

The association between sustained professional development and physics learning

William E. Lindsay,¹ Sari Widman,¹ and Matthew Garcia¹

¹*University of Colorado, School of Education, 249 UCB, Boulder, Colorado, 80309*

This study investigates the association between involvement with sustained professional development (PD) and students' physics learning for teachers engaging with the Physics through Evidence, Empowerment through Reasoning (PEER) curricular suite. PEER supports high school teachers attempting to align their instruction with the Next Generation Science Standards (NGSS) through collaborative engagement with curricular materials, sustained PD, and three-dimensional assessments. Using data collected from 21 teachers and employing hierarchical linear modeling, we examine whether variation in PD participation is predictive of students' gains on a conceptual examination of physics learning. Findings indicate that attending a PEER PD session was associated with a 1.46% positive difference in average gain scores, when controlling for teacher characteristics. PD attendance also explained 49.48% of between-teacher variation in conceptual gains. These results emphasize the efficacy of sustained PD for increasing student learning in NGSS-aligned courses.

I. INTRODUCTION

The recent publication of *A Framework for K-12 Science Education* and the subsequent implementation of the Next Generation Science Standards (NGSS) is a bold attempt to improve science education for all K-12 students [1, 2]. To reach this goal, the NGSS advocates for students to gain mastery over three dimensions of scientific knowledge: disciplinary-specific content ideas, cross-cutting concepts, and science and engineering practices. However, scholars of educational reform have identified many large and expensive reform efforts during the 20th century that failed to foster meaningful and sustainable change [3, 4]. As a possible explanation for this tendency, scholars point to a historical lack of systemic support for teachers attempting ambitious instructional change, including a lack of adequate preparation, material resources, and ongoing professional learning [5, 6]. Advocates of NGSS implementation have attempted to confront this dilemma by claiming that science teachers need ongoing support implementing the standards, including sustained professional development (PD) [7, 8].

One effort to provide reform supports is the Physics through Evidence, Empowerment through Reasoning (PEER) curricular suite. PEER engages teachers with sustained professional development that includes community collaboration to solve persistent problems of practice and is grounded in three-dimensional curricular materials and assessments. Initially based on the design principles of the Physics and Everyday Thinking curriculum [9], PEER has been collaboratively developed over the past eight years by researchers, teachers, and teacher educators at the University of Colorado Boulder. Awareness of the curricular suite has grown through word-of-mouth, conference presentations, and formal and informal connections to the PEER group.

Previous research has suggested that teacher participation with the PEER suite may be associated with increased conceptual gains for students when compared to classrooms using traditional pedagogies [10]. This finding aligns well with other studies of courses employing student-centered physics pedagogies [11-13]. However, research has not disentangled the effects of PEER components on student learning, including sustained PD participation.

The goal of this study is to investigate the relationship between PD participation and physics learning in courses taught by teachers engaging with the PEER curricular suite. To do so, we use student achievement data gathered from the courses of 10 teachers during the 2016-2017 school year and 11 teachers during the 2017-2018 school year. As participation with PEER has grown from a few initial pilot testers to over 60 current teachers, there has been variability in the amount of PD sessions teachers have attended. For example, early implementers of PEER have received more PD than later participants. More random variation in PD attendance has occurred due to scheduling conflicts, illness, or teaching schedule changes. The current study will take

advantage of this variability to analyze whether sustained PD is associated with increased student achievement. More specifically, we attempt to answer the following research questions:

1. *Is teachers' PEER PD participation predictive of students' gains in conceptual understandings of physics?*
2. *How does controlling for teacher characteristics affect the association between PEER PD participation and student gains?*

To explore these questions, we performed a descriptive analysis using hierarchical linear modeling (HLM) to account for the nesting of students' conceptual gains within teachers and to test the idea that teachers' exposure to PD may have an association with students' conceptual gains. While this analysis will not provide conclusive evidence that PD participation is causally linked to conceptual gains, it will provide first-order evidence that this may be the case.

II. THE IMPORTANCE OF SUSTAINED PD

3D-instruction requires facility with scientific knowledge building, content expertise, and disciplinary-specific instructional techniques [1, 15]. Unfortunately, many science teachers have neither expertise with scientific inquiry [16] nor physics-specific pedagogical content knowledge [8]. Moreover, many science teachers are trained by Schools of Education that prioritize learning educational theories over scientific research, often resulting in misunderstandings regarding scientific knowledge production [17].

