1-2021

Scaling Professional Development: Integrity of Implementation as a Measurement Approach

Brooke A. Whitworth

Lori Rubino-Hare

Nena E. Bloom

Scaling Professional Development: Integrity of Implementation as a Measurement Approach

Brooke A. Whitworth
Clemson University
Lori Rubino-Hare
Northern Arizona University
Nena E. Bloom
Northern Arizona University

Abstract

Professional learning about an innovative teaching method is a demonstrated way to improve teacher practices, and ultimately impact student learning. One way to scale up professional learning is a facilitator development model, in which professional learning and development (PLD) designers prepare facilitators to understand the innovation and they in turn, teach teachers. To understand the effectiveness of this model, identifying how facilitators implement the model with teachers is critical. As such, the Power of Data (POD) team scaled-up effective PLD by providing Facilitation Academies to teach others to facilitate POD Teacher Workshops (TWs). The expectation was that changes based on local contexts would occur; thus, we focused on Integrity of Implementation (IOI) rather than fidelity of implementation. A measurement approach for IOI was created to understand how facilitators (n=13) delivered TWs and how they retained or modified the program principles. Examples from this project illustrate how a robust IOI measurement approach based on a variety of data sources can inform the design of PLD experiences, enable understanding of teacher PLD experiences, and allow researchers to determine whether the scaled-up model is effective.

Introduction

Effective research-supported educational innovations are essential as there are continued concerns about student literacy in science (e.g., NASEM, 2021) and student preparation for (e.g., National Research Council, 2011; National Academy of Engineering & National Research Council, 2014) and interest in the future STEM workforce (e.g., Hidi & Renninger, 2006; Krapp, Hidi, & Renninger, 1992; Lent, Brown, & Hackett, 1994; Maltese & Tai, 2011; Nugent et al., 2015. STEM education reform has been tightly linked with teachers' professional learning and development (PLD) and it is widely accepted that PLD is at the center of educational reform (e.g. Borko, Koellner, & Jacobs, 2014; Desimone, 2009). However, prior to the pandemic the majority of PLD for teachers was often delivered through face-to-face workshops and is often costly and localized (Kennedy, 2016). If we view PLD as critical to educational reform, then it is necessary to identify methods to scale-up effective innovations and programs to benefit a larger number of teachers and students, while still holding to the core principles of the original innovation (Heck, Plumley, Stylianou, Smith, & Moffett, 2019). In order to maximize impact of effective PLD, it should be scaled-up by incorporating the knowledge gained from prior research to ensure success and to benefit more teachers and students. A scaled or scaled-up PLD takes an innovation, or PLD model, from its' initial context and "scales" it or implements in a variety of other settings and contexts – it is termed scaling because it grows from one setting to many (e.g., Borko, 2004). Identifying, promoting, and confirming that the effective aspects of an innovation are implemented is crucial to ensuring that the core principles of the original innovation are maintained during a scale-up.

One way to scale-up an innovation is to use a facilitator development model (Perry & Boylan, 2018), in which trained facilitators work with teachers in their home locales. However, when an innovation is implemented by those other than the designers, it becomes even more necessary to measure how it was implemented, otherwise it is impossible to attribute outcomes to the original innovation. While fidelity of implementation is used extensively in studying implementation in K-12 education settings (O'Donnell, 2008), and recently in studying implementation in higher education (Borrego et al., 2013; Stains & Vickrey

Keywords: *teacher professional learning and development*, *integrity of implementation*, *scaling*, *facilitator development*, *evaluation*

2017), little attention has been given to implementation measures in professional development scale-up studies.

The Power of Data (POD), a PLD model which helps teachers enhance existing instruction by incorporating the examination of geospatial data for a purpose, scaled-up nationally by providing Facilitation Academies (FAs) which taught facilitators how to provide Teacher Workshops (TWs). Previous participants in POD PLD increased their technological and pedagogical skills, confidence teaching with geographic information systems (GIS), and implemented lessons which provided opportunities for students to analyze data and make claims based on evidence (Claesgens et al., 2013). POD PLD resulted in both teacher and student improvement in content knowledge, spatial skills, and scientific reasoning skills (Claesgens et al., 2013). Because of the diversity of local sites and context during implementation, and to allow for facilitators to take ownership over implementation, exact fidelity of implementation of the TWs was not an expectation. Rather, the expectation was for facilitators to implement defined key principles of the PLD, an expectation we termed "integrity of implementation." To determine how facilitators implemented TWs, integrity of implementation (IOI) measures were constructed. This scale-up provides an illustrative case for how to construct a measurement approach for IOI.

Background: Development of the Power of Data Professional Learning and Development Model

PLD is a critical component for successfully developing practices for teachers that lead to effective integration of science and technology (Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). There is an established body of research demonstrating the characteristics of effective PLD (e.g. Desimone, 2009; Loucks-Horsley, et al., 2003; Luft & Hewson, 2014; Whitworth & Chiu, 2015): active learning, coherence, collective participation, content focus, duration, and expert facilitation. PLD standards unique to the practice of science teaching include active investigations of phenomena, interpretation of results, and sense-making (Jeanpierre, Oberhauser, & Freeman, 2005). Participants need opportunities for reflection and metacognition (e.g. Heller, Daehler, Wong, Shinohara, & Miratrix, 2012). Incorporating each of these design elements and conditions of PLD is more likely to result in increased teacher knowledge, changes in practices, and/or improvement in student achievement (Borko, 2004; Garet, Porter, Desimone, Birman, & Yoon, 2001).

