Research Article

Teaching Aquatic Science as Inquiry through Professional Development: [1]

Teacher Characteristics and Student Outcomes

Kanesa Duncan Seraphin 0000-0001-9961-5782 0000-0001-9961-5782[a,b]

George M. Harrison 0000-0002-6011-0063 0000-0002-6011-0063[a]

Joanna Philippoff[a]

Paul R. Brandon[a]

Thanh Truc T. Nguyen[a]

Brian Lawton[a]

Lisa M. Vallin[a]

[a] Curriculum Research & Development Group, University of Hawaiʻi at Mānoa, Honolulu, HI, USA

[b] University of Hawaiʻi Sea Grant College Program, Honolulu, HI, USA

[1] This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/tea.21403

Abstract

We present an inquiry-based, aquatic science professional development (PD) for upper-elementary, middle, and high school teachers and examine changes in student outcomes in light of participating teachers' characteristics and the grade band of the students. Our study lends support to the assertion that inquiry- and content-focused PD, paired with classroom implementation, can effectively improve student learning. Our findings indicate that students improved in their nature of science (NOS) and aquatic science content knowledge and that these changes depended in some ways on the participating teachers' characteristics and adherence to the program. The students' improvements were amplified when their teachers adhered more closely to the PD activities during their classroom implementation. The teachers' previous science PD experience and pre-PD understanding of inquiry-based teaching also explained some of the variability in student growth. In both NOS and content, students of teachers with less prior science-PD experience benefited more. Grade band also explained variation in student outcomes through interactions with teacher-characteristic variables. In high school, students of teachers with lower pre-PD inquiry knowledge appeared to learn more about NOS. Our results suggest that inquiry and content training through PD may minimize disparities in teaching due to inexperience and lack of expertise. Our study also demonstrates the value of PD that teaches a flexible approach to inquiry and focuses on underrepresented, interdisciplinary content areas, like aquatic science.

Keywords: inquiry, student outcomes, teacher characteristics, marine science, content knowledge, curriculum, professional development, program adherence, nature of science

Teaching Aquatic Science as Inquiry through Professional Development: Teacher Characteristics and Student Outcomes

Inquiry-based teaching is intended to engage students in experiential learning through investigations that are carefully sequenced and connected to previous experience. Inquiry pedagogy is rooted in the philosophy of constructivism, which postulates that knowledge is built incrementally through thinking, integration of ideas, and social interaction (American Association for the Advancement of Science, 1990). The value of inquiry to science teaching and learning in K–12 education is supported by a compelling body of research that demonstrates positive effects on student achievement, attitudes, critical thinking, process skills, problem solving, creativity, and vocabulary (e.g., Aulls & Shore, 2008; Curriculum Research & Development Group [CRDG], 2000; Cohen & Spillane, 1993; National Research Council, 2012; Wilson, Taylor, Kowalski, & Carlson, 2010; Wu & Hseih, 2006). For example, Minner, Levy, and Century's (2010) analysis of 138 studies showed a positive correlation between inquiry-based learning, students' conceptual understanding, and students' science content knowledge. An inquiry-based versus commonplace teaching experiment by Wilson et al. (2010) also lends support to the premise that inquiry-based teaching can lead to positive gains in student knowledge, reasoning, and argumentation.

At the heart of inquiry-based science teaching is the propagation of students' ability to do science and students' understanding of the nature of science (NOS) through (a) wondering and questioning, (b) planning and designing, (c) investigating and using data, (d) analyzing, and (e) communicating (e.g., review by Capps, Crawford, & Constas, 2012; NGSS, Appendix H, p. 96, connecting standards to NOS). However, in studies of inquiry, researchers report that teachers generally lack crucial knowledge related to its use (e.g., Lakin & Wallace, 2015). For example, teachers often equate inquiry with hands-on instruction, thereby failing to recognize the wide spectrum of approaches that highlight teachers as the facilitators of inquiry in the classroom (Capps et al., 2012; Seung, Park, & Jung, 2013; Rankin, 2000). Differences in teachers' perceptions of inquiry reflect the range of inquiry definitions held by education experts nationally and internationally (e.g., Abd-El-Khalick, BouJaoude, & Duschl, 2004). Thus, there is compelling support for PD programs that not only help teachers to more fully understand and implement inquiry but also help researchers to better understand how student learning is affected by changes in teachers' inquiry views.

Inquiry-based professional development (PD) interventions differ in structure and goals, including features such as the level of teacher participation in PD courses, level of flexibility in addressing scientific problems, and degree of focus on content versus pedagogy (Kennedy, 2016). The present study addresses a hybrid (face-to-face and online), year-long, 87-hour teacher PD that used an inquiry-based approach designed to approximate scientific practice with a strong focus on the content of physical, chemical, biological, and ecological aquatic science.

Intended Outcomes and Research Questions

We developed and implemented our PD under the assumptions that (a) inquiry-based PD would improve teachers' understanding of inquiry and scientific process, (b) aquatic science PD would improve teachers' aquatic science content knowledge, and (c) teachers' implementation of inquiry-based aquatic science activities would improve students' understanding of the NOS and increase their knowledge of aquatic science content.

This study addresses changes in student outcomes in light of teachers' participation in the PD. We sought to understand the variability in student changes with respect to grade band, experience and expertise of incoming teachers, and the degree to which teachers adhered to the program activities. Our research questions were

1) To what extent is teacher participation in the PD associated with student improvement in (a) NOS understanding and (b) knowledge of aquatic science content?

2) To what extent are student changes in the outcome variables (NOS understanding and knowledge of aquatic science content) moderated by (a) grade band (elementary, middle, and high school) and (b) teacher characteristics (i.e., pre-PD knowledge of aquatic science content and inquiry-based teaching, years of experience in teaching science, number of previous science PD courses taken, and adherence to activity guidelines when implementing activities in their classroom)?

Background and Framework

Professional Development

Inquiry-based PD that builds teachers' scientific process skills can increase teachers' understanding of science and confidence in their ability to teach science (Bencze, 2010; Bulunuz & Jarrett, 2010; Nadelson et al., 2013; Raino, 2008). When teachers participating in PD improve in their self-efficacy with science process skills, use of constructivist teaching practices typically increases (Alexander, Heaviside, & Farris, 1998; Beamer, Van Sickle, Harrison, & Temple, 2008). There is also evidence that sustained, content-focused PD can improve teaching practice and student learning (Desimone, 2009; Desimone & Garet, 2015). For example, Smith et al. (2007) reported that when teachers participated in PD lasting more than 35 hours, the disparity between teachers with and without formal science degrees became less of a factor in predicting use of inquiry-based teaching and hands-on activities.

Research also supports the claim that teacher PD can improve student achievement in science (e.g., Blank, de las Alas, & Smith, 2007; Roth et al., 2011). And yet, there are markedly few studies that examine the effects of teacher PD on student learning (Garet, Porter, Desimone, Birman, & Yoon, 2001). For example, Yoon, Duncan, Lee, Scarloss and Shapley (2007) reviewed studies with the potential to address the link between teacher participation in math or science PD and student gains. Their examination of more than 1,300 studies resulted in only 27 studies that focused on student outcome measures. Of these 27 studies, only 9 met Yoon et al.'s rigorous experimental design criteria. All nine studies were conducted at the elementary level. And, only two of the nine studies focused on science—underscoring the developmental nature of the PD research field and the need for more studies investigating the relationship between teacher participation in science PD and student outcomes across grade bands.

Teachings Science as Inquiry (TSI) Model

The Teaching Science as Inquiry (TSI) model formed the pedagogical foundation of our PD. TSI emphasizes the flexible and collaborative nature of scientific inquiry and guides teachers to help students recognize the dynamic aspects of NOS. TSI is similar to other inquiry models, like Bybee's 5 Es, in its use of phases in a learning cycle (e.g., Bybee et al., 2006). However, TSI is unique in its emphasis on fluidity between inquiry phases and its emphasis on multidirectional instruction (Pottenger, 2007; Seraphin, Philippoff, Kaupp, & Vallin, 2012). The five inquiry phases of the TSI model are (a) *initiation*, (b) *invention*, (c) *investigation*, (d) *interpretation*, and (e) *instruction*. *Initiation* is a phase of originating interest or developing a focus for inquiry. The *invention* phase entails problem solving and information gathering, such as creating a testable hypothesis or troubleshooting a procedural step. Students engage in *investigation* as they gather new knowledge by conducting tests or analyzing data. In *interpretation,* information is distilled through both internal and external reflection. *Instruction* is integral to each phase and includes communication from teacher to student, student to student, and student to teacher.

The TSI phases are intended to help teachers guide their students' learning and doing of science in an authentic way. Like other inquiry models, the TSI phases are purposefully connected in a cycle. However, the TSI cycle lacks numbers and arrows because progression through the TSI phases is not meant to be unidirectional (Figure 1A). The fluid, multidirectional nature of the phases reflects the reality of scientific practice—in contrast with the linear way that many teachers interpret the scientific method.

The promotion of authentic scientific processes during teaching and learning through fluid phases is a unique feature of the TSI model. A teacher may plan a lesson with *initiation* at the beginning, but *initiation* can, and should, repeat throughout the course of a learning progression as students experience anomalies or consider new information—in much the same way practicing scientists do. Teachers are thus encouraged to use questioning strategies to *re-initiate* and re-engage students throughout the course of a lesson.

The TSI model is also unique in its inclusion of *instruction* in the inquiry cycle. In TSI, the *instruction* phase surrounds and influences the other phases (Figure 1A). This creates an environment where *instruction* comes not only from the teacher as the research director but also from students as they *instruct* each other. The *instruction* phase thus recognizes the social aspect of learning that is central to students' construction of knowledge and understanding of NOS. The encompassing position of *instruction* in the TSI model also encourages teachers to recognize its importance and plan for instructional interaction in their lessons.