For practicing science teachers who have not received adequate training during pre-service preparation, PD has been offered as a possible solution [18]. However, not all PD programs are created equal. Specific to the NGSS, Wilson's synthesis of PD research for science teachers identified five key elements for effective PD: focusing on specific content, engaging teachers in active learning, enabling collective participation, coherence, and sufficient duration [19]. Wilson's emphasis on coherence and duration differs from historical approaches to PD, which were usually one-off sessions given by educational experts that focused on transmitting a particular practice or skill without sustained support or follow-up on the impact of PD [19].

Relatedly, Penuel and colleagues' HLM analysis of 454 teachers nested in 28 PD programs found that the length of PD programs was positively associated with the coherent implementation of inquiry-based curricular materials and increased student learning outcomes [20]. The length of PD was also identified in a meta-analysis conducted by Yoon and colleagues as a key predictor of increased student learning gains [21]. These findings align well with our hypothesis that the frequency of PEER PD participation may be linked to increased student learning gains.

III. PEER PD

The design principles of PEER PD are informed by what is currently considered best practices by the education research community [22]. For example, PD experiences are designed to help teachers induce both physics and pedagogical principles. To do so, PEER PD engages teachers with model building activities to learn about evidence, inferences, and knowledge production as a dynamic process. These activities mirror the same inquiry approach that guides PEER lessons, to help teachers understand the pedagogical principles underlying curricular supports. PEER PD occurs during a two-day summer institute and four five-hour sessions on Saturdays during the school year.

PD is also intentionally designed to meet the diverse needs of teachers as they gain more experience collaborating with PEER. For instance, experienced teachers identify common problems of practice that arise across inductive courses and complete investigative cycles to produce local instructional theories [23]. Some experienced teachers also attend trainings to become PD facilitators as the program grows. This sustained, evolving, and responsive commitment to pedagogical improvement through PD is a key component of the PEER approach and may support teachers in implementing student-centered pedagogies previously linked to increased student learning outcomes [10-13].

The current study does not employ a causal design and may produce biased results. For example, if teachers volunteer to attend more PD, their enthusiasm may be a better explanation for their students' higher conceptual gains than what they learned during PD. However, given that participation with PEER is voluntary, studied teachers were all eager to align their instruction with the NGSS and were all volunteering to attend PD. Consequently, we argue that PD attendance variation is mostly random, strengthening the validity of the connection between PD attendance and students' conceptual gains.

IV. STUDY CONTEXT AND DATA COLLECTION

Data collection occurred over the 2016-2017 and 2017-2018 school years as part of a larger study broadly examining the implementation of NGSS-aligned curricular materials in high school physics courses. During the 2016-2017 school year, researchers collected data from seven PEER teachers working in six schools ($N=201$ students) and three teachers using a more traditional approach ($N=131$ students). To recruit comparison teachers, we sent an email to a listserv of local physics teachers offering a small financial compensation for participation. Three teachers from high-achieving suburban schools responded. The inclusion of teachers with no experience teaching PEER and no attendance at PD increases the validity of our analysis, as they act as comparative cases. During the 2017-2018 school year, researchers collected data from the classes of 11 PEER teachers working in 11 different schools ($N=417$ students). We then collapsed the data from 2016-2017 and 2017-2018

into the single dataset used in this analysis. In the cases where teachers taught multiple sections of physics, we collapsed students from different sections into a single nested cluster linked to their teacher. Altogether, the study includes a sample of 749 students from 21 teachers.

Participating teachers were primarily responsible for gathering data from each of their students who consented/assented to participate with the research study. The sample of 749 students represents 85.6% of the students in participating teachers' physics courses, as 126 students declined research participation. There were no apparent patterns indicating why certain students declined or consented/assented to be involved in research. Student-level variables gathered included pre- (August) and post- (May) scores on a conceptual examination of physics knowledge, Likert-scaled measures of confidence and affinity towards physics, and demographic information. The current study uses students' pre- and post-scores on the conceptual exam of physics knowledge as the outcome of interest.

Researchers collected data on teacher experience, PEER participation, and demographics through a survey. Teacher-level variables included the number of PEER PD sessions they had attended, gender, years of experience teaching, years of teaching using PEER curricular materials, the level and discipline of degree they had received, and a self-reported Likert-scaled measure of their district's alignment with the NGSS. The current study uses frequency of PD attendance, years of experience teaching with PEER curricular materials, physics degree, and gender. Descriptive statistics for each variable included in our model are included in Table I.

