AN EXAMINATION OF THE LONG-TERM EFFECTS OF A TEACHER PROFESSIONAL DEVELOPMENT IN INQUIRY SCIENCE

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

EDUCATIONAL PSYCHOLOGY

AUGUST 2021

By

Joanna K. Philippoff

Dissertation Committee:

Katherine T. Ratliffe, Chairperson
Paul R. Brandon
George M. Harrison
Mark A. Murphy
Judy D. Lemus

Keywords: professional development, long-term effects, inquiry science, mixed methods,
content retention, pedagogy retention, science teacher education, marine science

ACKNOWLEDGEMENTS

I first want to express my gratitude to all of the teachers who took part in this study. Without your participation, this research would not have been possible. I suspect I learned much more with and from you than you from me.

To Kanesa Duncan Seraphin, who brought me onto the project that served as the basis for this dissertation—at a time in my life when I had lost my way. Thank you for helping me find it.

I would like to thank my colleagues in CRDG who inspire me daily. A special thank-you to my mentor, Thanh Trúc T. Nguyễn, my faculty cohort member Kara Suzuka, and directors Don Young, Kathy Berg, Barbara Dougherty, and Linda Venenciano.

Many thanks to the Educational Psychology department faculty at the University of Hawaiʻi at Mānoa for supporting me throughout this long journey. I would like to give a special thank you to my dissertation committee for supporting me along this journey, especially to my chair, Kathy Ratliffe, whose calm presence and encouragement pushed me along; to George Harrison, for always being gracious about answering my questions—including sifting through rather old data; and to Paul Brandon, who has been a true mentor in every sense of the word.

I would like to give a special shout out to the strong-willed wonderful women who I relied on for support and encouragement through this process—especially Anuschka Faucci, Lauren Kaupp, Annie Lee, and Cheryl Squair. I hope I can return the favor someday.

Finally, and most importantly, I want to express my gratitude to my family. This dissertation would not have been possible without your love and support. Thank you to my Hawaiʻi ʻohana the Vallin-Manuguids, especially Lisa Vallin. I truly do not think I could have done this without you. Thank you to my parents, Arlene and George Philippoff, my siblings, Andrea Jones and Karl Philippoff, and my Aunt Irma for supporting me throughout this process and lifting me up when I fell down. I want to give a special thank you to my mom, the eternal optimist, who never had any doubt that I could do this, and to my brother, for countless hours editing. I hope your now deep knowledge of teacher PD comes in handy someday. Finally, I want to thank my husband Ross Langston for hanging in there. In spite of everything—I seem to have made it.

DEDICATION

This dissertation is dedicated to my daughter,

Norah Philippoff Langston.

I love you more than you will ever know.

And what I really, really liked about TSI was they started out by saying, you know, like ‘Dude, water...Let's start with water.’ And if we start there, we will hit everything. You know, physics, chemistry, biology, philosophy. It was beautiful, it was an excellent idea. And it was really good in that it—while it may have been really annoying at the time when we were pressed for time, taking it apart and piece-by-piecing it out how the brain works with the—what did you call them?—the I's? Yeah, the phases. And then the modes... But, it makes you really think it through that if you start—just get started some one place, some topic, and really work it through, you can build multiple years, you know, each class, you could have a biology class based on water. And start there and move on. You could have a chemistry class, physics, religion, for heaven sakes. History, civilization, whatever. Human anatomy. It was a really, really good idea. It was a great idea. And I think for us, who have sort of an inherent propensity towards inquiry and being curious, the inquiry method makes sense. But it doesn't work for everyone, but it makes sense as the best way to get the most people interested. And I loved my TSI experience.

- Teacher, study participant

ABSTRACT

Although many studies have looked at changes in teacher knowledge, beliefs, and behavior over the course of professional development programs, long-term follow-up remains rare. In this case study of the long-term effects of a high-quality year-long professional development in inquiry science, we administered the same instruments to the same teachers ($N = 23$) before and after the professional development and 2.5 years after the professional development ended. We also interviewed the teachers about factors that enhanced or impeded PD implementation. The teachers demonstrated pre- to post-professional development gains on instruments examining inquiry-based teaching knowledge, content knowledge (marine science), and self-efficacy. Although declines were expected due to the lack of follow-up, results varied over instruments. The teachers' scores dropped significantly on all of the measures that showed a pre- to post-professional development gain, but remained significantly above pre-professional development baseline levels at the long-term follow-up time point. On the pedagogical content knowledge instrument the teachers' scores grew significantly from the post-professional development to the follow-up time point. Interviews revealed a sustained highly positive response to the professional development with self-reported gains in confidence, inquiry and content knowledge, and modest to substantial changes in teaching practice attributed to the intervention. While explicit use of professional development pedagogy and activities declined over time, aspects of the pedagogical approaches became embedded in teachers' practices. This case study emphasized the need for long-term support of teachers to sustain change over time and has implications for interpreting the outcomes of professional development.

CONTENTS

ACKNOWLEDGEMENTS ii

DEDICATION iii

ABSTRACT v

TABLES xiii

FIGURES xvi

CHAPTER 1: INTRODUCTION 1

Inquiry-based Teaching 1

Teacher PD 2

Problem Statement 4

Statement of Purpose 6

Rationale and Significance of the Study 7

Description of the PD 9

Teaching Science as Inquiry PD 9

PD Design 10

Theory of Change 12

Evaluation Results 13

Limitations 15

Selection Bias 15

Reliance on Self-report 16

Summary and Organization of the Study 17

CHAPTER 2: REVIEW OF THE LITERATURE 19

Constructivism 19

Historical Roots of Constructivism 20

Piaget..20
Vygotsky..21
Dewey..21
Modern Interpretations...22
Active Learning...22
Build on Prior Knowledge..22
Scaffold Instruction..23
Make Learning Relevant..23
Develop Self-regulation...23
Colloborative Learning..23
Teachers as Facilitators...24
Constructivism and Inquiry..24
TSI Pedagogical Framework...27
TSI and NGSS..29
Choice of Content—Aquatic Science..33
Components of Effective PD...35
Criticism of Components..37
Individual Experiences in PD..39
PD Framework...41
Bandura’s Theory of Reciprocal Determination..42
Content Knowledge..43
Pedagogical Content Knowledge...45
Self-efficacy...46
Importance of Envionment..46

Research Questions 47
CHAPTER 3: METHODS 50
Design 50
Limitations to the Design 54
Participants 56
Instruments 61
Quantitative Instruments 61
Inquiry Teaching Assessment 62
Pedgagogical Content Knowledge Assessment 64
Self-efficacy Assessment 65
Content Knowledge Assessment 66
Limitations of the Instruments 72
Content Knowledge Assessment Limitations 72
Qualitative Interview 74
Procedures 77
Analytic Methods 79
Research Question 1—Analysis of Quantitative Data 79
Preparatory Steps 79
Inquiry Teaching Assessment 80
Statistical Analyses 83
Credibility of Results—Comparisons Among Instument Scores 84
Research Question 1A—Examining Patterns inf the Quantitative Data 85
Research Question 2—Analysis of Interview Data 85

Validity Threats When Analyzing Qualitative Data 87
Research Question 3—Mixed Methods 88
Positionality 89
CHAPTER 4: RESULTS 92
Quantitative Assessment Results—Research Questions 1 and 1A 94
Inquiry Teaching Assessment 94
Content Knowledge Assessment 101
Pedagogical Content Knowledge Assessment 108
Self-efficacy Assessment 113
Effects of PD Participation After PD Under Study 120
Comparisons Between and Across Quantitative Assessments 121
In Summary—Results of Research Questions 1 and 1A 124
Research Question 1 126
Research Question 1A 127
Interview Qualitative Results—Research Question 2 127
Envionmental Factor 128
PD Embodied Adult Learning Principles 128
Personal Factors 131
Gained Pedagogical Knowledge 131
Gained Content Knowledge 134
Erosion of Pedagogical and Content Knowledge 136
Interaction Between Environmental and Personal Factors 138
PD Affirmed Teaching Philosophy 138
PD Increased Confidence 140

Personal Preferences and Agency 141
Curriculum Modifications for Students 142
Transfer of PD Principles Beyond “Focus’ Class 144
Interaction Between Environmental and Behavioral Factors 145
Student Engagment and Learning 145
PD Provided Resources for Implementation..147
Importance of Class Alignment 149
Content Alignment 149
Science Standards 151
Pacing Guides, Testing, and Time 153
Recognition, Resistance, and the Importance of Community..................................153
Interactions Between Personal and Behavioral Factors...156
Importance of Practice 156
Sharing PD Pedagogy and Curriculum 158
Behavioral Factors—Changes in Teaching Practice..158
Implementation of Specific PD Activities 166
In Summary—Results of Research Question 2 168
Mixed Methods Analysis—Research Quesiton 3 170
Self-reported Levels of PD Behaviors 171
Teaching Change Relationships...176
Mixed Method Joint Display Tables...177
In Summary—Results of Research Question 3 180
CHAPTER 5: DISCUSSION...183
Quantitative Instrument Constructs—Research Question 1 184

Understanding of Inquiry-based Teaching 185
Content Knowledge 186
Pedagogical Content Knowledge 190
Self-efficacy 191
Relationship Between Pedagogical Content Knowledge and Self-efficacy 191
Effect of Grade Level, Experience, and Subsequent PD Participation—
Research Question 1A 182
Interview Findings—Research Question 2 193
Interactions Between Personal and Behavioral Factors 193
Interactions Between Environmental and Personal Factors 195
Interactions Between Environmental and Behavioral Factors 198
TSI and NGSS 200
Mixed Methods—Research Question 3 201
Limitations 203
Suggestions for Future Research 205
Implications 208
Conclusion 209
APPENDIX A: Instructions for Rating the Inquiry Teaching Assessment Responses 211
APPENDIX B: Inquiry Teaching Assessment 213
APPENDIX C: Classroom Instruction Questionnaire 221
APPENDIX D: Teacher Content Assessment 225
APPENDIX E: Content Knowledge Assessment Blueprint 237
APPENDIX F: Interview Slides 238
APPENDIX G: Examinations of Non-Normality 241

REFERENCES 243

TABLES

Table 2.1 The Modes of Inquiry Addressed in TSI 32

Table 3.1 Study Design: Instrument Administration 51

Table 3.2 Research Questions, Instruments, Representations, and Statistical Approaches 54

Table 3.3 Teacher Characteristics: Cohort, Years Teaching Science, and Grade Band 58

Table 3.4 Teacher Characteristics: Discipline Taught 60

Table 3.5 Description and Source of Instruments of Constructs on the Follow-up Questionnaire 62

Table 3.6 Descriptive Statistics of the Inquiry Teaching Assessment 63

Table 3.7 Descriptive Statistics of the Pedagogical Content Knowledge Assessment and Self-Efficacy Assessment 66

Table 3.8 Teacher Content Knowledge Assessment Results 67

Table 3.9 Alignment of Follow-up Content Knowledge Assessment Items to PD Content 71

Table 3.10 Follow-up Content Knowledge Assessment Descriptive Statistics 72

Table 3.11 Follow-up Aquatic Science Content Knowledge Assessment Descriptive Statistics 79

Table 4.1 Descriptive Statistics of the Inquiry Teaching Assessment 95

Table 4.2 Paired-samples t-Test Results Between Different Time Points of the Inquiry Teaching Assessment 98

Table 4.3 Independent-samples t-Test Results of the Score Changes Between Grade Bands on the Inquiry Teaching Assessment 101

Table 4.4 Descriptive Statistics of the Content Knowledge Assessment 102

Table 4.5 Paired-samples t-Test Results Between Different Time Points of the Content Knoweledge Assessment 104

Table 4.6 Independent-samples t-Test Results by Grade Bands on the Content Knowledge Assessment 106

Table 4.7 Descriptive Statistics of the Pedagogical Content Knowledge Assessment 109

Table 4.8 Wilcoxon Signed-Rank Test Results Between Different Time Points of the Pedagogical Content Knowledge Assessment 111

Table 4.9 Independent-samples t-Test Results by Grade Bands on the Pedagogical Content Knowledge Assessment....113

Table 4.10 Descriptive Statistics of the Self-efficacy Assessment....114

Table 4.11 Paired-samples t-Test Results Between Different Time Points of the Self-efficacy Assessment....117

Table 4.12 Independent-samples t-Test Results by Grade Bands on the Self-efficacy Assessment....120

Table 4.13 Independent-samples t-Test Follow-up Instrument Results Between the Teachers With and Without Significant PD Experiences After the PD Under Study....121

Table 4.14 Score Correlations Within Instruments Across Different Time Points....122

Table 4.15 Score Correlations Across Instruments and Different Time Points....124

Table 4.16 Summary of Results and Interpretation of Data Sources by Research Questions....125

Table 4.17 Adult Learning Principles and Core PD Features....129

Table 4.18 Pedagogical Concepts the Teachers Thought were Significant in the Follow-up Interview....133

Table 4.19 Follow-up Interview Self-reported Teaching Changes....162

Table 4.20 Frequency of PD Activities Mentioned by the Teachers in the Follow-up Interview....167

Table 4.21 Quantitative Summary of Interview Results Using Themes Derived from Bandura’s Theory of Reciprocal Determination....169

Table 4.22 Descriptive Statistics of the Likert-scale Interview Questions....172

Table 4.23 Score Correlations Between Instruments at the Follow-up Time Point and the Teachers’ Application of PD Practices and Implementation of PD Activities....175

Table 4.24 Independent-samples t-Test Results Comparing Grade Level with Self-reported Implementation....175

Table 4.25 Independent-samples t-Test Results of Instruments, Self-reported Knowledge and Behaviors, and Extent of Teaching Change....178

Table 4.26 Influence of Environmental Factors on Changes in Teaching Practices....179

Table 4.27 Influence of Environmental Factors on Self-reported Frequency of Implementation of PD Activities....181

FIGURES

Figure 1.1 The TSI Aquatic PD Project Structure 11

Figure 1.2 TSI Aquatic PD Theory of Change 13

Figure 2.1 How Components of the TSI Framework Relate to Scientific Inquiry 28

Figure 2.2 The TSI Phase Diagram....................31

Figure 2.3 Bandura's Triadic Reciprocation Model 43

Figure 3.1 Data Collection Over the Course of the PD and at the Follow-up 51

Figure 3.2 Possible PD Instrument Score Trajectories: Stable, Increasing, or Eroding 52

Figure 3.3 Flow Chart of Convergent Mixed Methods Design Illustrating Collection, Analysis, and Merging of Qualitative and Quantitative Data 53

Figure 4.1 Histogram of the Inquiry Teaching Assessment Scores at Different Time Points 96

Figure 4.2 Boxplot of the Inquiry Teaching Assessment Scores at Different Time Points 97

Figure 4.3 Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Inquiry Teaching Assessment....................100

Figure 4.4 Histogram of the Content Knowledge Assessment Scores at Different Time Points 102

Figure 4.5 Boxplot of the Content Knowledge Assessment Scores at Different Time Points 103

Figure 4.6 Scatterplot Relationship Between the Changer in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Content Knowledge Assessment 107

Figure 4.7 Histogram of the Pedagogical Content Knowledge Assessment Scores at Different Time Points 109

Figure 4.8 Boxplot of the Pedagogical Content Knowledge Assessment Scores at Different Time Points 110

Figure 4.9 Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Pedagogical Content Knowledge Assessment 112

Figure 4.10 Histogram of the Self-efficacy Assessment Scores at Different Time Points....................115

Figure 4.11 Boxplot of the Self-efficacy Scores at Different Time Points.................................115

Figure 4.12 Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Self-efficacy Assessment.............118

Figure 4.13 Synthesis of Quantitative Instrument Results..126

Figure 4.14 Venn Diagram of Relationship Between PD Content Knowledge and Teaching Relevance 136

Figure 4.15 Synthesis of Interview Findings Overlaid on Bandura's Theory of Reciprocal Determination 170

Figure 4.16 Histogram of Extent of Application of PD Practices to Teaching 172

Figure 4.17 Histogram of Implementation of PD Activities...173

Figure 4.18 Synthesis of Self-Reported Behavior Relationships with Personal Constructs...176

Figure 4.19 Overview of Mixed Methods Results...182

CHAPTER 1

INTRODUCTION

This chapter begins with a discussion of inquiry-based teaching and the importance of, and concerns about, teacher professional development (PD). Next, I provide the problem statement, the statement of purpose, and the rationale and significance of this study. I then describe the PD under study including its design, theory of change, and original evaluation results. Finally, I discuss the limitations and describe the organization of this study. The purpose of this study is to determine if gains in the scores on instruments administered pre-to-post PD are sustained over time, and the factors that may have enhanced or hindered classroom behaviors tied to PD practices 2.5 years after the end of a year-long inquiry-based PD.

Inquiry-based Teaching

Inquiry-based teaching develops students' understandings of scientific concepts and processes while engaging them in experiential investigations. Inquiry pedagogy has been shown to improve students' understanding of scientific practices, increase students' conceptual knowledge, enhance their critical thinking and problem-solving abilities, and keep students engaged in and interested in the learning material (e.g., Armbruster et al., 2009; Aulls & Shore, 2008; Cohen & Spillane 1993; Luckie et al., 2012; Maeng et al., 2020; Marshall et al., 2017; Minner et al., 2010; National Research Council [NRC], 2012; Wilson et al., 2010).

Although science educators have been advocating for teaching science in an inquiry-based manner for years, teachers are still hesitant to implement inquiry in their classrooms (e.g., Fitzgerald et al., 2019). Inquiry-based teaching unfortunately still represents a paradigm shift for many educators. This is, in part, because teachers have trouble defining and articulating what inquiry means and designing inquiry-based lessons for their classrooms (Fitzgerald et al., 2019).

Thus, there are a number of misunderstandings about inquiry which can prevent or lead to flawed implementation (Reiser, 2013; Schraw et al., 2006). One way to address these misconceptions and train teachers to effectively practice scientific inquiry in their classrooms is through teacher PD (Smith, et al., 2007).

Teacher PD

Research in the field of educational effectiveness has consistently shown that teachers are the most critical in-school influence on student learning (e.g., Nye et al., 2004; Rivkin et al., 2005). There is also a general consensus that PD is essential to building teachers' capacities and improving educational outcomes (e.g., Gulamhussein, 2013; Wei et al., 2010). In order to systematically implement educational reforms, reform movements need to galvanize teachers. Teacher PD, when effective, is a key component in the ultimate success or failure of reform efforts.

Although PD has been recognized as a lynchpin to achieving educational progress, its benefits are still more theoretical than concrete. Each call for reform over the past few decades has yet to be systematically implemented (Hargreaves & Goodson, 2006; NRC, 2007). While there is evidence that PD can enhance teacher instructional practices (Borko, 2004; Desimone et al., 2002; Garet et al., 2001; Porter et al., 2000) and support student learning and achievement (Akiba & Liang, 2016; Blank & de las Alas, 2009; Shaha et al., 2015; Yoon et al., 2007), meta-analyses have shown that most PD does not have a clear effect on teacher and/or student outcomes (Garet et al., 2016; Guskey & Yoon, 2009; Harris & Sass, 2011; Hill et al., 2013; Pellegrini et al. 2021; Rudolph, 2014; Supovitz & Turner, 2000). Studies on inquiry-based PD mirror these findings; research shows that it enhances teachers' understanding of science, self-efficacy in their ability to teach science, and use of inquiry-based teaching practices (Akerson &

Hanuscin, 2007; Alexander et al., 1999; Beamer et al., 2008; Bencze, 2010; Bulunuz & Jarrett, 2010), but the research on the overall effects on teacher and student learning is mixed (Cairns & Areepattamannil, 2019; Slavin et al., 2014). Critics have pointed to the ambiguity in PD effectiveness, and often considerable costs, as being a poor return on investment (Jacob & McGovern, 2015; Shear & Penuel, 2010; Whitehead, 2010).

The conflicting evidence about the overall effects of PD is compounded by the lack of high-quality PD (Banilower et al., 2013; Desimone et al., 2002). For example, a survey of teachers at four schools in one district found that teachers were inundated with "a barrage of PD experiences" (Linn et al., 2010, p. 680), but there was little evidence that these "experiences" had any material effect within the classroom. A recent study comparing science teacher PD in 2012 and 2018 found little evidence of improvement in quality over time, describing the overall number of quality PD offerings and rates of participation as "troubling" (Smith, 2020, p. 603). It is perhaps no wonder then that teachers have, in general, continued to be largely dissatisfied with PD (e.g., Darling-Hammond et al., 2009; Penuel et al., 2007). This is unfortunate as to accomplish the student outcome goals desired by educational reformers "the burden of change falls almost entirely on the shoulders of teachers. But with so little support, the demands are monumental" (Smith, 2020, p. 608).

Teacher adoption of reforms is complex and difficult to assess (e.g., Spillane et al., 2002; Windschitl, 2002), and there is uncertainty about how PD affects teacher learning and instructional practice (Kennedy, 2016). But, as Guskey (2000) noted, "one constant finding in the research literature is that notable improvements in education almost never take place in the absence of professional development" (p. 4). While lots of PD does not guarantee success, minimal PD virtually guarantees failure (Pellegrini et al., 2021). Thus, the empirical study of PD components, including their design, theories of change, promoted strategies, and the assessment of their outcomes continues to be critical. However, effectively evaluating PD is difficult, as in addition to the many design factors at play (e.g., Curriculum Research & Development Group [CRDG], 2012; Darling-Hammond et al., 2017; Desimone, 2009; Loucks-Horsley et al., 2012; Sancar et al., 2020; Timperley et al., 2007), context is important (e.g., Banilower et al., 2007; Hargreaves & Goodson, 2006).

Problem statement

Many studies have shown pre-post PD gains, and there is general agreement about the components of effective PD—even if the specifics and relative importance of them are contested (e.g., Hill et al., 2020; Kennedy, 2016; Lynch et al., 2019; Sims & Fletcher-Wood, 2021). Most PD evaluations, however, stop collecting data upon the conclusion of the PD program and so cannot investigate longer-term effects (Desimone & Stuckey, 2014; Hargreaves & Goodson 2006). Although there is little argument that effective PD should produce lasting change in teacher knowledge, most studies on PD effectiveness are essentially based on post-PD snapshots and fail to capture the extended story of how teachers adopt, adapt, or abandon PD practices. This is in part due to funding cycles (Avalos, 2011; Kennedy, 2016; Odden et al., 2002). But successful PD, as measured by near-term teacher learning, does not necessarily lead to enduring changes in teacher practice (Fixsen et al., 2005). At the post-PD time point, teachers often have yet to implement the program without substantial assistance; thus, program evaluations may be overly optimistic (Liu & Phelps, 2020). Program effects are not fixed, especially in the context of an educational landscape that is constantly in flux. PD studies that neglect the time-varying environmental considerations in which teachers operate may be contributing inaccurate, or at least incomplete, evidence to the PD literature (Liu & Phelps, 2020). But the lack of long-term studies means there is little understanding of what factors contribute to longitudinal PD success and how best to measure it (Desimone & Stuckey, 2014).

Successful PD should lead to sustained teacher instructional change and continued use of PD practices, complemented by sustained gains in student learning and understanding. Ideally, evaluations of PD should assess both post-PD and long-term outcomes to address not just the initial success of an intervention but its durability (Scher & O'Reilly, 2009). However, while assessing student outcomes immediately as well as long-term is ideal, this is often not feasible. Since teachers influence student learning outcomes, if the changes in teacher behavior that led to student gains are maintained, then some measure of the previously reported student gains should be retained. Thus, PDs should continue to monitor participants, on some level, after the PD has ended to determine if, and how long, they can maintain pre- to post-PD gains (Kennedy, 2016).