The overall intent of TSI is to enable teachers to create classrooms that function as a community of scientists—where students learn science by engaging in the practice of science. Students are expected to learn scientific practices such as those outlined in the Next Generation Science Standards (NGSS): (a) asking questions; (b) collecting, analyzing, and interpreting data; and (c) communicating findings. By using the TSI phase diagram to reflect on their instruction, teachers can develop a deeper, richer understanding of scientific inquiry (Seraphin et al., 2012). As part of this PD and subsequent classroom implementation, both teachers and students participated in TSI-based activities and then reflected using the TSI phase diagram (Figure 1B).

This type of purposeful reflection is a central component of TSI and provides an opportunity for dialogical thinking and internalization of specific NOS elements—that scientific investigations move in non-linear fashion, use a variety of methods, are open to revision, and require human input.

TSI uses ten modes of inquiry to promote a spectrum of inquiry practice, from open-ended to guided inquiry. The TSI modes of inquiry include *curiosity*, *description*, *authoritative knowledge*, *experimentation*, *product evaluation*, *technology*, *replication*, *induction*, *deduction*, and *transitive knowledge* (Table 1). Teachers are encouraged to integrate multiple TSI modes into their daily teaching in order to build students' inquiry skills and NOS understanding. The TSI modes help to free teachers from the misconception that inquiry-based teaching needs to be hands-on all the time. The inclusion of the *authoritative knowledge* mode, for example, gives teachers explicit permission to provide direct instruction, in combination with student-lead investigations, within the context of an inquiry classroom (Seraphin et al., 2012). The modes also prompt students to treat science investigations as human endeavors that involve creativity and revision based on interaction with evidence.

Aquatic Science Content

Our PD was designed with a focus on aquatic science—a fundamental, yet underrepresented content area in U.S. curricula, standards, and classrooms. On a national level, ocean and aquatic sciences have been marginalized in K–12 education (Schoedinger, Cava, & Jewell, 2006), largely because they were absent from the National Research Council's 1996 National Science Education Standards (Strang, 2008). Even in Hawai‘i, an island state where this PD was conducted, only about 20% of public high school students enroll in marine science classes. And, the vast majority of the students who take marine science in Hawai‘i are among the lower performing science and math students (with Geometry or Algebra 1 as their highest math course; Osumi & Huang, 2015). In other words, marine science is not a priority area of study even in Hawai‘i.

This lack of emphasis on aquatic science is unfortunate considering how vital understanding the ocean is to our future and to students’ overall scientific literacy (Strang, deCharon, & Schoedinger, 2007). In addition to covering more than 70% of our planet, the ocean affects every aspect of human life, from climate and food security to tourism and economic and social stability. Although the NGSS (2013) have made strides to rectify deficiencies, aquatic science content is still underrepresented compared to other content, such as space science (Strang et al., 2015).

Aquatic science thus provided a meaningful, uniting topic for our PD, for which most teachers (and students) have had little formal training. Moreover, because aquatic science is inherently interdisciplinary, it was an appropriate topic for PD targeting teachers from different grade bands and subject areas. The theme of aquatic science also served as a place-based anchor for our island state, helping the teachers in our PD to connect their students’ school learning with everyday life.

Teacher Characteristics

It is widely accepted that teachers play a central role in bringing educational change into the classroom, which in turn provides the general rationale for teacher PD (Hewson, 2007). With inquiry-based PD, teachers’ characteristics likely influence the way curricula and pedagogy are enacted, which in turn affects student experiences and learning (Puntambekar, Stylianou, & Goldstein, 2007). Research on the effects of teacher participation in inquiry-based PD on student outcomes has examined, among other variables, participants’ teaching experience, knowledge of content, understanding of scientific inquiry, and degree of adherence to the program (e.g., Liu, Lee, & Linn, 2010; McNeill, Pimentel, & Strauss, 2013). These studies contributed to decisions about which teacher covariates to include in quasi-experimental studies on teacher PD (Murnane & Willett, 2011) and have the potential to maximize teachers', and thus students', learning experiences by enabling better PD planning (Fullan, 2001).

Content knowledge. A compelling collection of studies has shown that student achievement is directly influenced by teachers' content knowledge (e.g., Hill, Rowan, & Ball, 2005; Kanter & Konstantopoulos, 2010). Teachers with higher content knowledge tend to have students with higher content scores (e.g., Diamond, Maerten-Rivera, Rohrer, & Lee, 2014; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Kanter & Konstantopoulos, 2010). For example, in their study of Grade 5 teachers, Diamond et al. (2014) reported that teachers' content knowledge explained 6% of the variability in students' scores. However, there are also questions about the differential effects of teacher content knowledge on student learning, including how content knowledge interacts with other teacher variables and how teacher content knowledge operates differently at different grade levels (Abell, 2007).

Pedagogical content knowledge of inquiry. Teachers' beliefs about and knowledge of scientific inquiry constitute another set of teacher characteristic variables. Studies looking at teachers' inquiry beliefs tend to address perspectives about how to enact inquiry in the classroom. This can be represented by teachers' self-efficacy in teaching inquiry (McNeill et al., 2013) or by teachers' beliefs about the value of inquiry-based teaching, such as in Liu et al. (2010), who found positive associations between these beliefs and student learning. Teachers' familiarity with the scientific process is also correlated with their ability to facilitate scientific research with their students (e.g., Wee, Shephardson, Fast, & Harbor, 2007).

However, teachers' knowledge of scientific inquiry does not consistently predict their teaching practice. For example, Bartos and Lederman (2014) found little congruence between teachers' inquiry knowledge and their classroom practice. Researchers have postulated, therefore, that a more informative characteristic is teachers' understanding of what inquiry looks like in the classroom (Lakin & Wallace, 2015; Schuster, et al., 2007). Two instruments, the Pedagogy of Science Inquiry Test (Lakin & Wallace, 2015; Schuster, et al., 2007) and one of its derivations, the Inquiry Teaching Assessment (Harrison & Vallin, 2012) have been used to measure that understanding. In effect, these instruments measure pedagogical content knowledge of inquiry, which is likely to be less susceptible to self-report biases and better targeted to classroom practice than measures of scientific inquiry knowledge.

Teaching experience. Reviews by Rice (2010) and the REL Midwest Reference Desk (2015) support the predictive power of teaching experience on student achievement, although there are exceptions. For example Liu et al. (2010) and McNeill, Pimentel, and Strauss (2013) failed to find any effect of teaching experience on student outcomes. The divergence between these studies suggests that further examination PD participants' teaching experience is worthwhile—not only for understanding the effect of PD in light of participants' experience but also for understanding whether other teacher characteristics interact with, or depend on, years of teaching experience.

PD experience. Similar to years of teaching experience, teachers' prior science PD participation constitutes an important teacher characteristic because PD often constitutes the bulk of teachers' situational experience with interdisciplinary content such as aquatic science (Liu et al., 2010). Prior PD may also play a role in how content is taken up in a new PD. For example, in Supovitz and Turner's (2000) analysis of the variables predicting inquiry-based teaching practice, the teachers with no PD experience used inquiry-based teaching practices less frequently (by .40 of a standard deviation) than the average teacher (whose prior PD experience averaged between 1 and 19 hours). Thus, the potential for prior PD to have a differential effect on teacher practice makes this an important variable to consider in researching PD effects.

Adherence. In addition to knowledge and experience, teachers' adherence to the PD is important. Adherence constitutes a primary component of fidelity of program implementation and is considered to be a valuable predictor of student achievement (e.g., Wallace, Blase, Fixsen, & Naoom, 2008). Adherence to program goals and curriculum guidelines during classroom implementation has been shown to be a main factor affecting student outcomes (e.g., Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003; Summerfelt, 2003). For example, McNeill et al. (2013) found, in their study of 22 teachers in a year-long inquiry PD, that students of teachers who adhered more strongly to the curriculum had higher science problem-solving assessment scores than students of teachers who did not adhere. O'Donnell and Lynch (2008) similarly concluded that increased fidelity of implementation to curriculum instructional strategies results in increased student achievement gains.

Adherence to program implementation is also an important component in determining whether the observed outcomes are a function of the intervention (Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008). A positive relationship between adherence and measured PD outcomes lends causal support to the interpretation that teacher and/or student gains are a result of the PD. For this reason, adherence is an important measure to consider when assessing the effect of PD.

Design and Methods

Description of the PD

We developed and implemented a year-long teacher PD. The PD was conducted in a four-part, modular structure with workshop-style experiences, in-person meetings, and online follow-up—punctuated by teachers' classroom implementation of activities with students. The PD was prepared and conducted as an Institute of Education Sciences development grant (Goal

2), with five teacher cohorts on different islands in the state of Hawaiʻi: Cohort 1 (Oʻahu) in 2010–2012, Cohorts 2 and 3 (Maui and Hawaiʻi island) in 2011–2012, and Cohorts 4 and 5 (Oʻahu II and Kauaʻi) in 2012–2013.

Modularized PD Structure. Teachers in the PD were exposed to TSI and NOS themes throughout the school year within the context of four themed modules: physical, biological, chemical, and ecological aquatic science (Table 2). The spacing of the modules was intended to provide a temporally accessible yet sustained PD format, coupling intense workshops with extended classroom implementation, reflection, and interaction. The PD began with an introductory meeting (3 hours) followed by four modules, each consisting of (a) a two-day in-person workshop (16 hours); (b) an in-person follow-up training (3 hours); (c) teachers' classroom implementation with students; and (d) a synchronous, face-to-face online follow-up (2 hours), for a total of 87 facilitator-teacher contact hours. The modules were united by an asynchronous online learning community built into our curriculum website (exploringourfluidearth.org). The website housed aquatic science curricula, which included activities, associated student and teacher materials, and an interactive teacher community.