In this study, the POD team (n=5), consisting of geologists, GIS experts, and science educators, developed the POD PLD model to align with effective practices for quality PLD and to embody the most current research and expertise of what works in STEM learning. The model helps educators build the technological and pedagogical skills necessary to design data-driven learning activities and GIS tools using an instructional framework called Geospatial Inquiry. Guskey and Yoon (2009) found through a review of the PLD literature that at least 30 contact hours were required to achieve positive effects and that follow-up activities benefit teachers; the POD PLD was designed to be delivered in person over 35 hours in five sessions. PLD is coherent when it is integrated into a program of teacher learning and aligns with the national, state, district, and/or school standards and policies (Desimone, 2009). In the POD PLD model, teachers identify concepts that can be enhanced via the exploration of geospatial data and are provided a template for designing their own lessons (Whitworth et al., 2020). Doing so allows them to align what they have learned with standards in their district and to consider how best to implement it within their own contexts, for their own students, providing some coherency for the PLD. POD facilitators also provided follow-up activities and support as teachers went back to classrooms to then implement these lessons; this further supported coherency as the program was implemented.

The POD PLD model was designed to be delivered by the POD team. However, in response to a call to examine broader implementation of interventions that have demonstrated evidence of impact, the POD team garnered funding to scale-up this innovation via a facilitation development model. When using a facilitation development model, there are multiple tradeoffs to consider. For example, by using this model the POD team did not have the monetary costs of traveling to multiple sites to implement the PLD model, nor did the team have to give up multiple weeks to engage in the implementation. Thus, the POD team had to entrust the implementation of the POD PLD to other trained facilitators who may or may not implement the POD PLD as designed. Although the POD team attended to recommendations for effective PLD when designing the PLD, the team recognized that improving inquiry-based instructional behaviors can be difficult, and that PLD is highly contextual. PLD must be differentiated and responsive to local participants (Desimone & Garet, 2015). We would need to remain flexible in our approach to measuring implementation how the PLD model was translated by others as they delivered Teacher Workshops.

Purpose

The purpose of this paper is to document how one project approached the scaling of PLD and examined the IOI of this scale-up in multiple sites and contexts. We describe how we designed and scaled-up the POD PLD model to support IOI by employing literature from a variety of sources. We considered what is known about scaling innovations, the characteristics of effective PLD, and examined the literature on IOI. We provide a summary of development, then turn our attention to the merits and limitations of our approach to measuring IOI for interpreting downstream effects such as impacts on teacher and student learning, and for upstream implications such as making modifications to teacher and facilitator PLD. We describe implications of IOI for designers and developers who are scaling PLD and for researchers and evaluators who are identifying study outcomes and interpreting results.

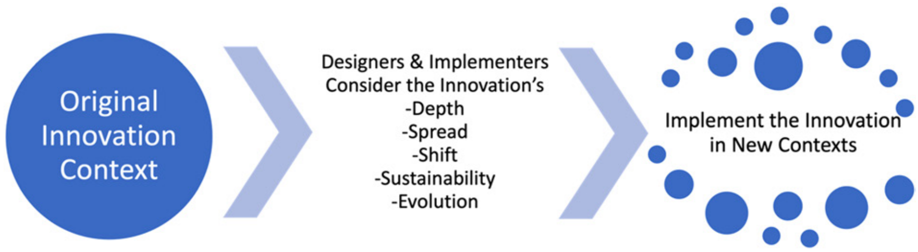

Figure 1. Conceptual Framework for Scaling-up PLD based on Dede and Rockman (2007)

Conceptual Framework – Scaling-up Educational Innovations

Dede and Rockman (2007) suggest a conceptual framework with five dimensions to consider when taking an educational innovation that works in one setting to other contexts (Figure 1). These dimensions include *depth, spread, shift, sustainability, and evolution*. This involves understanding and maintaining the sources of effectiveness and sources of leverage while *shifting* ownership of the innovation to the users, making changes that will enable it to thrive in less than ideal conditions (Dede & Rockman, 2007). In scaling-up an innovation, it is important to consider these five dimensions before the innovation is placed in new hands. While all the dimensions were considered prior to the scaling-up of the POD PLD, *shift, sustainability,* and *evolution* were dimensions we found particularly related to how facilitators were scaling-up the innovation. These dimensions indicate how facilitators are modifying the innovation, how the context necessitates changes to the innovation, and how the innovation is changing in the hands of new implementers.

Considering Fidelity of Implementation

When working with large-scale studies and considering how to evaluate the implementation of an innovation as it is scaled-up, researchers often measure the fidelity of implementation (O'Donnell, 2008). Doing so is important for several reasons: to avoid Type III error (attributing observed changes to the innovation/intervention when they truly are not/conclusions that the observed findings were due to the innovation rather than the way it was implemented), to help explain the downstream successes and failures of implementation, to allow researchers to identify changes that were made by the implementer and how those changes impacted results, and to illuminate the feasibility of implementing an innovation as designed (Dusenbury, Brannigan, Falco, & Hansen, 2003). Traditionally, fidelity of implementation is defined as, "the degree to which teachers and other program providers implement programs *as intended by the program developers*" (Dusenbury et al., 2003, p. 240). In the literature, IOI is similarly defined as, "the degree to which an intervention is *implemented as planned*" (Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993, p. 254). There is a subtle difference between these definitions, but when one more deeply examines the definitions of fidelity and integrity, the differences become clearer. Fidelity is defined as, "accuracy in details, exactness" and integrity defined as "firm adherence to a code of especially moral or artistic values; incorruptibility" (Merriam-Webster, 2019).

In our view, integrity represents more of what we hope facilitators will do with our work as they conduct TWs. We do not expect facilitators to be exact in their implementation, but rather to hold firm to the POD Principles and keep those at the center of their work as they make TWs their own. Given these terms are often interchangeable in the literature, we choose to refer to it as IOI. Aligning with an integrity perspective moves the focus from "exact replication" to implementation that meets the learning goals of the program and supports students in achieving the intended outcomes (Penuel, Phillips, & Harris, 2014).