> Studies that are coterminous with the PD itself cannot tell us whether teachers merely comply with program recommendations as long as they have to, whether they continue building on the program's ideas over time, or whether, perhaps, they revise the advice so severely that its original meaning is lost. (Kennedy, 2016, p. 973)

If sustainability of PD effects is to be prioritized, and teacher change and growth are recognized as long-term processes, there is still the question of when researchers should follow-up with teachers. As Ford (1994) noted, "the point in time at which changes in behavior are measured can have a major impact on the conclusions drawn about whether transfer has occurred" (p. 23). Implementation of new instructional practices tends to be a gradual process; shifts may occur over several years (Fixsen et al., 2005; Guskey, 2002a). Researchers need to balance waiting until teachers are no longer struggling with understanding an innovation, but are refining it and personalizing it, with the accumulation of contextual and environmental factors over time that may affect program implementation. For example, teachers took 3–5 years to reach a level of mastery in a study of a middle-school science program (Young, n.d.), and researchers studying the effects of an early childhood PD recommended waiting three years to follow-up on PD effects (Bierman et al., 2013). In practice, however, the timing of follow-up tends to vary widely, at least partly because no standards currently exist in the literature. In general, longer-term follow-ups tend to be fortuitous and ad-hoc (e.g., Shaharabani & Tal, 2017; Padua, 2018); while those done closer to the end of programs are incorporated into the original PD research design (e.g., Goldschmidt & Phelps, 2010; Gore et al., 2017). However, it is clear that following-up with teachers several years after a PD is completed grounds implementation in the realities of teachers' practice. Findings from longitudinal research may help to elucidate why most PD has insignificant effects (e.g., Garet et al., 2016; Guskey & Yoon, 2009; Harris & Sass, 2011; Hill et al., 2013; Pellegrini et al. 2021; Rudolph, 2014; Supovitz & Turner, 2000; Yoon et al, 2007), and what factors are important for teachers to continue to implement the program after PD supports have been removed (Kennedy, 2016). Findings may also lead researchers to be more realistic about the strengths, and limitations, of PD, as "we need to ensure that PD promotes real learning rather than merely adding more noise to [the teachers] working environment" (Kennedy, 2016, p. 974).

Statement of Purpose

PD program stakeholders, including facilitators and funders, value interventions that have enduring effects beyond the project. While pre-post measures are ubiquitous—and important to understanding content and pedagogy uptake—only by following teachers after projects end can the effectiveness of PD programs to enact lasting changes in education be measured (e.g., Fixsen et al., 2005; Kennedy, 2016). However, long-term monitoring of changes in teacher practice as a result of their participation in PD is rare (e.g., DeMonte, 2013; Yoon et al., 2007). "Because studies usually last only one year, we do not have a good understanding of the ebb and flow of teacher learning or the way that implementation of new instruction continues into year 2, year 3, and beyond as a result of high-quality professional development" (Desimone & Stuckey; 2014, p. 472). The limited research examining PD participants longitudinally represents a gap in the literature that warrants additional exploration. This study examined the extent to which teachers who participated in a year-long inquiry-based science PD program sustained gains on measured constructs and changed their teaching practices 2.5 years after the conclusion of the PD.

Rationale and Significance of the Study

There is a limited, but growing, number of empirical studies that have investigated the long-term effectiveness of PD. These studies used one or more of the following data collection methods: self-report surveys (e.g., Eylon & Bagno 2006; Havice et al., 2018; Lydon & King, 2009; Supovitz et al., 2000), interviews (e.g., Deglau & O'Sullivan, 2006; Shaharabani & Tal, 2017; Gaikhorst et al., 2017; Knapp & Peterson, 1995; Padua, 2018), classroom observations (e.g., Penner-Williams et al., 2019), and quantitative assessments (e.g., content knowledge or self-efficacy; Goldschmidt & Phelps, 2010; Heller et al., 2012; Liu & Phelps, 2020). Of the studies that used multiple methods, researchers most commonly paired classroom observations with one of the other data sources (e.g., Antoniou & Kyriakides, 2013; Bierman et al., 2013; Franke et al., 2001; Gore et al., 2017; Murray et al., 2018; Wolf & Peele, 2019).

Long-term PD studies that have looked at changes in knowledge over time have focused more on elementary teachers (e.g., Goldschmidt & Phelps, 2010; Heller et al., 2012; Sandholtz & Ringstaff, 2016), thus there is a need to examine how secondary teachers' knowledge and practices evolve over time in response to PD (although see Clary et al., 2018; Lydon & King, 2009; Supovitz et al, 2000). While the number of teachers who participated in this study is small from a quantitative perspective (*N* = 23), this sample size is large compared to other long-term PD studies that have relied on interviews (although see Franke et al. 2001; Knapp & Peterson, 1995; Sandholtz & Ringstaff, 2016). In interviews, teachers have the opportunity to fully describe their current practices. Luft and Hewson (2014) considered studying how teachers put their learning into practice important because "by understanding the boundaries of context and teacher learning, there is the potential to design [PDs] that better support teacher learning in science" (p. 921). In addition, understanding how context affects teacher implementation of the PD can highlight not only what additional support mechanisms need to be in place, but also how to position them temporally.

Teachers implement PD programs differently in part because they attend to and draw upon different program components to support their own unique instructional needs (Luft & Hewson, 2014). These diverse implementations interact with and affect teachers' knowledge and beliefs (e.g., Desimone, 2009; Guskey, 2002b). Reporting that teacher beliefs or understandings changed as a result of participating in a PD can be very informative. However, scores on instruments neglect the complexity inherent in project implementation. Interviews can elucidate the reasons for the variation in PD implementation over time and collect information on factors that enhanced or impeded implementation of PD components, which can shed light on what "seeds of change" planted by the PD have taken root and, more generally, how program experiences led to program outcomes.

The map of teacher learning is made up of strategies that were tried and abandoned and those that were maintained and modified. The extent of maintained and modified strategies and the level of coherence of tools and approaches years after the [PD programs] are evidence of teacher learning. (Shaharabani & Tal, 2017, p. 1050)

This study adds insight into the stability of changes to teachers' understandings, beliefs, and instructional practice years after their completion of an intensive PD program. Specifically, the findings of this research study add to the research base on the longitudinal effects of PD on educational change at the teacher level. Results from this study contribute to our understanding of the barriers to teacher behavioral change and extend the conversation about what constitutes PD success.

Description of the PD

Teaching Science as Inquiry PD

This study followed the participants of a PD that utilized the Teaching Science as Inquiry (TSI) model as its pedagogical foundation, which is detailed in Chapter 2. The TSI Aquatic PD was funded through an Institute for Educational Sciences development grant. The objectives of the PD were to (a) increase the teachers' content knowledge in aquatic science, (b) improve the teachers' understanding of scientific inquiry and pedagogical content knowledge needed to create classrooms that function as a community of scientists, and (c) improve the teachers' self-efficacy in using inquiry pedagogy. Ultimately, as with all PD, the goal was to improve student learning. The TSI Aquatic PD incorporated many of the features considered essential to effective science teacher PD programs. It incorporated active learning, infused pedagogy with content, was a conducted over an extended time (one year), modeled and provided high-quality instructional materials, included multiple opportunities for implementation and reflection, and involved collaborative learning and discussion (e.g., Cohen & Hill, 2001; CRDG, 2012;

Desimone, 2009; Garet et al., 2001; Loucks-Horsley et al., 2010; Penuel et al., 2007; Scher & O'Reilly, 2009; Wei et al., 2009; Yoon et al., 2007).

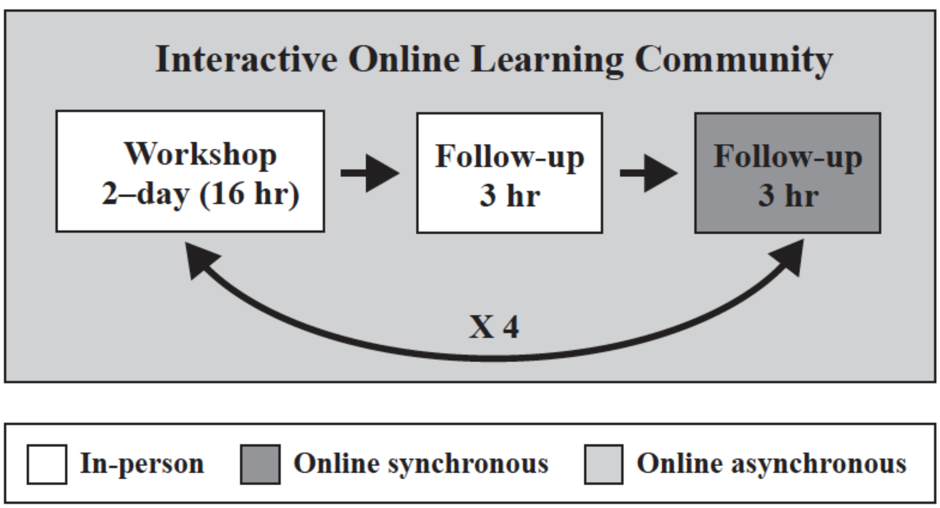
The importance of skilled practitioners in the delivery of PD programs has gotten increasing attention (CRDG, 2012; Lydon & King, 2009; Lynch et al., 2019; Shaw et al., 2018). A review of PD programs by Kennedy found that many of the more effective ones "were offered by individuals or groups who had long histories of working with teachers, were very familiar with teachers and with the problems they face, and based their programs on their own personal experience and expertise" (2016, p. 973). PD facilitators are also responsible for establishing a supportive environment where teachers feel safe to be vulnerable in their learning (Darling-Hammond et al., 2020; Podolsky et al., 2019). The facilitators of the PD under study had extensive expertise in the subject area and were well-trained in TSI pedagogy. Each of the four facilitators had graduate-level degrees in science, three had extensive experience providing science PD, and two were former secondary science teachers.

PD Design

The TSI Aquatic PD intervention was delivered through an iterative, modular cycle of teacher workshops and classroom implementation. PD material was organized into four thematic modules around (a) physical, (b) chemical, (c) biological, and (d) ecological aquatic science. Thus, the PD was designed to help the teachers navigate the massive, interconnected, three-dimensional ocean system by building foundational knowledge that spanned from water's molecular structure to studying large-scale patterns in marine environments. Each module consisted of a two-day workshop (16 hours), in-person follow-up (3 hours), and an online synchronous follow-up (2 hours) where the teachers shared their implementation of PD activities including successes, modifications, and considerations for the future. The synchronous workshops were embedded in an asynchronous online learning community (Figure 1.1). Modules were spread over the course of an academic year, resulting in an immersive inquiry experience with sustained interaction—but in an accessible format that allowed the scaffolding of content and pedagogy. The PD totaled 87 face-to-face contact hours (21 per module, plus a 3-hour introductory meeting); asynchronous participation in the online learning community involved at least an additional eight hours of interaction with PD materials as estimated by project personnel. (For a complete description of the TSI Aquatic PD structure, see Duncan Seraphin, 2014).

Figure 1.1

The TSI Aquatic PD Project Structure



Note. The TSI Aquatic PD project structure encompassed four modules of in-person, online synchronous, and online asynchronous components.

A key aspect of the TSI Aquatic PD was that facilitators modeled and engaged the teachers in the TSI inquiry approach as students. The PD provided the teachers the opportunity to dive much more deeply into the subject matter than that which bound the lessons, situating the content within broader core ideas, and allowed time for in-depth discussions of the pedagogy employed during the activities.

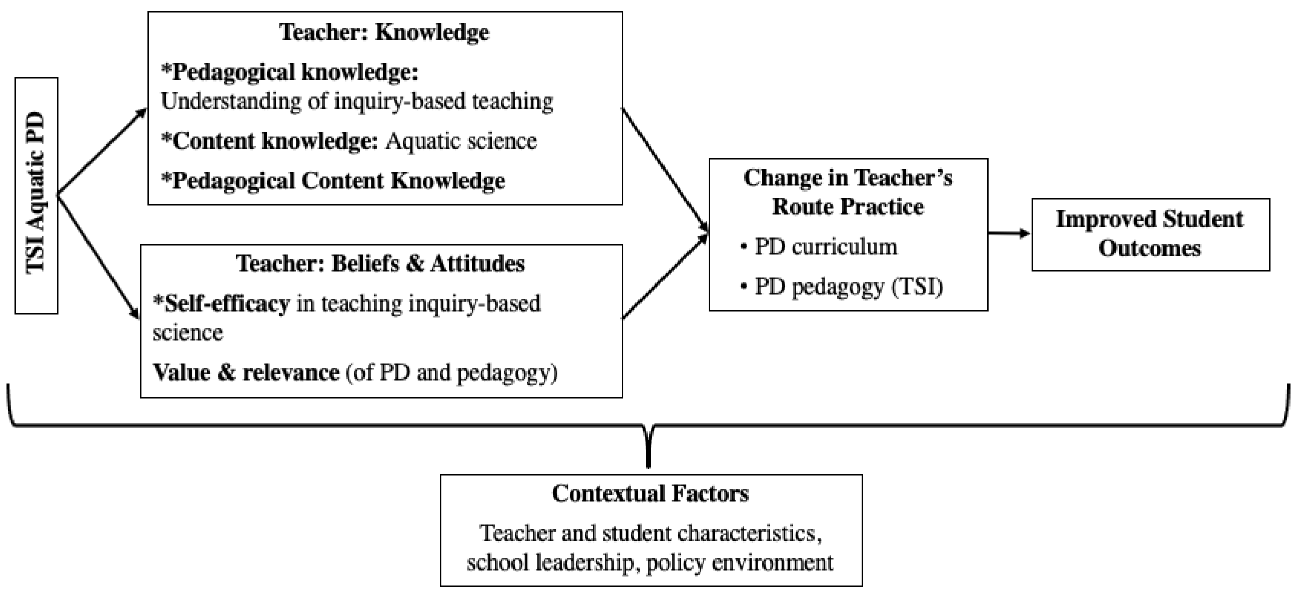
The teachers were required to implement a minimum of three activities from each of the four modules, for a total of at least 12 activities over the course of the project. The teachers were given the supplies to implement these activities. The activities were part of *Exploring Our Fluid Earth* (exploringourfluidearth.org), an inquiry-based curriculum founded on TSI pedagogical principles. The *Exploring Our Fluid Earth* website housed the TSI Aquatic PD online learning community, which the teachers were required to post and interact on as part of the PD. In addition to the facilitators modeling and explicitly teaching TSI pedagogy, the teachers used the pedagogy to plan and reflect on activities and were required to explicitly teach the pedagogy to their students. Requiring implementation of PD activities provided the teachers an opportunity to employ the content and pedagogy productively, develop confidence with their delivery, assess student responses, and share this information with their peers, all of which are important for advancing teachers' understanding and influencing their beliefs about the value of the innovation (Abrami et al., 2004; Deglau & O'Sullivan, 2006; Gorozidis & Papaioannou, 2014; Guskey, 2002b).

Theory of Change

The theory of change for the TSI Aquatic PD is represented in Figure 1.2. Though this figure simplifies a complex process and is not all-inclusive, it represents a current understanding of the literature and is largely similar to other models of effective PD (e.g., CRDG, 2020; Desimone, 2009; Guskey, 2002b; Kennedy, 2016; Loucks-Horsley et al., 2010; Luft & Hewson, 2014; Lynch et al., 2019). Several studies have provided empirical evidence for the relationships among these features (e.g., Desimone & Garet, 2015; Garet et al., 2001, Kennedy, 2016). The TSI Aquatic PD theory of change includes teacher knowledge and beliefs, classroom practice or behavior, the context or environment, and student outcomes. The arrows represent the reciprocal relationships among these variables in mediating the effects of the PD on students' achievement. While the connections among these constructs are not unidirectional, the intended outcome of the PD is for changes in the teacher behavior to positively affect student learning. The quantitative constructs assessed in this follow-up study are denoted by stars in Figure 1.2; they emphasize the teachers' knowledge and self-efficacy. All other variables were assessed through the follow-up teacher interviews.

Figure 1.2

TSI Aquatic PD Theory of Change



Note. Stars denote the quantitative constructs assessed in the follow-up study.

Evaluation Results

While this study focused on long-term teacher outcomes, it is important to situate these results in the context of the teacher and student outcomes found pre- to post-PD. The PD was conducted statewide in Hawaiʻi over five cohorts of secondary teachers ($N = 63$). An evaluation of the teachers in the final two cohorts ($N = 28$) indicated there was a statistically significant effect of the PD on the teachers' self-efficacy, content knowledge, and inquiry-based teaching knowledge and on their students' content knowledge and understanding of the nature of science (Duncan Seraphin, 2014; Duncan Seraphin et al., 2017). In Chapter 3 I go into more detail about the instruments used to assess the teachers and pre-post PD outcomes as they constitute the comparison time points for this study.

The students whose teachers implemented the PD activities with higher levels of fidelity, based on constructs outlined in Dusenbury et al. (2003), showed greater improvements in their content knowledge and understanding of the nature of science (Duncan Seraphin et al., 2017). High-school students appeared to learn more about the nature of science if their teachers had lower pre-PD knowledge of inquiry-based teaching, and students of the teachers who had attended less prior science PD benefited more (Duncan Seraphin et al., 2017). The evidence of student learning, however nuanced, found in the original PD evaluation is important as many PDs do not include student measures. As other studies have shown that it can take time for new interventions to influence student achievement, student gains are not always captured in post-PD measurements (Heller et al., 2012; Kennedy, 2016). As the PD under study was known to affect student gains over the course of a year, I postulated that *sustained* changes in the teachers' practice would result in sustained student learning gains. Overall, the quantitative findings suggested that the PD minimized gaps due to inexperience and lack of understanding of inquiry-based science teaching (Duncan Seraphin et al., 2017).

The teachers were interviewed immediately following the PD. The teachers perceived the PD to be valuable and relevant to their teaching practice, especially the science content and community building features (Duncan Seraphin, 2014). They reported success in teaching the PD activities and improving in the quality of their implementation of TSI pedagogical aspects over the course of the project. However, there were divergent opinions among the teachers about the value of the TSI pedagogy—aspects of which were reported as difficult to bring into their classrooms (Duncan Seraphin, 2014).

Limitations

Although the PD that is the focus of this research can be considered an exemplary project on a number of measures, it also had some significant drawbacks. These flaws include that it was a single project, with a small sample size, and did not include a comparison group. This limits the degree to which the quantitative findings can be generalized. Growth as measured by scores on PD instruments pre-to-post are already difficult to directly attribute to interventions, so changes in scores on instruments years later, without a comparison group, are even more tenuously tied to the PD under study. Additional limitations related to the methods are discussed in Chapter 3. In the following section I focus on the overarching limitations of selection bias and the reliance on self-reports—decisions that were made as part of the original PD recruitment and evaluation that affected this study.

Selection Bias

One threat to internal validity is that the participants volunteered for this study as well as the original PD intervention—thus there is likely a self-selection bias in effect for teacher characteristics that may affect PD findings (Desimone, 2009). Because the teachers applied to participate, they represent a more motivated group of teachers who may be primed to reflect, adjust, and improve their practice (Kennedy, 2016). Measured instructional growth may be related to the general growth orientation of this population rather than the PD itself. This is particularly true for this PD as it was not tied to school initiatives. Thus the teachers enrolled in the PD likely because they were interested in aquatic science and/or inquiry teaching; they probably do not reflect a representative sample of all of the secondary science teachers in Hawaiʻi. Rather, the teacher population in this study represents an exemplar caste, a best-case scenario. However, it is worth noting that because the PD was not top-down or coercive, the teachers did not *have* to implement PD activities or change their practices after the PD ended. Thus, this study is a glimpse into the long-term effect of teachers' implementation of *voluntary* PD.

Reliance on Self-report

Observation protocols allow researchers to examine and evaluate how teachers translate their internal representations of inquiry in their classrooms (Lee et al., 2016). Observations are generally considered the gold standard of measurement for teacher practice and allow for concurrent validation of teachers' claims on questionnaires or in interviews. However, they require extensive resources and the practices being studied must be typical enough to occur in most instructional periods (Desimone et al., 2002). Observations can also be perceived as intimidating by participants and may lead the teachers under scrutiny to take more time to prepare for the observed activities than usual (a.k.a., the Hawthorne effect). Therefore, much PD program research, including this study, relies on teacher self-reported measures (e.g., Garet et al., 2001).

An obvious limitation of self-report measures is the likelihood of potential discrepancies between what people say they do and what they actually do, which may result in variables that reflect social desirability in addition to actual behavior. A number of studies have found that teacher-level self-report variables are likely biased (e.g., Hansen & McNeal, 1999; McFadden, 2019; Scher & O'Reilly, 2009). For example, some studies have shown evidence of discrepancies between teachers' perceptions of their practice, as indicated in interviews and on questionnaires, and their instructional practice as evidenced in classroom observations (Bartos & Lederman, 2014; Hill & Erickson, 2019; Lakin & Wallace, 2015; Lee et al., 2004). Teachers may overestimate their use of instructional strategies due to subject-expectancy effects. If teachers find a PD valuable, they may report their behavior changed, or be particularly attuned to small changes, because they think it was supposed to change. In contrast, other researchers have found evidence that teacher self-reports are correlated with practice (Desimone, 2009; Smithson & Porter, 1994), particularly when asked about current practices (Porter et al., 1993). Thus, there is a relationship between self-reported classroom practices and quantitatively measured knowledge and beliefs, but its exact nature is disputed (Crawford, 2007; Richardson et al., 1991; Yang et al., 2020b). Self-reports of practice are generally understood to be useful, but should be interpreted with caution. Questionnaires continue to be an important assessment tool because, compared to other methods, they are cost-effective and allow for easier comparisons between participants and analysis of trends over time. In this dissertation, I recognize the inherent biases of self-report measures, but nonetheless rely on them to make claims about teachers' levels on constructs the self-report instruments were intended to measure.

Summary and Organization of the Study

The PD under study adhered closely to a number of best PD practices and was successful based on post-PD evaluation findings. Because this PD can be considered an exemplary outlier compared to the PD most teachers routinely experience, the long-term results are indicative of the persistence of constructs and practices of high-quality PD. PD that is less well-designed and occurs over a shorter duration would likely have more marginal results.

This dissertation is organized into five chapters, a reference list, and appendices. This chapter, Chapter 1, introduces the study and provides an overview of the research, including a problem statement and the significance of the study. Chapter 2 presents a literature review that situates this study in the greater research context and concludes with the research questions. In Chapter 3, I provide a detailed description of the research methods, and in Chapter 4, I present the research findings. In the last chapter, Chapter 5, I discuss the significance of the research findings and raise some questions for future research.

CHAPTER 2

REVIEW OF THE LITERATURE

Literature in the areas of inquiry and PD is vast. In the following review I narrow my focus to four areas relevant to this study: (a) constructivism—the historical roots of inquiry; (b) the TSI framework, the pedagogical foundation of the PD under study, and a justification for the choice of PD content (aquatic science); (c) the components of effective PD; and (d) Bandura's Theory of Reciprocal Determination, the guiding theoretical framework for the analysis of data and the interpretation of results. In this last section I focus on the personal and environmental factors that affect implementation of PD practices as related to this study. I end this chapter with my research questions.

Constructivism

The pedagogy utilized in the PD under study, TSI, has its roots in constructivism. Although not a pedagogical theory (Airasian & Walsh, 1997), constructivism has core assumptions that have important implications for instruction and curriculum design, including that of PD.

Constructivism is both a theory of knowledge and learning; it is an epistemological perspective on the nature of knowledge as well as an explanation for how knowledge is acquired (Airasian & Walsh, 1997). The theory of learning contends that individuals construct their own understandings and meanings through interactions with their surroundings. Constructivism highlights the importance of context in the acquisition and refinement of skills and knowledge (Schunk, 2012). New knowledge is created by interactions between learners' preconceptions and beliefs, which have been derived from prior experiences, and new ideas and situations. In this way, constructivism describes how one attains, develops, and uses cognitive processes (Bredo, 1997).

Historical Roots of Constructivism

The idea that learners construct their own understandings has deep historical roots (e.g., Socrates in ancient Greece). However, constructivism as an educational theory was not formalized until the 20th century; the works of Jean Piaget, Lev Vygotsky, and John Dewey heavily influenced its development. Although these theorists have been situated under the modern constructivist umbrella post-hoc, their ideas are considered foundational because they helped to shift perspectives from a deficit model of education, where knowledge was transmitted from teacher to student and passively accumulated, to one in which knowledge is the result of active learning and engagement. Their legacies have diverged and converged over time as subsequent researchers have built upon their ideas. This complex development has contributed to the difficulty in developing a concise description of constructivist theory. Different constructivist schools differ in the extent to which they ascribe the construction of knowledge to social and environmental interactions and the extent to which they think the mental models that learners develop reflect reality (e.g., Bredo, 1997; Schunk, 2012). That so many of Piaget's, Vygotsky's, and Dewey's constructivist approaches to education are as relevant today as they were nearly a century ago is as much a testament to the abilities of these forward-thinking individuals as it is a reflection of the challenges associated with enacting change in education.