Our program incorporated many of the features research has shown to constitute effective PD, including a focus on content, sustained implementation and follow-up, activities that authentically demonstrated science knowledge and skills, and embedded opportunities for implementation and reflection (e.g., Banilower, Heck, & Weiss, 2007; Dunst, Bruder, & Hamby, 2015; Lumpe, Czerniak, Haney, & Beltyukova, 2012; Yoon et al., 2007; Smith et al., 2007).

Pedagogy and Content in the PD. Each PD module (physical, biological, chemical, and ecological aquatic science) included a series of activities intended to provide a cohesive set of content to be learned through inquiry. The TSI pedagogy in the activities was explicitly modeled by facilitators and discussed with the teachers during the workshop components of the PD. The teachers were then directed to include TSI activities in their classrooms and to scaffold their instruction with the critical TSI components, like phases and modes (Table 2).

Teachers taught a total of 12 TSI-based activities over the course of the year with their students (three activities per module). Two activities per module focused on foundational aquatic science content (e.g., density, chemical properties of water, and evolutionary connections). One activity per module focused on NOS and inquiry, building from structured to open-ended inquiry (e.g., practices of scientists, phases of inquiry, and modes of inquiry). The content was aligned to the state of Hawaiʻi Content and Performance Standards III because NGSS did not begin rollout in Hawaiʻi until Fall 2016. Table 2 shows the aquatic science content goals by module. (Detailed module agendas are in CRDG, 2014, pp. 118–124, and activities are available online: exploringourfluidearth.org.)

In the workshop component of the PD, facilitators presented the goals and foci of the activities. The teachers then participated in activities as students and discussed the processes and practices they had engaged in. To support implementation with their own students, teachers were provided with teacher guides and supplies for the activities. The teachers' classroom implementation was tracked via activity reflection logs and a post-PD questionnaire. Teachers also shared their implementation experiences during the online follow-ups. In addition, the teachers were required to post reflections of their implementation experiences to the online learning community portion of the curriculum website.

Participants

The participants in our PD were teachers in private and public schools (including public charter schools) from throughout the state of Hawaiʻi. The schools were in urban and rural settings; some teachers' classrooms were allocated to special education students. The work conducted for Cohorts 1–3 (41 teachers) was preparatory for the final two cohorts, at which time the materials had been revised through iterative testing and were sufficiently mature to conduct a pilot study. Thus, the present study examines Cohorts 4 and 5 (located on the islands of Oʻahu and Kauaʻi), in the 2012–2013 school year. In these two cohorts, 28 teachers (90% of the 31 who enrolled) completed the PD and all of the research instruments. Of the 632 students enrolled in Cohort 4 and 5 teachers' classes, 530 (84%) provided consent, and 413 (78% of the consenting students) completed both the pre- and post-PD student instruments, for an average of 15 students per class (*SD* = 8.11).

Variables, Teacher Characteristics, and Instruments

Due to the developmental nature of the project we were only able to recruit a relatively small number of teachers, and no comparison teachers were included in the study. Thus, we used a pre-post single-group design, with explanatory covariates, which is appropriate for examining outcomes for a small sample of participating teachers. We selected instruments based on their utility in addressing the research questions, their availability, and their development feasibility. As with any set of social-science measurements, the instruments bear some degree of construct underrepresentation because they sampled larger domains of the targeted constructs. The development of many of the instruments was paired with the development of the PD; we added or adapted items and prompts as the program and its intended outcomes became more refined. The instruments are detailed in CRDG, 2014 (pp.61–98 and pp.125–148), where we also provide complete descriptions of instrument development, reliability, and validity evidence.

Teacher variables. We examined the differential effect of six variables on student outcomes. The variables included grade band as well as five teacher characteristics. Four of the teacher variables were related to teacher experience and expertise; one variable was a measure of fidelity of implementation.

The teacher experience and expertise variables were (a) the number of years the teacher had taught science, (b) the number of science PD courses the teacher had taken in the previous five years, (c) a composite of the teachers' pre-PD scores on aquatic science content (across the four content areas), and (d) the teachers' pre-PD score on the Inquiry Teaching Assessment (a vignette-based performance assessment adapted from Schuster et al., 2007; Harrison & Vallin, 2012). The fidelity of implementation variable was a measure of adherence, taken multiple times throughout the year-long PD. The instruments measuring the constructs of content knowledge, inquiry knowledge, and adherence were developed and pretested during the first three cohorts. We describe their development and operationalization in Table 3.

We collected background information from the teachers on a pre-PD questionnaire (CRDG, 2014, Appendix C). The teachers in Cohorts 4 and 5 taught a range of subjects, from core courses like physics, chemistry, and biology to specialty classes like marine science, clinical health, and human anatomy. The teachers taught high school (n = 11), middle school (n = 15), and elementary school (n = 2). Although we designed the PD for middle and high school teachers, we permitted the entry of the two elementary-school teachers to increase recruitment on Kauaʻi, which is a less populous island (Table 4). We thus included the elementary data, but our analysis is primarily focused on middle and high school.

Teachers varied in their PD and teaching experience. The number of science PD courses taken during the previous 5 years ranged from none (n = 10) to 5 or more (n = 6), with the remaining 12 teachers having taken between 1 and 4 science PD courses in the past 5 years (Table 4). The number of years teaching science ranged from less than 2 years (n = 6) to more than 10 years (n = 10), with the remaining 12 teachers having between 7 and 10 years of science teaching experience (Table 5).

Adherence was the only self-report moderator in our analysis (Table 3). Adherence can be measured directly (i.e., through observations) or indirectly (i.e., self-report measures such as teacher logs). Observations are not only expensive, requiring observer training and multiple samples of classroom lessons, but they also tend to invite subject-expectancy effect bias (O'Donnell, 2008; Sanchez et al., 2007). On the other hand, there is evidence that teacher self-reports of adherence are well correlated with practice, particularly when teachers use logs for reporting content and instructional strategies (e.g., Desimone, 2009; Mullens et al., 1999 Rowan & Correnti, 2009). By using multiple self-report logs over time and across lessons, we were able to demonstrate that teachers in our PD were consistent in their reporting of their adherence levels (Cronbach's alpha = .86).

Student variables. In Table 6, we provide descriptions of the two student-outcome variables used in our analysis: understanding of NOS and knowledge of aquatic science content. The student instruments were administered before and after the TSI activities had been implemented. In Table 7, we include the descriptive statistics of these student variables by grade band. As expected, high school students had higher starting scores than middle and elementary students on both NOS and content.

Analysis

The students, when taken as a whole without considering their grade band or the characteristics of their teacher, improved from pre- to post-test on both outcomes (Table 7). To address Research Question 1 (assessing the extent of student growth based on teacher participation in PD), we tested the statistical significance of gains by conducting a multilevel regression, with time as the only predictor, for each of the two student-level outcomes (NOS and content). The multilevel aspect accounts for the clustering of students within classrooms by adjusting the standard error.

To address Research Question 2 (assessing the influence of grade and teacher on student growth), we added moderating variables to the multilevel regression analyses. Grade band, teacher characteristics (number of years teaching science, number of prior science PDs taken, understanding of inquiry teaching, and knowledge of aquatic science pre-PD), and teacher adherence were used as moderating variables. We examined all possible higher-order interactions among the variables (e.g., grade band by number of science PD courses taken) and retained only those interactions that contributed to the regression (i.e., with $p < .05$ or better fit statistics). All main-effect predictors were retained regardless of contribution. To facilitate the meaning of the intercept, we centered the interval-scale moderating variables and coded the ordinal-scale moderating variables so that a value of zero on each variable represented its mean or mode. Tables 3 and 4 display these centered and coded values.

Results

Our research questions concerned changes in students' understanding of the NOS and their knowledge of aquatic science content, including the degree to which these changes were contingent upon the moderating variables of grade band, indicators of teacher experience and expertise, and teachers' adherence. Table 8 displays the results of the multilevel regressions with these two outcome variables. The first two regressions shown in Table 8 (one for NOS and one for content) address the change in the students' scores over time regardless of any moderating variables. That is, these include only the intercept and time variables, where the estimate of the effect of time is the post minus the pre for all of the students. The second two regressions in Table 8 include the moderators for the second research question, which addresses the effect of teacher characteristics.

Nature of Science Assessment Results

The NOS assessment results indicated that regardless of grade band or teachers' characteristics, after accounting for the clustered nature of the data, the change in the students'

NOS scores was weak but statistically significant (estimated change = +0.06 logits, $p < .05$, Hedges' $g = 0.10$). That is, the student gains in NOS were about one-tenth of a standard deviation. When the grade band and teacher variables were accounted for (and centered at, or coded at, zero), the pre-to-post increase in students' scores was about three times higher (estimated change = +0.21 logits, $p < .01$, Hedges' $g = 0.32$). That is, over the course of teachers' implementation, a typical student in a typical class tended to have an improved understanding of the NOS by about one-third of a standard deviation.

The estimated improvement in students' NOS understanding depended on the number of prior science PD courses their teachers had taken ($p < .01$). For the students in classes taught by teachers who had not taken any science PD courses in the previous five years, students' improvement was estimated at 0.32 of a standard deviation. The students' improvement was lower when the teacher had more PD experience. For example, when teachers had taken one or two previous science PDs, the estimated gain among their students was 0.14 score points (in logits), which is about 0.22 of a standard deviation. For teachers who took three or four science PD courses in the past five years, the estimated gain was less, at 0.08 score points (or 0.12 of a standard deviation), when all other variables were held constant.