As designers of PLD consider the scaling of their model, it is essential to determine how IOI will be measured and evaluated as the innovation is scaled-up. Researchers must determine what approach or framework they will use and then consider what instruments and measures will be used to best evaluate integrity in this way. We argue IOI of the original model is a more realistic expectation and allows for changes based on local contexts. Taking an IOI approach allows researchers to recognize the difficulty of scaling-up an innovation and accepts the natural *shifts* and *evolution* that occur in making an innovation *sustainable* (Dede & Rockman, 2007).

Measurement Approach of IOI

There are multiple frameworks and approaches in the literature that refine or combine models for measuring IOI (e.g. Carroll et al., 2007; Century et al., 2010). Dane and Schneider (1998) suggest there are five different ways researchers should typically measure "program integrity": 1. Adherence to the designed program, 2. Exposure (amount of time or program received), 3. Quality of delivery, 4. Participant responsiveness, and 5. Program differentiation (whether critical program features are present). It is often recommended that researchers collect data on all five of these measures; however, it is not evident if all five need to be present for an innovation to accomplish its designed goals (Dane & Schneider, 1998). These five dimensions can be divided into two categories, structure and process, and often are, as researchers seek to measure integrity (Mowbray, Holter, Teague, & Bybee, 2003). Structure refers to the design of an innovation or program and would include measures of adherence and exposure. Process refers to how a program or innovation is delivered which would include quality of delivery and program differentiation. Participant responsiveness would be included in both of these categories (O'Donnell, 2008).

Another approach to implementation that is sometimes combined with the above frameworks is the critical component approach; Hall & Hord, 1987. The critical component approach (Hall & Hord, 1987) suggests that an innovation consists of defining features or components that must be measured to determine if integrity of a program exists or not. Stains & Vickrey (2017) also describe the importance of identifying critical components of an intervention, including both structural (expected organizational features) and process components (expected implementation features) (from Century, Rudnick & Freeman, 2010), in their work to develop a fidelity of implementation framework for evidence-based instructional practices in higher education. They go on to suggest a process for using a fidelity of implementation framework for both efficacy and effectiveness studies.

The team chose to measure IOI by examining three areas: whether the critical components of the original innovation are present, in what doses, and delivered with what measure of quality (e.g., Heck et al., 2012). Below we describe our study, and then present a case of how we conceptualized and measured IOI in our study for illustrative purposes and to present an argument for this approach.

Scaling the POD PLD Model

In scaling the POD PLD model, the team attempted to anticipate how the successful PLD model might *shift* and *evolve* in the hands of others, what critical components must be maintained or the *depth* of the innovation that would be retained, and what could be modified (Whitworth et al., 2020). Additionally, the team needed to identify how to communicate core elements of the PLD model so it might best be shared with future facilitators. To do so, the POD Team analyzed how the PLD model was enacted in the effective Teacher Workshops that were delivered by the designers in the past as well as the current literature in science education, GIS education, and teacher PLD. This resulted in defining effective science teaching with GIS through the process of *Geospatial Inquiry*: asking and answering a question through the analysis and communication of data that are linked to a geographic location on, above, or near Earth (Rubino-Hare et al., 2016a). These data are often represented visually via maps. The goal of POD Teacher Workshops (TWs) is to enable secondary teachers to enhance existing courses via effective *Geospatial Inquiry* (Whitworth et al., 2020).

Additionally, the team identified seven principles of the PLD that would be critical for scaling (Rubino-Hare et al., 2016a, p. xix). These principles emphasize how *Geospatial Inquiry* supports learning and promotes student-centered, responsive teaching. They were based on literature reviews and a theoretical framework of situated cognition, which suggests learning is contextualized and created as individuals interact with their environment to achieve a goal (Brown, Collins, & Duguid, 1998). The principles and *Geospatial Inquiry* formed the conceptual basis for and drove the development of the content of the POD Teacher Workshops (TWs) and core elements of POD TWs.

Six session components that prepare teachers with technological and pedagogical skills and knowledge necessary to integrate *Geospatial Inquiry* into existing courses were identified and woven throughout the 35-hour program (Figure 2). Scaffolds and tools that manifest POD principles and support teacher learning including a *Geospatial Inquiry* learning cycle, frameworks for supporting geospatial analyses (Mitchell, 1999) and a *Geospatial Inquiry* Lesson template (Rubino-Hare et al., 2016a) were developed and incorporated into a Facilitation Guide and an accompanying Teacher Guide to enable facilitators to implement Teacher Workshops with integrity to the model (Whitworth et al., 2020).

POD Facilitation Academies - Designing for Flexibility

Effective PLD is generally characterized by expert facilitation (Borko et al., 2014). An expert facilitator must have knowledge of not only the content pedagogy, but also of how to effectively lead PLD. Therefore, potential facilitators were required to apply to participate in the in-person POD Facilitation Academies. Facilitators were chosen based on their knowledge of science, their GIS technical skills, their experience teaching with GIS, learning with GIS, teaching others to teach with GIS and facilitating adult learners. By taking this approach, we ensured the selection of expert facilitators for POD TWs (Whitworth, et al., 2020).

In the Academy, facilitators experienced the POD TW they would ultimately facilitate in-person so they could broaden their understanding and knowledge of the content (Loucks-Horsley et al., 2003). Facilitators need time to practice and reflect upon ways to support adult learners (Loucks-Horsley, et al., 2003; Perry & Boylan, 2018); facilitators debriefed the facilitation moves (Garmston & Wellman, 1999) that were modeled by the POD Team during the

Session 1	Session 2	Session 3	Session 4	Session 5
Intro to POD	Geospatial Inquiry	Geospatial Inquiry	Geospatial Inquiry	Implications for Teaching with Geospatial Inquiry
Geospatial Inquiry	Designing a Geospatial Inquiry	Implications for Teaching with Geospatial Inquiry	Implications for Teaching with Geospatial Inquiry	
Pedagogical Moves		Career Spotlight	Career Spotlight	Designing a Geospatial Inquiry
Geospatial Inquiry	Career Spotlight	Designing a Geospatial Inquiry	Pedagogical Moves	
		Pedagogical Moves	Designing a Geospatial Inquiry	Celebration
Metacognition	Metacognition	Metacognition	Metacognition	Evaluation

Figure 2. Order of Sessions and Components in Teacher Workshops

delivery of the FA such as negotiating working agreements, responsiveness to participants, and varying grouping techniques to promote active participation and productive discussion.