TABLE I. Descriptive statistics for model variables

Variable	Mean	SD	Min	Max
<u>Student-Level (N=749)</u>				
Pre %	29.80	12.91	0	88.57
Post %	51.64	21.48	0	100
Gain Score %	21.88	18.45	-40	91.43
<u>Teacher Level (N=21)</u>				
PD Frequency	6.81	4.69	0	14
Years PEER Exp	1.48	1.21	0	4
Female %	42.85			
Physics Degree %	38.10			

The content of the conceptual exam was drawn from a variety of items that have had evidence of content and face validity gathered for their use in introductory physics courses. This includes items from the Force Concept Inventory [24] and the Physics and Everyday Thinking Conceptual Exam [9]. Altogether, the exam included 60 items. Cronbach's Alpha for the post-exam equaled .91 for students from the 2016-2017 sample, and .88 for students from the 2017-2018 sample, indicating a high degree of assessment reliability.

Gain scores were calculated for each student by subtracting the percentage of correct answers on the pre-exam by the percentage of correct answers on the post-exam. On average, students gained 21.88% from their pre-score to their post-score. Based on conversations with physics education researchers who had used items from the Force Concept Inventory and Physics and Everyday Thinking Conceptual Exam when teaching introductory physics courses, this average gain was typical for students who had not previously taken physics [25]. The average gain score had a standard deviation of 18.45, indicating high variability.

The variable of PD frequency represents the number of full-day summer PD sessions or half-day Saturday PD sessions that participants reported attending. The average number of PD sessions attended equaled 6.81 with a standard deviation of 4.69. The three comparison teachers attended zero PD sessions, while the minimum PD frequency for PEER teachers equaled three. The variability of PD participation is closely tied to when teachers began partnering with the PEER project, alongside more random variability when teachers did not attend PD because of scheduling or other conflicts. Some of the earliest participants with PEER also used PEER curricular materials without associated PD as PD was not yet offered. This led to variability in teachers' years of experience using PEER curricular materials.

V. METHODOLOGY

We performed a descriptive analysis using HLM. Using two-level hierarchical linear models for this analysis allows us to predict how much of the variation in students' gain scores is explained by teacher characteristics, including the PD attendance frequency, versus how much is explained by student characteristics. Level-1 of the model describes students, while level-2 describes teachers.

We ran three two-level models with student gain scores as our outcome. We first ran an unconditional model to understand the variability of conceptual gains. Then we added our primary predictor of interest at level-2, to see whether there was an association between frequency of PEER PD and conceptual gains, as well as to determine how much conceptual gain variance could be explained by PD attendance. In Model 3, we controlled for potentially confounding teacher-level variables to see if the association found in Model 2 held. We present this final model below.

$$L1: \text{GAINS}_{ij} = \beta_{0j} + r_{ij} \quad r_{ij} \sim N(0, \sigma^2)$$

$$L2: \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{FreqPD}_j) + \begin{pmatrix} \gamma_{02}(\text{YearsExp}_j) \\ \gamma_{03}(\text{PhysDegree}_j) \\ \gamma_{04}(\text{Female}_j) \end{pmatrix} + u_{0j}$$

$$u_{0j} \sim N(0, \tau_{00})$$

(1)

For each model, the outcome variable GAINS_{ij} represents the gain score of student i taught by teacher j . For Model 1, because we do not include any predictor variables at level-2, γ_{00} represents the average gain score for students in a typical teacher's course. We also used the parameter τ_{00} from Model 1 to determine the variance in gain scores between teachers.

In Model 2, we added the level-2 predictor variable FreqPD_j to the unconditional model. We centered this predictor variable around the grand mean, so that the intercept, γ_{00} , represented the average gain score for a student whose teacher participated in an average amount of PD. Accordingly, γ_{01} represents the average change in students' gain scores, associated with the amount of PD attended by teacher j . If γ_{01} is statistically significant it suggests that the student gain score variation is not due to chance alone, but that the amount of PD attendance was a significant predictor of student gain scores. We also compared the variance parameter τ_{00} to that of Model 1, to understand how much gain score variance is explained by PD attendance.

Lastly for Model 3, we introduced a vector of potentially confounding teacher-level variables that includes the dummy variables Female_j and PhysDegree_j , and the continuous variable YearsExp_j . We then compared the results of this final model with Model 2 to see if the addition of potentially cofounding variables altered the findings.