The estimated improvement in students' understanding of the NOS was also contingent upon the teacher's degree of adherence when implementing activities ($p < .05$). When the teachers adhered more closely to the activities, their students had stronger pre-to-post gains compared to students in classes with teachers at the average amount of adherence. For example, when the teachers adhered more closely by one standard deviation unit, their students' gain on pre-to-post NOS was estimated at 0.29 score points, which is 0.45 of a standard deviation on the NOS assessment.

Additionally, the estimated change in student NOS scores was contingent upon grade band, which in turn was also contingent upon how much inquiry teaching knowledge their teacher had prior to participating in the PD. This is represented by the significant grade-band-by-inquiry-teaching-assessment interaction in the multilevel regression (Table 8), as illustrated in Figure 2 (a plot of the interaction between these two variables on the predicted changes in students' NOS scores). One way to interpret this interaction is to examine high school students' scores in light of their teachers' pre-PD Inquiry Teaching Assessment scores. The regression predicts that high school students taught by teachers with lower prior Inquiry Teaching Assessment scores (by one standard deviation) improved by 0.28 score points, whereas their counterparts taught by teachers with the average pre-PD inquiry teaching knowledge improved by 0.09 score points. This differential effect comes out to about 0.28 of a standard deviation.

For students in middle school, the regression predicts a different outcome, in which students' changes in NOS scores seem to have depended very little on the teachers' pre-PD inquiry-based science teaching knowledge. For elementary students, the regression predicts weaker NOS improvements when their teacher has a lower Inquiry Teaching Assessment score, which is the opposite of the predicted trend with high school students. The plot of observed student change scores is consistent with this prediction from the multilevel regression even though the plot does not control for the other variables in the way that the regression does. (Figure S1, in the supplemental materials, displays the observed teacher-level data of these two interacting variables for all three grade bands.)

No other possible interactions were significant. Additionally, the other teacher variables, including teachers' pre-PD science content knowledge and the number of years they had taught science, had no detectable moderating effect on students' NOS growth.

Content Assessment Results

Similar to what was observed with the NOS assessment, the change in students' content assessment scores was positive and significant when ignoring all variables besides time (estimated change = +0.26 logits, $p < .01$, Hedges' $g = 0.35$). As shown in Table 8, when the school and teacher variables were included and held constant, the change was stronger (estimated change = +0.47 logits, $p < .01$, Hedges' $g = 0.63$).

Also similar to the results in the NOS outcome, the students whose teachers adhered closely to the activity guidelines had stronger gains from pre-to-post than their counterparts in classes taught by teachers with the average amount of adherence. For example, students in classes whose teachers adhered more closely by one standard deviation were estimated to increase in their content scores by 0.59 score points (Hedges' $g = 0.79$). This is higher than the estimated improvement of 0.47 score points among students in classes with an average amount of adherence.

The number of previous science PD courses taken by a teacher was estimated to have a moderating effect on the students' gain, but this effect was further contingent upon the grade band. This is observed by the significant grade-band-by-number-of-science-PDs-taken interaction in the multilevel regression ($p < .01$) and is illustrated by Figure 3 (a plot of the interaction between these two variables on the predicted changes in student content scores). The predicted change in content scores among high school students whose teachers had taken no previous science PDs were higher than their counterparts whose teachers had taken multiple science PDs. This effect is also predicted to exist at the middle school grade band, but to a weaker extent. For elementary students, the regression predicts that content improvements will be higher in classes taught by teachers with more previous PD experience, which is opposite the predicted trend with high school students. This result should be taken with caution, however, because the regression was estimated with only two elementary school classes. (Figure S2, in the supplemental materials, displays the observed means of the student content scores per teacher's class across all three grade bands.) No other possible interactions were significant, nor did the regression identify any other teacher-level variables with a detectable moderating effect on student growth.

Discussion

We developed a year-long, inquiry-based PD that provided 87 contact hours in a four-part, modular structure with repeated exposure to content and pedagogy, punctuated by teachers' implementation of activities in their classroom. The PD focused on aquatic science content and covered physical, chemical, biological, and ecological processes. The proximal goal of the PD was for teachers to become successful facilitators of inquiry and the scientific process in the context of aquatic science. The purpose of this study was to examine whether the distal goal of the PD was achieved—that is, to examine whether students improved in their aquatic science content knowledge and in their understanding of the NOS—and whether the magnitude of the changes in these two student-level constructs depended on grade band, teacher experience, teacher background knowledge and understanding, or teachers' adherence to implementation of the program in the classroom.

Student gains in the context of teacher gains. The premise of effective PD is that participation will improve teachers' content and/or pedagogical knowledge, which will lead to gains in student outcomes (Desimone, 2009). And yet, we know that teachers come to PD with a range of experience, content knowledge, and teaching environments (Desimone & Garet, 2015). It is important, therefore, for researchers to assess teacher gains prior to addressing the relationship between teacher and student change (Diamond et al., 2014). In an earlier evaluation report (CRDG, 2014, pp.99–114), we showed that the teachers in the TSI PD improved their understanding of inquiry-based teaching and their aquatic science content knowledge pre- to post-participation. These teacher level changes were strong; teacher improvement in understanding of inquiry teaching had an effect size of Hedges' $g = 1.24$, and the teacher content assessments (pre-post per module) had improvements that ranged in effect size from $g = 0.66$ to 1.00.

The results of the present study indicate that the students had statistically significant gains in their aquatic science content knowledge and NOS understanding. The overall pre-to-post gain across all students on the NOS outcome was weak, with $g = 0.10$, but statistically significant, at $p < .05$. On the aquatic science content outcome, the gain was moderate, at $g = 0.35$ for all students. When the student outcomes were estimated in a typical middle school class, the effects were stronger, with NOS at $g = 0.32$ and content at $g = 0.63$.

Teacher adherence and student gains. An important consideration in assessing whether the observed outcomes are a function of the intervention is the degree to which teachers deliver a program as intended (Mowbray et al., 2003; O'Donnell, 2008). Thus, although we acknowledge that some of the observed student growth may have been due to normal maturation, our finding that the students' improvements were amplified when their teachers adhered more closely to the activity guidelines and pedagogical practices supports the plausibility of a causal link between the PD and student gains. And, insofar as the teachers' experiences in the PD transferred to their classroom teaching (adherence to TSI strategies when implementing activities), the changes in students' knowledge and understanding were likely due, at least in part, to the teachers' participation in the PD.

Improved aquatic science content knowledge was likely due to the PD. Aquatic science was chosen as a focus because of its importance to scientific literacy, its historical and current under-representation in K–12 curriculum (Schoedinger et al., 2006; Strang, 2008; Strang et al., 2015), and its place-based connection to Hawai‘i. The increase in both teacher and student content scores pre-to-post point toward the likelihood that the PD improved teacher content knowledge and provided a mechanism for student content gains (e.g., Hill et al., 2005; Kanter & Konstantopoulos, 2010). Because aquatic science content was not part of the standard curricula, student increases were likely due to the program rather than to normal maturation. This supports the premise that under-represented, interdisciplinary content may be improved through PD that serves a range of grade bands and course topics simultaneously. This is an important logistical consideration when attempting to increase students' exposure to topics like aquatic science.

More prior PD was related to lower student growth. Research has demonstrated a predictive relationship between the time spent in a PD course and improved student performance (e.g., Blank et al., 2007; Shaha, Glassett, & Ellsworth, 2015; Smith et al., 2007). Our study supports the ability of long-term PD to improve teacher and student knowledge (the TSI PD was 87 hours and spanned a full school year), but our findings suggest that the effect of multiple PD courses may be monotonically additive rather than positive. The amount of student growth in both NOS and aquatic science content was inversely related to the teachers' prior levels of science PD experience; the PD was more beneficial to students whose teachers had participated in less prior science PD. In other words, teachers in our program who had previously taken multiple PD courses did not seem to provide as great of a benefit to their students. This finding was somewhat unexpected because other research suggests that teachers' content knowledge and pedagogical knowledge improves with increased exposure to PD (e.g., Desimone, 2009; Guskey, 2002). We therefore predicted a cumulative effect of PD and expected that teachers with more PD experience would be better able to incorporate TSI and aquatic science into their teaching practice, resulting in larger student gains.

It is possible, though, that the teachers' previous PD training attenuated the effects of the TSI PD. In most new programs, participants experience some degree of cognitive dissonance when they are asked to question their current behavior and beliefs (Fullan, 2001; Roehrig, Kruse, Kern, 2007). And, when teachers' beliefs and behaviors are well established, they may be less likely to incorporate PD innovations into their teaching practice (which is consistent with the view that experienced teachers can be less receptive to educational reform; Prawat, 1992). Alternatively, a characteristic particular to the teachers who took multiple PD courses may have affected student outcomes. For example, Kennedy's (2016) analysis of PD programs suggests that disparity in motivation to attend PD has significant effects on teacher and student outcomes—overwhelming effects of the length and intensity of PD. Indeed, it is possible that some teachers in our program were taking PD primarily to accumulate course hours, rather than taking PD primarily because of a desire to learn.

These results should be interpreted with caution because we used a single question to measure this demographic variable. More research is needed to reveal whether teachers with more PD experience are less likely to adopt innovative PD pedagogy and practices into their classroom practice and whether this has an effect on student outcomes. If further research finds similar results, it would suggest that PD interventions could benefit by considering ways to enhance uptake with PD-seasoned teachers and match PD offerings with teachers' motivation.