The FA included the examination of resources and the underlying program *principles*, which were intended to deepen facilitators' understanding of the design of the TW and how each session component of the TW helps teachers provide opportunities for students to engage in *Geospatial Inquiry*. The team shared rationale for PLD design decisions such as decreasing the technological demands in favor of using simple GIS analysis tools in powerful ways. Facilitators also had time to plan and customize their TWs within the constraints of the principles. They were told they had freedom to make changes as long as these changes were in agreement with overall POD principles and were aligned to the goals identified for teachers within each of the components and sessions. For example, the context for *Geospatial Inquiry* was earth and environmental science, but this could be modified as long as teachers were provided opportunities to examine geospatial data for a specific purpose, use GIS as a tool to explore patterns and relationships, critically analyze and interpret data, creatively select and display appropriate data as evidence to support or refute a claim, and complete one full cycle of *Geospatial Inquiry* as an adult learner. Drawing attention to these design elements helped increase the likelihood that IOI would be achieved, and the implementation would be successful as facilitators began leading TWs (Whitworth et al., 2020). Following the Academy, participants participated in a professional learning community via synchronous online meetings to support continued improvement (Learning Forward, 2011) as well as frequent emails and individual meetings with the POD PLD designers as needed.

Within one year of completing the FA, facilitators were required to implement their first POD TW, provide feedback, and within a second year they were required to implement the second iteration of a POD TW (Whitworth et al., 2020). Since the goal was to design for flexibility and scale-up, and the designers expected facilitators to adapt materials for their unique contexts, a Design-Based Research (The DBR Collective, 2003) approach was employed. Through informal and formal feedback from facilitators after implementation of two FAs and subsequent TWs, the design team made modifications that provided support to facilitators so they could effectively adapt the innovation with integrity (Whitworth et al., 2020). To determine whether the POD PLD model was effective when implemented at scale, a rich measure of IOI was necessary.

Participants

Participants (n=15) of the FA were PLD specialists, college faculty, school STEM education specialists, and others whose primary professional goal was to support teachers. Applicants were selected as participants based on whether they had prior experience leading PLD for secondary teachers, held high geospatial technology (GST) skills as measured through an assessment, and whether they had experience implementing projects with students. If facilitators had prior experience, scored highly on a GST assessment, and had experience implementing projects with students, they were offered the opportunity to attend the FA. Participants attended a week-long FA to: 1) Learn through *Geospatial Inquiry* similar to a "typical" TW; 2) Develop an understanding of the POD Principles essential for facilitating successful TWs; and 3) Receive access to the Facilitation Guide, data, and other online materials to support implementation of at least two TWs. Within one year of completing the Academy, 13 facilitators implemented a TW, each with 10-15 teachers in their own local setting representing nine unique states across the country. Two of these TWs were delivered by a team of two facilitators working as a pair.

IOI Measurement Approach

There were a variety of approaches and modifications to delivery of POD TWs. We needed a way to analyze and measure these changes. We modified an approach provided by Heck, Chval, Weiss, and Ziebarth (2012) to develop an IOI measurement approach. Our approach included logistic integrity (following lesson sequence and timing), enactment integrity (adhering to lesson purpose, objectives, goals, principles, and the level of modifications; adherence and program differentiation), and quality (adhering to high quality facilitation moves as identified by principles; quality of delivery and participant responsiveness). These categories encapsulated the various ways we felt facilitators might implement differently and allowed us to examine and categorize the different ways facilitators were modifying the TWs.

Logistic Integrity Scores

Logistic integrity scores measured facilitators' adherence to the sequence of activities within a daily session, and the amount of time (exposure) spent on activities as suggested in the TW. The data sources were the implementation logs facilitators completed after each day of the TW. For instance, for each daily session, facilitators were asked to compare how much time was spent on the session with how much time was recommended for the session in the manual. Data from these logs were used to determine a mean adherence rating for the daily session sequence, and a mean activity exposure rating for the daily session, each measured on a three-point scale. These means were totaled to provide each facilitator with a logistic integrity score, on a scale of 0-6.

Enactment Integrity Scores

Enactment integrity scores measured facilitators' adherence to the lesson purpose, principles and objectives, as well as the level of modifications they made during implementation. The data sources were the daily implementation logs facilitators completed. For instance, on the implementation forms, facilitators were asked to indicate if they taught explicit steps of the workshop components, made modifications or extensions as compared to how the session components were presented in the TW manual, and asked to identify which POD principles were present in the session. Data from implementation logs were utilized to create a daily session adherence/level of modifications rating on a three-point scale. These ratings were summed to provide each facilitator with an enactment integrity score, on a scale of 0-15.