Piaget. Piaget proposed that cognitive development encompasses four principal factors: biological maturation, experience with the natural environment, experience with the social environment, and "equilibrium" (Schunk, 2012). Equilibrium refers to the need for people to maintain an optimal state of consistency between their internal cognitive structures and the external environment. When learners encounter information that does not align with their current worldview, the process of equilibrium resolves this mental conflict through assimilation, by fitting new information into existing cognitive structures, and/or accommodation, by changing internal structures to align with the new ideas (Duncan, 1995). Thus, equilibrium is the central, motivating force behind cognitive development and reflects the outcomes of cognitive dissonance. Piaget's theory forms the basis of cognitive constructivism because he proposed that information is not passively received, but internally processed in accordance with current mental structures and capacities.

Vygotsky. Vygotsky, one of the founders of social constructivism, stressed the relationships between interpersonal, cultural-historical, and individual factors in the construction of knowledge (Schunk, 2012). Social interactions are considered critical for learning because they stimulate cognitive development—learning does not occur in isolation. Psychological tools, such as language, mediate and facilitate learning and characterize it as a joint activity—both the learner and teacher construct knowledge. Because knowledge is socially constructed, learning is situated in specific contexts and differs among individuals. Vygotsky is well known for positing the idea of the "zone of proximal development" (1978), the difference between what an individual can do on his or her own and what they can do with assistance from others. This emphasizes the importance of interactions between learners and more competent individuals, including both peers and teachers, and the role of prior knowledge in cognitive development.

Dewey. Dewey spearheaded the progressive reform movement in education at the turn of the 20th century. Although today many scholars consider his ideas to align with constructivist principles, Dewey has been primarily associated with the philosophy of pragmatism (Pulliam & Van Patten, 2007), which prioritizes action and problem solving over philosophical debates on the nature of knowledge (Patton, 2014). Dewey thought that education should be student-centered. He advocated for purposefully building on students' interests to make learning more meaningful; his pedagogical approach involved learners addressing genuine problems through a cycle of scientific inquiry (Pulliam & Van Patten, 2007). Although important, Dewey was clear that student engagement could not, by itself, be equated to learning. The teacher was critical; they structured activities so students could relate new information to prior experiences (Dewey, 1938). Navigating the balance between addressing content knowledge and embedding learning in meaningful contexts defines Dewey's educational philosophy (Windschitl, 2002).

Modern Interpretations

Constructivism's learner-centered assumptions include "the belief that all learners construct their own knowledge and understandings from their experiences, that this knowledge development is incremental, and that the knowledge we hold in common is developed and clarified through interactions with others" (CRDG, 2001). The following are instructional principles associated with constructivist learning environments (e.g., Duncan Seraphin et al., 2012; Hmelo-Silver et al., 2007; Prawat, 1992; Richardson, 2003).

Active Learning. Students learn by doing. Constructivist classrooms engage students in activities that facilitate discovery and sense-making. Learning is an active process where students are encouraged to observe, predict, analyze, and interpret information and phenomena. Learning takes place in authentic, real-world environments and is organized around meaningful problems and goals.

Build on Prior Knowledge. Learning involves building on or modifying existing ideas. Preconceptions shape learning because they filter everything a learner experiences.

Constructivist teachers are aware of current student conceptions and they connect new information to prior experiences because "new conceptions are only learnt and retained if they can be fitted or joined to already existing knowledge" (Sewell, 2002, p. 24).

Scaffold Instruction. Constructivist classrooms carefully sequence activities to allow students to construct increasingly complex knowledge structures. Big ideas are studied from multiple perspectives and time is allowed for students to reflect on and clarify their own ideas. These learning experiences support student competence by connecting concepts in ways that allow them to develop sophisticated frameworks, which encourages knowledge use and transfer.

Make Learning Relevant. The construction of knowledge requires effort and persistence. This is partially because of the need to create dissatisfaction with existing cognitive conceptions before new knowledge can be deeply integrated. Constructivist classrooms build on student curiosities and interests and situate learning in relevant contexts to make it more meaningful.

Develop Self-regulation. Students develop their own knowledge. Thus, in constructivist classrooms, students are taught metacognitive skills. Metacognitive processing refers to the ability to understand, monitor, evaluate, and ultimately control one's cognitive processes, skills, and learning environment (Brown, 1987; Flavell, 1976; Jacobs & Paris, 1987; NRC, 2001). Thus, students learn to set goals as well as how to monitor and evaluate their learning. This helps them to develop ownership of their learning and become actively involved in constructing their own understandings.

Collaborative Learning. Learning is a social activity that involves engaging with others. Students in constructivist classrooms work in small groups to discuss and share their ideas and engage in class discussions where they develop consensus about these ideas. A diversity of perspectives and understandings is encouraged and respected.

Teachers as Facilitators. In constructivist classrooms, teachers serve as guides and facilitators of learning. Teachers structure activities in ways that allow learners to become actively involved in constructing their own interpretations. In order to do this effectively, teachers need to understand what students are thinking; they do this through questioning and other formative assessment strategies. Instructors tailor their teaching in response to student perspectives, challenging student misconceptions or supporting weak reasoning. Listening to their students' points-of-view also allows teachers to capitalize on student interests and pose problems of emerging relevance. Teacher feedback helps students reflect on and revise their ideas. In order for these interactions to occur, teachers need to create a safe, nonjudgmental classroom culture that encourages students to express their ideas, ask questions, and make mistakes. Although the teacher should know where the class is going, have an idea where students started, and monitor progress towards learning goals, the actual course will depend on factors that may not be readily accounted for in planning. Building conceptual understanding does not follow a linear path.

Constructivism and Inquiry. Although educators were implementing constructivist strategies before the term was coined, these efforts were concentrated in pockets of reform. However, in the early 1990s, constructivist-informed teaching and learning gained widespread interest in the education community (e.g., Richardson, 2003). This renewed interest was due in part to the concurrent standards-based education movement, which drew attention to the continuing gap between education in practice and the vision outlined in reform documents. But teacher adoption of reform practices is difficult (e.g., Windschitl, 2002). Constructivist instructional strategies appear deceptively simple to put into practice. There are a number of barriers to the implementation of constructivist strategies—including curricula that cover too many topics, assessments that emphasize memorization over deeper understanding and reasoning skills, and the lack of a sustained, supportive community to help guide teachers through the often difficult process of not only changing conceptions about teaching and learning but then translating these ideas into meaningful changes in practice. This has often led to the incomplete application of constructivist strategies. Rather than allowing students to construct their own understandings, many teachers still have the underlying positivist belief that they are responsible for transmitting knowledge about the world to their students. Thus, although they may engage students in activities, relate information to the real-world, and have students work in groups, they still view their primary job as organizing knowledge about the world in the most rational way possible and presenting it to their students rather than focusing on student understanding (Prawat, 1992; Smith, 2020).

In part to clarify the implementation of constructivist strategies in science classrooms, science education moved away from using the term constructivism to emphasizing the term inquiry in standards documents in the 1980s and 90s (e.g., AAAS, 1990; 1993; NRC, 1996). The term 'inquiry' in science education, which can be thought of as a branch of constructivism, refers not only to the "*abilities* students should develop to design and conduct scientific investigations [emphasis added]," but also to the "*understandings* they should gain about the nature of scientific inquiry [emphasis added]" (NRC, 2000, p. xv). Teaching science through inquiry emphasizes the role of students in developing their own ideas, through concrete experiences, to form complex understandings of scientific concepts and the scientific process.

However, knowledge of inquiry instruction does not necessarily translate into inquiry-based teaching (Abd-El-Khalick et al., 1998). Just like with constructivism, the term ‘inquiry’ came to have varied meanings. Further, it is difficult for teachers not familiar with the real-world practice of science to successfully facilitate inquiry teaching (NRC, 2000; Penuel & Fishman, 2012; Wee et al., 2007; Zembal-Saul et al., 2002). This may be because the concept of scientific inquiry is so closely connected to the work of professional scientists. Although there are universal consistencies across scientific disciplines, to those less familiar with the process of science the range of fields, approaches, and methodologies can be difficult to distill to the essential components of inquiry. This has led to oversimplifications. For example, teachers often equate inquiry with hands-on instruction (Capps et al., 2012; Rankin, 2000; Seung et al., 2014). Teachers need good role models for what inquiry-based teaching looks like, especially those who have low science teaching efficacy beliefs and limited pedagogical content knowledge in implementing inquiry-based instruction (Fitzgerald et al., 2019).

In 2012 the National Research Council replaced the term ‘scientific inquiry’ with ‘scientific practice,’ as the construct of inquiry had become fraught by misunderstandings and miscommunication (NRC, 2012). The science and engineering practices in the Next Generation Science Standards (NGSS) present a more crystallized vision of what students should do while engaged in science and can serve as pedagogical approaches for teachers (Halawa et al., 2020; NGSS Lead States, 2013). Although the NGSS eschews both the terms inquiry and constructivism, and other researchers continue to use phrases like ‘reform-based,’ the differences in terminology obscure universal consistencies. Most researchers and practitioners ascribe to some version of constructivism, even if not always made explicit in writing, because “rather than talk about how knowledge is acquired, they speak of how it is constructed" (Schunk, 2012, pg. 139).

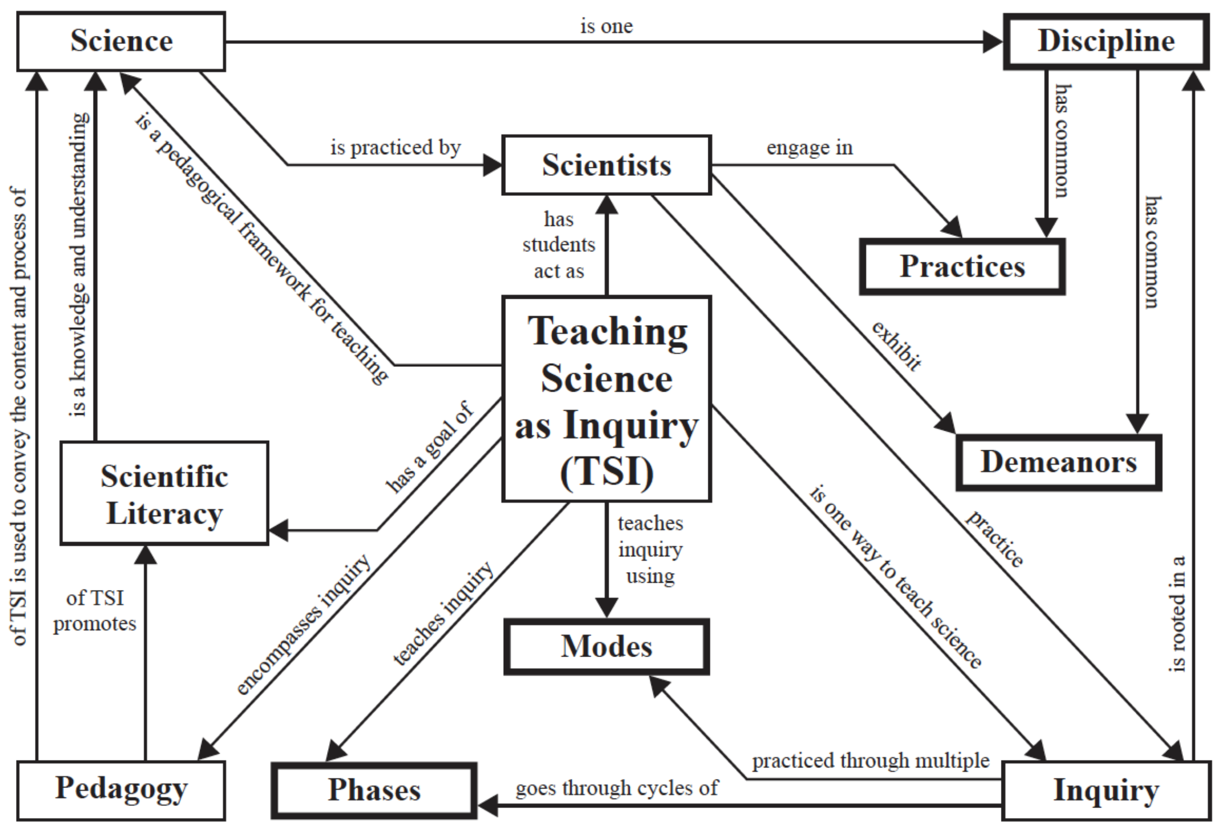
TSI Pedagogical Framework

The foundations of TSI are embedded in early constructivist ideas. TSI was developed at the University of Hawaiʻi at Mānoa's College of Education Curriculum Research & Development Group (CRDG; Pottenger, 2007; Pottenger & Berg, 2006; Pottenger et al., 2007). In addition to the ideas advocated by educational theorists like Bruner (e.g., 1966), Vygotsky (e.g., 1978), and Dewey (e.g., 1938) TSI is based on well-known instructional models (e.g., Bybee et al., 2006) and the disciplines of knowledge (King & Brownell, 1966). A central premise of TSI is that learning, including that done at the professional level through PD, is best accomplished through authentic application of knowledge and skills. This means teaching science the way science is practiced, with a focus on the nature of science and the collaborative nature of scientific inquiry.

TSI was developed to address one of the difficulties surrounding inquiry-based pedagogy—that there is a range of inquiry definitions (Abd-El-Khalick et al., 2004; Settlage, 2007) and a lack of models of inquiry-based instruction. The TSI framework makes inquiry accessible by explicitly defining components of good inquiry teaching, thus making the practice of inquiry-based teaching more comprehensible. Figure 2.1, a visual representation of the TSI framework, was used to introduce and guide teachers though the components in the PD. Aspects of the figure, and relationships, were introduced piece by piece as the components were introduced, modeled, and utilized over the course of the PD.

Figure 2.1

How Components of the TSI Framework Relate to Scientific Inquiry



Note. TSI components and their relationships to science and the inquiry process. TSI components are in bolded boxes. This figure was used with the teachers in the PD, components and their relationships were introduced gradually over the course of the PD, building to this representation. Modified from Duncan Seraphin et al. (2014).

In a TSI classroom, teachers and students are linked as part of a disciplinary community. Students are expected to act as scientists by exhibiting the demeanors of scientists and engaging in the practices of science (Duncan Seraphin & Baumgartner, 2010; Duncan Seraphin et al., 2017). Equal importance is placed on learning about the scientific process (e.g., what scientists do) and the scientific content (e.g., what scientists know). Deep understanding of both scientific content and process helps students build scientific habits of mind (Handler & Duncan, 2006).

Two of the main components of the TSI framework are the *phases*, a cycle of learning and instruction that illustrates how the inquiry process is multidirectional (Figure 2.2), and the *modes*, which emphasize the flexibility of science and the variety of ways to engage in scientific inquiry (Table 2.1). Like other learning cycles, the TSI phases are represented in a circular model (e.g., Bybee et al., 2006). Unlike other learning cycles, TSI rejects the existence of a fixed, step-by-step progression through the cycle. The TSI phase diagram depicts this by providing an area for each phase to connect with each of the other phases, illustrating the interconnected nature of inquiry. The TSI modes of inquiry lay out a spectrum of inquiry practice, from guided to open-ended inquiry. To build students' understanding of the nature of science, teachers are encouraged to use multiple modes of inquiry in their lessons. The TSI PD also emphasized the practice of metacognition (Brown, 1987; Flavell, 1976; Jacobs & Paris, 1987; NRC, 2001). Fostering student self-regulation and intentional learning is a difficult task, and metacognitive skills are difficult to report and assess because they often develop in the absence of conscious reflection (Schraw et al., 2006). However, studies have documented that the explicit teaching of metacognitive strategies can improve student learning and be an effective PD practice (Darling-Hammond et al., 2020). In summary, the TSI phases and modes give teachers and their students a common language to describe and communicate about the scientific process to make their thought processes more explicit (Duncan Seraphin et al., 2013).

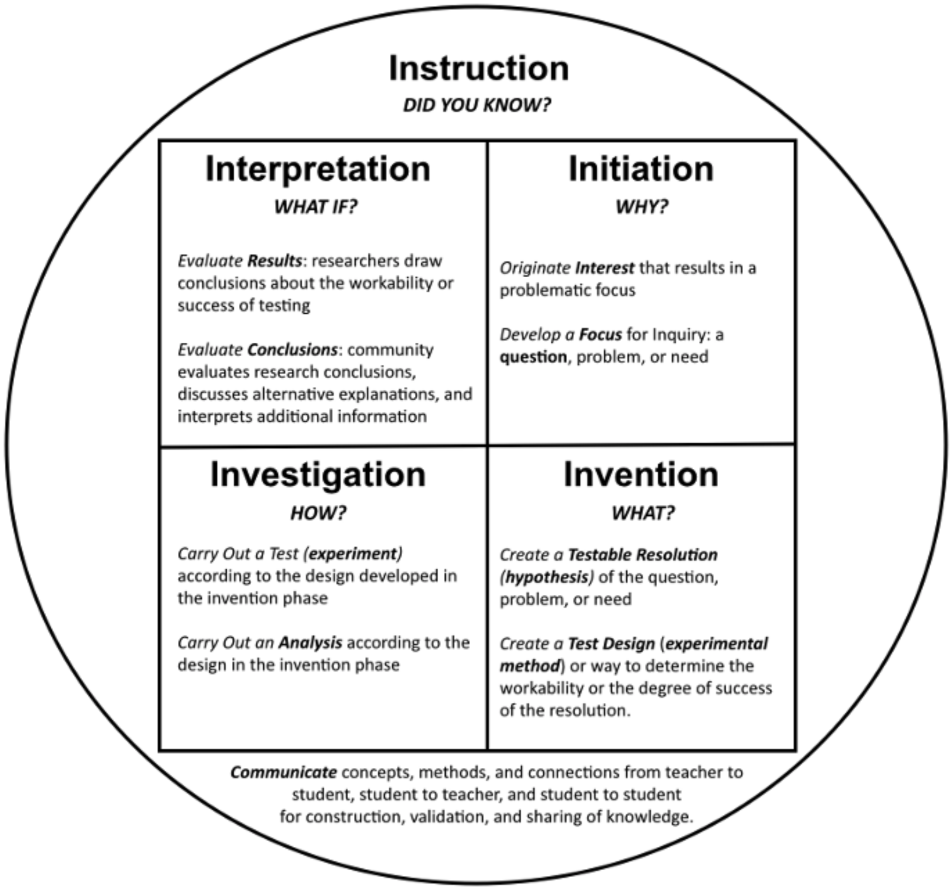
TSI and NGSS

The components of TSI are directly aligned with the NGSS (NGSS Lead States, 2013). The emphasis on *initiation* in TSI emphases connecting to students' interests and experiences and how scientific investigations begin with a question. Its corollary in NGSS is beginning investigations with a phenomenon around which to structure an inquiry (NGSS Lead States, 2013; Penuel et al., 2020). The prominence of the *instruction* phase highlights the importance of social interaction and discourse in science teaching and learning and is a reminder that teachers must plan for interaction, including the use of effective questioning strategies, in their lessons (Reiser, 2013). While the teacher acts as research director in a TSI classroom, knowledge is socially constructed; both teachers and students can hold and communicate knowledge. The fluidity of the TSI *phases* accentuates that there is no single sequence of steps in a scientific investigation and serves to remind teachers that scientific knowledge, while reliable and durable, changes. The NGSS refer to this aspect of the nature of science as "scientific knowledge is open to revision in light of new evidence" (NGSS Lead States, 2013, Appendix H, p. 5).

The TSI phase diagram serves to remind teachers that *investigation*, in all its forms, is only one of the phases; equal emphasis is placed on asking questions, developing explanations, and obtaining and communicating information. Teachers tend to overemphasize investigations (e.g., collecting data), as opposed to asking questions and developing explanations for phenomena. Using the phase diagram as a planning tool may help address this issue. In the PD the teachers were also taught to use the phases as a metacognitive tool to plan and reflect on their path through the process of inquiry. The importance of metacognition to the conceptual shifts in the NGSS is gaining prominence (e.g., Monroe & Brunsell, 2019). The explicit incorporation in TSI of attributes of the *discipline* of science aligns with research that has shown that understanding the structure of a discipline, including how it has a "code of patterns" and other organizational regularities, can help people learn more efficiently (Darling-Hammond et al., 2020, p. 115). The TSI *demeanors* of scientists draw attention to how scientific knowledge is influenced by scientists' attitudes and beliefs and that creativity plays an important role in knowledge generation. The NGSS refer to this as “science is a human endeavor” (NGSS Lead States, 2013, Appendix H, p. 5).

Figure 2.2

The TSI Phase Diagram



Note. The TSI phases illustrate the fluid nature of scientific inquiry; the phases are connected but without prescribed 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 each of the other phases. From Duncan Seraphin et al. (2017).

Table 2.1

The Modes of Inquiry Addressed in TSI

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

Note. From Duncan Seraphin et al. (2017).

The TSI modes emphasize that scientists employ an array of investigatory methods. *Experimentation* is just one of the ten modes. The modes counteract the chronic misunderstanding that NGSS calls for "hands-on" experiences (Lin, 2020), tunnel vision that misrepresents the full range of activities practiced by scientists in the real world (NGSS Lead States, 2013; Penuel et al., 2007; Smith, 2020). The *authoritative knowledge* mode elevates the role of the instructor, in part to grant teachers permission to, at the appropriate time, explain science concepts to students. This addresses the misconception that the student-centered instruction exposed in the NGSS equates to open-ended inquiry (Fitzgerald et al., 2019; Halawa et al. 2020). Appropriately integrated explicit instruction enhances inquiry-based learning (Darling-Hammond et al., 2020).

The NGSS represent a call for science education reform. Previous calls for reform, including the progressive movement championed by Dewey and the curriculum development movements of the 1950s and 60s championed by Bruner (e.g., 1966), were never systematically implemented, and had little overall effect on teaching practices (Rudolph, 2014). Although the NGSS are sophisticated standards that go beyond establishing what students should learn by outlining a conceptual shift in teaching, they are complex and their enaction is stymied as they are written and thus static (Friedrichsen & Barnett, 2018; Fulmer et al., 2018). Without PD to explain the NGSS and help translate them into teacher practice, they may go the way of other educational reforms and not be implemented in schools in a meaningful way. Connecting the NGSS with nature of science concepts may help teachers develop more coherent conceptions of the scientific process and enhance their implementation of the standards (Bartos & Lederman, 2014). The PD that is the focus of this study was implemented before NGSS was rolled out in Hawaiʻi. However, TSI strategies may have supported alignment with the NGSS.

Choice of Content—Aquatic Science

A main tenet of TSI is that inquiry instruction is more salient when taught through, and embedded in, content for context. On a national level, aquatic sciences have been marginalized in K–12 education, largely due to their absence in the NRC's 1996 *National Science Education Standards* (Strang, 2008). This has resulted in an underrepresentation of the ocean in curriculum nationally (Hoffman & Barstow, 2007), a finding that is mirrored internationally (Fauville et al., 2018), and has resulted in a populace that is largely ocean illiterate (NRC, 2007; Steel et al., 2005; Steel, 2006; Belden et al., 1999). Although one reason for the deficit of ocean content in K–12 education may be due to the difficulty in accessing the ocean, even in Hawaiʻi, an island state, only about 7% of high school students take courses in marine science (L. J. Kaupp, personal communication, June 15, 2021). This is unfortunate considering how engaging the ocean has been shown to be as a context for learning (Lambert, 2001).

Critical understanding of the ocean is essential to our future and thus to the broader goal of enhancing scientific literacy (Strang et al., 2007). The ocean affects every aspect of human life— from climate and food security to tourism and economic and social stability. Although the NGSS have made strides to rectify this aquatic content deficiency (NGSS Lead States, 2013), historical inattention has "resulted in a generation of Americans largely ignorant of the importance of the ocean" (Carley, n.d.); this generation gap includes many current teachers (Boubonari et al., 2013; Plankis & Marrero, 2010).

The functioning of the marine environment is rooted in the interplay between physical, chemical, biological, ecological, and social processes (Larkin et al., 2015). Thus, using aquatic science as a cohesive content umbrella in a PD allowed not only for the coverage of underrepresented and poorly understood concepts, but also emphasized the interdisciplinary nature of science and made the PD accessible to a wide variety of educators. Although studies on effective PD recommend focusing on a single content area (Desimone et al., 2002; Garet et al., 2001; Penuel et al., 2007; Yoon et al., 2007); the modular design of the TSI Aquatic PD allowed for each of these areas to be focused on individually, as well as for more broader connections to be made between disciplines. The NGSS emphasizes more integrated content (NGSS Lead States, 2013), and the importance of learning about connections among the different fields of science in PD was recently ranked highly by 100+ physics education stakeholders (e.g., scientists, teachers, and educational researchers) from 42 countries (Kranjc Horvat et al., 2021). Integrating biological, chemical, geological, and physical characteristics of the ocean can increase students' content knowledge (Lambert, 2006), in part due to the coherent nature of marine science (Lambert, 2001).

