL interventions ($p = .100$) with a positive treatment coefficient (+2.73) and effect size with 95% confidence interval, $\delta_\tau = .275$ [-0.09, 0.64]. Students of teachers who experienced online STeLLA were not significantly different at posttest from students of teachers who experienced face-to-face content deepening PL, but the effect size was notable with a large confidence interval.

Table 4. Model-based results for models 1-4

Analysis	Fixed Effect	Coefficient	Std. error	*t*-ratio	Approx. *d.f.*	*p*-value
Model 1 – F2F STeLLA v Online STeLLA Confirmatory Analysis	Intercept $\beta_{0,} \gamma_{00}$	60.68	0.75	80.72	63	<.001
	Level 2					
	TRT, γ_{01}	-2.56	1.53	-1.68	63	0.098
	Mean Pretest, γ_{02}	0.24	0.14	1.64	63	0.106
Model does not include demographic variables	**Level 1**					
	Pretest, β_{1j}, γ_{10}	0.65	0.04	17.88	1235	<.001
Model 2 – F2F Content Deepening v Online STeLLA Exploratory Analysis	Intercept $\beta_{0,} \gamma_{00}$	56.23	0.82	68.53	39	<.001
	Level 2					
	TRT, γ_{01}	2.73	1.62	1.69	39	.100
	Mean Pretest, γ_{02}	0.11	0.16	0.708	39	.483
Model does not include demographic variables	**Level 1**					
	Pretest, β_{1j}, γ_{10}	0.63	0.05	12.31	701	<.001
Model 3 – F2F STeLLA v Online STeLLA Exploratory Analysis	Intercept π_0, $\beta_{00,} \gamma_{000}$	60.79	0.87	69.93	42	<.001
	Level 3					
	Pct Undsrvd, γ_{001}	-0.21	0.08	-2.78	42	.008
	Pct FRL, γ_{002}	0.07	0.05	1.54	42	.132
Model includes demographic variables at school level (Level 3)	Pct ELL, γ_{003}	0.13	0.05	2.66	42	.011
	Level 2					
	TRT, $\beta_{01,} \gamma_{010}$	-0.93	1.92	-0.49	19	.633
	Mean Pretest, $\beta_{02,} \gamma_{020}$	0.19	0.14	1.43	19	.169
	Level 1					
	Pretest, π_1, β_{10}, γ_{100}	0.65	0.04	18.19	1188	<.001
Model 4 – F2F Content Deepening v Online STeLLA Exploratory Analysis	Intercept π_0, $\beta_{00,} \gamma_{000}$	56.02	0.68	82.05	27	<.001
	Level 3					
	Pct Undsrvd, γ_{001}	-0.28	0.05	-5.40	27	<.001
	Pct FRL, γ_{002}	0.17	0.38	4.44	27	<.001
	Pct ELL, γ_{003}	0.18	0.03	5.69	27	<.001
Model includes demographic variables at school level (Level 3)	**Level 2**					
	TRT, $\beta_{01,} \gamma_{010}$	4.14	1.47	2.82	10	0.018
	Mean Pretest, $\beta_{02,} \gamma_{020}$	0.15	0.19	0.80	10	0.444
	Level 1					
	Pretest, π_1, β_{10}, γ_{100}	0.63	0.05	11.45	669	<.001

In the original CRT, students of teachers who experienced face-to-face STeLLA outperformed students of teachers who experienced face-to-face content deepening PL ($\delta_\tau = .52$; $p < .001$; Taylor et al., 2017). We reran the original CRT analysis with the smaller dataset (including only those teachers using the water cycle unit in the spring) and nearly replicated the original results, in which students of teachers experiencing face-to-face STeLLA outperformed students of teachers experiencing face-to-face content deepening PL with an estimate of the treatment coefficient (+5.12) and an effect size, $\delta_\tau = .48$; $p = .005$).