Quality Integrity Scores

Quality integrity scores measured the extent to which the facilitator supported adult learning, met the goals of the session as stated in the TW manual, and used appropriate facilitation moves (e.g. supporting academically productive talk and building rapport, best practices for adult learners) during the TW. Data sources included the videos of TWs and post-workshop surveys from teachers who participated in the workshops. Facilitators were asked to video-record a purposefully selected seven sections of the TW across the five sessions. The segments were selected to enable researchers to examine a variety of research-based factors for supporting adult learning, such as how facilitators framed the learning, supported teachers in sense making, how they emphasized the principles, and how they wrapped-up the learning. Facilitators then sent the video files to the researchers after the TW. To identify quality, researchers looked for whether facilitation strategies supported adult learning and whether session goals were met. Researchers noted the identified goals and POD principles for each of the seven sections, as identified in the TW manual. Two researchers observed implementation of the TWs using the videos, discussed what they saw, and assigned a quality of delivery rating on a three-point scale with three indicating highest use of strategies that were likely to enhance learning, improve upon resources of the session, and/or result in meeting most of the session's goals. Lower scores indicated that few facilitation strategies that support learning were used, or few session goals were met.

As another source to assess the quality of the PLD, researchers analyzed teachers' post TW surveys for their perception of facilitator quality. Ratings of teacher satisfaction with the PLD for preparing them to understand and be prepared to teach with *Geospatial Inquiry* and scores of teacher satisfaction with the level of facilitator geospatial technology skill and overall effectiveness as a PLD facilitator, all on a five-point Likert scale, were added. The video observation score was weighted higher than the teacher response score, as teachers might be likely to be satisfied with PLD whether it aligned with intended design features or not. The video score and the teacher survey rating were summed to assign each facilitator with a quality score, on a scale of 0-14.

Analysis of the IOI Data

We analyzed the IOI data in a number of ways, depending on the purpose and use of the data.

Developing Facilitator IOI Profiles

Profile scores were developed for each facilitator. The logistic integrity, enactment integrity, and quality integrity scores (described above) were converted to percentages to make comparison across different scales easier. These were then entered into a bar graph to allow comparison of profiles. As profiles were compared, similar profiles were grouped together so patterns could be more easily visualized and analyzed. Analyzing the data in this way allowed us to identify patterns and to group the types of modifications we were seeing more easily. A sample is provided (Figure 3) for demonstration purposes.

Average IOI Score of Facilitators

The average facilitator ratings on each IOI construct were calculated (Table 1). Results indicated facilitators had a range of logistic integrity ($M = 4.4$, $SD = 1.0$), an average of 73% adherence to the measure, and enactment integrity ($M = 9.8$, $SD = 3.5$), an average of 65% adherence to the measure. However, facilitators had high quality of integrity scores ($M = 12.6$, $SD = 1.3$), an average of 90% adherence to the measure. Analyzing this information provided early information to the researchers and the development team about how well the facilitators were able to follow the program model. Because part of the application process was to provide evidence of prior experience delivering PLD, it is not surprising that most facilitators exhibited high quality facilitation moves. Although not part of the IOI analysis framework, post-workshop interviews with facilitators (described below) illuminated a variety of reasons they modified

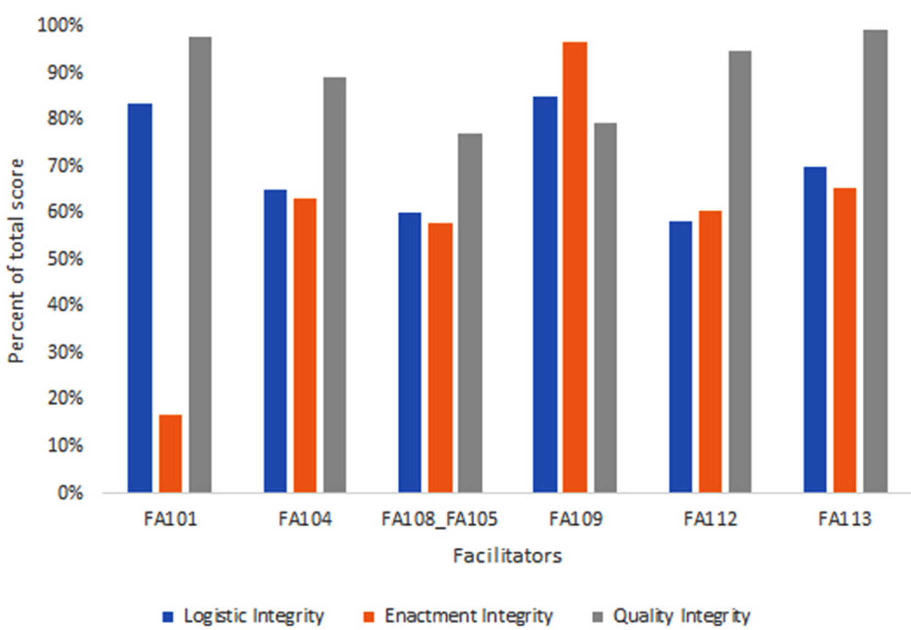

Figure 3. Examples of IOI profiles.

Table 1. Average Facilitator IOI Scores (n=13)

	Logistic Integrity (0-6)	Enactment Integrity (0-15)	Quality (0-14)
Mean	4.4	9.8	12.6
Standard Deviation	1.0	3.5	1.3
Percentage of total score	73%	65%	90%

their workshops. It is useful to triangulate IOI scores with qualitative feedback. This feedback can help to explain the varying IOI scores, why facilitators modified sessions, and may help to identify supports and/or challenges facilitators encountered in *shifting* the innovation to their context. Some challenges to enactment and logistic integrity in the POD PLD were due to facilitators offering the TW online or through a hybrid designed course that changed how they chose to implement and modify the model.