Perhaps most importantly, marine science and the ocean has particular relevance to Hawaiʻi. The ocean is meaningfully and substantially connected to the everyday issues and challenges of teachers and students living on an island, and is thus a rich source of relevant phenomena to ground inquiry investigations (NGSS Lead States, 2013). Place-based education connects students directly with their local environment, boosts student achievement, and results in an educated populace able to make better-informed decisions about environmental practices and sustainable resource use (e.g., Lovell et al., 2009). Using aquatic science as a cohesive content umbrella in a PD allowed for the coverage of underrepresented and poorly understood concepts and emphasized the interdisciplinary nature of science. Research on the effectiveness of aquatic science PD has been limited (although, see Goodale & Sakas, 2019 and Wulff & Johannesson, 2019). This study aims to add to the literature on this topic.

Components of Effective PD

One of the main complexities in empirically examining PD is that programs are wildly dissimilar, making it difficult to determine and compare program effectiveness (Koellner & Jacobs, 2015; Wei et al., 2009). However, some general recommendations have emerged: (a) apply and leverage adult learning theory (andragogy, e.g., Knowles, 1978), and (b) include characteristics found to be reflective of effective PD. Of note in designing teacher PD, Knowles (1978) suggested that adult learning environments be experiential (e.g., incorporate active learning) and focused on problems of practice (e.g., problem-centered). One of the difficulties in effecting change in teachers' behavior as a result of PD is that it is hard to improve upon established practices—even if those practices are ineffective. Slavin explained this challenge in a blogpost:

> For example, if as an adult you took a course in tennis or golf or sailing or bridge, you probably noticed that you learned very rapidly, retained what you learned, and quickly improved your performance in that new skill. Contrast this with a course on dieting or parenting. The problem with improving your eating or parenting is that you already know very well how to eat, and if you already have kids, you know how to parent. You could probably stand some improvement in these areas, which is why you took the course, but no matter how motivated you are to improve, over time you are likely to fall back on well-established routines, or even bad habits. (Slavin, 2019)

Over time teachers develop routines and settle into patterns. If teachers run into difficulties enacting PD strategies they may settle back into their routines—*unless* the strategy solves a problem of practice and/or results in considerable student learning gains (Barron & Darling-Hammond, 2008; Darling-Hammond et al., 2020; Guskey, 2000; Klingner et al., 1999; Windschitl et al., 2012; although see Copur-Gencturk & Papakonstantinou, 2016). Adult learning theory also emphasizes that teachers have prior knowledge and expertise in addition to differing values, identities, and interests that they bring to PD (Sutton & Levinson, 2001). PD needs to recognize and leverage this experience and knowledge to enhance teacher motivation and buy-in (Loucks-Horsley et al., 1996; Opfer & Pedder, 2011).

In the 1980s and 90s the literature on teacher PD was coalescing as interest in teacher learning intensified. A considerable amount of research began at this time, and still continues, focused on describing the characteristics indicative of "best" PD practices—those associated with transformations in teacher practice which led to student learning gains. Adult learning principles are reflected in these lists of effective PD design features. These features of high-quality, effective PD are widely cited and include: active learning and collaboration; the modeling of effective practice; iterative opportunities for professional experimentation, reflection, and feedback; a content-focus (i.e., utilization of content when teaching pedagogy as opposed to being content-agnostic); sufficient duration and intensity (including sustained support); coherence with school and district initiatives and priorities (practice-based); to use or align with evidence-based curriculum; and a theoretical base of teacher learning and change (e.g., CRDG, 2012; Darling-Hammond et al., 2017; Desimone, 2009; Loucks-Horsley et al., 2010; Penuel et al., 2007; Sancar et al., 2021; Supovitz & Turner, 2000; Timperley et al., 2007; Wei et al., 2009; Yoon et al., 2007). In a review of the literature specifically on inquiry-based PD, Capps et al (2012) found a general alignment with the above features but emphasized also providing authentic inquiry experiences. Thus, the instructional strategies outlined above that guide constructivist learning are hypothesized to apply to students of all ages—children and adults. While the specifics of what each of these characteristics include or encompass varies by researcher, because these principles were generated from extensive literature reviews and meta-analyses, they have become influential. Over the last two decades these factors have become widely adopted as best practices and are frequently cited as core or critical PD features, although typical teacher PD experiences rarely incorporate most of these (Linn et al., 2010).

Criticism of Components

While the recommendations for designing PD to be reflective of adult learning principles are uncontested, research on programs containing some or all of the "key" PD features has yielded disappointing results (e.g., Arens et al., 2012; Bos et al., 2012; Garet et al., 2008; Garet et al., 2011; Givvin & Santagata, 2011) and there is little empirical evidence on the extent to which these features are associated with enhanced teaching and student learning gains (Asterhan & Lefstein, 2020; Desimone et al., 2002; Kennedy, 2016; Kowalski et al., 2020). Indeed, in the PD literature the emergent consensus is on the general lack of clarity about the features of PD that improve educational outcomes (e.g., DeMonte, 2013; Hill et al., 2013; Slavin et al., 2014). In general, critics have argued that these features are surface-level, and do not capture the mechanisms that explain how and why PD is, or is not, effective (Hill et al., 2020; Kennedy, 2016). Recent studies have also contended that the evidence underpinning the generation of these principles is flawed (e.g., Sims & Fletcher-Wood, 2021). Lynch et al. note that this is "not because the studies on which the extensive reviews are based were weak or badly executed studies per se, but because most of them were never designed to test the question of PD effectiveness" (2019).

This is not to say that "key" PD features are wholly ineffective, rather, additional research on these components is needed to determine which are more effective, or best in combination, or work best with different teachers in different settings (Kennedy, 2016; Lynch et al., 2019). Factors unrelated to PD design characteristics can also influence the effectiveness of PD. In 2020, a meta-analysis by Kowalski et al. found that, in addition to design factors, effect size estimates for teacher PD outcomes varied by research design (e.g., sample size, measurement tools), study context, and what aspect(s) of teaching and learning the PD was aiming to address (see also Desimone, 2009). Conceptual issues can also confound measurements—such as lack of clarity in the definition of key constructs, theoretical frameworks, and/or theory of change (e.g., Gore et al., 2017). Most likely, effectiveness depends on how PD features are enacted (Kennedy, 2016).

The literature on PD duration provides a good example of how researchers are taking a second look at a well-cited core PD component. For a long time, PD core feature lists were clear that long-term interventions had a greater impact on teacher practice and thus student learning (e.g., Banilower et al., 2007; Darling-Hammond et al., 2009; 2017; Garet et al., 2001; Scher & O'Reilly, 2009; Supovitz & Turner, 2000; Yoon et al., 2007), and short PD episodes were unlikely to be effective (e.g., Lydon & King, 2009). The hypothesis was that longer duration PD allowed educators time to try and reflect on new strategies, a key PD practice. However, recent reviews by Lynch et al. (2019) and Kennedy (2016) found no evidence of a positive association between PD duration and program impacts. Lynch et al (2019) postulates that this is due to the limited number of studies that meet review criteria, which limit the generalizability of the findings, and the decisions researchers make when they bin and analyze contact hours. There are PD programs that have a relatively short duration but moderate to large effects on student achievement (e.g., Clements & Sarama, 2007). There are also examples of intense PD that showed little or no impacts on teacher practice and/or student learning (e.g., Devlin-Scherer et al., 1998). While PD duration probably does have an effect, more contact hours cannot override poor PD content and strategies (Kennedy, 2016). If the PD is not useful, more hours will not help.

Individual Experiences in PD

Regardless of the nature of the PD, if the teacher exists within a context that prevents adequate translation of what was learned into the classroom, then the PD, and all of the associated expenses, were for naught (Fitzgerald et al., 2019). One of the root causes of unsuccessful attempts to advance inquiry-based teaching is a lack of insight into the teacher characteristics and environmental factors involved in the process of behavior change (Guskey, 1986; Wee et al., 2007). Many measures of change related to PD address teacher variables like attitudes, knowledge, or beliefs without trying to understand the constraints and affordances of teachers' contexts (Allen & Penuel, 2015; Honig, 2006). Furthermore, in a recent review by Hill & Erickson (2019), few studies were found that incorporated teachers' perspectives on program implementation.

> We rarely found teachers' insights into typical barriers to implementation, typical difficulties working with ideas from professional development or instructional materials, and typical reasons in which implementation varied, qualitatively, from what the authors of the interventions intended. Improving program implementation at scale cannot occur without a more nuanced understanding of these issues. (Hill & Erickson, 2019, p. 596)

Educators are active participants in the process of program implementation (Muijs et al., 2014). However, teachers react differently to educational innovations (Van Eekelen et al., 2006). Some embrace innovation, some are resistant to change, and some try new approaches but are dissatisfied and return to their previous practices (Larrivee, 2000). Understanding their adoption and adaptation pathways is crucial to understanding program success, as teacher successes and failures are key to program implementation.

The likelihood of everyone learning something from any given PD opportunity is high—but what people learn is highly individualized based on what they bring into the PD and what happens after they leave (Bell & Gilbert, 1996; Desimone et al., 2002; Franke et al., 1998; Minor et al., 2016). Different teachers have different outcomes. For example, Johnson et al. (2007) found that teachers in a PD all implemented the PD strategies differently based on differences in their beliefs. It stands to reason that if meta-analyses show that PD is largely ineffective overall—that it can be effective for some teachers, but not others, resulting in an average of no effect. To adapt a metaphor from Slavin (2018), consider the population of teachers in PD as constituting three groups. One group of teachers already know the content before the PD. Another group of teachers is not going to learn or apply the PD content—regardless of the training. These two groups are unlikely to show any effects of the PD on content. However, the third group of teachers, those who do not know the content prior to the PD *and* for whom the content is applicable, is the only group likely to demonstrate gains as a result of taking the PD. But this third group of teachers' gains are dampened by the other groups. In other words, average effects may be masking substantial individual variation.

In summary, many variables are involved in whether or not a teacher makes a substantial and enduring shift in their practice. Features or programs that "work" in one setting or with one group of teachers may not work in another (Loucks-Horsley et al., 2010; Short & Hirsh, 2020; Starkey et al., 2009). Evaluating a teacher PD program requires a thorough understanding of the myriad of factors influencing teacher's instructional practices and students' learning.

PD Frameworks

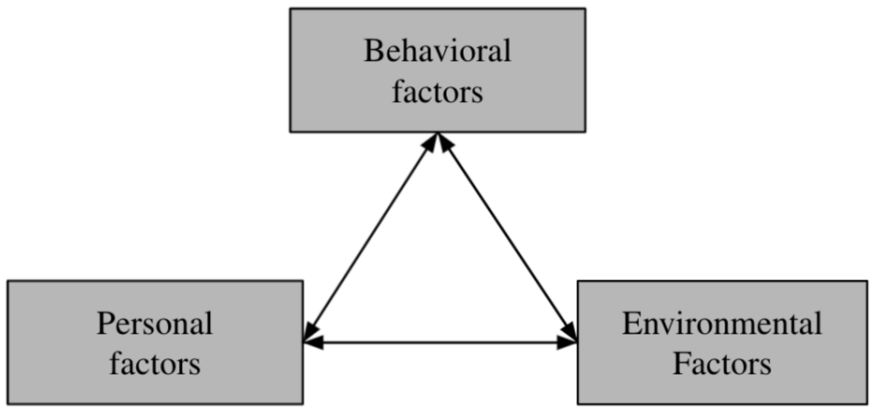
PD designs vary in accordance with resources and the problems of practice they are designed to address. Thus, their hypothesized mechanisms for effecting change vary as well. But most researchers agree that effective PD occurs when there are gains in teachers' knowledge, changes in their beliefs and attitudes, and changes in their practices; a combination that results in improved student outcomes (e.g., Wei et al., 2009). Indeed, these are the core components of most PD theories of change. However, the mechanisms and direction of change varies across models. For example, Desimone's (2009) highly cited model of effective PD indicates PD may increase teacher knowledge and change teacher attitudes and beliefs toward teaching, thereby improving teacher classroom practice, which in turn results in increased student learning and achievement. Thus, change in beliefs leads to change in practices. In other models (e.g., Guskey, 2002b), changes in practice lead to student learning, which then leads to changes in the teachers' beliefs. Though the direction of this connection between beliefs and practice is emphasized differently in different frameworks, they are all simplifying complex systems and relationships. Regardless of the initial steps, the relationship between these variables is reciprocal, reinforcing, and heightened by professional experimentation and reflection (e.g., Clarke & Hollingsworth, 2002; Deglau & O'Sullivan, 2006; Loucks-Horsely et al., 2010; Fishman et al., 2003; Remillard, 2005).

Bandura's Theory of Reciprocal Determination

I utilized Bandura's (Bandura 1977, 1978, 1983a, 1983b, 1989; Bandura & National Institute of Mental Health, 1986) Theory of Reciprocal Determination to explore the findings of this study and situate it within existing literature (Figure 2.3). This triadic model offers a visual representation of the reciprocal interactions among an individual's behavior, personal experiences, and environmental influences. The components of most PD theories of change can be aligned to this model. Personal factors include teacher characteristics like experience, beliefs and attitudes, content and pedagogical knowledge, efficacy levels, and motivation. Environmental factors include the influences of the PD, school and community, the influence of students, and the availability of resources. For example, the more teachers perceive students learn and are engaged (environmental factors), the more likely they are to implement a behavior; this success can enhance their self-efficacy (personal factor, Guskey, 2002b).

Figure 2.3

Bandura's Reciprocal Determination Model



Note. Factors of reciprocal (triadic) determinism in Social Learning Theory (Bandura, 1977).

The focus of this study is on interactions that support or suppress behavior related to the implementation of PD activities and pedagogy, including environmental factors that enhanced or hindered implementation and the cognitive factors that were measured in the original PD evaluation. These personal factors, discussed below, were understanding of inquiry-based teaching (pedagogical knowledge), content knowledge, pedagogical content knowledge, and self-efficacy. As pedagogical knowledge is generally considered to be a subset of pedagogical content knowledge (Shulman, 1986), it is not discussed separately.

Content Knowledge. Several studies have explored the relationship between PD and teacher content knowledge; most of these have found a positive association (e.g., Garet et al., 2001; Heller et al., 2012; Supovitz & Turner, 2000, Yang et al., 2020b). Teachers who have science training (e.g., undergraduate or graduate degrees in science), and thus are likely to have higher levels of content knowledge, are more likely to use inquiry-based teaching strategies (Kolbe & Jorgenson, 2018; Smith et al., 2007). This may be because teachers who have less experience in science have conflicting views of the nature of science. Akerson and Hanuscin (2007) found that teachers who had little experience with the scientific process could view scientific knowledge as both a fixed body of knowledge and as subjective and tentative. As teachers' understanding of inquiry increased by participating in an inquiry-focused PD, they became more aware of the nature of science and incorporated aspects of it into their instruction (Akerson & Hanuscin, 2007).

However, while PD focused on subject-matter content is important, it must be in conjunction with pedagogical instruction, as PD solely focused on content is not linked to student achievement (Herrington et al., 2016; Yang et al., 2018). Similarly, when inquiry knowledge is considered separately, even if teachers view this pedagogical approach favorably, it has not been shown to translate into classroom practice (Abd-El-Khalick et al., 1998). PD programs that focus just on pedagogy have not found a positive effect on student achievement (Pellegrini et al., 2021). PD that addresses both content and pedagogy together are more effective (Duncan Seraphin et al., 2017; Heller et al., 2012; Lynch et al., 2019; Scher & O'Reilly, 2009; Smith et al., 2007), as challenges to PD implementation have been linked to gaps in both content and pedagogical content knowledge (Akuma & Callaghan, 2019; Kyriakides et al., 2017). However, studies evaluating long-term retention of PD gains in content and pedagogical knowledge are rare, and have shown mixed results (Clary et al., 2018; Sandholtz et al., 2016; St. Clair et al., 2020). This study aims to add to this knowledge base.

As teachers with less subject-matter and inquiry experience are more likely to follow the structures and sequences in their textbooks (Gess-Newsome & Lederman, 1993), quality curriculum materials are an important guide, and crutch, to teachers' implementation of educational innovations (McNeill, 2009). However, just supplying resources is insufficient in transforming teacher practice, as teachers often find it difficult to discern strategies for engaging students in scientific inquiries from materials alone (Alozie et al., 2010). Thus, curriculum-focused PD, centered around new materials and the content and pedagogy that underpins them, and that models instructional practices, is particularly effective at improving student achievement (Cohen & Hill, 1998; Southerland et al., 2016; Lynch et al., 2019).

Pedagogical Content Knowledge. Pedagogical content knowledge is used to refer to the knowledge of how to teach content in a way that enhances student learning (Shulman, 1986). Cochran et al. (1993) described this as "a type of knowledge that is unique to teachers, and is based on the manner in which teachers relate their pedagogical knowledge (what they know about teaching) to their subject matter knowledge (what they know about what they teach)" (p. 4). According to Magnusson et al. (1999) and Park and Oliver's (2008) frameworks, components of pedagogical content knowledge include: (a) orientation toward teaching science; (b) knowledge of curriculum; (c) knowledge of learners; (d) knowledge of instructional strategies; and (e) knowledge of assessment. Gess-Newsome (1999) found that a strong science background is related to pedagogical content knowledge levels in science teaching, and studies have shown that pedagogical content knowledge builds with teaching experience (Lee et al., 2007; Magnusson et al., 1999; although see Rowan et al., 2001). However, empirical studies show both positive (Fishman et al., 2003; Garet et al, 2011; Supovitz & Turner, 2000) and insignificant (Drits-Esser et al., 2017) relationships between teacher pedagogical content knowledge and student learning outcomes. This discrepancy may be due to the multidimensional nature of pedagogical content knowledge, the measures used to assess it, and the wide range of content knowledge in science (Yang et al., 2020a).

Self-efficacy. Self-efficacy describes an individual's belief about their ability to perform the behaviors needed to reach desired goals and is related to people's effort and persistence. "People with high assurance in their capabilities approach difficult tasks as challenges to be mastered rather than as threats to be avoided" (Bandura, 1995, p. 11). Self-efficacy is a major component for understanding motivation, self-confidence, and anxiety as they pertain to teaching and learning (Bandura, 1997; Shahzad & Naureen, 2017; Zimmerman & Schunk, 2006). Of relevance to this study, self-efficacy centers around a teacher's beliefs about their ability to influence student science learning and their ability to enact PD practices (e.g., Bandura, 1977; Settlage et al., 2009). Self-efficacy is essential to PD. Without self-efficacy, teachers will not implement innovations (Lakshmanan et al., 2011).

As teachers accumulate mastery experiences, their confidence and self-efficacy increases (Bandura 1997; Padua, 2018; Ross, 1994). However, the link between self-efficacy and teacher practice is often more nuanced. For example, Herrington et al. (2016) found that two different self-measures of self-efficacy were not indicative of teachers' classroom practices. Although interview data and observation of teachers in their classrooms indicated enactment of inquiry-based practices had increased, their scores on one measure of self-efficacy remained stable or increased and their scores on another measure decreased. They attributed this in part to teachers' developing a more sophisticated understanding of inquiry-based teaching over the course of the PD, and thus being more critical of their practice (Herrington et al., 2016).

Importance of Environment. A number of environmental factors have been shown to both support and hamper PD implementation. PD is most successful when administrative support is clearly evident, when resources to implement the PD are provided including curriculum and supplies, and when innovations are aligned with policy initiatives at all levels—from school culture to standards and other student learning outcomes (Banilower et al., 2007; CRDG 2012; Cohen & Hill, 2001; Desimone, 2009; Johnson et al., 2007; Klingner, 2004; Linn et al, 2010; Luft & Hewson, 2014). Students and peers are also important aspects of the environment that affect behavior. When teachers see students benefiting from a practice, they are more likely to implement it (Fullan, 1987; Klingner et al., 1999; Lynch et al, 2019; Guskey, 2002b). If teachers participate alongside their colleagues in a PD, they can benefit from informal discussions to troubleshoot issues, share successes and adaptations, and begin to take educational ownership of the innovation (Coburn, 2003; Datnow & Stringfield, 2000; Lynch et al., 2019). Conversely, a lack of supportive teacher networks can impede innovative implementation efforts (Coburn et al., 2012). Additional common barriers to implementing PD include time restrictions, including mandated curriculum pacing and a lack of time for planning; excessive administrative loads; a lack of resources; and testing and other evaluative constraints (American Federation of Teachers, 2017; Blumenfeld et al., 2000; Buczynski & Hansen, 2010; Fitzgerald et al., 2019; Hill & Chin, 2018; Murrill et al., 2013; Pomerantz & Condie, 2017). Thus, even if a PD is of high-quality, there are several potential environmental hurdles and barriers to successful implementation

Research Questions

This is a mixed-methods case study that used a delayed post-test design (Shadish et al., 2002). The purpose of this study was to examine the persistence of PD changes on the teachers' practice, beliefs, and understanding of PD pedagogy and content 2.5 years after the completion of the program. The following three research questions guided this study:

1. To what extent do the teachers' survey response levels on the constructs (a) understanding of inquiry-based teaching, (b) aquatic science content knowledge, (c)

pedagogical content knowledge, and (d) self-efficacy, as operationalized by their respective instruments, change 2.5 years after the PD?

   a. A sub-question was: To what extent are variables that were significant in the original PD evaluation (i.e., teacher level, experience, subsequent PD participation) related to scores on the instruments?

2. What personal and environmental factors do the teachers report enhanced and/or hindered their long-term implementation of PD pedagogical strategies and concepts?
3. What are the relationships and/or patterns that describe the degree to which the teachers' long-term implementation aligned to personal, environmental, and behavioral interactions?

Question one entailed examining whether there was a statistically significant effect of the PD on instruments measuring constructs considered essential to the PD's theory of change 2.5 years after the PD. To answer this question, I tested the significance of the change scores across three time points for each of four instruments. The sub-question explored if the variables found to be significant in the original PD evaluation (Duncan Seraphin et al., 2017) affected the teachers' scores in this follow-up study. In other words, this question asked if the same patterns that were present in the original PD evaluation were present in this study. The second question addressed the extent to which the teachers reported the PD changed their practice and was explored in the teacher interviews. After the PD, the teachers who chose to continue to implement PD components did so by choice—what affected those choices? Question three was addressed in a convergent mixed-methods analysis. Using a combination of quantitative techniques and joint display tables and figures, I explored the relationships and patterns between the quantitative and qualitative data guided by the factors and interactions in Bandura's Theory of Reciprocal Determination.

CHAPTER 3

METHODS

This chapter begins with a description of the study design and the research participants. Next, I provide an overview of the instruments and the procedures by which they were administered to the participants. I then describe the methods used to analyze the data for each research question. Lastly, I discuss my own positionality with respect to this study.

Design

This follow-up study involved administering four quantitative instruments, (a) an Inquiry Teaching Assessment, (b) a content knowledge assessment, (c) a pedagogical content knowledge assessment, and (d) a self-efficacy assessment, and one qualitative instrument (an interview) to the participants at least 2.5 years after a year-long PD course. Three of the four quantitative instruments used in this study were identical to those utilized in the PD, the fourth was an abbreviated version of the original assessment. This resulted in a one-group pretest–posttest design with a second, follow-up, posttest. According to Shadish et al. (2002), this is represented as: $O_1 \quad X \quad O_2 \quad O_3$. In this setup, X represents the treatment (PD intervention), O_1 represents the pretest observation (prior to the PD; August 2012), and O_2 and O_3 represent the two posttest observations (after the PD). The first posttest (O_2) occurred immediately following the conclusion of the PD (June 2013), while the second posttest (O_3) was administered at least 2.5 years after O_2 (Jan—May 2016) (see Figure 3.1 and Table 3.1). In addition to the quantitative instruments, interviews were conducted at point O_3. While iterative testing using the same instrument can influence results due to participant familiarity with it (i.e., the testing effect, Shadish et al., 2002), the quantitative instruments were administered far enough apart (10 months between O_1 and O_2; 2.5 years between O_2 and O_3) to limit this particular validity threat.