VI. FINDINGS

The fixed and random effects of each model are included in Table II and Table III.

TABLE II. Fixed effects (SE)

Parameter	Model 1	Model 2	Model 3
Intercept, γ_{00}	20.83*** (1.66)	20.86*** (1.29)	23.18*** (1.90)
PD Frequency, γ_{01}		1.14** (0.30)	1.46*** (0.33)
Years PEER Exp, γ_{02}			-0.93 (1.23)
Physics Degree, γ_{03}			-1.26 (2.49)
Female, γ_{04}			4.46* (2.25)
Model Characteristics			
Variance (β_{0j})	281.11	280.83	280.99
ICC (%)	15.47	8.42	8.47

*p<.05, **p<0.01, ***p<.001

TABLE III. Random effects

Parameter	Model 1	Model 2	Model 3
Level 1 Variance, σ^2	281.11	280.83	280.99
Level 2 Variance, τ_{00}	51.45	25.81	25.99
L2 Var. Explained (%)		49.83	49.48

The first model is helpful for identifying the unconditional variance of student gain scores. As seen in Table II, the coefficient on γ_{00} is statistically significant ($p < 0.001$), suggesting an average gain score of 20.83% for any student i nested in teacher j . Table II also shows an intra-class correlation coefficient for Model 1 = 15.47%, indicating that 15.47% of the student gain score variance occurs between teachers.

Knowing that 15.47% of the student gain score variation is between teachers, we then sought to understand if a portion of the between-teacher variance in student gain scores is associated with PEER PD. Table II shows the coefficient on γ_{00} is again statistically significant ($p < 0.001$), suggesting a continuance in associated student gain scores between Model 1 and Model 2. The coefficient on γ_{00} suggests an average associated difference in gain scores of 20.86% for any student i nested in teacher j . Furthermore, the coefficient on γ_{01} , which represents teacher PD attendance, is statistically significant ($p < 0.001$). This suggests that attending an additional PD is associated with a 1.14% difference in student gain scores. Furthermore, Table III shows the level-2 between-teacher variance for Model 2, $\tau_{00} = 25.81$, is smaller by about half of the level-2 between-teacher variance of Model 1, suggesting that adding PEER PD participation to the model reduced the between-teacher variation by 49.83%.

Table II also shows the fixed effects of Model 3, which expands on Model 2 by controlling for teacher level factors. The coefficient on γ_{00} is again statistically significant ($p < 0.001$), suggesting a perseverance in associated student gain scores across Models 1 through 3. In Model 3, the coefficient on γ_{00} suggests an average associated difference in conceptual understandings of physics of 23.18% for any student i nested in teacher j . The coefficient on γ_{01} is statistically significant ($p < 0.001$), suggesting that attending an additional PD is associated with an average gain score difference of 1.46% for any student i nested in teacher j .

Table III shows the level-2 between-teacher variance for Model 3, $\tau_{00} = 25.99$, is again smaller by about half of the level-2 variance of Model 1, suggesting that adding PEER PD participation reduced the between-teacher variation by 49.48%. This is consistent with the findings for Model 2, suggesting a lack of substance to controlling factors. Like Model 2, Model 3 shows a lower intra-class correlation coefficient than Model 1, indicating less between-teacher variation. Therefore, we can assume the lower observed

variation is associated with teachers' PD participation, and that the teacher factors we suspected might play a part explaining the variation difference are not significant.

VII. CONCLUSION

There is evidence to suggest a positive relationship between teachers' sustained participation in PEER PD sessions, and gains in students' conceptual physics knowledge. Among sample classrooms, teacher PD participation plays a substantial role explaining the between-teacher gain score variation. Attending a PD session was also associated with a 1.46% positive difference in students' average gain scores, indicating an association between PD and student learning within our sample. Considering that the average student-level gain score in our sample was 21.88%, an average 1.46% positive difference in students' average gain scores associated with teachers attending one additional PEER PD session is fairly substantial.

Because we lack a true control group, the claims we can make from this analysis are limited. Nonetheless, the results of this study are encouraging, as they suggest that sustained PD is a valuable support for teachers implementing NGSS-aligned physics instruction. Based on this finding, we contend that further study into the relationship between sustained PD and students' learning gains in NGSS-aligned physics courses is a worthwhile pursuit.