High school students' scores depended on teachers' pre-PD inquiry pedagogical knowledge. In the high school grade band (and in the elementary grade band in the opposite direction), the degree of student growth in NOS depended on teachers' prior understanding of inquiry-based teaching. In contrast, the middle school students' changes in NOS scores depended very little on the teachers' pre-PD, inquiry-based teaching knowledge. This result was unexpected. A previous study examining both grade band and teachers' inquiry teaching (Liu et al., 2010) reported positive main effects of both variables. In their study, high school students outperformed middle school students, and students of teachers with high inquiry-teaching beliefs outperformed their counterparts on student posttests of science-integration knowledge. Liu et al. (2010) did not, however, report testing whether these two variables interacted. In other words, they did not examine whether there was a differential effect of teachers' inquiry beliefs by grade band. Additionally, it may be that Liu et al.'s measure of teachers' beliefs represents teachers' attitudes toward the intervention, whereas our measure of teachers' inquiry teaching understanding required teachers to articulate how inquiry manifests in the classroom, which is a different construct than belief. To our knowledge, ours is the first study examining the moderating effect of pre-PD pedagogical content knowledge of inquiry-based teaching on students' NOS and content scores.

Teacher characteristics less influential in middle school. The finding that grade band had a moderating effect on student achievement may be due to the TSI PD being targeted more toward middle school teachers. Because the majority of PD participants had lower starting content knowledge than anticipated, the PD was modified (during cohorts 1–3) to better accommodate the lower level of content knowledge among our incoming teachers. We had hoped this focus on lower-level content and inquiry skills would provide an opportunity for high school teachers and students to review and dig deeper into foundational knowledge—even if the content was not as challenging. However, our revised focus resulted in less content rigor in the PD workshop components, which may in turn have led to less of an increase in high school student scores as compared to middle and elementary school students. High school students had higher starting content knowledge scores than middle and elementary school students. And, high school students had less content knowledge gains—a finding that was amplified by the number of PD courses teachers had taken (more prior science PD, less gains). Similarly, high school students did not improve as much as middle school students in NOS—a finding amplified by teachers' prior inquiry knowledge (more pre-inquiry knowledge, less gains).

Overall, students benefited more from their teachers' PD participation when the teachers had less prior experience. When the teacher experience and expertise variables were considered together (years of teaching experience, incoming inquiry teaching knowledge, prior content knowledge, and previous PD experience), we saw larger gains among students of teachers with less experience and expertise. These results are supported by Jones and Eick (2007), who postulated that teachers with low content knowledge can succeed when PD provides structure for implementing inquiry in the classroom. In other words, less prepared teachers may be able to overcome their starting deficits by using the pedagogy taught in the PD and implementing the TSI-based activities. This supports the hypothesis that inquiry-based PD, paired with curriculum and classroom implementation, can effectively improve the teaching and learning of interdisciplinary topics like aquatic science across subjects and grade bands. Smith et al.'s study (2007), which showed that science-focused PD could decrease disparities between teachers and increase teachers' use of inquiry, lends further credence to our assertion. However, larger and more representative sample sizes are needed to provide additional evidence to support this generalization, especially in light of our lack of a control group and the relatively few other studies that have examined the effects of PD on the content knowledge of students—particularly in upper grade bands (Diamond et al., 2014).

One possible counter-explanation to our interpretation is that students who started out lower had more room to grow, which could constitute a regression-toward-the-mean effect rather than an effect due to the PD. There was weak evidence of this, however, because the students who scored below the mean at the pre-PD time point did not necessarily grow more. The main effect estimates (i.e., when not interacting with time) of two of the four teacher variables measuring experience and expertise were negative (though not significant), so by sheer chance we would expect the pre-to-post change to be positive. However, when testing the variables' interactions with time, the estimates for these parameters were negative (and significant), which suggests that these teachers did not transfer the benefits of the PD to their students as much as those who arrived with less experience and expertise.

Another compelling explanation is that the teachers with higher starting inquiry teaching understanding, or the teachers who had taken more PD, were simply less apt to adhere to the TSI PD activities. In other words, the TSI PD may have contradicted their prior understanding to an extent that they were unwilling to implement it (e.g., Caton, 2014). This alternative hypothesis is not supported, however, because the interaction effects between adherence and each of these two teacher variables (starting inquiry teaching understanding and number of prior science PDs) were not statistically significant (at least insofar as our adherence variable was able to capture actual adherence). There may be unobserved variables explaining this mechanism, but the simplest explanation is that the PD benefited students who otherwise would not have received as much inquiry experiences as students in classes with teachers having higher prior inquiry knowledge or more prior PD.

Conclusions and Future Directions

Our findings provide support to the growing body of evidence that combined inquiry- and content-based PD can lead to meaningful gains in student content and NOS knowledge. With regard to the extent to which students gained in their NOS understanding and content knowledge, we found weak but significant NOS gains (g = 0.10) and moderate content gains (g = 0.35). When considering moderating covariates, these estimates were stronger. In the middle school grade band, which the PD was geared toward, and in classes taught by the typical teacher, estimated student gains were higher, with an effect size of g = 0.32 for NOS and g = 0.63 for content.

Teachers' prior knowledge and previous PD experiences, in conjunction with adherence, are important for understanding effects of PD and for planning future PD. All else being equal, and with the exception of the elementary grade band, the teachers in our program who began with less PD experience and less starting inquiry teaching knowledge passed on more gains to their students. This is opposite of the trend that we would expect in a standard comparison condition, and as such it supports the causal effect of the PD on student gains. Our study further implies that inquiry pedagogy and content training through PD may serve to minimize disparities in teaching and learning due to less experienced or knowledgeable teachers.

Our study supports prior research that teachers' increased adherence, to pedagogy and activity guidelines, results in higher student gains. Our finding, that increasing pedagogical knowledge of teachers with low prior inquiry understanding helped to bolster students' NOS scores, supports the assertion that inquiry is an important component in teaching and learning about the process of science. Our study is also valuable in its examination of the TSI model, which is unique in its interpretation of the NOS through (a) flexibility of pathways through the scientific process and the learning cycle, through the TSI phases, and (b) promotion of multiple modes of learning—including authoritative and teacher-directed instruction in the context of inquiry. Our findings support further research on PD that addresses the complexity of scientific practice, including the inquiry process. Furthermore, our focus on aquatic science content for teachers of varied disciplines provides an important example of how PD can be used to bring under-represented content into the classroom.

In summary, our study supports the use of flexible inquiry models as the basis of PD pedagogy and suggests areas for further investigation, such as the interaction between prior teacher preparation and grade band. However, with a relatively small sample size and the absence of a control group, there were limitations to our study. Future research is needed to better understand the underlying mechanisms. Our teacher preparation variables, for example, included measures of teachers' inquiry teaching and associated understanding of teaching scientific process skills but not an explicit measure of teachers' understanding of NOS, which is known to vary among teachers (Lederman, 2007). Teacher NOS may interact with grade band or other teacher variables, which could result in different findings than those presented here. Another future consideration is the way that previous PD experience is operationalized, including the nature of those PD experiences and the degree to which teachers are open to questioning their existing knowledge. Additionally, further research on the use of the TSI modes is warranted. For instance, a more detailed examination of how teachers' characteristics affect their understanding and implementation of the authoritative mode may inform future inquiry-based PD. Our findings, and continuing research in how teachers' adherence articulates with their pre-PD characteristics, will be valuable for determining covariates for subsequent quasi-experimental designs, which are an important in preparation for larger PD studies. Overall, the patterns observed in this paper will serve PD program planning and research efforts to better understand how inquiry instruction affects learning.

References

Abd-El-Khalick, F., BouJaoude, S., & Duschl, R. (2004). Inquiry in science education: International perspectives. Science Education, 88, 397–419.

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 1105–1150). Mahwah, NJ: Routledge.

Adams, R. J., Wu, M. L., Haldane, S. A., & Xun, S. X. (2012). ConQuest (Version 3.0.1) [computer software]. Camberwell, VIC, Australia: Australian Council for Educational Research.

Alexander, D., Heaviside, S., & Farris, E. (1998). Status of education reform in public elementary and secondary schools: Teachers' perspectives. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

American Association for the Advancement of Science. (1990). Science for all Americans. New York, NY: Oxford University Press.

Aulls, M. W., & Shore, B. M. (2008). Inquiry in education (Vol. 1). New York, NY: Lawrence Erlbaum.

Ayala, C. C. (2005, October). Development and validation of an inquiry science student achievement. Paper presented at the annual meeting of the American Evaluation Association, Toronto, Canada.

Banilower, E. R., Heck, D., & Weiss, I. (2007). Can professional development make the vision of 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, 375–395.

Bartos, S. A., & Lederman, N. G. (2014). Teachers' knowledge structures for nature of science and scientific inquiry: Conceptions and classroom practice. Journal of Research in Science Teaching, 51, 1150–1184. doi: 10.1002/tea.21168

Beamer, T., Van Sickle, M., Harrison, G., & Temple, G. (2008). Lasting impact of a professional development program on constructivist science teaching. Journal of Elementary Science Education, 20(4), 49–60.

Bencze, J. (2010). Promoting student-led science and technology projects in elementary teacher education: Entry into core pedagogical practices through technological design. International Journal of Technology & Design Education, 20(1), 43–62. doi:10.1007/s10798-008-9063-7

Blank, R. K., de las Alas, N., & Smith, C. (2007). Analysis of the quality of professional development programs for mathematics and science teachers: Findings from a cross-state study. Washington, DC: Council of Chief State School Officers.

Brandon, P. R., Taum, A. K. H., Ayala, C. C., Young, D. B., Gray, M. E., Speitel, T. W., Pottenger, F. M. (2007). Final report of a phase-I study of the effects of professional development and long-term support on program implementation and scaling up. Honolulu, HI: University of Hawai‘i at Mānoa, Curriculum Research & Development Group. Retrieved from http://manoa.hawaii.edu/crdg/wp-content/uploads/SCUP_final_report.pdf

Bulunuz, N., & Jarrett, O. (2010). The effects of hands-on learning stations on building American elementary teachers' understanding about Earth and space science concepts. Eurasia Journal of Mathematics, Science & Technology Education, 6(2), 85–99.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). The BSCS 5E Instructional Model: Origins and effectiveness. (pp. 1–80). Colorado Springs, CO: Biological Sciences Curriculum Study. Retrieved from : https://uteach.wiki.uml.edu/file/view/UTeach_5Es.pdf/355111234/UTeach_5Es.pdf

Capps, D. K., Crawford, B. A., & Constas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. Journal of Science Teacher Education, 23, 291–318.