Figure 3.1

Data Collection Over the Course of the PD and at the Follow-up

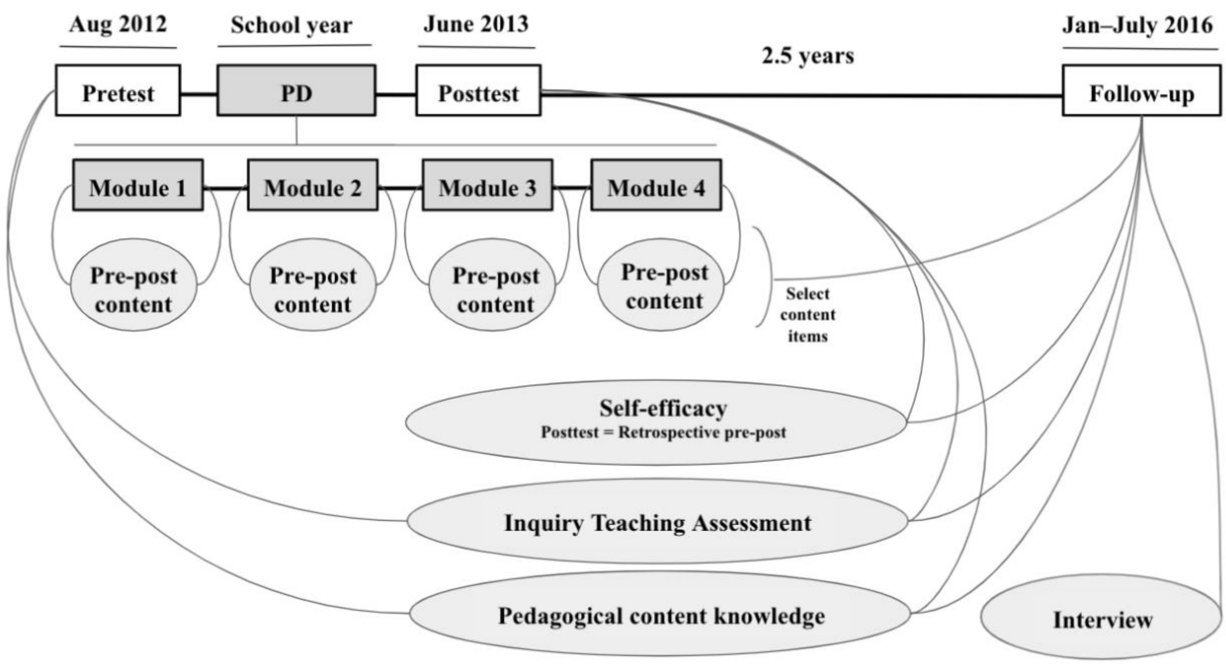


Table 3.1

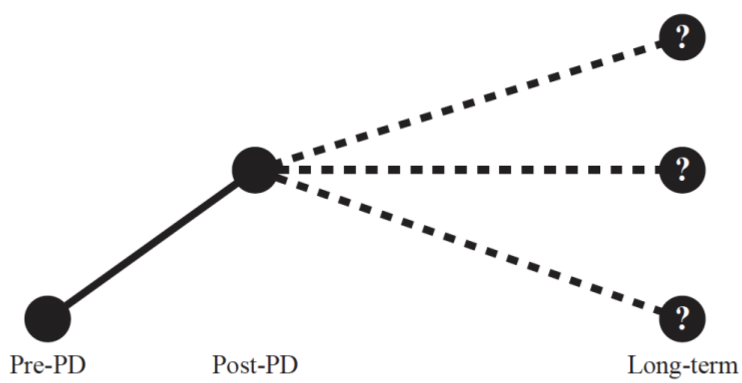
Study Design: Instrument Administration

Instrument	Pretest (O1, prior to PD)	Posttest (O2, after PD)	Follow-up (O3, 2.5 years after PD)
Inquiry Teaching Assessment	X	X	X
Pedagogical content knowledge assessment	X	X	X
Self-efficacy assessment		X (Retrospective pre-post)	X
Content knowledge assessment	Pretest and posttest before and after each of the four PD modules		X
Interview			X

For the second, delayed, posttest (O_3, the focus of this study), the quantitative and qualitative data collection instruments were administered sequentially—the four quantitative assessments first, followed by the interview. Administering the same instruments that were employed by the PD was a straightforward way to investigate how the teachers' post-PD scores had changed over time (Figure 3.2). The quantitative assessments addressed personal factors, while the interviews provided information about how interactions between personal, environmental, and behavioral factors affected teacher practices longitudinally.

Figure 3.2

Possible PD Instrument Score Trajectories: Stable, Increasing, or Eroding

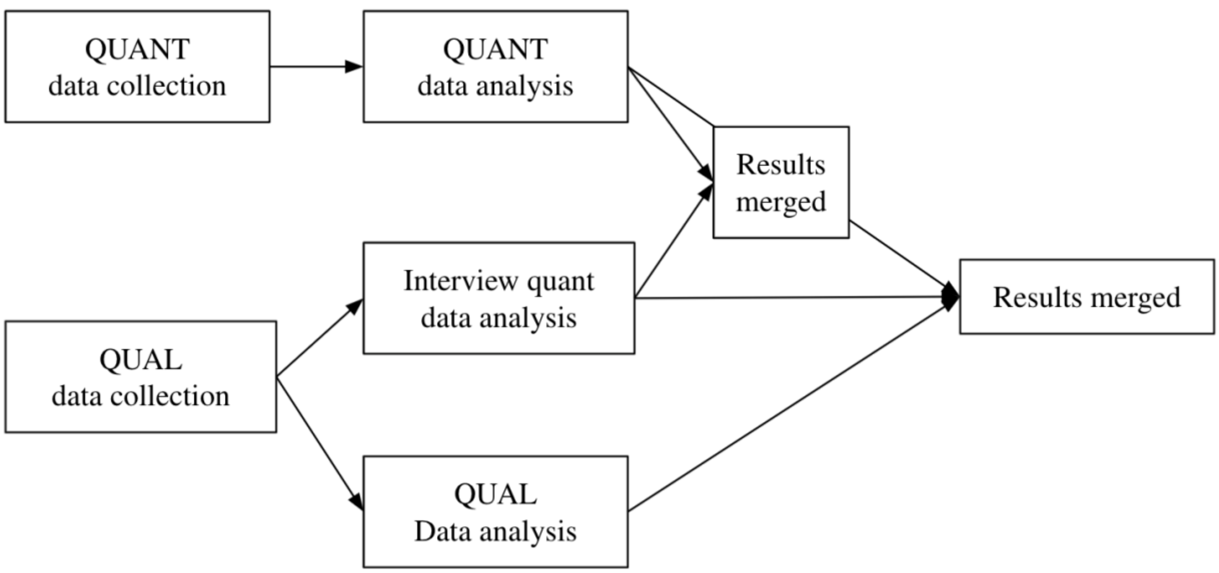


Note. All four quantitative instruments utilized in this study had pre- to post-PD gains, three of these gains were statistically significant.

Although the quantitative instruments were administered first, the data were not analyzed prior to the interviews, and thus did not inform their design or administration. The quantitative and qualitative data can thus be considered distinctly collected; they were also analyzed separately before the results were merged, a convergent mixed method design (Creswell & Plano Clark, 2018). This allowed me to make direct comparisons between the teachers' perspectives (collected through interviews) and traditional PD instruments. This permitted me to "give voice to the participants as well as report statistical trends" (Creswell & Plano Clark, 2018, p. 72) and is illustrated in Figure 3.3. As part of the interview, I asked the teachers to respond to three close-ended questions. I examined these questions using traditional quantitative techniques and merged these data separately with the quantitative data, the instruments' scores, before all the data were comprehensively analyzed.

Figure 3.3

Flow Chart of Convergent Mixed Methods Design Illustrating Collection, Analysis, and Merging of Qualitative and Quantitative Data



Note. Close-ended questions administered during the interview were examined using quantitative methods and merged with the quantitative instruments' scores prior to combining all the data.

An overview of the research questions, data-collection instruments, representations (i.e., figures), and statistical approaches is in Table 3.2. The analytic methods section of this chapter details the statistical approaches by research question.

Table 3.2

Research Questions, Instruments, Representations, and Statistical Approaches

Research Questions	Instruments				Interview		Represent-ations	Statistical approaches
	Inquiry Teaching Assessment	Content knowledge	Pedagogical content knowledge	Self-efficacy	QUAL info	QUAN info		
Q1. To what extent do the teachers' survey response levels on the constructs (a) understanding of inquiry-based teaching, (b) aquatic science content knowledge, (c) pedagogical content knowledge, and (d) self-efficacy, as operationalized by their respective instruments, change 2.5 years after the PD?	X	X	X	X			Histograms, boxplots, graphical display overview	Reliability scales, descriptive statistics, paired-samples *t*-tests, correlations
Q1A. To what extent are variables that were significant in the original PD evaluation (i.e., teacher level, experience, subsequent PD participation) related to scores on the instruments?	X	X	X	X			Scatterplots, graphical display overview	Paired-samples *t*-tests; correlations
Q2. What personal and environmental factors do the teachers report enhanced and/or hindered their long-term implementation of PD pedagogical strategies and concepts?					X		Graphical display overview	
Q3. What are the relationships and/or patterns that describe the degree to which the teachers' long-term implementation aligned to personal, environmental, and behavioral interactions?	X	X	X	X	X	X	Histograms, joint display tables and figure	Paired-samples *t*-tests; correlations

Limitations to the Design

This study design has no control group, thus gains or losses cannot be attributed solely to this PD. Because there is no control group, a history threat to internal validity exists (Shadish et al., 2002). Events that co-occurred with the treatment (PD) or that occurred in the 2.5 years after the PD could produce changes in the observations and thus affect both the results and their interpretation. Similarly, a maturation effect, referring to any changes in scores that would occur naturally and are unrelated to effects of the treatment, cannot be teased apart with this design (Shadish et al., 2002).

One of the challenges of convergent mixed methods research designs is the difficulty that can arise if the quantitative and qualitative results do not comport (Creswell & Plano Clark, 2018). Although these discrepancies can provide new insights into the topic under investigation, the nature of the design means they are not easily resolved by making adjustments to subsequent data collection. In addition, it can be challenging to merge textual and numerical data in a meaningful way, especially if the concepts addressed by each are different (Creswell & Plano Clark, 2018).

Structural corroboration involves the use of multiple types of data to arrive at an interpretation. Gathering data from more than one source reduces the risk of chance associations and systematic biases. However, the teacher interviews were the only source of information about classroom behavior. Originally, an instrument designed to measure adherence, an important fidelity of implementation variable used to evaluate PD effectiveness (e.g., Hill & Erickson, 2019), was to be administered as part of the follow-up (posttest). However, during pilot testing of the interviews, it was apparent the teachers had difficulty determining when, and how often, they implemented activities. They also referred to the activities by a wide variety of names. Asking the teachers if they had implemented an activity may have led to inaccurate responses if the teachers either failed to recognize the activity as described or did not accurately recall their implementation frequency. Due to these issues, the decision was made *not* to have the teachers complete a questionnaire concerning their implementation of PD activities. Although the interview was the only data source used to examine behavior at the follow-up time point, I corroborated my interpretations using multiple teacher perspectives.

Participants

The PD engaged five cohorts of Hawai‘i teachers, statewide, between 2009 and 2013. Cohorts 1–3 (41 teachers) were preparatory for the final two cohorts (Cohorts 4 and 5)—meaning, the curriculum and evaluation instruments were revised and updated based on the findings derived from the first three cohorts. These adjustments were relatively minor (see Duncan Seraphin, 2014 for instrument development details). The teachers were recruited into the PD through email listservs and flyers. As part of their participation in the research portion of the program they were eligible for a stipend of $1,200 ($300/module). The teachers also had the opportunity to accrue up to 12 Hawai‘i Department of Education PD course credits over the course of the PD. The PD took place on Fridays (substitute teachers were provided), Saturdays, and Wednesday evenings. This follow-up study was not conceptualized until after the PD was completed.

The teachers who participated in Cohorts 4 and 5, which took place during the 2012–2013 school year on O‘ahu and Kaua‘i, were invited to participate in this follow-up study. I recruited the teachers in December 2015 and January 2016. To be eligible for this study, the teacher participants had to have (a) completed the PD, (b) finished all of the evaluation instruments, and (c) taught in a classroom for at least one full year after the intervention ended. The first two inclusion criteria ensured that participants had received the full intervention and would have easily comparable instrument scores. The last criterion ensured that they had the opportunity to implement PD materials with students after the PD had ended.

Of the 28 teachers in Cohorts 4 and 5, 25 were eligible for this study. The other three teachers had left the classroom within a year of the PD. Of the eligible teachers, one did not participate for health reasons, and one declined to participate. Thus, the analytic sample included 23 teachers, representing 82% of the PD participants. Twenty-two teachers completed all of the follow-up study instruments (one teacher declined to be interviewed). Of 15 eligible Oʻahu teachers, 14 (93%) participated in this study. Of 13 eligible Kauaʻi teachers, 9 (69%) participated in this study (Table 3.3). The teachers were offered an incentive of $100 to participate in this follow-up study. Neither the teachers' participation nor their responses had any bearing on their professional evaluations or eligibility for PD course credit.

Demographic information, such as grade(s) and discipline(s) taught and number of years teaching science, was collected upon admission into the PD. The only additional demographic information collected for the follow-up study was gleaned from an interview question asking study participants: "What subject(s) and grade(s) have you taught since you participated in the TSI PD? At what schools?"

Approximately a third of the teachers in the PD and in this follow-up study had taught science for eleven or more years, about a quarter had taught science between seven and ten years, and about 20% had taught science for two or fewer years (Table 3.3). Most of the teachers who participated in the PD and in this follow-up study were middle-school teachers (Table 3.3). Only one participant was an elementary-school teacher (5th grade), although three of the four sixth-grade teachers in the PD, and one in this follow-up study, were housed in elementary schools and taught all subjects. Four of the five teachers who had participated in the PD, but did not participate in this follow-up study, were high-school teachers; the other was a middle-school teacher. Thus, the high-school teachers made up a smaller percentage of the follow-up study than the PD (Table 3.3). The influence of teacher experience appears to be more substantial at the beginning of a teachers' career; over time the benefits of experience level off (Gawlik et al., 2012; Kini & Podolsky, 2016). Fifteen of the 23 teachers (65%) were "experienced," defined as having five or more years of classroom experience prior to the PD. Seven of the 22 teachers (32%) were "novices," defined as having three or less years of classroom teaching experience at the start of the PD. None of the teachers in this study entered the PD with exactly four years of teaching experience.

Table 3.3

Teacher Characteristics: Cohort, Years Teaching Science, and Grade Band

Characteristic	PD teachers ($N = 28$)		Follow-up teachers ($N = 23$)	
	n	%	*n*	%
Cohort				
O‘ahu	15	53.6	14	60.9
Kaua‘i	13	46.4	9	39.1
Years teaching science (prior to PD)				
≤ 2	6	21.4	5	21.7
3–6	5	17.9	4	17.4
7–10	7	25.0	6	26.1
≥ 11	10	35.7	8	34.8
M	9.3		9.3	
SD	7.0		7.1	
Grade band taught				
Elementary school (grade 5)	1	3.6	1	4.3
Middle-school (grades 6–8)	16	57.1	15	65.2
High-school (grades 9–12)	11	39.3	7	30.4

Note. The follow-up teachers were a subset of the PD teachers, not a separate sample.

Teachers were recruited from all science disciplines into the PD (Table 3.4). During the PD, the teachers taught many classes but had only one "focus" class where they had to implement PD activities. They could have chosen to implement activities in their other classes as well, but this was optional. In the follow-up interviews, the teachers listed every subject they were currently teaching as there was no longer any "focus" class. The fifth-grade teacher taught all subjects. Three of the teachers who were categorized as teaching general science in the PD, and one in this follow-up study, were sixth-grade teachers housed in elementary schools and taught all subjects. The teacher who did not participate in the interview, but did complete all of the quantitative instruments, was also a sixth-grade teacher housed in an elementary school. Two of the teachers taught in special education classrooms. More teachers reported teaching life science (middle school) or biology (high school) at the follow-up (n = 10, 46%) than during the PD (n = 5, 18%).

Between the end of the PD and the start of this follow-up study many of the teachers switched discipline(s), including three teachers who became exclusively math and/or English teachers for a year before switching back to teaching science. Many had moved on from positions they had held during the PD. Four of the teachers were teaching in, or had taught, in different states (California, Arizona, New York, and Montana); two had moved into administration positions; and one had left the classroom for an informal environmental education position. Four of the teachers had changed grades and disciplines taught each year since the end of the PD.

Table 3.4

Teacher Characteristics: Discipline Taught

Characteristic	PD teachers ($N = 28$)		Follow-up teachers ($N = 23$)[a]	
	n	%	*n*	%
Science discipline(s) taught[b]				
Chemistry	1	3.6	2	9.1
Earth/Space science	7	25.0	6	27.3
Engineering/Robotics			3	13.6
Life science/Biology	5	17.9	10	45.5
Marine science	6	21.4	7	31.8
Science (misc/general)[c]	6	21.4	7	31.8
Physics/Physical	3	10.7	3	13.6
Non-science discipline(s) taught				
All	4	14.3	2	9.1
Math			2	9.1
English			4	18.1
Other[d]			2	9.1
Resource teacher or administrator[e]			2	9.1

Note. [a] Follow-up discipline information was only collected from 22 teachers as one teacher declined to participate in the interview. [b] During the PD the teachers taught many classes but only had one "focus" class where they had to implement PD activities. Thus, percentages for science discipline taught during the PD add up to 100. Percentages in the follow-up column do not add up to 100 as many of the teachers were teaching multiple disciplines. [c] Miscellaneous/general science answers included "science," "AVID," "elementary school science," and "inquiry" in the follow-up. [d] Other responses included "tutor" and "technology." [e] One teacher took an administrative position and one took on a science resource teacher role after spending at least a year in the classroom after the PD.

Instruments

Quantitative Instruments

The use of multiple instruments (four quantitative, one qualitative) allows for the integration of information from different data sources to help validate the findings. The four quantitative instruments utilized in this study were also used in the PD evaluation. These instruments were originally chosen or developed to address the PD's goals and evaluate the extent to which the teachers improved their (a) pedagogical content knowledge, (b) inquiry-based science teaching knowledge, (c) self-efficacy in teaching science, and (d) aquatic science content knowledge between the beginning and end of the project. The instruments would thus provide a rough estimate of the effect of the PD on these aspects of the teachers' science teaching. The purpose of the instruments in this follow-up study is similar; to estimate the degree to which the teachers changed on these scales from the end of the PD to 2.5 years later. A description of these constructs, the rationale for their inclusion, and the source of the instruments is described in Table 3.5. Scores on these instruments were validated through an iterative process of administration over multiple cohorts during the original project evaluation. For a full explanation of the development of these instruments and the score validation procedure, see Duncan Seraphin (2014). This follow-up study takes advantage of the applicability of the validity evidence detailed in the PD evaluation (Duncan Seraphin, 2014).

Table 3.5

Description and Source of Instruments of Constructs on the Follow-Up Questionnaire

Construct	Description of construct and rationale for inclusion	Source of instrument
Content knowledge	Content knowledge is declarative knowledge, or factual information. In this PD, content knowledge was focused on aspects of physical, chemical, biological, and ecological aquatic science. Effective PD is grounded in content (e.g., Garet et al., 2001; Luft & Hewson, 2014).	Science teacher content assessments (Duncan Seraphin, 2014)
Inquiry-based science teaching knowledge	Knowledge of inquiry-based science teaching is a type of procedural knowledge—knowing how to perform a task, in this case how to teach in an inquiry-based manner. PD is a format through which to address inquiry-based teaching misconceptions and train teachers to effectively implement inquiry in their classrooms (e.g., Smith et al., 2007). Teachers' conceptions of inquiry affect the likelihood that teaching practices advocated in inquiry-based PD are implemented (Lotter et al., 2007).	*Inquiry Teaching Assessment* (Duncan Seraphin, 2014), modified from the *Pedagogy of Science Inquiry Teaching Test* (Schuster et al., 2006)
Pedagogical content knowledge	Pedagogical content knowledge refers to the knowledge of how to teach a particular topic and encompasses knowledge of instructional strategies, conceptual models, and/or classroom activities (Scarlett, 2008). In science teacher PD, pedagogical content knowledge may be able to bridge the gap between content and pedagogical knowledge (Magnusson et al., 1999).	*Pedagogical Content Knowledge Scale* (Duncan Seraphin, 2014), modified from Scarlett (2008)
Self-efficacy in teaching science	Self-efficacy is a measure of one's belief in one's ability to complete tasks (Bandura, 1997), in this case to teach in an inquiry-based manner. Teachers with higher self-efficacy are associated with higher levels of student achievement (see review in Ross & Bruce, 2007) and are more likely to use practices advocated by inquiry-based science PD in their classrooms (Lakshmanan et al., 2011)	*Self-efficacy in Science Teaching Scale* (Duncan Seraphin, 2014), modified from Ayala (2005), Brandon et al. (2007) and Lawton (2005)

Inquiry Teaching Assessment. The Inquiry Teaching Assessment was designed to measure the teachers' understanding of the nature of inquiry-based science teaching in the classroom. Modified from the *Pedagogy of Science Inquiry Teaching Test* (Schuster et al., 2006), the instrument includes both selected-response (i.e., multiple choice) and constructed response (i.e., short essay) prompts following seven short science-teaching vignettes (Appendix B). Although each vignette included a multiple-choice question, these responses were not analyzed. Rather, the teacher's justification for each of their responses was measured. In the PD evaluation, the teachers' responses on both the pre- and post-PD assessments were compiled into a single dataset. This was done to reduce possible rater bias that could have occurred if the raters knew which time point the responses came from and to reduce variability due to rater drift that may have occurred if the ratings were given on two separate occasions. Two trained raters used a rubric with a nine-point scale to rate the responses. Ratings were Rasch analyzed using a multi-facet rating-scale model to estimate the teachers' scores (Andrich, 1978; Linacre, 2010). The two raters' internal consistency was high, with a Rasch reliability estimate of .95. Other descriptive statistics are provided in Table 3.6. The teachers' scores yielded a gain of 1.24 standard deviation units from pre- to post-PD ($p < .01$).

Table 3.6

Descriptive Statistics of the Inquiry Teaching Assessment

Time	*N*	*M*	*SD*	Min	Max
Pre	28	-1.01	1.24	-2.85	2.50
Post	28	0.53	1.22	-1.82	2.65

Note. $N = 28$. These values are from the PD evaluation report (Duncan Seraphin, 2014). Results are in logits based on scores from the multi-facet rating-scale model (Andrich, 1978; Linacre, 2010) under a Rasch analysis framework, with both the pretest and posttest scores modeled together to ensure scores at the two time points were on the same scale. Each teacher was rated on a nine-point performance assessment scale. There were two raters. The variability of the rater facet was approximately zero ($M = 0.00$) at both time points. The Rasch reliability of the person separation was .95. All seven items fit the model (within +/- 2.0 standardized units), with point-biserial correlations ranging from .72 to .84. The effect of the pre- to post-gain was strong (Hedges' $g = 1.24$, $p < .01$).

Pedagogical Content Knowledge Assessment. The *Pedagogical Content Knowledge Scale*, hereafter referred to as the pedagogical content knowledge assessment, was developed to measure self-reported changes in this construct in science teachers who participated in PD (Scarlett, 2008). The instrument was based on a transactional model of pedagogical content knowledge and was developed to evaluate a PD program utilizing the same pedagogical framework (TSI) as the PD under study. It grouped six constructs addressing pedagogical content knowledge into three clusters—contextual (assessment and relevance), content (inquiry), and pedagogy (collaboration, discussion, and facilitation). The original instrument had 43 five-point Likert-type scale items; seven of these were removed when it was modified for this PD. Thus, the final instrument had 36 items. These changes were made because the PD facilitators thought the original instrument did not reflect shifts in the theoretical domain as defined and expressed in the PD. In the revised instrument 50% (18 of 36) of the items were in the "pedagogy" cluster, 28% were in "context," and 22% were in "content." As the pedagogical cluster was composed of items addressing collaboration, discussion, and facilitation, this instrument leaned more towards assessing classroom management strategies than other aspects of pedagogical content knowledge.

Pre- and post-PD, the teachers were prompted to select a response indicating how often they, or the students in their classes, engaged in certain behaviors. While the items were not TSI-specific, they were considered reflective of the overall intent of the PD. The instrument had high internal-consistency reliability, with Cronbach's alphas of .87 (pre) and .92 (post). The descriptive statistics for the scale are shown in Table 3.7. The pre- to post-PD gain was only approximately a quarter of a standard deviation unit ($p > 0.05$). The instrument is in Appendix C (part 1).