These analyses suggest that students of STeLLA Online teachers performed at a level between students of teachers experiencing face-to-face content deepening PL and students of teachers experiencing face-to-face STeLLA PL, but not significantly different from either group. We found our initial results unsatisfying and uninformative. We decided to conduct additional exploratory analyses to consider how demographic differences between groups might explain the relative impact of teacher PL experiences on student achievement.

Further exploratory analyses: including demographic covariates. As shown in Table 1, students of teachers participating in online STeLLA had different demographic characteristics from students of face-to-face STeLLA teachers and students of face-to-face content deepening teachers. Both face-to-face conditions were situated in schools that were Whiter and wealthier than schools participating in online STeLLA. The original CRT study was unable to secure individual student demographic data, so all relevant demographics were obtained from publicly available databases (NCES and SEDA) at the school level. The new analytic model for these exploratory analyses included a level 1 student pretest covariate, level 2 treatment and mean pretest score by teacher covariates, and level 3 school demographics, including the percent of students from groups (such as Black and Latinx) underserved in science; the percent of students receiving free or reduced-price lunch, and the percent of students learning science in an unfamiliar language context.

When comparing face-to-face STeLLA to online STeLLA while controlling for demographic characteristics, we found a smaller estimate of the treatment coefficient (-0.93) and a smaller treatment effect size $\delta_\tau = -0.10$ than what was found in our focal analysis, with students in the face-to-face STeLLA intervention still demonstrating higher achievement. The 95% confidence interval around the effect size, however, is larger [-0.50, 0.30] reflecting a good deal of uncertainty in the estimate. The original differences in student achievement between the online and face-to-face STeLLA interventions may be explained, in part, by the different demographic contexts in the schools.

We repeated the analysis comparing student outcomes for the online STeLLA group to the face-to-face content deepening group. The treatment coefficient was larger than in the original comparison (4.14) with an effect size of $\delta_\tau = 0.44$ and a 95% confidence interval more clearly favoring the students in the STeLLA Online intervention [0.13, 0.74].

Conclusions

Video-based analysis-of-practice models are often complex and intensive, requiring skilled facilitation to support teachers in noticing key elements of instruction captured on video. It also requires high levels of trust: teachers don't engage in constructive critique of other teachers' videos unless a high degree of trust has been established (Beisiegel, Mitchell, & Hill, 2018). Nevertheless, these data suggest it may be possible to construct a fully online version of a complex video-based analysis-of-practice PL model that retains some of its impact on students while providing greater accessibility for teachers and at a lower cost for districts. This study suggests it may be possible for PL developers and providers to create more cost-effective and scalable online versions of complex PL models.

A limitation to the study is the lack of random assignment. It is unclear whether teachers in each modality preferred that modality. Another limitation relates to timing of the online STeLLA study. The online PL began in the summer of 2020. Students receiving instruction from teachers in the online STeLLA PL model were learning science during the most disrupted school year we have ever witnessed. We restricted student data collection to those students who learned science in a face-to-face setting (similar to the experience of students in the original CRT of the face-to-face STeLLA PL). That is, while we wanted to test whether *teachers* could learn in an online format, we did *not* intend to test whether students could learn in an online format. Further study is needed.

References

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice based professional development programs. Zdm, 43(1), 175-187.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568-589.

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.

Roth, K. J., Wilson, C. D., Taylor, J. A., Stuhlsatz, M. A., & Hvidsten, C. (2019). Comparing the effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science. *American Educational Research Journal*, *56*(4), 1217-1253.

Seago, N. M., Jacobs, J. K., Heck, D. J., Nelson, C. L., & Malzahn, K. A. (2014). Impacting teachers' understanding of geometric similarity: results from field testing of the Learning and Teaching Geometry professional development materials. *Professional Development in Education*, *40*(4), 627-653.

Sun, J., & Van Es, E. A. (2015). An exploratory study of the influence that analyzing teaching has on preservice teachers' classroom practice. *Journal of teacher education*, *66*(3), 201-214.

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. (2012). Examining the development of a teacher learning community: The case of a video club. *Teaching and Teacher Education*, *28*(2), 182-192.

Zhang, M., Lundeberg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education, 27*(2), 454-462