Caton, M. (2014). To what extent does professional development influence constructivist science teaching in classrooms?: A comparative study of education in the United States of America and Finland. (Master's Thesis). Available at Digital Commons @Brockport database. (Paper 404) Retrieved from http://digitalcommons.brockport.edu/cgi/viewcontent.cgi?article=1413&context=ehd_theses

Cohen, D. K., & Spillane, J. P. (1993). Policy and practice: The relationship between governance and instruction. In S. H. Fuhrman (Ed.), Designing coherent education policy: Improving the system (pp. 35–95). San Francisco, CA: Jossey-Bass.

Curriculum Research & Development Group (CRDG). (2000). FAST: A summary of evaluations. Honolulu, HI: University of Hawai‘i at Mānoa.

Curriculum Research & Development Group (CRDG). (2014). Accessible professional development for teaching aquatic science inquiry: Final report. Honolulu, HI: University of Hawaiʻi at Mānoa. Retrieved from http://manoa.hawaii.edu/crdg/wp-content/uploads/TSI-CRDG-final-report-Secured-Final-Version.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, 23–45.

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

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

Diamond, B. S., Maerten-Rivera, J. Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. Journal of Research in Science Teaching, 51, 635–658.

Dunst, C. J., Bruder M. B., & Hamby D. W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. Educational Research and Reviews, 10(12), 1731–1744.

Duncan Seraphin, K., Philippoff, J., Parisky, A., Degnan, K. & Papini Warren, D. (2012). Teaching Energy Science as Inquiry: reflections on professional development as a tool to build inquiry teaching skills for middle and high school teachers. Journal of Science Education and Technology, 1-17. doi 10.1007/s10956-012-9389-5

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. Health Education Research Theory and Practice, 18, 237–256.

Fullan, M. (2001). Leading in a culture of change. San Francisco, CA: Jossey-Bass.

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal, 38, 915–945. doi: 10.3102/00028312038004915

Guskey, T. R. (2002). Professional development and teacher change. Teachers and Teaching: Theory and Practice, 8(3/4), 381–391.

Harrison, G. M., Duncan Seraphin, K., Philippoff, J., Vallin, L. M., & Brandon, P. R. (2015). Comparing models of nature of science dimensionality based on the Next Generation Science Standards. International Journal of Science Education, 37, 1321–1342. doi: 10.1080/09500693.2015.1035357

Harrison, G. M., & Vallin, L. M. (2012, October). Developments in measuring teachers' knowledge and practice in teaching science as inquiry. Paper presented at the annual meeting of the American Evaluation Association, Minneapolis, MN.

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. Journal of Research in Science Teaching, 49, 333–362.

Hewson, P. W. (2007). Teacher professional development in science. In S. K. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 1179–1203). Mahwah, NJ: Routledge.

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

Jones, M. T., & Eick, C. J. (2007). Implementing inquiry kit curriculum: Obstacles, adaptations, and practical knowledge development in two middle school science teachers. Science Education, 91, 492–513. doi: 10.1002/sce.20197

Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. Science Education, 855–887. doi 10.1002/sce.20391

Kennedy, M. M. (2016). How does professional development improve teaching? Review of Educational Research, 86, 945–980.

Lakin, J. M., & Wallace, C. S. (2015). Assessing dimensions of inquiry practice by middle school science teachers engaged in a professional development program. Journal of Science Teacher Education, 26, 139–162. doi: 10.1007/s10972-014-9412-1

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 831–879). Mahwah, NJ: Lawrence Erlbaum.

Liu, O. L., Lee, H. S., & Linn, M. C. (2010). An investigation of teacher impact on student inquiry science performance using a hierarchical linear model. Journal of Research in Science Teaching, 47, 807–819. doi: 10.1002/tea.20372

Llewellyn, D. (2002). Inquire within: Implementing inquiry-based science standards. Thousand Oaks, CA: Corwin Press.

Lumpe, A., Czerniak, C., Haney, J., & Beltyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. International Journal of Science Education, 34, 153–166.

McNeill, K. L., Pimentel, D. S., & Strauss, E. G. (2013). The impact of high school science teachers' beliefs, curricular enactments and experience on student learning during an inquiry-based urban ecology curriculum. International Journal of Science Education, 35, 2608–2644. doi: 10.1080/09500693.2011.618193

Minner, D. D., Levy, A. J., & Century J. (2010). 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, 474–496.

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

Mullens, J. E., Gayler, K., Goldstein, D., Hildreth, J., Rubenstein, M., Spiggle, T., Walking Eagle, K., & Welsh, M. (1999). Measuring classroom instructional processes: Using survey and case study field test results to improve item construction (Working Paper 1999-08). Washington, DC: Policy Studies Associates, Inc. Retrieved from http://files.eric.ed.gov/fulltext/ED434144.pdf.

Murnane, R. J., & Willett, J. B. (2011). Methods matter: Improving causal inference in educational and social science research. New York, NY: Oxford University Press.

Nadelson, L. S., Callahan, J., Pyke, P., Hay, A., Dance, M., & Pfiester, J. (2013). Teacher STEM perception and preparation: Inquiry-based STEM professional development for elementary teachers. Journal of Educational Research, 106, 157–168. doi:10.1080/00220671.2012.667014

National Research Council. (2012). A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. Washington D.C.: National Academy Press.

National Research Council. (2005). National science education standards. Washington, D.C.:

National Academy Press.
NGSS Lead States. (2013). Next generation science standards: for states, by states, volume 1: the standards—arranged by disciplinary core ideas and by topic. Washington, DC: The National Academies Press. Retrieved from http://www.nextgenscience.org/next-generation-science-standards
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.
O'Donnell, C. L., & Lynch, S. J. (2008, March). Fidelity of implementation to instructional strategies as a moderator of science curriculum effectiveness. Paper presented at the annual meeting of the National Association of Research in Science Teaching, Baltimore, MD. doi: 10.1.1.331.4222
Osumi, J., & Huang, A. (2015, April). How Prepared are HIDOE Graduates for STEM Majors? Paper presented at the HICAN Conference: Navigating the College, Career, Community Readiness Definition, Kapolei, HI.
Pottenger, F. M. (2007, January). Inquiry and disciplinary natural science teaching. Paper presented at the Hawaiʻi International Conference on Education, Honolulu, HI.
Prawat, R. S. (1992). Teachers' beliefs about teaching and learning: A constructivist perspective. American Journal of Education, 100, 354–395.
Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. The Journal of the Learning Sciences, 16, 81–130.
Raino, A. (2008). Developing the classroom as a "figured world". Journal of Educational Change, 9, 357–364.
Rankin, L. (2005). Lessons learned: Addressing common misconceptions about inquiry. In: Inquiry: Thoughts, views, and strategies for the K-5 classroom. Foundations: A monograph for professionals in science, mathematics, and technology education. Arlington, VA: National Science Foundation. Retrieved from https://www.nsf.gov/pubs/2000/nsf99148/.
REL Midwest Reference Desk. (2015). Teacher Experience and Student Achievement. Naperville, IL: American Institutes for Research. Retrieved from http://www.relmidwest.org/sites/default/files/RDR2015_01_QP10201638_Teacher%20Experience%20and%20Studetn%20Achievement.pdf
Rice, J. (2010). The impact of teacher experience: Examining the evidence and policy implications. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.urban.org/sites/default/files/alfresco/publication-pdfs/1001455-The-Impact-of-Teacher-Experience.pdf
Roehrig, G. H., Kruse, R. A., & Kern, A. (2007). Teacher and school characteristics and their influence on curriculum implementation. Journal of Research in Science Teaching, 44, 883–907.
Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. Journal of Research in Science Teaching, 48(2), 117–148. doi: 10.1002/tea.20408
Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. Educational Researcher, 38, 120–131.
Sanchez, V., Steckler, A., Nitirat, P., Hallfors, D., Cho, H., & Brodish, P. (2007). Fidelity of

implementation in a treatment effectiveness trial of Reconnecting Youth. Health Education Research, 22, 95–107.

Schoedinger, S., Cava, F., & Jewell, B. (2006). The need for ocean literacy in the classroom. The Science Teacher, 73, 44–47.

Schuster, D., Cobern, W. W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2007). Assessing pedagogical content knowledge of inquiry science teaching: Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry. In D. Deeds & B. Callen (Eds.), Proceedings of the National STEM Assessment Conference (pp. 247–265). Washington, DC: Drury University. Retrieved from http://www.openwatermedia.com/downloads/STEM(for-posting).pdf

Seraphin, K. D., Philippoff, J., Kaupp, L., & Vallin, L. (2012). Metacognition as means to increase the effectiveness of inquiry-based science education. Science Education International, 22, 366–382.

Seung, E., Park, S., & Jung, J. J. (2014). Exploring preservice elementary teachers' understanding of the essential features of inquiry-based science teaching using evidence-based reflection. Research in Science Education, 44, 507–529.

Shaha, S. H., Glassett, K. F., Ellsworth, H. (2015). Long-term impact of on-demand professional development on student performance: A longitudinal multi-state study. Journal of International Education Research, 11, 29–34.

Smith, T. M., Desimone, L. M., Zeidner, T. L., Dunn, A. C., Bhatt, M., & Rumyantseva, N. L. (2007). Inquiry–oriented instruction in science: Who teaches that way? Educational Evaluation and Policy Analysis, 29, 169–199.