Self-efficacy Assessment. The *Self-Efficacy in Science Teaching Scale*, hereafter referred to as the self-efficacy assessment, was adapted from earlier scales (Ayala, 2005; Brandon et al., 2007; and Lawton, 2005) by adding items that addressed components of TSI pedagogy. The 15 six-point Likert-type scale items prompted the teachers to select a response indicating their degree of ability with respect to specific teaching practices. It was administered according to a retrospective pre-post design. In this approach, the teachers were asked to think about their levels of self-efficacy after having completed the PD, and then asked to reflect on their self-efficacy before participating in the PD. The rationale behind this administration choice was that the teachers would not be expected to be familiar with TSI pedagogy prior to the PD, thereby precluding them from making informed judgements of their self-efficacy. Retrospective pre-post administration designs are recommended when an intervention will change the respondents' understanding of the questions, and thus invalidate their pretest responses (Aiken & West, 1990). There are concerns about the role of confirmation bias with this type of administration, but research has shown that when this bias is present, traditional pretest-posttest comparisons can result in an underestimation of program effects (Cantrell, 2003; Pratt et al., 2000).

The self-efficacy assessment had high internal-consistency reliability, with Cronbach's alphas of .97 (pre) and .96 (post). The descriptive statistics for the scale are shown in Table 3.7. The teachers gained approximately 1.50 standard deviation units ($p < .01$) from pre- to post-PD. The instrument is in Appendix C (part 2).

Table 3.7

Descriptive Statistics of the Pedagogical Content Knowledge Assessment and Self-Efficacy Assessment

Instrument	Time	N_{items}	*M*	*SD*	se_M	Cronbach's α	Hedges' *g*	*p*-value
Pedagogical content knowledge assessment [a]	Pre	36	3.84	0.36	0.07	.87	0.24	.21
	Post	36	3.93	0.36	0.07	.92		
Self-efficacy assessment [b]	Pre	15	3.52	0.99	0.19	.97	1.50	< .01
	Post	15	4.79	0.67	0.13	.96		

Note. $N = 28$. These values, the results of a repeated measures general linear model, are from the PD evaluation report (Duncan Seraphin, 2014).

[a] Response categories ranged from *1* (never) to *5* (always). [b] Response categories ranged from *1* (low ability) to *6* (high ability).

Content Knowledge Assessment. The only quantitative instrument modified for the follow-up study was the selected-response aquatic science content knowledge assessment. In the PD, a content knowledge assessment was administered prior to and after each of the four PD modules. A total of 141 content questions were administered twice (pre and post) over the course of the school year (module 1 = 30 items, module 2 = 40 items, module 3 = 37 items, module 4 = 34 items, see Duncan Seraphin, 2014 for details). The results of these assessments by module are detailed in Table 3.8. Six of 141 items originally administered were excluded after examining how the item functioned and conducting distractor analysis. In the original PD evaluation, as displayed in Table 3.8, only items with a difficulty index of .75 or less on the pretest were included in the analysis. As administering all 141 of the content items was unrealistic for this follow-up study, I developed a new assessment composed of a subset of the original items.

Table 3.8

Teacher Content Knowledge Assessment Results

Instrument	$N_{teachers}$ [a]	N_{items}	M [b]	SD	se_M	Min [b]	Max [b]	KR-20	Hedges' g
Module 1 Pre	31	29	19.39 (66.9%)	5.33	0.96	9 (31.0%)	29 (100%)	.83	0.66
Module 1 Post	31	29	22.68 (78.3%)	4.58	0.82	14 (48.3%)	29 (100%)	.82	
Module 2 Pre	29	38	25.14 (66.1%)	7.15	1.33	10 (26.3%)	37 (97.4%)	.87	0.76
Module 2 Post	29	38	30.10 (79.2%)	5.88	1.09	18 (47.4%)	37 (97.4%)	.86	
Module 3 Pre	28	36	19.71 (54.7%)	5.62	1.06	9 (25.0%)	29 (80.6%)	.80	0.76
Module 3 Post	28	36	24.00 (66.7%)	5.58	1.05	9 (25.0%)	34 (94.4%)	.81	
Module 4 Pre	28	32	15.25 (47.8%)	4.39	0.83	9 (28.1%)	25 (78.1%)	.66	1.01
Module 4 Post	28	32	19.96 (62.5%)	4.90	0.93	13 (40.6%)	28 (87.5%)	.75	

Note. These values, the results of a repeated measures general linear model, are from the PD evaluation report (Duncan Seraphin, 2014). All pre-post differences were significant ($p < .0001$).

[a] Although 31 teachers started the PD, only 28 teachers completed it. [b] Percentages are percent-correct scores.

In order to develop a new assessment, I first assigned each of the original 141 items a content area (e.g., density or sampling). Then I determined in which activities the content was addressed. Each of the four modules had "mandatory" content activities that teachers were required to implement (one in modules 1–3, two in module 4) and two "choice" content activities in which teachers could choose which of the activities they would implement (with the exception of module 4, which did not have any choice content activities). Most of the PD time was spent on the mandatory and choice activities, and the teachers were given the supplies to implement these activities. Since the other activities were less likely to have been implemented by the teachers, I made the decision to focus on the content covered in the mandatory and choice activities, reflecting what the teachers had the most opportunity to learn. This criterion eliminated 58 of the original 141 content knowledge items (41%).

I then calculated item-difficulty indexes. Although these were estimates, I used them to make decisions about which items to select for the new version of the instrument. However, the overall size of the sample was a limitation. Thus, I do not know if item-difficulty values were due to item functioning or sampling error as high accuracy and precision cannot be assumed when using small samples.

I first calculated the item-difficulty index on the pre-assessment and then the item-difficulty-index change (pre-to-post PD gains) for each content knowledge item. If the item difficulty index on the pretest was high this meant there was a ceiling effect—most of the teachers knew the content coming into the PD. Sixteen items (11% of the original 141 items) were eliminated because they showed a ceiling effect, they had an item-difficulty index on the pretest greater than or equal to .85. If the item-difficulty-index change was small, it could either mean the PD did not affect content knowledge assessed by the item, or that the item was flawed in some way (i.e., written poorly, or was measuring a different construct). Sixteen items (11% of the original 141 items) were eliminated because they had small item-difficulty-index changes, their item-difficulty-index changes between the pretest and posttest were less than or equal to .1. An item's sensitivity to growth can also be confounded with the effectiveness of the PD to cause growth. Thus, I do not know if the items I removed had small item-difficulty-index changes due to flaws with the item, content misalignment, or PD effectiveness. I only kept items that showed growth over the course of the intervention—but I do not know if this was due to the PD, or something else, as I did not look at the items' sensitivity to change with another population. Thus, I may have made spurious conclusions about the effects of the PD on content knowledge items.

In summary, items were removed if (a) the item was not aligned to classroom activities the teachers had to implement (58 items removed, 41% of the original 141 items), (b) the item-difficulty index on the pretest was greater than or equal to .85 (16 items removed, 11% of the original 141 items), or (c) the item-difficulty-index change between the pretest and posttest was less than or equal to .1 (16 items removed, 11% of the original 141 items). Applying these criteria decreased the list of potential content knowledge items from 141 to 51.

I then looked at item discrimination—an estimate of how well an item functions at distinguishing between people on the low and high ends of the scale. I calculated point-biserial correlations, which are correlations between dichotomous (0,1) posttest scores on each item and the total raw score, and item-total correlations on posttest items. The item-total correlation is an adjusted point-biserial correlation whereby the item response is correlated with the total score after that item has been removed. For these calculations I used the total scores from all four content areas (modules 1–4—physical, chemical, biological, and ecological aquatic science) as opposed to using the scores from the item's content area (e.g., just chemical aquatic science). Focusing only on content categories with more than five items, I removed those with the lowest point-biserial and item-total correlations. Thus, if a content category (e.g., density) was associated with seven items, I removed the two with the lowest point-biserial and item-total correlations—even if they had high item-difficulty index changes. Thus, I retained items with lower item-difficulty index changes whose scores were more strongly correlated to the overall test score. This process of elimination removed ten items, five of which had high item-difficulty index changes (0.3 and above). It preserved items that were not correlated well to the overall test, but which addressed content categories with three or less items. For example, the electrolysis items had particularly low correlations to overall test scores but were retained because an activity on electrolysis was a choice classroom activity in module 2. After these adjustments, the final number of items on the follow-up content knowledge assessment was 41. The final content knowledge assessment is in Appendix D; item-difficulty index changes, pre-values on pretest, point-biserial correlations, and item-total correlations are in Appendix E. Table 3.9 shows the number of items per module and how items on the final content knowledge assessment aligned to the PD content.

Module 1 includes one item aligned to an overarching PD goal rather than a specific activity—that Earth has one ocean composed of many ocean basins, as opposed to distinct oceans (Table 3.9). This highlights how the world ocean is connected (e.g., through its water circulation and biological activity), and has important conservation implications.

The descriptive statistics and reliability estimates (KR-20) for the 41-item follow-up instrument, hereafter referred to as the content knowledge assessment, based on the original PD evaluation pre-PD (pretest) and post-PD (posttest) scores are shown in Table 3.10. Reliability estimates were .86 for the pretest and .88 for the posttest; this is higher than the average internal-consistency reliability estimate for the eight original assessment administrations ($N = 8$, $M = .80$, $SD = 0.07$). The raw score gains (pre-PD - post-PD scores) ranged from 1 to 18 points. The percentage correct ranged from 20–80% on the pretest and from 39–98% on the posttest. There was a mean gain of 23.8 percentage points from pretest to posttest ($M_{pre} = 51.7\%$; $M_{post} = 75.5\%$).

Table 3.9

Alignment of Follow-up Content Knowledge Assessment Items to PD Content

Module	Content	Alignment	N_{items}
1 (Physical)	One ocean	Overarching PD goal	1
	Density	Mandatory & Choice activities [a]	5
	Moon/Tides	Choice activity	3
		Module 1 Total	9
2 (Chemical)	Water properties	Mandatory activity	5
	Electrolysis	Choice activity	3
	Solubility/Conductivity	Choice activity	3
		Module 2 Total	11
3 (Biological)	Scientific language	Mandatory activity	4
	Fish	Choice activity	3
	Evolution	Choice activity	3
		Module 3 Total	11
4 (Ecological)	Sampling	Mandatory activity	5
	Abundance	Mandatory activity	5
		Module 4 Total	10
		Grand Total	41

[a] Density was the content of both the mandatory activity and one of the choice activities in module 1.

Table 3.10

Follow-up Content Knowledge Assessment Descriptive Statistics

Time	$N_{teachers}$	N_{items}	M [a]	SD	se_M	Min [a]	Max [a]	KR-20
Pretest	28	41	21.21 (51.7%)	7.48	1.41	8 (19.5%)	33 (80.5%)	.86
Posttest	28	41	30.96 (75.5%)	6.79	1.28	16 (39.0%)	40 (97.6%)	.88

[a] Percentages are percent-correct scores.

Limitations of the Instruments. To examine changes in the teachers' understanding and use of constructs over time, I adopted, with some revisions, the same instruments utilized in the PD evaluation. Although the instruments were developed iteratively and evaluated for validity, three out of four were originally developed in-house, which can produce larger than average effect sizes (Hill et al., 2020; Wolf, 2021), and the theoretical underpinnings of some of the scales (i.e., self-efficacy and pedagogical content knowledge) have shifted since their conceptualization. However, the benefit to implementing the same instruments in the follow-up study, even though they have some flaws, is that the scores can be readily compared on the same scale across time points. Because the quantitative instruments used in the current follow-up study are the same or very similar to the original instruments used to evaluate the project, the intended interpretation and use of the instruments did not change between administrations. The change in context due to the temporal shift in administration of the instruments was examined through interviews.

***Content Knowledge Assessment Limitations*.** The content knowledge assessment likely does not meet the assumption of unidimensionality. However, the original four module content assessments also likely did not meet this assumption. The content items were written to assess the content addressed in each module, which were roughly grouped by discipline area (physical, chemical, biological, and ecological aquatic science). However, several topics were not closely tied to the discipline module in which they were addressed (e.g., scientific language in the biological module and data interpretation and plankton in the ecological module). The scientific language items aligned to a mandatory implementation activity and were thus kept in the final revision of the content assessment. However, the data interpretation and plankton questions were removed as these questions were not tied to mandatory or choice PD activities. Content topics covered in each module shared fundamental underlying concepts but may not have been grouped together by other researchers (e.g., density and moon/tides into physical aquatic science). This presents some challenges to interpreting the scores of the content knowledge assessment(s), in both the original PD evaluation and this follow-up study, because they do not reveal whether teachers' gains or losses are due to one dimension or the combination of dimensions. In the final version of the follow-up instrument the number of items per module was about the same (between 9 and 11), resulting in approximately the same representation of each of the four dimensions in the total test score.

Another limitation to the content knowledge assessment is that I do not know if the items I removed had small item-difficulty-index changes due to flaws with the item, content misalignment, or PD effectiveness. Only items that showed growth over the course of the PD intervention were utilized, but I do not know if this growth was due to the PD, or something else, as there was no comparison group. This weakens the validity of claims about the teachers' changes in content knowledge from this instrument. This limitation means that inferences drawn from this instrument should be interpreted with caution.

Qualitative Interviews

The main purpose of the teacher interviews was to obtain a more complete picture of the phenomena under study (Creswell & Plano Clark, 2018). I utilized a semi-structured interview protocol to provide the detailed insights and rich contextual information needed to understand the challenges and successes the teachers experienced implementing PD components. The interviews focused on (a) eliciting and probing information about the nature and degree of each teacher's implementation of classroom practices targeted by the PD intervention, (b) situating the PD within the context of each teacher's overall learning and practice, and (c) gathering contextual information about factors that enhanced or hampered the use of PD components.

Of the 23 teachers who completed the quantitative instrument portion of the study, 22 (96%) participated in the interview. Because the limited scope of the interview resulted in the collection of a relatively small amount of information per person, there was a risk of insufficient data collection. However, the number of teachers interviewed was higher than many other long-term PD studies (e.g., Padua, 2018; Shaharabani & Tal, 2017), thus even if the interviews were limited in duration the number of people surveyed hopefully provided some level of data saturation.

I carefully created interview questions following standard instrument development procedures. First, a list of PD and implementation variables was generated through (a) reviewing TSI pedagogy and PD documents, (b) examining the post-PD interview prompts for items to re-use or modify, (c) consulting with PD project and evaluation team members, and (d) examining PD literature. I then operationally defined the variables and used them to identify potential items from existing PD evaluation instruments, as well as generate new items. The final interview protocol consisted of five sections:

1. Practical significance—of PD content and TSI pedagogy,
2. Application—the extent to which the participants felt they applied what they learned in the PD and what enhanced or hindered this application,
3. Implementation—the extent to which the participants implemented PD activities (including reasons for implementation, modifications to activities, and what enhanced or hindered implementation),
4. Teaching Practice—if the participants thought their practice had changed and why, and
5. Professional trajectory—how the PD fit into the participants' overall learning and professional growth.

In addition, the teachers were asked about their current positions (e.g., teacher, administrator) and answered questions regarding where and what they taught (i.e., school(s), grade level(s), and subject area(s)) since the PD ended). The teachers were also asked to briefly describe any significant PD experiences they had had since the end of the PD (which would subsequently be dichotomously coded as yes or no) and answer one Likert-scale question about their current implementation of PD activities and one forced-choice question with ordered categorical data responses about their current application of PD practices.

I conducted verbal cognitive interviews with two teachers who were not eligible to participate in this study, but whose PD participation meant they were familiar with TSI pedagogy and terminology. These interviews were used to determine whether the instrument was functioning well, to inform and refine the interview questions, and to identify where edits to the protocol needed to be made. I followed a think-aloud protocol with some post-protocol probing, following guidelines according to Beatty and Willis (2007). I revised the interview protocol and questions after each teacher participated. For example, the intent of the interview questions was to gather information about the teachers' implementation of PD components and self-perceived pedagogical shifts after the completion of the PD 2.5 years ago. However, the verbal cognitive interviews indicated that it was difficult for the teachers to differentiate between changes that occurred in their teaching practice concomitant with the PD and any changes thereafter. This may have been because the teachers participating in the verbal cognitive interviews had left the classroom after the PD and thus focused more on past or future teaching as opposed to current teaching. Regardless of the reason, I updated the wording of the questions to elicit clearer answers. In addition, the verbal cognitive interviews indicated that the teachers were not confident about whether they implemented specific PD activities during the PD or after the PD ended, and sometimes got the PD activities confused with similar activities. For example, in the PD we modeled an inquiry-based fish-printing activity to examine form and function. One of the teachers said they did fish printing in their class prior to the PD, during the PD, and after the PD. I did not know if this teacher was implementing the PD version of the activity after the PD, or had reverted to their prior version of the activity. I updated question prompts about implementation to take these memory lapses into account. Lastly, the teachers who participated in the verbal cognitive interviews often wanted to share how specific activities were implemented in their classrooms and how they were received by students—even if the questions did not explicitly ask for this information. As I considered this information indicative of the teachers' perceptions of students' learning and engagement, I did not reword question prompts to lead to more succinct answers.

To control for researcher bias, one of the members of the PD evaluation team conducted the interviews. The interviewer had extensive experience and training in conducting semi-structured interviews. While key questions were predetermined to allow for systematic data collection and enhance objectivity, the interviewer also had the flexibility to be conversational and respond to situational factors. The interviewer aimed to cultivate an environment where the teachers felt comfortable saying what they thought—whether critical or supportive of the PD. The interviewer made the purpose of the study very clear to the participants—we were not researching or evaluating *them*. Rather, their responses, together with others, would help us understand the overall effect of the *program*. The interviewer stressed that the teachers' responses were confidential, and that we appreciated their forthright and honest answers (see Appendix F for interview slides).

Procedures

One of the original evaluators of the PD corresponded with the participants and administered all of the instruments, thus decreasing the threat of social desirability to internal validity (King & Bruner, 2000). By not conducting the interviews and not observing the teachers in their classrooms, I sought to minimize any alteration of the participating teachers' behavior through Hawthorne or novelty effects (Onwuegbuzie & Leech, 2007).

An initial email invitation to participate in the follow-up study was sent in late December 2015. Subsequent invitation emails were sent during the following weeks. With the exception of one potential participant, all of the teachers were successfully contacted. After the participants signed the study consent form, we administered the instruments in tiers. Only once a participant had completed an instrument, were they sent the next instrument in the sequence. First, they were sent the Inquiry Teaching Assessment, then the content knowledge assessment, and finally the self-efficacy and pedagogical content knowledge assessments, which were bundled together as a single instrument. This bundling was also done in the PD. Instruments were administered in the same order as in the PD. This order was chosen to prioritize the Inquiry Teaching Assessment, which takes the longest to complete (Table 3.11) as it prompts the participants to respond to open-ended questions. The subsequent instruments took progressively less time to complete. All instruments were administered asynchronously online through SurveyMonkey; this platform was also used in the PD evaluation and was thus familiar to the participants.

When all of the quantitative instruments had been completed, the participants were scheduled for interviews. With the exception of three interviews, all were conducted through Blackboard Collaborate. This platform was used extensively during the PD, as well as for post-PD interviews, and was thus familiar to the participants. All of the online interviews were recorded for later transcription. Three interviews were conducted in-person at the request of the participants; for two of these, the audio was recorded. One interview was conducted in-person during a participant's class. As students were in the room, the interview could not be recorded; instead, the interviewer took detailed notes. The final interview took place in July 2016; thus, the data collection for this study took approximately seven and a half months.

Table 3.11

Follow-up Aquatic Science Content Knowledge Assessment Descriptive Statistics

Instrument	Order in sequence	Administration dates	*M* (min)	*SD* (min)	Range (min)
Inquiry Teaching Assessment	1	1/1–3/2/2016	49:03 [a]	23:16 [a]	17:15–100:50 [a]
Content knowledge assessment	2	1/9–6/13/2016	27:40	11:34	09:43–59:07
Pedagogical content knowledge & self-efficacy assessments	3	1/11–6/09/2016	9:34 [b]	4:50 [b]	05:18–21:51 [b]
Interview	4	1/21–7/15/2016	44:16	14:20	14:00–75:00

Note. $N = 23$ for all instruments except the interview where $N = 22$.

[a] This does not include three responses that lasted over three hours. Seven responses that were included were over an hour. Longer response times indicate the participants likely responded to the questions over multiple sessions. [b] This does not include two responses that took over thirty minutes.

Analytic Methods

Research Question 1—Analysis of Quantitative Data

Preparatory Steps. For each of the quantitative instruments, I calculated the reliability of the scales, and analyzed the items using descriptive statistics at each time point (pre-PD, post-PD, and follow-up). In preparation for running paired-samples *t*-tests, I examined four statistical assumptions (a) the variable is on an interval or ratio scale, (b) the respondents are independent, (c) the data are approximately normally distributed, and (d) there are no outliers. Each of the four variables (the Inquiry Teaching Assessment, content knowledge, pedagogical content knowledge, and self-efficacy scores) are on an interval scale; the Likert-type scale assessments because of the number of effective scale points (five or more for each scale; Rhemtulla et al., 2012). Regarding the assumption of independence, seven of the 23 teachers (30%) were from the same three schools on Kaua‘i because of the limited number of schools on this largely rural island. Two of these teachers were spouses. I do not know if the teachers socialized in similar circles, but based on their interactions at the beginning of the PD, it appeared as if those from different schools did not know each other. However, the nature of the data collection (online) precluded the teachers from assisting each other in completing the assessments. And while some of the teachers may have inadvertently violated this assumption, the teachers stood little to gain from helping each other. Thus, I am assuming the respondents are independent enough. For the assumptions having to do with normality and outliers, I examined score distributions for each time point with histograms and boxplots. The statistical assumption of normality was examined using the Shapiro-Wilks test of normality on each of the respondents' difference scores between time points on each of the four variables (e.g., the difference between the post-PD and follow-up time points). This resulted in 12 tests (three difference scores for each of the four variables). Where non-normality was found, the differences were examined with histograms. If the non-normality did not appear to be severe, it was stated as a limitation. If the non-normality was strong, and due to outlier(s) that affected the mean and caused the variance to expand, I used the Wilcoxon signed-rank test (a non-parametric analogue to the paired *t*-test) instead. Additional details concerning the preparatory steps for the analysis of the Inquiry Teaching Assessment follows.

Inquiry Teaching Assessment. Two raters coded each of the participants' constructed-response Inquiry Teaching Assessment items using a holistic rubric developed in the PD. Rubric scores ranged from -1 to +3 in half point increments, equating to a 9-point scale. The same raters were used for this study as in the PD. One of the raters was on the PD evaluation team and had expertise in teaching and assessing adults, while the other rater was a state-level science administrator who was one of the PD facilitators.

I asked the raters to assess two dimensions on the Inquiry Teaching Assessment: (a) the teacher's understanding of what inquiry looks like in the classroom, ranging from misconceptions to deep and comprehensive understanding, and (b) the relevance of the information provided to justify their responses, including how much written material the teachers' provided for the raters. For the first dimension, I asked raters to focus on the indicators that suggested the teacher understood what teaching science as inquiry looked like in a classroom, rather than on the correct usage of TSI terminology. One teacher might, for example, have used the TSI term "initiation phase," while another teacher might have provided a description that showed they understood what the initiation phase was, even if they did not use those exact words in their description. Alternatively, a teacher whose response indicated they did not understand what TSI pedagogy looked like in practice might have used TSI terminology in their response. To address the second dimension, the rubric included statements such as "There is some evidence that…" or "There is strong evidence that…". I instructed the raters to give a score of zero when not enough information was given in the response to make a definitive judgment. A zero score could also represent a balance between a teacher's understanding and misconceptions about inquiry-based teaching and what it looks like in practice. I instructed the raters to give negative scores when there was evidence that a teacher's misconceptions more than offset their understanding of what inquiry-based teaching looked like.