Use of IOI Scores for Formative Changes to POD PLD Model

We determined the importance of providing specific feedback to facilitators on which model elements could be provided online and which should be face-to-face, as well as which should be synchronous or asynchronous, though we were constrained by data we needed to collect for the research. For example, some sessions had to be implemented in person so we could collect videos for observations. We also made minor modifications to the TW based upon the logistic and enactment scores. For example, we changed the timing of a few sessions because multiple facilitators reported it took significantly more or less time than suggested, and we simplified some directions. These changes were shared via updated Facilitation and Teacher Guides and were explained in an online meeting with facilitators so they could use the modified Guides in successive TWs. We did not want to ask facilitators to implement a TW that was drastically different from what they experienced in the FA, so some elements, such as the sequence of activities, were not changed. In a subsequent version, we changed the sequence of the activities. The changes were implemented in a second FA attended by a new cohort of facilitators (n = 13). Another change involved our recruitment of facilitators. In Cohort 1, most facilitators attended the FAs as individuals, but some came in teams. Those who came with a partner reported benefits of planning and delivering TWs together, so we encouraged teams to attend FAs in the second cohort. The logistic and enactment scores also informed more significant edits to the final TW and accompanying Facilitation Guide. For example, when we realized the suggested enactment of a session was difficult for multiple facilitators, we made changes accordingly. Results from the IOI scores have continued to inform our work. For instance, the designers recently developed and delivered an online version of POD PLD based on the results of this study, using the IOI scores to determine which parts of the program needed to be synchronous.

Use of IOI Scores in Larger Data Analysis

Using an IOI score in analysis allows researchers to have a better understanding of what implementation factors are important in a scale-up of an innovation, and how implementation factors affect resulting outcomes. For example, as part of the analysis of the POD PLD scale-up, researchers are determining the skills, understandings and preparedness of teachers who participate in the TWs in a future study. Facilitator implementation choices in the PLD, as well as other variables, are hypothesized to influence teacher outcomes. Therefore, to understand the relationship between these variables and facilitator implementation factors, in a larger study comprising an additional cohort of facilitators, many of whom taught multiple TWs, IOI scores are included in analysis. Using Poisson regression models, results indicated high logistic integrity, but not enactment integrity or quality integrity, influenced GST skills. This indicates facilitator logistic integrity may not be a factor when trying to improve teachers' GST skills scores. This information illuminates that low logistic integrity may not be important for positive teacher outcomes but ensuring facilitators of PLD have a deep understanding of the principles (enactment integrity) and an ability to implement high quality PLD (quality integrity) may be of more importance when scaling-up and implementing PLD Using an IOI score in analysis allowed us to better understand teacher outcomes that resulted from the scale up of our program.

Reasons for Varying IOI

Beyond quantitative data calculated to create the IOI measurement approach, qualitative data from the participants can be examined to provide researchers a richer understanding of implementation, providing an actor-oriented perspective (Penuel, Phillips, & Harris, 2014) and a window into how and why participants implement the way they do. To illustrate this concept, we share some insights gleaned from Facilitator open-ended responses on surveys as well as quotes from post-TW facilitator interviews about implementation choices and challenges.

Oftentimes facilitators made decisions to modify activities based on their own beliefs about what was important to emphasize and their learning goals for their teacher participants. For example, one facilitator stated,

> We spent more time on designing their own inquiry because I feel that is the most important part of the POD workshop. I allowed the participants more time to work on their geospatial designs and think about how ArcGIS could be implemented. (FA112)

Other times changing the delivery mode from in person to online impacted timing: "some of it was modified just because it was going into an online format…" (FA114). And another facilitator shared:

I wrote down the time allocations for [the overall agenda] based on the guidance in the manual and then I kind of went through those ahead of the workshop to parse out what I could have the teachers do ahead of time and reduce time for those activities. We didn't cut anything out. We just kind of condensed things or they did some of the reading ahead of time or watched some of the videos and then we have mentioned that and by the way you watched that ahead of time. (FA113)

Facilitators also made decisions because of a need to differentiate the PLD for their particular context:

More hands-on exercises and experience in ArcGIS Online is needed for almost all teachers. Initially the teachers struggled with learning ArcGIS Online. Some were getting frustrated. There is no way I could have progressed through the workshop without providing additional instruction and time to learn ArcGIS Online. (FA109)

This facilitator felt the need to provide more time to work with the technology before working through other aspects of the PLD. In addition, several comments provided insight on the usefulness of the components and materials provided: "just too much for teachers to process. It's a lot of stuff to get through with teachers - it feels complicated. We ran out of time to do everything." (FA104). Another facilitator shared,

The teachers really needed the time to work on Geospatial Inquiry. I think the Implications for Teaching portion helped…. I think having task cards that show how to buffer and do additional analysis would have helped with their creation of their presentations/story maps…. They really like the [GIS] task cards. I thought examining student work and discussing it was really helpful for them. (FA113)

Taken together these qualitative results help explain why there were varying IOI scores and what might have been influencing some of the variance singular we saw in the data.

Discussion

Having a measure of IOI has allowed us to better understand teacher outcomes that result from scaling-up PLD. For instance, we learned that teachers who attended workshops provided by facilitators with high logistic integrity scores tended to score higher on GST performance assessments. Studying this construct also allowed us to confirm some of our conceptual assumptions, that the innovation will *shift* and *evolve* in the hands of users (Dede & Rockman, 2007), and that not all changes will negatively affect teachers' development of skill or preparedness in the program. In scaling the POD PLD model, we also encountered a variety of issues and learned many lessons. Here we discuss and share those challenges and lessons we think are most pertinent for others examining IOI as they scale-up their own innovations.

Data Collection

One of the largest issues encountered in measuring facilitator implementation of the POD PLD model, as might be expected, was the collection of data from facilitators and teachers in the workshops. Despite tying stipends to the completion of data submission, many facilitators failed to turn in complete data sets (i.e. videos of TWs for scoring observations for the Quality rating) for the implementation of TWs. One reason was a shift in delivery mode which made it difficult to collect data through the means we had designed. Although the original model was designed for face-to-face TWs, facilitators reported it was often difficult for teachers to commit to a five-day, 35 hour, in person workshop, particularly in rural geographic locations. Therefore, they chose to implement in a variety of other contexts including face-to-face over a series of weekends, all online, and/or hybrid workshops where some portions were attended online and others in-person. This significantly changed how we could collect data to measure IOI. Thus, our research plan lacked a means to collect observation data for online components rather than via videos of in-person implementation. Desimone (2009) suggests that other methods, such as well-designed teacher surveys, can be used to measure certain features of teacher learning experiences, such as the quality of the PLD, and may be more consistently able to capture this construct than observations. Hybrid and blended PLD will likely grow in importance, thus having appropriate methods to collect implementation, and other research data, in these environments is a critical consideration to identify how implementation changes in these environments.