Strang, C. (2008). Education for ocean literacy and sustainability: Learning from elders, listening to youth. Current: The Journal of Marine Education, 24, 6–10.

Strang, C., deCharon, A. & Schoedinger, S. (2007). Can you be science literate without being ocean literate? Current: The Journal of Marine Education, 23, 7–9. doi: 10.5281/zenodo.30563

Strang, C., Halversen, C, Holland, K, Schoedinger, S, Weiss, E., & Payne, D. (2015). Alignment of Ocean Literacy Framework to the NGSS. Berkeley, CA: Ocean Literacy Network. Retrieved from: http://oceanliteracy.wp2.coexploration.org/next-generation-science-standards-2/

Summerfelt, W. T. (2003). Program strength and fidelity to evaluation. Applied Developmental Science, 7, 55–61.

Supovitz, J. A., Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. Journal of Research in Science Teaching, 37, 963–980.

Wallace, F., Blase, K., Fixsen, D., & Naoom, S. (2008). Implementing the findings of research: Bridging the gap between knowledge and practice. Alexandria, VA: Educational Research Service.

Wee, B., Shepardson, D., Fast, J., & Harbor, J. (2007). Teaching and learning about inquiry: Insights and challenges in professional development. Journal of Science Teacher Education, 18, 63–89.

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. Journal of Research in Science Teaching, 47, 276–301.

Wu, H. K., & Hsieh, C. E. (2006). Developing sixth grader's inquiry skills to construct explanations in inquiry-based learning environments. International Journal of Science

Education, 28, 1290–1313.
Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Supplementary Information linked to the online version of the paper at Wiley-Blackwell

- Figures S1, S2

Table 1. The Modes of Inquiry Addressed in TSI (modified from Duncan Seraphin, Philippoff, Parisky, Degnan, & Papini Warren, 2012)

Mode *Inquiry learning through use of*	**Description** *Search for new knowledge*
Curiosity	through informal or spontaneous probes into the unknown or predictable
Description	through creation of accurate and adequate representation of things or events
Authoritative knowledge	through discovery and evaluation of established knowledge via artifacts or expert testimony
Experimentation	by testing predictions derived from hypotheses
Product Evaluation	about the capacity of products to meet valuing criteria
Technology	through construction, production and testing of artifacts, systems, and techniques
Replication	through duplication; testing the repeatability of something seen or described
Induction	in data patterns and generalizable relationships—a hypothesis finding process
Deduction	in logical synthesis of ideas and evidence—a hypothesis making process
Transitive knowledge	in one field by applying knowledge from another field in a novel way

Table 2. Content by Module in the TSI PD

Module	**Aquatic science content**	**TSI focus**
1 Physical	Investigate the influence of density, wind, waves, tides and the ocean floor on global ocean circulation	Use TSI phases and modes to reflect and become more metacognitive about teaching and practices of science

2 Chemical	Build an understanding of the water molecule and how the unique chemical properties of water have important physical and biological implications	Guide students through TSI phases and modes to enhance learning and understanding of NOS
3 Biological	Explore aquatic diversity—focusing on structure, function, and evolutionary connections between organisms	Practice using inquiry questioning strategies to lead classroom learning through TSI phases
4 Ecological	Apply physical, chemical, and biological principles to the investigation of an aquatic environment	Transfer use of TSI pedagogy to teachers' own lessons

Table 3. Methods for Collecting Data for Operationalizing Teacher Variables Examined in the Study[a, b]

Variable	Description	Development	Administration and scoring
Inquiry: Teachers' understanding of the nature of inquiry-based science teaching in the classroom	A performance assessment, with seven short science-teaching vignettes with multiple-choice and constructed-response prompts	A revision of the Pedagogy of Science Inquiry Teaching Test (Schuster, Cobern, Applegate, Schwartz, Vellom, & Undreiu, 2007); multiple-choice items serve as prompts for the constructed-response items. Conducted multiple reviews with content experts; had four project personnel pilot test the items; modified and revised the items after each review.	Administered online as a pre (before Module 1) and as a post (after Module 4). Scored with a nine-point holistic-rating rubric. Pre and post responses were scored in a single data file with multi-facet Rasch modeling. Rasch reliability = .95.
Content: Teachers' knowledge of aquatic-science content	Four multiple-choice content assessments (*N* items ranged from 29 to 40) administered prior to each module aggregated into a composite percent-correct score	Identified major topics and prepared test blueprints; identified existing items and wrote others; conducted multiple iterative reviews and revisions; conducted content expert reviews and item functioning analyses using the responses from earlier cohorts of teachers (Cohorts 1–3).	Administered online at the beginning of each module. Each test was scored as percent-correct per module; the KR-20 reliabilities were .83, .87, .80, and .66 for Modules 1 through 4, respectively. The composite variable was the mean across the four assessments, with reliability (α) = .91.

Adherence: The extent to which teachers' conducted all of the steps in the PD activities (an aspect of fidelity of implementation)	Four aggregated, Likert-scale items on an evaluation questionnaire given at the end of the PD and nine items on a teacher-reflection log completed after each activity was conducted	Prepared four, six-point self-report items (*1* = not at all, *6* = completely) and nine, five-point self-report items (*1* = not at all, 5 =very much) asking about the degree to which the teachers implemented various aspects of the TSI activities. Conducted multiple reviews and pilot-tested the items with a small sample of teachers.	Administered the four items at the conclusion of each cohort and administered the nine items after the teachers completed each activity that they implemented. After transforming the six-point scale items to five-point scales, averaged the responses to the 13 items across multiple reflection instruments. Reliability (α) = .86.

[a]In addition to the variables described here, we examined demographic variables measured with three items on a background questionnaire administered when the teachers began the PD. One question asked about the teachers' school levels ("What grade(s) are you currently teaching?"), another asked about the total number of years they had taught science to date ("How many *years* have you been teaching *science*?"), and the third asked how many science PD courses they had taken in the previous five years ("About how many *science* professional development courses have you participated in over the *past five years*?").
[b]Full descriptions of the development, administration, and scoring of all the instruments are freely available online in CRDG, 2014, Chapters IV and V.

Table 4. Frequency of Teacher Responses on the Ordinal-scale Moderating Variables

Variable	Numeric coding	*n* teachers
Grade Band		
Elementary (Grade 5)	-1	2 (7%)
Middle (Grades 6–8)	0	15 (54%)
High (Grades 9–12)	1	11 (39%)
Number of science PD courses previously taken		
None	0	10 (36%)
One or two	1	4 (14%)
Three or four	2	8 (29%)
Five or more	3	6 (21%)

Note. Results are from a demographic questionnaire presented to the 28 participating teachers (15 on O‘ahu and 13 on Kaua‘i). The prompts were "What grade(s) are you currently teaching?" and "About how many *science* professional development courses have you participated in over the *past five years*?" The numeric coding was used in the multilevel regression.

Table 5. Descriptive Statistics of the Interval-scale Moderating Variables

Variable	Data	*M*	*SD*	Min	Max
Years Teaching: Number of years teaching science	Number of years, as reported by the teachers in a pre-PD demographic questionnaire	9.25 (0.00)	6.96 (6.96)	1 (-8.25)	30 (20.75)
Inquiry: Inquiry teaching assessment pre-PD (see Table 3)	Logit scores from a multi-facet rating-scale Rasch model of performance ratings Pre-PD	-1.01 (0.00)	1.24 (1.00)	-2.85 (-1.48)	2.50 (2.82)
Content: Aggregate of teacher pre-aquatic-science content assessment (see Table 3)	Mean of the four, percent-correct scores on the content assessments (preceding each module)	58.74 (0.00)	14.63 (1.00)	32.12 (-1.77)	84.49 (1.80)
Adherence: Teacher adherence to activity implementation during PD (see Table 3)	An index ranging from 1 to 5 based on a composite of four post-PD Likert-scale items and nine log items (post-implementation for 12 activities), with higher numbers representing higher adherence	3.52 (0.00)	0.47 (1.00)	2.76 (-1.63)	4.59 (2.31)

Note. The results are for the 28 teachers (15 on O‘ahu and 13 on Kaua‘i) who completed the project. The numbers in parentheses are the descriptive statistics of the centered (and standardized, for the latter three) scores, which were used in the multilevel regression.

Table 6. Methods for Collecting Student Variables Examined in the Study

Variable	Description	Development	Administration & Scoring
Nature-of-science: Students' understanding of the nature-of-science	A set of 18 multiple-choice and 10 semantic-differential-scale items	The semantic-differential-scale items were developed in an Interagency Education Research Initiative grant project (Brandon, et al., 2007) using NOS domains from Ayala (2005). The multiple-choice items were subsequently developed based on those content domains and NOS descriptions in National Research Council (2005). The items were content-reviewed, pilot-tested, and iteratively revised—including a trial with 350 Grades 6–12 students in an affiliated laboratory school and field-tests in Cohorts 2 and 3. Rasch analysis was used to identify further revisions. After the NGSS (2013) were released, we collected content evidence from five judges about each item's match with the eight NOS themes (in Harrison, et al., 2014).	Administered as a paper-and-pencil instrument pre and post to 413 students. The instrument interspersed the nature-of-science items and the aquatic-science content items. Responses were scored in a multidimensional Rasch analysis, with nature-of-science as one dimension and aquatic-science content as another dimension. Rasch reliability estimates with the nature-of-science assessment were pre = .74 and post = .80; with the content assessment, they were pre = .37 and post = .55.
Content: Students' aquatic-science content knowledge	A 17-item assessment addressing the most important content areas of each of the four modules	The most important aquatic science content areas were identified based on the PD module content. Rasch analysis was used to identify major revisions based on item field-testing with Cohorts 1–3; the instrument with those revisions was administered to Cohorts 4–5.	

Note. The instrument is available online in CRDG (2014, pp. 331–342) along with full descriptions of the development, administration, and scoring of the instrument's variables (CRDG 2014, pp. 84, 88–91).