To examine and account for any drift in raters' scoring over time, between when the original PD responses were rated (2013) and when the follow-up study responses were rated

(2016), the post-PD responses from seven teachers were embedded in the follow-up responses and re-coded in 2016. These responses served to anchor the follow-up scores so they were on the same scale as the scores from the PD evaluation. I chose the seven teachers from a stratified random sample of high, middle, and low scoring participants on the post-PD administration of the Inquiry Teaching Assessment. High-scoring teachers scored at least one logit above the mean score, while low-scoring teachers scored at least one logit below the mean score. These were estimated with the multi-facet rating-scale Rasch model in the original study. Two teachers were chosen from each of the "high" ($N = 9$) and "low" ($N = 6$) categories; three teachers were chosen from the "middle" category due to the number of individuals in this category ($N = 13$). I assigned all of the teachers' responses a random identification number so the raters could neither identify the teacher nor determine response time period (post-PD or follow-up).

Before rating the full set of responses, the raters practiced on 15 items in a practice set. After individually rating each of the practice responses, the raters shared their scores and discussed any discrepancies to make sure they were in agreement about how to score responses in accordance with the rubric's criteria. After discussion, the raters were within 0.5 point of each other on all of the items (9-point scale). Instructions, training procedures, and the rubric for rating the Inquiry Teaching Assessment constructed-response items are included in Appendix A. On average, the two raters scored the follow-up items 0.2 points lower on the holistic rubric at the follow-up time point. Thus, either the raters were more stringent in their scoring at the follow-up or the teachers did worse. However, this rater variability was accounted for in the Rasch model because all time period measures were modeled together through the use of anchoring items.

Ratings were analyzed using a multi-facet rating-scale model under a Rasch analysis framework to estimate the teachers' scores (Andrich, 1978; Linacre, 2010). In the analysis, the two raters were coded as two new raters at the follow-up time point because of the time span between the PD evaluation and follow-up scoring (2.5 years) and to account for their slight rater drift. Their ratings were linked to the earlier (pre- and post-PD) dataset by anchoring the re-rated post-PD test scores, which were fixed to be of the same ability as measured in the post-PD assessment. Only six of the seven teachers were used as anchors, because in an earlier analysis of the data, one teacher had a very high misfit (a z-score of 2.90), which may otherwise have skewed the follow-up time point scores. The new follow-up time point data was estimated based on the anchored data.

Statistical Analyses. I conducted paired-samples *t*-tests or Wilcoxon signed-rank tests, if assumptions for normality were not met, to evaluate the difference between scores at each time point. I first compared pre- to post-PD scores from the original evaluation population ($N = 28$ teachers) to this study's sample ($n = 23$) for each instrument to confirm that the results in the smaller sample were reflective of the original evaluation (Duncan Seraphin, 2014). I hypothesized that the statistical change in scores from pre- to post-PD for each instrument from the initial evaluation of the PD would hold, even with the loss of five teachers from the PD to the follow-up. Additional comparisons were then made between the scores at each time point for each instrument. I hypothesized that scores would decline, but not return to the baseline, for the Inquiry Teaching Assessment, and content knowledge and self-efficacy assessments. For the pedagogical content knowledge assessment, I hypothesized that there would be no change in scores across any of the time points, as no statistical change pre- to post-PD had been found in the PD evaluation. Due to the number of participants in this study, Hedges' *g* was used to determine effect sizes. If I used the Wilcoxon signed-rank test the effect size was calculated by dividing the absolute standardized test statistic by the square root of the number of pairs ($r = z / \sqrt{N}$)

Because I conducted repeated tests on outcomes that were likely to be highly correlated (i.e., the pre- to post-PD outcome on the Inquiry Teaching Assessment will likely be highly correlated with the pre-PD to follow-up outcome on the same instrument), I followed the suggestion of What Works Clearinghouse (2020), which recommended applying the Benjamini-Hochberg correction to account for multiple comparisons (Benjamini & Hochberg, 1995). The Benjamini-Hochberg adjustment corrects the “multiplicity” that can occur when testing more than one hypothesis within a domain, potentially leading to a greater likelihood of type 1 errors—when one mistakenly rejects the null hypothesis and concludes that a difference is significant when it is not (What Works Clearinghouse, 2020). For all Benjamini-Hochberg adjustments I used the false discovery rate of 0.05.

Credibility of Results—Comparisons Among Instrument Scores. To examine the credibility of the results for the first research question, I standardized all the quantitative instrument scores for each time point through the use of z-scores to normalize differences and more easily allow for comparisons across instruments’ scores. I then correlated estimates within and between instruments’ scores for each time point. None of the quantitative instruments were designed to measure the same underlying construct. Generally, if test scores on instruments purporting to measure different constructs are dissimilar, this would be evidence of discriminant validity. However, because constructs such as self-efficacy and pedagogical content knowledge are considered mediating factors with respect to instructional practice, I anticipated that the scores on these instruments would be positively correlated. Within instrument score positive correlations are evidence that the instruments were consistently measuring the same construct over time.

Research Question 1A—Examining Patterns in the Quantitative Data

I examined the extent to which variables that were significant in the original PD evaluation including teacher grade level (middle- or high-school), years of teaching experience, and participation in subsequent PDs (Duncan Seraphin, 2014; Duncan Seraphin et al., 2017), related to scores on the instruments. I first calculated the correlation between score changes and experience to see if the number of years of science teaching experience prior to the PD was related to their score gains or losses on each instrument. I then conducted independent samples *t*-tests to evaluate if there were differences in the magnitude of change scores between time points on each instrument and the grade band that teachers taught (i.e., middle- or high-school). I hypothesized that scores between the teachers of each grade band would be similar for the Inquiry Teaching Assessment, pedagogical content knowledge assessment, and self-efficacy assessments. The middle-school teachers' content knowledge assessment scores had risen more over the course of the PD (Duncan Seraphin et al., 2017), which I hypothesized was in part because middle-school teachers (licensed K–8) generally have less scientific coursework than high-school teachers (licensed 9–12). However, I did not think there would be a difference in the magnitude of the decline in content knowledge assessment scores between these groups of teachers post-PD. Finally, I conducted independent-samples *t*-tests to evaluate the differences between the groups of teachers who did, and did not, self-describe significant PD experiences after the PD under study and their scores on the follow-up instruments. I hypothesized that scores would be similar between these groups.

Research Question 2–Analysis of Interview Qualitative Data

I analyzed the interviews after analyzing the quantitative data. After transcribing the interviews, I coded them according to the three main factors comprising Bandura's Theory of Reciprocal Determination—environmental, personal, and behavioral contexts of learning. Examples of how I categorized these factors follow:

- Environmental: Resources (e.g., curriculum and/or supplies), school context and peer communities, class alignment (in terms of content, standards, and class time), and references to student learning and engagement
- Personal: references to previous knowledge, what the teachers said they learned or did not learn, connections with beliefs and preferences, confidence levels, and references to years of experience and grade level(s) taught
- Behavior: incorporation of aspects of the PD into the teachers' praxis, and transfer and sharing of the PD beyond their classroom

This initial categorical coding allowed me to pay particular attention to identifying environmental and personal factors that influenced the implementation of PD practices.

Within each of these broad categories, I then coded using a more inductive approach rooted in grounded theory and recursive thematic analysis (Corbin & Strauss, 2014). I identified themes through open coding in relation to my research questions and the PD pedagogical framework. At this stage I also looked for evidence of variables identified in the PD literature (e.g., adherence, participant responsiveness, and evidence of transfer) and focused on coding interview data that were associated with the constructs addressed in the quantitative instruments (e.g., content knowledge and pedagogical knowledge). In subsequent analyses, I focused on emergent topics and linking factors (i.e., personal and environmental factors that enhanced or hindered behavior). I also identified examples to illustrate the degree and kinds of changes the teachers made in their teaching practice due to the PD. Codes were continually reviewed and grouped, and re-grouped, into categories (Merriam & Tisdell, 2015). When appropriate, I differentiated results by teacher experience (novice vs. experienced) and teacher grade level (middle vs. high school) as these are important personal factors that can influence findings.

When possible, I quantized qualitative data—a numerical transformation that allows the data to be analyzed statistically. While data reduction and the use of percentages to describe qualitative data can oversimplify complex results, transforming qualitative data into counts allowed me to statistically analyze these variables with the quantitative data and elucidate the personal experiences of the teachers in joint display tables.

I organized the qualitative results according to themes that emerged indicating interactions between environmental factors, behavior factors, and personal factors according to Bandura's Theory of Reciprocal Determination.

Validity Threats When Analyzing Qualitative Data. I followed up on surprises in the data and closely examined outliers. Constantly looking for validity threats is one of the key components of the *modus operandi* method, which considers the implications of these threats rather than seeking to control them (Onwuegbuzie & Leech, 2007). Threats to theoretical validity when analyzing qualitative data "occur when a researcher does not collect or pay attention to discrepant data, or does not consider rival explanations or understandings of the underlying phenomena" (Onwuegbuzie, 2002, p. 12–13). In addition to utilizing an audit trail, I used respondents' quotes in my interpretations to illustrate themes and findings to address the implications of this validity threat. I also considered rival explanations and continually modified emerging hypotheses through the use of negative (discrepant) case analysis.

Confirmation bias is only a threat “when there exists at least one plausible rival explanation to underlying findings that might be demonstrated to be superior if given the opportunity” (Onwuegbuzie, 2002, p. 18). To generate and assess rival explanations, I searched for evidence to support competing conclusions and critically considered divergent patterns, paying particular attention to negative or disconfirming cases. Outlier cases were examined closely to determine if they broadened understandings about a pattern, were exceptions to the pattern, or provided evidence for rival explanations that could lead to revisions or rejections of emerging hypotheses.

Research Question 3–Mixed Methods

Although the qualitative and quantitative results were analyzed separately, I also considered them together using a convergent mixed methods design. The integration involved the merging of the results from each of the datasets so that comparisons could be made and a more complete understanding could emerge than could have been provided by either the quantitative or qualitative results alone. The theoretical framework of Bandura’s Theory of Reciprocal Determination was used to guide the merging of the data.

I first examined relationships in the change scores on each instrument using scatterplots. I examined these plots for evidence of patterns, for example if the teachers clumped together in meaningful ways (e.g., by teaching level, experience, or discipline). I also estimated the correlation between the pre-to-post PD change scores and the post-to-follow-up change scores to determine if gains in scores over the course of the PD were related to losses occurring after the PD.

The interviews included two forced-choice items that asked the teachers to report on their current levels of PD-related behavior. Item-level responses were analyzed and examined using quantitative methods, including the use of descriptive statistics and histograms. I estimated the correlation between the self-reported behavior measures to test my hypothesis that they were related. I also estimated the correlation between the teachers' scores on the instruments at the follow-up time point with self-reported behavior changes. I hypothesized that, with the exception of the content knowledge assessment, the scores on the Inquiry Teaching Assessment, pedagogical content knowledge assessment, and self-efficacy assessments would be positively correlated with behavioral changes. Independent-samples *t*-tests were conducted to explore the difference between grade-level groups and self-reported behaviors.

I assumed that the intervals between the single-item Likert-scale responses and ordered categorical data were equivalent for statistical purposes when interpreting the results of these self-reported behavior questions. However, the magnitude of differences between categories was most likely interpreted differently by each teacher respondent (Harrison, 2020). For the purposes of this study, these scores are only one measure of the teachers' behavior; they serve as a starting point for the more detailed information provided in the interviews.

I compared the teachers' descriptions of the extent of their teaching changes with their scores on the quantitative instruments and their self-reported behaviors using independent-samples *t*-tests. Finally, to further examine the relationships between the quantitative and qualitative data I constructed two joint display tables anchored with two measures of the teachers' behavior change to show how the results from the two datasets were congruent or contradictory. One measure of behavior change was culled from the teacher responses to a forced-response question and one measure was culled from the teachers' descriptions of the extent to which their teaching had changed in the interviews. Behaviors were corroborated with emergent environmental factors from the interviews. I synthesized these results in a graphical joint display of mixed methods inferences.

Positionality

Onwuegbuzie defined confirmation bias as "the tendency for interpretations and conclusions based on new data to be overly congruent with *a priori* hypotheses" (2002, p. 18). A researcher's first line of defense in avoiding this threat is to approach data analysis with neutrality. This involves being aware of, and transparent about, your own perspective as a researcher and your theoretical predispositions. Bias cannot be fully eliminated, but it is important to be aware of your own preferences so you can control or counter them by being conscientiously open-minded and disinterested.

I was the project manager and one of the facilitators of the original PD. In addition to being mindful of my own prejudices throughout this study, I took precautions to establish impartiality in this follow-up study. Some of these steps were detailed in the previous sections, including the use of an evaluator to conduct the interviews and the use of external raters to code the responses to the Inquiry Teaching Assessment instrument. I also consulted extensively with colleagues and gave myself time to immerse myself in the data in order to be receptive to multiple and unanticipated perspectives, and to allow myself to be creative with the inductive process of discovery. In addition, the collection of rich, thick data in the form of verbatim interview transcripts "minimizes confirmation bias by facilitating the testing of emerging theories" (Onwuegbuzie & Leech 2007, p. 244).

While my close connection to the design and implementation of this PD could be seen as a limitation, in that I might be too invested in the results to examine the data dispassionately, my intimate knowledge of the program can also be viewed as an asset. As an "embedded researcher"

I was able to apply my in-depth understanding of the program and its context when interpreting the data (e.g., Sprague, 2005). The process of considering both quantitative and qualitative evidence to synthesize findings highlights the role of the researcher in using their professional judgment to attend to and consider multiple types of practical and theoretical evidence, for both proposed and rival hypotheses, and to make an overall evaluative judgment. I am uniquely qualified to do this synthesis.

Perhaps most importantly, I think the time that has passed between the PD intervention and this study has made me more critical of the PD and has heightened my curiosity about the PD's effectiveness. Similarly, although respondents were not anonymous, the level of social desirability should have receded as the participants had not been involved in the project for some time. Spending time away from the project, maintaining a conceptual framework, and continually keeping research questions in mind enhances objectivity and reduces bias (Onwuegbuzie & Leech, 2007). Overall, I feel that the benefits of my participation as the lead researcher on this study outweighed the risks.

CHAPTER 4

RESULTS

In this chapter I present the results of the quantitative instruments and the qualitative interviews. I then consider these results together in a convergent mixed-methods analysis. At the end of each substantive section I present a graphic synthesis of the results in that section.

I first present the quantitative data. The quantitative results are presented by assessment (i.e., all the results associated with the Inquiry Teaching Assessment are presented together). These results include the descriptive statistics for each assessment at each of the three time points (pre-PD, post-PD, and follow-up) and illustrations of how the scores shifted over time with histograms and boxplots. I also share the reliability of the scales and the results of the paired-samples *t*-tests or Wilcoxon signed-rank tests between time points. These analyses address the first research question (statistical differences in scores by time point). I also share the results of independent-samples *t*-tests between grade bands on each instruments' scores and correlations between scores and teacher experience for each assessment. These analyses address the first research question sub-question (relationship of variables to scores on instruments). Lastly, for each assessment, I present my examination of the patterns of gains or losses between time points using scatterplots. This is one of the analyses that address the third research question (patterns in implementation). Thus, research question 1, as well as aspects research questions 1A and 3, are addressed by quantitative instrument.

I then present the results from the forced-choice interview question about participation in significant PD after the PD under study. I share the results of the independent-samples *t*-tests between the scores on each of the assessments at the follow-up time point of the teachers who attended significant additional PD and those who did not. This concludes the presentation of the results for research question 1A. I end this section by presenting the correlations between each of the instruments' scores and between time points for the same variable. Thus, the correlations exploring the credibility of the quantitative constructs are addressed after the findings of the quantitative assessments are presented.

I then present the results of the qualitative interview data to address the second research question (factors that enhance or hindered implementation). Themes that emerged from the qualitative interview are structured and presented according to Bandura's Theory of Reciprocal Determination. I begin this section by presenting results that correspond with the quantitative instruments and the emergent theme of the importance of high-quality PD. I then focus on the interactions among environmental, behavioral, and personal factors. I end this section by describing the teachers' self-perceived changes in their teaching practice (behavioral changes) as a result of the PD and the frequency of their implementation of PD activities.

At the end of this chapter I address the bulk of the third research question (patterns in implementation). I first present the remainder of the quantitative data, two forced-choice interview questions, including the descriptive statistics for each question. For the two self-reported implementation questions I illustrate the score distributions with histograms. I also present correlations between the implementation questions and the scores on the quantitative instruments. Finally, I merge the quantitative and qualitative data by comparing the teachers' descriptions of the extent of their teaching changes with scores from the quantitative instruments and self-reported behaviors using independent-samples *t*-tests. To further examine the relationships between the quantitative and qualitative data I share two joint display tables that show how the results from the two datasets were congruent or contradictory. The tables were structured based on two measures of behavior change—one self-reported and one based on the teachers' descriptions of the extent of their behavior change due to the PD. Behaviors were aligned with quantitated qualitative data concerning emergent environmental factors extracted from the interviews. This chapter ends with a synthesis of the mixed methods results in a graphical joint display.

Quantitative Assessment Results—Research Questions 1 and 1A

Four instruments were administered to the participants at three time points, (a) at the beginning of the PD (pre-PD), (b) after completion of the PD (post-PD), and (c) at least 2.5 years after the completion of the PD (follow-up). This section describes the findings of the quantitative assessments. As described above, for each instrument results will be presented that address research question 1, as well as aspects research questions 1A and 3. A summary of the results associated with research questions 1 and 1A ends this section.

Inquiry Teaching Assessment

Table 4.1 presents the descriptive statistics for the Inquiry Teaching Assessment at different time points. Results are in logits based on scores from a multi-facet rating-scale model under a Rasch analysis framework (Andrich, 1978; Linacre, 2010). All measures from each of the time points were combined into one set and modeled together to ensure they were on the same scale. The Rasch reliability index was .94; this reliability index is higher than it would be for any single time point. The person-separation index, or how well a set of items is able to separate persons measured, was 4.10. This implies that respondents were able to be reliably separated into four different levels. A person-separation index of 1.50 is considered acceptable, 2.00 is good, and 3.00 is excellent (Duncan et al., 2003). A chi-square test of independence, using a fixed estimate as the two raters utilized in this study were experts in scoring this construct, showed there was some rater variation, χ^2 (2, N = 819) = 52.5, $p < .01$.

Table 4.1

Descriptive Statistics of the Inquiry Teaching Assessment

Time	*M* [a]	*SD*	se_M	Min	Max
Pre-PD	-0.83	1.29	.27	-2.85	2.50
Post-PD	0.48	1.31	.27	-1.82	2.65
Follow-up	-0.08	1.27	.26	-2.05	3.17

Note. *N* = 23. Results in logits based on scores from a multi-facet rating-scale model under a Rasch analysis framework. All time period measures were modeled together to ensure they were on the same scale. The Rasch reliability index was .94.

The average at the pre-PD time point was -0.83 logits; this equates to approximately a score of 0.5 on the Inquiry Teaching Assessment rubric (a 9-point scale that ranged from -1 to +3, see Appendix D) which reads: "There is some evidence that the teacher understands what teaching science as inquiry looks like in the classroom, or there is evidence that the teachers' knowledge appears to be developing." The average at the post-PD time point was .48 logits; this equates to approximately an average score of 1.1 on the Inquiry Teaching Assessment rubric. The scale for 1.0 reads:

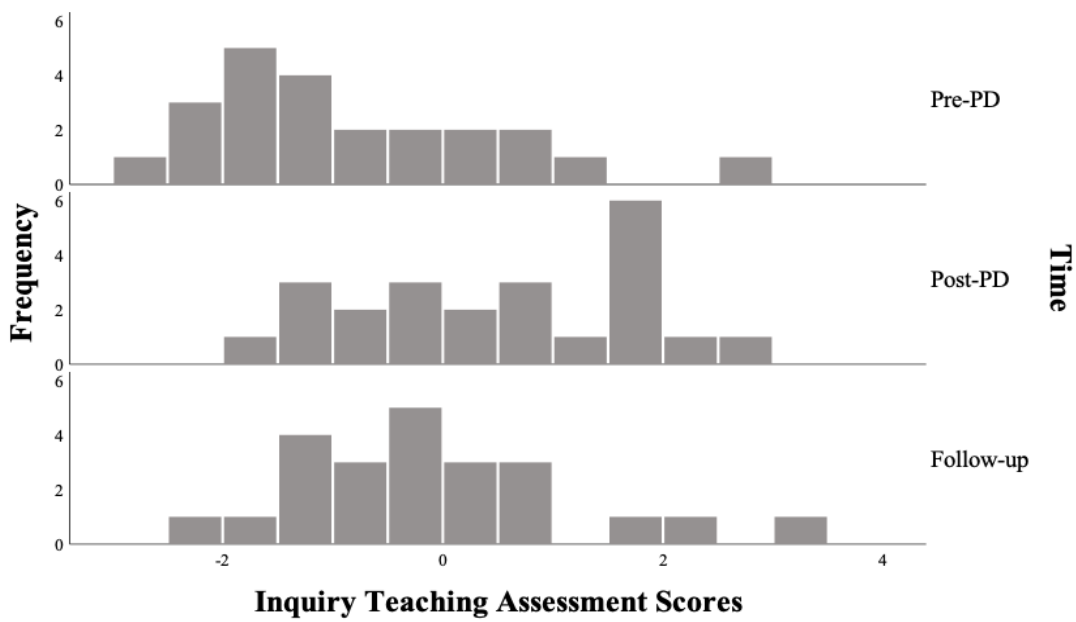
> There is good evidence that the teacher has a fairly good breadth or depth of understanding of what teaching science as inquiry looks like in the classroom. The response may include mostly accurate and somewhat precise descriptions of what inquiry-based instruction would look like. There is no evidence of misconceptions.

The average at the follow-up time point was -0.08 logits; this equates to approximately an average score of 0.9 on the Inquiry Teaching Assessment rubric. Thus, the description of the teachers' scores, 1.0 on the Inquiry Teaching Assessment rubric, did not change from the post-PD time point.

Inquiry Teaching Assessment scores at each time point are illustrated in a histogram (Figure 4.1) and boxplot (Figure 4.2). From these depictions, you can see that the scores on this instrument skew toward the higher end of the scale before the PD, indicating a large proportion of the teachers had little understanding of what inquiry-based teaching looked like in a classroom prior to the PD. Scores became more evenly distributed after the PD and at the follow-up time point. Both the histogram and boxplot illustrate there is a clear outlier at the follow-up time point.

Figure 4.1

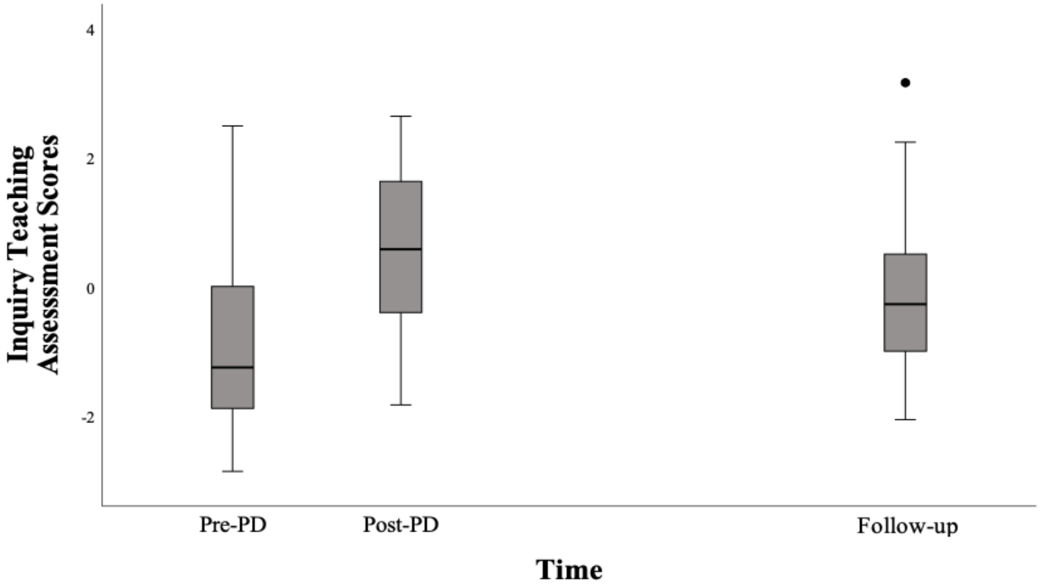
Histogram of the Inquiry Teaching Assessment Scores at Different Time Points



Note. $N = 23$. Results in logits based on scores from a multi-facet rating-scale model under a Rasch analysis framework. All time period measures were modeled together to ensure they were on the same scale.