Depending on the scale-up of the innovation, it may be a challenge to collect data at distance. Collecting teacher and student data at a distance was difficult. We relied on the facilitators to help us with the data collection and despite stipends being tied to data collection, we had great difficulty in obtaining the data. If we could do it over, we would utilize a learning management system where all facilitators and teachers enrolled and completed aspects of the course and data collection through the online system. Doing so would allow us to directly monitor the completion of workshops, data, and lessons.

Additionally, we intended each of the 15 facilitators to teach two TWs for 15 teachers each; however, in response to the feedback received from the initial cohort, we modified the FA to encourage teams of facilitators to deliver TWs together. This change resulted in lower numbers of teachers, and ultimately students, from whom we could collect data to determine the downstream impacts of the PLD model. This is a tradeoff we accepted when choosing the DBR model.

Implications for Stakeholders

Our work has led us to several recommendations for researchers, evaluators, and PLD designers as they take their innovation to scale and study outcomes. Additionally, we provide some insight for educators and items to consider.

Examining the outcomes using an IOI measurement approach has been illuminating. Program design and support embedded in the FAs and explicitly taught to facilitators, such as having a defined model, principles and critical PLD components, have enabled IOI (Whitworth et al., 2020). Nevertheless, some unexpected outcomes arose. For example, a few facilitators chose to deliver their Workshops in hybrid and online settings and others delivered them over a period of several months even though the program was designed to be delivered face-to-face over five consecutive days. Approaching the research from a design-based perspective has enabled flexibility during these times.

For Researchers and Evaluators

Rather than assuming that implementation was as intended by developers, having a measure of IOI promotes validity of study findings (Stains & Vickrey, 2017). IOI also allows researchers and evaluators to more deeply examine certain aspects of the innovation scale-up, such as what educators modify and what they keep, as well as whether different methods of delivery are equally effective for identification of and implementation of critical program components. Part of this understanding, in our study, came from the logistic and enactment integrity scores and examining closely where *shifts* were made and identified in the implementation logs. qualitative data collected provided additional insight into why *shifts* were made by facilitators. It may also be that the learning outcomes impact what part of IOI is most important when scaling. As in our research we found facilitator logistic integrity was positively correlated with teacher GST skills; however, if our focus had been on different learning outcomes, we may have had a different result. Researchers should also carefully consider implications for collecting data at scale, when there is no direct contact with the participants. A learning management system may be a solution for this challenge; however, the training time to use a learning management system may also need to be considered. If this option is selected it may be important to identify one with a high ease of use and/or familiarity for many.

For PLD Designers

Sustainability requires the innovation design be strong enough to tolerate negative shifts in context (Dede & Rockman, 2007), retaining its core characteristics for success yet flexible enough to adapt and modify as needed. As described above, the POD Team worked with facilitators and continued to develop resources so TWs could be delivered to meet the needs of individuals in particular contexts.

Having a measure of IOI is critical to determine what is important to maintain in the model, and what is not. For example, when low IOI is measured, taking an actor-oriented approach (Penuel et al., 2014) can help designers better understand reasoning for modifications, identify potential problems and make adjustments so materials can support higher integrity when implemented under less than ideal conditions. We have learned through examining the facilitator IOI scores, to maintain the effectiveness of the model, designers must ensure facilitators both understand the underlying principles of the PLD model and the core program components and implement these in the TWs. We have also learned that the order of component implementation in the workshop (logistic integrity) is more critical for skill development. With this understanding, even with modifications, educators will maintain the principles integral to the program (Krajcik, Blumenfeld, Marx, & Soloway, 1999).

Ensuring this deep understanding also allows for delivery of PLD in both online and hybrid environments, which may decrease costs by not requiring the use of space and/or hands-on materials and result in higher teacher participation. In the event that online PLD happens, as per a proposed spectrum of online PLD (Cheng & Hanuscin, 2012), designers can suggest a mix of asynchronous activities, such as lesson planning or practicing GIS, and synchronous interactions in the form of webinars or break-out rooms to allow a more face-to-face authenticity to the online environment (e.g. discussions).

Finally, although we attempted to provide support for facilitators via frequent communication, they likely required more. Facilitators sought out a community of practice and reached out to one another for ideas. More resources and space could have been built into our model. We recommend designers who scale innovations through a facilitation development model attend more closely to the needs of facilitators (Perry & Boylan, 2018).

Conclusion

The examples from the scaling of the POD PLD model have demonstrated how a robust IOI measurement approach utilizing a variety of sources of data can inform the design of PLD experiences. It also enables us to better understand facilitation in a scale-up project and the resulting teacher PLD experiences, which in turn enables researchers to determine whether the scaled-up model is effective. Through this example, we hope others can find ways to define and develop their own IOI measures for projects they are scaling-up. In future research, we hope to gain a deeper understanding of how the IOI measurement approach can be used in the scaling-up of other projects and in other K-12 and higher education settings.

Acknowledgements

This material is based upon work supported by the National Science Foundation under DRL 1513287. Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

References

Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J.E. (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses. *Journal of Engineering Education*, 102, 394–425.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation science, 2*(1), 40-49.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199-218.