As stated in our literature review, not all PD is created equal. If the sustained PD offered to teachers implementing PEER is indeed effective, further study is needed to determine what characteristics of the PD make it successful. Meanwhile, our current analysis leads us to believe that sustained PD may be a valuable support for teachers implementing PEER, and a worthwhile intervention for teachers attempting to reform their physics instruction. As this conclusion also aligns well with the implementation suggestions outlined by the NGSS, alongside best practices for supporting teachers currently identified by the education research community, we are cautiously optimistic that efforts to provide sustained PD aligned with curricular materials have been worthwhile.

ACKNOWLEDGEMENTS

The writers wish to thank participating students and teachers, alongside NSF Grant #937941 for project funding.

[1] National Research Council, *A Framework for K-12 Science Education* (National Academy Press, 2012).
 [2] NGSS Lead States, *Next Generation Science Standards: For States, By States* (National Academy Press, 2013).
 [3] V.K. Otero and D.E. Meltzer, 100 years of attempts to transform physics education, *The Physics Teacher* **54**, 9 (2016).

[4] R.D. Anderson, Reforming science teaching: What research says about inquiry, *Journal of Science Teacher Education* **22**, 7 (2002).
 [5] R. Elmore, Getting to scale with good educational practice, *Harvard Educational Review* **66**, 1 (1996).
 [6] D. Tyack and W. Tobin, The "grammar" of schooling: Why has it been so hard to change, *American Educational Research Journal* **31**, 3 (1994).

- [7] E.M. Furtak, Confronting dilemmas posed by three-dimensional classroom assessment: Introduction to a virtual issue of *Science Education*, *Science Education* **101**, 5 (2017).
- [8] S.L. Pruitt, The next generation science standards: The features and challenges, *Journal of Science Teacher Education* **25**, 2 (2014).
- [9] F. Goldberg, V.K. Otero and S. Robinson, Design principles for effective physics instruction: A case from physics and everyday thinking, *American Journal of Physics* **78**, 12 (2010).
- [10] W. Lindsay, V.K. Otero and S. Belleau, PEER suite: A holistic approach to supporting inductive pedagogy implementation, in *Physics Education Research Conference Proceedings, 2018*, PER Conference (Washington D.C., 2018).
- [11] R.R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 1 (1998).
- [12] D.E. Meltzer and R.K. Thornton, Resource letter ALIP-1: active-learning instruction in physics, *American Journal of Physics* **80**, 6 (2012).
- [13] D.D. Minner, A.J. Levy and J. Century, Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002, *Journal of Research in Science Teaching* **47**, 4 (2010).
- [14] R. Stephen and B. Anthony, *Hierarchical Linear Models* (Sage Publications, 2002).
- [15] W.R. Penuel, C.J. Harris and A.H. DeBarger, Implementing the next generation science standards, *Phi Delta Kappan*, **96**, 6 (2015).
- [16] M. Windschitl, Folk theories of “inquiry:” How preservice teachers reproduce the discourse and practices of an atheoretical scientific method, *Journal of Research in Science Teaching* **41**, 5 (2004).
- [17] N.G. Lederman, *Handbook of Research on Science Education* (Routledge, 2013).
- [18] E.R. Banilower, D.J. Heck and I.R. Weiss, Can professional development make the vision of the standards a reality? The impact of the national science foundation’s local systemic change through teacher enhancement initiative, *Journal of Research in Science Teaching* **44**, 3 (2007).
- [19] S.M. Wilson, Professional development for science teachers, *Science* **340**, 6130 (2013).
- [20] W.R. Penuel, B.J. Fishman, R. Yamaguchi and L.P. Gallagher, What makes professional development effective? Strategies that foster curricular implementation, *American Educational Research Journal* **44**, 4 (2007).
- [21] K.S. Yoon, T. Duncan, S.W. Lee, B. Scarloss and K.L. Shapley, *Reviewing the Evidence on How Professional Development Affects Student Achievement* (Regional Educational Laboratory Southwest, 2007).
- [22] B.J. Reiser, What professional development strategies are needed for successful implementation of the Next Generation Science Standards, in *Invitational Research Symposium on Science Assessment, 2013*, (Washington D.C., 2013).
- [23] C. Lewis and R. Perry, Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning, *Journal for Research in Mathematics Education* **48**, 3 (2017).
- [24] D. Hestenes, M. Wells and G. Swackhamer, Force concept interview, *The Physics Teacher* **30**, 3 (1992).
- [25] D. Meltzer & B. Van Dusen, personal communication, October 26, 2018.