Table 7. Descriptive Statistics of Student Assessment Scores in Logit Values by Grade Band

Grade Band	*N* students	Instrument	Time	*M*	*SD*	Min, Max
Elementary School	23	NOS[a]	Pre	0.58	0.52	-0.27, 0.95
			Post	0.71	0.49	-0.21, 1.49
		Content[b]	Pre	-0.62	0.72	-1.66, 1.27
			Post	-0.04	0.76	-1.66, 1.27
Middle School	208	NOS	Pre	0.58	0.56	-0.64, 2.86
			Post	0.70	0.69	-0.72, 3.36
		Content	Pre	-0.46	0.66	-2.15, 1.66
			Post	-0.09	0.82	-2.15, 4.19
High School	182	NOS	Pre	0.78	0.65	-1.30, 2.46
			Post	0.78	0.76	-0.76, 4.76
		Content	Pre	-0.07	0.70	-2.15, 1.66
			Post	0.02	0.91	-2.93, 2.93
All	413	NOS	Pre	0.67	0.61	-1.30, 2.86
			Post	0.73	0.71	-0.76, 4.76
		Content	Pre	-0.30	0.71	-2.15, 1.66
			Post	-0.04	0.86	-2.93, 4.19

Note. These data are from 2 elementary school teachers, 15 middle teachers, and 11 high school teachers. The scores are logits from a partial-credit Rasch model (using ConQuest by Adams, Wu, Haldane, & Xun, 2012) after anchoring the item difficulties to be the same at both time points; a score of zero logits represents a student with knowledge that matches the average difficulty of the test items. A difficult test will have a mean score that is negative whereas an easy test will have a mean score that is positive; change over time is observed by subtracting the pre from the post.

[a]NOS = Nature of science assessment scores, comprising 28 items (18 multiple choice; 10 functioning Likert-scale items).

[b]Content = Aquatic science content assessment scores, comprising 17 multiple-choice items that measured knowledge of physical, chemical, biological, and ecological science concepts and knowledge in the context of aquatic science.

Table 8. Results of Multilevel Models of Student Assessments in Nature of Science (NOS) and Content

	NOS			Content		
Parameter	Estimate	*SE*	*t*	Estimate	*SE*	*t*
Models without moderators						
Intercept	0.62**	0.05	11.83	-0.37**	0.06	-6.48
Time[a]	0.06*	0.03	2.06	0.26**	0.04	6.77
Models with grade band and teacher characteristics as moderators						
Intercept	0.66**	0.08	7.83	-0.38**	0.08	-4.66
Time[a]	0.21**	0.05	4.35	0.47**	0.06	7.92
Grade band[b]	0.12	0.10	1.23	0.27*	0.10	2.78
Time*Grade band[b]	-0.12*	0.06	-2.03	-0.07	0.10	-0.70
No. years teaching science[c]	0.01	0.01	0.60	0.01	0.01	1.01
Time*No. years teaching science[c]	0.00	0.01	-0.34	0.00	0.01	0.26
No. science PD courses taken[d]	-0.08	0.05	-1.80	-0.07	0.04	-1.63
Time*No. prior science PD courses taken[d]	-0.07**	0.03	-2.61	-0.08*	0.04	-2.28
Inquiry teaching assessment[e]	0.05	0.06	0.86	0.07	0.06	1.16
Time*Pre-PD inquiry teaching assessment[e]	0.02	0.04	0.46	-0.02	0.04	-0.50
Aggregate content score[e]	-0.03	0.08	-0.40	-0.03	0.08	-0.39
Time*Pre-PD aggregate content score[e]	0.05	0.05	1.17	0.10	0.06	1.78
Teacher adherence[e, f]	0.01	0.06	0.17	0.04	0.06	0.73
Time*Teacher adherence[e, f]	0.09*	0.04	2.43	0.12**	0.04	2.71
Time*Grade band*Pre-PD inquiry teaching assessment	-0.20**	0.05	-3.87	—	—	—
Time*Grade band*No. sci. PD courses taken	—	—	—	-0.18**	0.06	-2.84

Note. Data were from 413 students (those who took both the pre and the post assessment) in Cohorts 4 and 5. Scores are on a Rasch-modeled logit scale after anchoring the item difficulties to be the same at both time points. The NOS and aquatic science content scores were modeled as separate but related dimensions, based on a partial credit model to account for polytomous and dichotomous data. Time was Level 1, Student was Level 2 (random intercepts and slopes), and Teacher (random intercepts) was Level 3. With the exception of adherence, the predictors were pre-PD.
* = $p < .05$; ** = $p < .01$.
[a] Time was coded 0 for pre and 1 for post. [b] Grade band was coded -1 for elementary, 0 for middle, and 1 for high school. [c] This predictor was centered. [d] Number of science PD courses taken was coded 0 for zero courses, 1 for one or two courses, 2 for three or four courses, and 3 for five or more courses. [e] These predictors were standardized.
[f] The adherence fidelity of implementation variable was a composite score from teachers' lesson reflections and responses on a post-PD questionnaire.

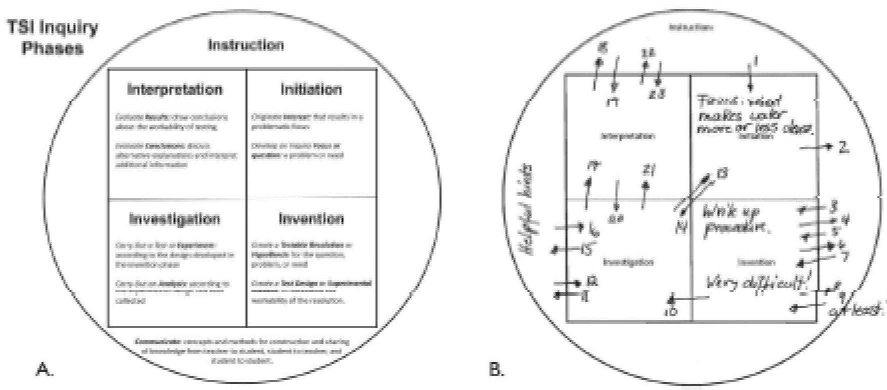

Figure 1. (A) The TSI phases illustrate the fluid nature of scientific inquiry; the phases are connected but without arrows in order to emphasize that researchers, teachers, and students move back and forth between phases throughout the inquiry process. The instruction phase encircles the other phases, linking it to the each of the other phases. (B) A teacher's reflective TSI phase diagram from an implemented classroom lesson on density (http://manoa.hawaii.edu/exploringourfluidearth/physical/density-effects/densitytemperature- and-salinity/activity-density-bags). The arrows show the multi-directional path of students between the phases and the repeated role of instruction in inquiry. (This figure is modified from Seraphin et al., 2012.)

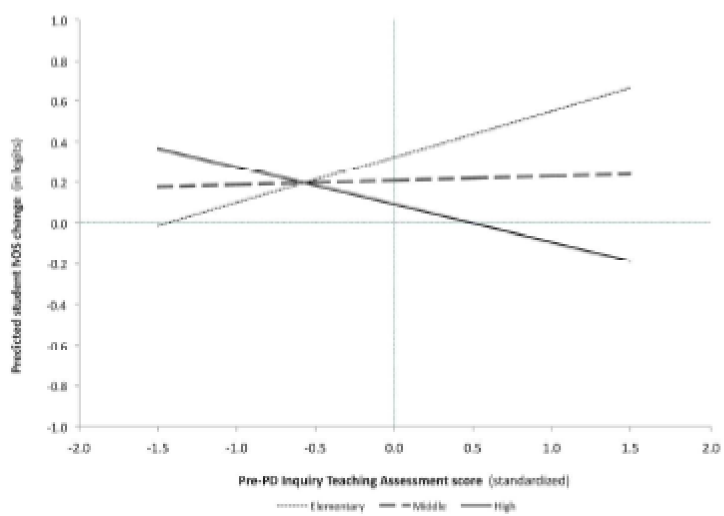

Figure 2. Prediction lines of student NOS change scores, by grade band. Predictions are based on teachers' previous inquiry teaching knowledge, displaying the interaction when all other independent variables are at zero.

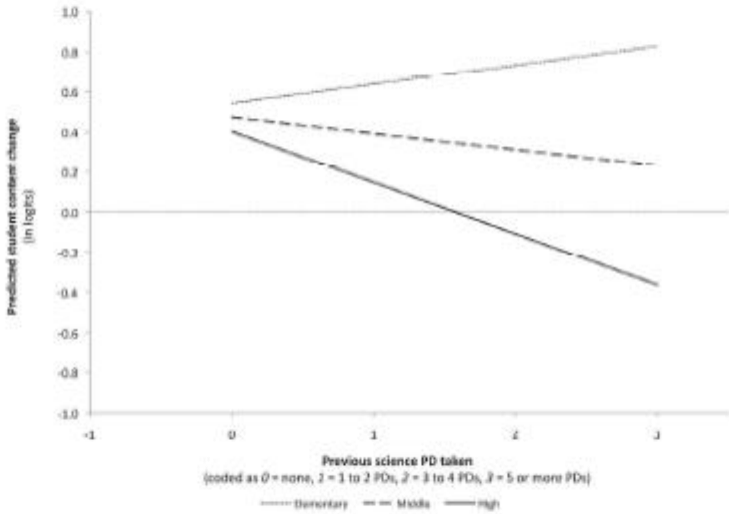

Figure 3. Prediction lines of student content change scores, by grade band. Predictions are based on the number of previous science PDs taken by the teacher in the past five years (0 = none, 1 = 1 to 2 PDs, 2 = 3 to 4 PDs, 3 = 5 or more), displaying the interaction when all other independent variables are at zero.