Claesgens, J., Rubino-Hare, L., Bloom, N., Fredrickson, K., Henderson-Dahms, C., Menasco, J. Sample, J.C. (2013). Professional development integrating technology: Does delivery format matter? *Science Educator, 22*(1), 10-18. https://www.nsela.org/assets/Professional%20Development%20Integrating%20Technology-%20Does%20Delivery%20Format%20Matter.pdf

Dane, A. V. & Schneider, B. H. (1998) Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review, 18*, 23–24.

Dede, C. & Rockman, S. (2007, Spring). Lessons learned from studying how innovations can achieve scale. *Threshold,* 4-10, 16.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003, April). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256. Retrieved from: https://doi.org/10.1093/her/18.2.237

Garmston, R., & Wellman, B. (1999). *The Adaptive school: A sourcebook for developing collaborative groups.* Norwood, MA: Christopher-Gordon Publishers, Inc.

Gresham, F.M., Gansle, K.A., Noell, G.H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22*, 254–272.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process.* State University of New York Press.

Heck, D.J., Chval, K.B., Weiss, I.R., & Ziebarth, S.W. (2012). Developing measures of fidelity of implementation for mathematics curriculum materials enactment. In D.J. Heck, K.B. Chval, I.R. Weiss, & S.W. Ziebarth (Eds.), *Approaches to Studying Enacted Mathematics Curriculum.* Charlotte NC: Information Age Publishing.

Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teacher's classroom practices. *Journal of Research in Science Teaching, 42*, 668-690.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*, 945-980. DOI: 10.3102/0034654315626800

Krajcik, J.S., Blumenfeld, P.C., Marx, R.W., & Soloway, E. (1999). Instructional, curricular, and technological supports for inquiry in science classrooms. In J. Minstrell & E. V. Zee (Eds.), Inquiry into inquiry science learning and teaching, Part 3 (pp. 83-315). Washington, DC: American Association for the Advancement of Science Press.

Lawrenz, F., Huffman, D., Appeldoorn, K., & Sun, T. (2002). *CETP core evaluation, classroom observation handbook.* Minneapolis, MN: CAREI.

Learning Forward. (2011). *Standards for professional learning.* Oxford, OH.

Loucks-Horsley, S., Love, N., Stiles, K.E., Mundry, S. & Hewson, P.W. (2003). *Designing professional development for teachers of science and mathematics (2nd Edition).* Thousand Oaks, CA: Corwin Press.

Luft, J. A. & Hewson, P. W. (2014). Research on teacher professional development programs in science. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*, (pp. 899-910). New York, NY: Routledge.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science education*, *95*(5), 877-907.

National Academy of Engineering and National Research Council. (2014). STEM integration in K-12 education: Status, prospects, and an agenda for research. Washington, DC: The National Academies Press. https://doi.org/10.17226/18612.

National Academies of Sciences, Engineering, and Medicine (NASEM). (2021). Call to action for science education: Building opportunity for the future. Washington, DC: The National Academies Press. https://doi.org/10.17226/26152.

National Research Council. 2011. Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics. Washington, DC: The National Academies Press. https://doi.org/10.17226/13158.

Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. Review of Educational Research, 78, 33-84.

Penuel, W. R., Phillips, R. S., & Harris, C. J. (2014). Analysing teachers' curriculum implementation from integrity and actor-oriented perspectives. *Journal of Curriculum Studies*, *46*, 751-777.

Perry, E. & Boylan, M. (2018). Developing the developers: Supporting and researching the learning of professional development facilitators. *Professional Development in Education, 44*(2), 254-271.

Petcovic, H.L., Cervenec, J, Cheek, K., Dahl, R., & Price, N. (2018). Research on elementary, middle, and secondary earth and space sciences teacher education. In St. John, K (Ed.) (2018). *Community Framework for Geoscience Education Research*. National Association of Geoscience Teachers. https://doi.org/10.25885/ger_framework/4

Rubino-Hare, L., Evans, E., Manone, M., Palmer, A., & Sample, J. C. (2016a). Power of data teacher guide. Teacher guide for power of data teacher workshops.

Rubino-Hare, L., Whitworth, B., Bloom, N., Claesgens, J., Fredrickson, K., Henderson-Dahms, C. & Sample, J.C. (2016b). Persistent teaching practices after geospatial technology professional development. *Contemporary Issues in Technology Education, 16*(3). https://citejournal.org/volume-16/issue-3-16/science/persistent-teaching-practices-after-geospatial-technology-professional-development/

Singer, J., Marx, R. W., & Krajcik, J. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist, 35*, 165–178.

Stains, M. & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education*

The Design-Based Research Collective. (2003). Design-Based Research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Wallace, M. R. (2009). Making sense of the links: Professional development, teacher practices, and student achievement. *Teachers College Record, 111*, 573-596.

Whitworth, B. A. & Chiu, J. L. (2015). Professional development and teacher change: The missing leadership link. *Journal of Science Teacher Education, 26,* 121-137. DOI: 10.1007/s10972-014-9411-2

Whitworth, B. A., Rubino-Hare, L., Bloom, N. E., Walker, M. C.*, & Arendt, K.* (2020). Scaling professional development: Preparing professional development providers to lead Power of Data Teacher Workshops. *International Journal of Science Education, 42,* 1-24. DOI: 10.1080/09500693.2019.1699975

Correspondence Details:
[1]bawhit@olemiss.edu; ORCID: https://orcid.org/0000-0002-3944-291X; @bawhit41
[2]*@lhare24; www.linkedin.com/in/lori-rubino-hare*

Short biographical notes on all authors:

[1]Brooke A. Whitworth is an Associate Professor of Science Education at Clemson University in Clemson, SC.

[2]Lori Rubino-Hare is a Professional Development Coordinator at Northern Arizona University in Flagstaff, AZ.

[3]Nena E. Bloom is the Assistant Director of Evaluation and Assessment Programs at Northern Arizona University in Flagstaff, AZ.