Figure 4.2

Boxplot of the Inquiry Teaching Assessment Scores at Different Time Points



Note. $N = 23$. Results are in logits based on scores from a multi-facet rating-scale model under a Rasch analysis framework. All time period measures were modeled together to ensure they were on the same scale.

As a preparatory step for evaluating the difference between scores at each time point, the Shapiro-Wilks test of normality was used to examine the statistical assumption of normality on each of the respondents' difference scores between each time point (e.g., the difference between the pre-PD and post-PD time points). None of the three sets of difference scores (pre-PD to post-PD, post-PD to follow-up, and pre-PD to follow-up) violated the assumption of normality (Appendix G). Therefore, I conducted paired-samples *t*-tests.

First, I explored the change in Inquiry Teaching Assessment scores pre- to post-PD for the sample of 23 teachers in order to understand if the results from this subsample aligned with what was found with 28 teachers in the prior PD evaluation (Duncan Seraphin, 2014). Row 1 of Table 4.2 indicates that there was an increase in scores by 1.31 logits from pre- to post-PD that was statistically significant ($p < 0.01$). This result demonstrates that the change in Inquiry

Teaching Assessment scores between these periods in this sample is similar to what was observed in the initial evaluation (Duncan Seraphin, 2014); confirming that this subsample is reflective of the original results. I then compared post-PD to follow-up scores to determine if there was a decrease in scores on the Inquiry Teaching Assessment after the conclusion of the PD. Row 2 of Table 4.2 indicates that the scores decreased by 0.55 logits from the post-PD to the follow-up time point, a statistically significant decline ($p < 0.05$). Finally, I compared the pre-PD and follow-up scores to determine if the follow-up scores remained higher than the pre-PD baseline scores. Row 3 of Table 4.2 indicates that there was a statistically significant increase in scores by 0.76 logits from the pre-PD to the follow-up time point ($p < .05$). Applying the Benjamini-Hochberg adjustment to the paired-samples *t*-tests did not alter the statistical significance of the results. Thus, although approximately half of the gains in scores on the Inquiry Teaching Assessment from pre- to post-PD were lost by the follow-up, they were still significantly higher than the pre-PD scores.

Table 4.2

Paired-samples t-Test Results Between Different Time Points of the Inquiry Teaching Assessment

Time 1	Time 2	Estimate	*SE*	*t*(22)	*p*	Statistical conclusion ^{a}	95% CI	Hedges' *g*
Pre-PD	Post-PD	1.31	0.31	4.16	< .001	Sig**	[0.66, 1.97]	0.99
Post-PD	Follow-up	0.55	0.28	2.01	.028	Sig*	[-1.12, 0.02]	0.43
Pre-PD	Follow-up	0.76	0.34	2.22	.018	Sig*	[0.05, 1.47]	0.59

Note. $N = 23$.

$^{*}p < .05$. (one-tailed) $^{**}p < .01$. (one-tailed)

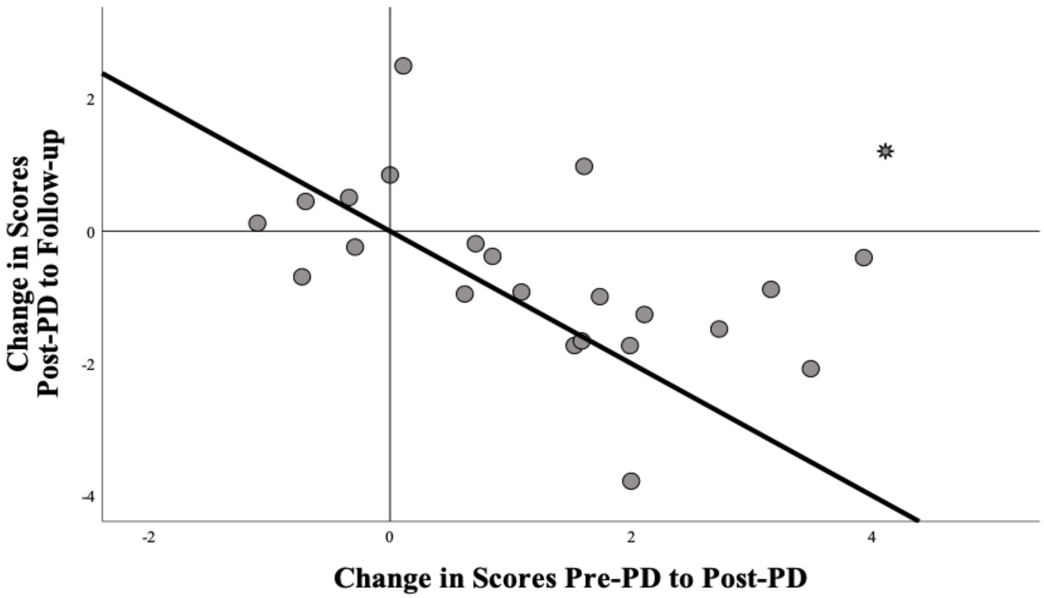
^{a} Statistical conclusion after applying Benjamini-Hochberg adjustment

Figure 4.3 shows a scatterplot of the relationship between the change in scores on the Inquiry Teaching Assessment from pre- to post-PD along the x-axis, and the change in scores from the post-PD to the follow-up along the y-axis. Thus, a positive x-coordinate indicates that an individual had a positive gain in score from pre- to post-PD, while a negative x-coordinate indicates a decrease in score from pre- to post-PD. Similarly, a positive y-coordinate indicates an individual had a positive gain in score from post-PD to the follow-up, while a negative y-coordinate indicates a decrease in score from post-PD to follow-up. Most of the participants (n = 14, 60.8%) are in the lower right quadrant, indicating their scores increased from pre- to post-PD but then decreased from post-PD to the follow-up. There was a negative correlation between the pre- to post-PD gains and the post-PD to follow-up losses, $r(22) = -.34$, $p = .115$ (two-tailed). This correlation became significant when the outlier in the upper right corner of Figure 4.3 was removed from the analysis, $r(21) = -.52$, $p = .013$ (two-tailed). The bold line in Figure 4.3 has a slope of negative one and passes through the origin (y = -x). The individuals falling close to this line had pre- to post-PD gains approximately equivalent to their follow-up losses. Compared to their score pre-PD, these individuals' scores showed no or very little change on this instrument at the follow-up.

There was no significant correlation between experience (number of years teaching prior to the PD) and scores on the Inquiry Teaching Assessment at the pre-PD, $r(22) = .42$, $p = -.179$ (two-tailed), and follow-up time points, $r(22) = .23$, $p = .282$ (two-tailed). There was also no significant correlation between experience and score changes from pre- to post-PD, $r(22) = .31$, $p = .145$ (two-tailed), or post-PD to follow-up, $r(22) = .04$, $p = .85$ (two-tailed). Removing an outlier teacher with 30 years of teaching experience did not change the significance of these results.

Figure 4.3

Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Inquiry Teaching Assessment



Note. $N = 23$. Results in logits based on scores from a multi-facet rating-scale model under a Rasch analysis framework. All time period measures were modeled together to ensure they were on the same scale. The line has a slope of negative one. The individuals falling near this line had pre- to post-PD gains approximately equivalent to their follow-up losses.

I conducted independent-samples *t*-tests to evaluate the difference in magnitude between the middle- and high-school teachers' change scores pre- to post-PD and post-PD to follow-up to see if the PD intervention had a differential effect on inquiry-based teaching knowledge based on participants' grade bands (Table 4.3). For this comparison, the one elementary school teacher (5th grade) was excluded. Row 1 of Table 4.3 indicates that there was a difference in change scores between the two groups of 0.56 logits pre-PD, which was not statistically significant. Rows 2 and 3 of Table 4.3 show that the difference in change scores pre- to post-PD (1.22 logits), and post-PD to follow-up (0.04 logits) was similarly not significant. Thus, it is unsurprising that at the follow-up time point (Row 4) the 0.09 logit difference in scores between the high-school and middle-school teachers' scores was not significant.

Table 4.3

Independent-samples t-Test Results of the Score Changes between Grade Bands on the Inquiry Teaching Assessment

Scores	Middle-school teachers (*n* = 15)		High-school teachers (*n* = 7)		Est.	*SE*	*t*(20)	*p*	95% CI	Hedges' *g*
	M	*SD*	*M*	*SD*						
Pre-PD	-0.98	1.17	-0.42	1.61	0.56	0.60	0.94	.360	[-1.82, -2.10]	0.41
Change scores: Pre- to post-PD	1.62	1.44	0.40	1.34	1.22	0.65	1.89	.073	[-0.13, 2.57]	1.35
Change scores: Post-PD to follow-up	-0.55	1.08	-0.51	1.90	0.04	0.63	0.06	.950	[-1.36, 1.28]	0.03
Follow-up	0.09	1.17	-0.53	1.53	0.62	0.59	1.04	.309	[-0.62, 1.85]	0.46

Content Knowledge Assessment

Table 4.4 presents the descriptive statistics for the content knowledge assessment at different time points. On the pre-PD assessment, the mean score was 21.4 out of 41 (52%). This low baseline percent-correct score was by design, as the instrument was modified (as described in Chapter 3) to more clearly detect changes over time. As expected, scores rose at the post-PD time point, and then fell at the follow-up. Content knowledge assessment scores at each time point are illustrated in a histogram (Figure 4.4) and boxplot (Figure 4.5). From these depictions, you can see that the scores on this instrument are fairly evenly distributed at the pre-PD and follow-up time points, but skew to the higher end of the score range at the post-PD time point. There is a low outlier at the post-PD time point.

Table 4.4

Descriptive Statistics of the Content Knowledge Assessment

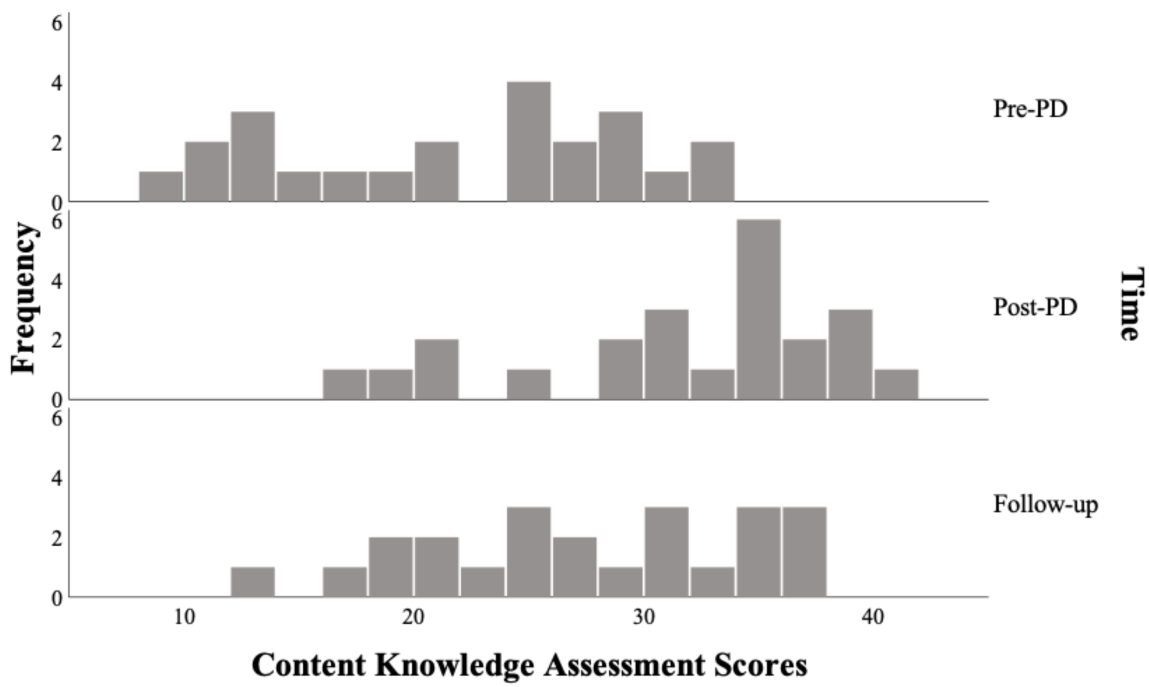
Time	*M* [a]	*SD*	se_m	Min [a]	Max [a]	KR-20
Pre-PD	21.39 (52.2%)	7.49	1.56	9 (22.0%)	33 (80.5%)	.86
Post-PD	31.22 (76.1%)	6.90	1.43	16 (39.0%)	40 (97.6%)	.89
Follow-up	26.87 (65.6%)	6.95	1.45	12 (29.3%)	36 (87.8%)	.86

Note. 41 item test with 23 teachers, KR-20 = reliability estimate

[a] Percentages are percent-correct scores.

Figure 4.4

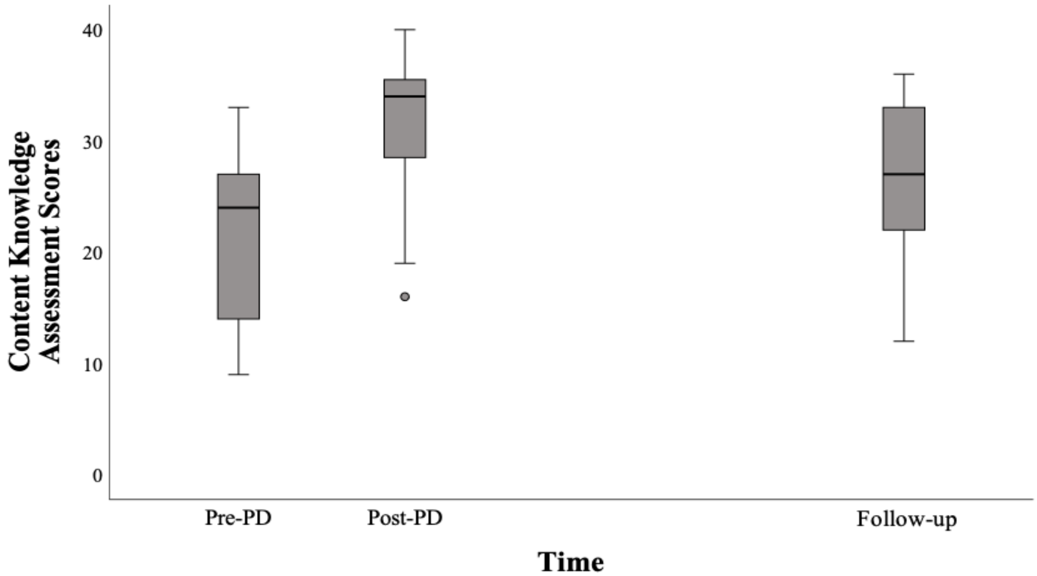
Histogram of the Content Knowledge Assessment Scores at Different Time Points



Note. $N = 23$.

Figure 4.5

Boxplot of the Content Knowledge Assessment Scores at Different Time Points



Note. $N = 23$.

The three sets of difference scores on the content knowledge assessment (pre-PD to post-PD, post-PD to follow-up, and pre-PD to follow-up) were normally distributed, as accessed by the Shapiro-Wilks test (Appendix G). Therefore, I conducted paired-samples *t*-tests to evaluate the difference between scores at each time point (Table 4.5). First, I explored the change in content knowledge assessment scores pre- to post-PD for the sample of 23 teachers in order to determine if the results from this subsample aligned with what was found with the teachers in the PD evaluation (Duncan Seraphin, 2014). Row 1 of Table 4.5 indicates that there was an increase in scores by 9.83 from pre- to post-PD, a statistically significant gain ($p < 0.01$). This result demonstrates that the change in content knowledge assessment scores between these periods in this subsample along, with a smaller set of items, is similar to what was observed previously (Duncan Seraphin, 2014). I then compared post-PD scores to follow-up scores to examine if there was a decrease in scores on the content knowledge assessment since the end of the PD.

Row 2 of Table 4.5 indicates that there was a statistically significant decrease in scores (-4.35, $p < 0.01$) from the post-PD to the follow-up time point. Finally, I compared the pre-PD and follow-up scores to determine if the follow-up scores were higher than the pre-PD baseline scores. Row 3 of Table 4.5 indicates that there was a statistically significant increase in scores by 5.48 from the pre-PD to the follow-up time point ($p < .01$). Statistical conclusions at these levels remained after applying the Benjamini-Hochberg adjustment. Overall, the teachers retained about 44% of pre- to post-PD content knowledge gains. Although about half of the gains in scores from pre- to post PD were lost by the follow-up, they were still significantly higher than the pre-PD scores.

Table 4.5

Paired-samples t-test Results Between Different Time Points of the Content Knowledge Assessment

Time 1	Time 2	Estimate	*SE*	*t*(22)	*p*	Statistical conclusion [a]	95% CI	Hedge's *g*
Pre-PD	Post-PD	9.83	0.91	10.80	< .001	Sig**	[7.94, 11.71]	1.34
Post-PD	Follow-up	4.35	0.71	6.10	< .001	Sig**	[-5.82, -2.87]	0.70
Pre-PD	Follow-up	5.48	0.74	7.39	< .001	Sig**	[3.94, 7.01]	0.75

Note. $N = 23$.

**$p < .01$. (one-tailed)

[a] Statistical conclusion after applying Benjamini-Hochberg adjustment

There was a significant negative correlation between pre-PD scores and pre- to post-PD scores on the content knowledge assessment, $r(23) = -.42$, $p = .045$ (two-tailed). Thus, the higher the teachers' pre-PD content knowledge score, the smaller their score increase was likely to be between the pre-PD and post-PD time points. Or, stated more simply, the more content knowledge the teachers had prior to the PD the less knowledge they gained over the course of the PD.

I conducted independent-samples *t*-tests to determine if differences in content knowledge change scores were in part driven by content knowledge differences between the middle- and high-school teachers. Row 1 of Table 4.6 indicates that the middle-school teachers had significantly less content knowledge than their high-school counterparts on the pre-PD content knowledge assessment (-9.3, $p < .01$). Row 2 of Table 4.6 shows that the middle-school teachers gained significantly more on the content knowledge assessment pre- to post-PD than the high-school teachers (5.47, $p < .01$). Similarly, row 3 of Table 4.6 shows that the middle-school teachers lost significantly more in the post-PD to follow-up period (-3.9, $p < .05$). Thus, the middle-school teachers came into the PD with less aquatic science content knowledge than the high-school teachers; over the course of the PD their scores increased more than the high-school teachers' scores and after the PD their scores decreased more than the high-school teachers' scores. However, at the follow-up time point the difference between the middle-school and high-school teachers' content knowledge assessment scores (-7.75) was not statistically different. Applying the Benjamini-Hochberg adjustment did not change these levels of significance. There was no correlation between the teachers' level of experience and scores on the pre-PD content knowledge assessment, $r(23) = -.28$, $p = .19$ (two-tailed), or score changes on the assessment from pre- to post-PD or post-PD to follow-up, both $r(23) = .24$, $p = .28$ (two-tailed). Removing the outlier teachers with 30 years of experience did not change the significance of these correlations.

The teachers that started with high pre-PD content knowledge assessment scores (> 66%, $n = 6$) only lost, on average, 1.5 points between the post-PD and the follow-up time points. The teachers that started with low pre-PD content knowledge assessment scores (< 37%, $n = 7$) lost, on average, 3.9 points. The teachers who got between 37% and 66% correct (n = 11) on the pre-PD content knowledge assessment lost, on average, 6.4 points. Thus, the teachers who had low content knowledge scores pre-PD lost less content knowledge than those who had average pre-PD content knowledge scores. In other words, the teachers with low content knowledge scores retained more of their gains post-PD to follow-up than the teachers with average content knowledge scores.

Table 4.6

Independent-samples t-Test Results by Grade Bands on the Content Knowledge Assessment

Scores	Middle-school teachers (n = 15)		High-school teachers (n = 7)		Est.	*SE*	$t(18)$	*p*	Stat. con.[a]	95% CI	Hedges' *g*
	M	*SD*	*M*	*SD*							
Pre-PD	18.53	7.12	27.85	4.14	9.32	2.41	3.86	.001	Sig**	[4.27, 14.38]	1.42
Change scores: Pre- to post-PD	11.47	3.98	6.00	2.89	5.47	1.69	3.24	.002	Sig**	[1.95, 8.97]	1.42
Change scores: Post-PD to follow-up	-5.47	3.29	-1.57	2.07	3.90	1.36	2.86	.010	Sig*	[-6.74, -1.05]	1.25
Follow-up	24.53	6.44	32.29	5.59	7.75	2.83	2.73	.188	Not Sig	[-13.65, -1.83]	1.22

*$p < .05$. (one-tailed) **$p < .01$. (one-tailed)

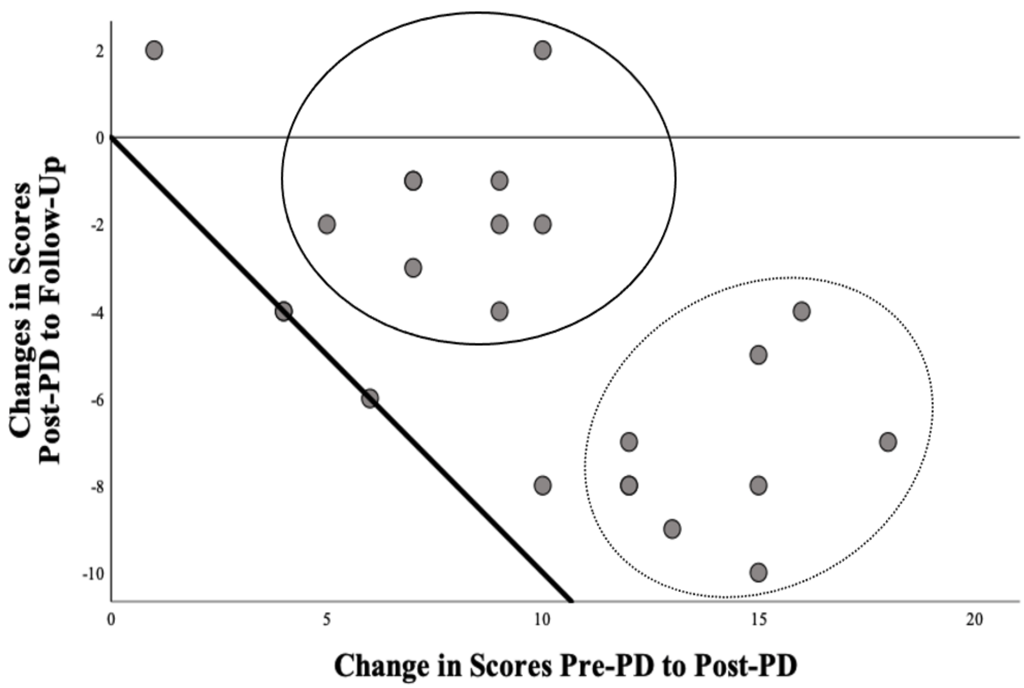
[a] Statistical conclusion after applying Benjamini-Hochberg adjustment

Figure 4.6 shows a scatterplot of the relationship between the change in scores on the content knowledge assessment from pre- to post-PD and the change in scores from the post-PD to the follow-up time point. Everyone in the study gained from pre- to post-PD; there are no negative x-coordinates. There was a significant negative correlation between the pre- to post-PD gains and the post-PD to follow-up losses, $r(22) = -.61$, $p = .002$ (two-tailed). This means that the more the teachers gained pre- to post-PD, the more they lost by the follow-up. The line in Figure

4.6 has a slope of negative one and goes through the origin (y = -x). The individuals falling close to this line had pre- to post-PD gains approximately equivalent to their follow-up losses. Although three of the teachers received the same score at the pre-PD and follow-up time points, none of the teachers scored less on the follow-up than the pre-assessment.

Figure 4.6

Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Content Knowledge Assessment



Note. $N = 23$. The line has a slope of negative one. The individuals falling near this line had pre- to post-PD gains approximately equivalent to their follow-up losses. The individuals in the solid circle had moderate pre- to post-PD gains that were maintained over time. The individuals in the dotted circle had substantial pre- to post-PD gains and moderate losses over time.

The individuals in the solid circle in Figure 4.6 had moderate pre- to post-PD gains that were maintained over time. The individuals in the dotted circle had substantial pre- to post-PD gains and moderate losses over time. The teachers in the solid circle (n = 7, M = 10.9 years, SD = 5.05), were slightly more experienced than those in the dotted circle (n = 9, M = 8.7 years, SD = 6.74)—but only by about two years. The teachers within the solid circle had marginally higher pre-PD content knowledge assessment scores (n = 7, M = 23.6, SD = 9.50) than those in the dotted circle (n = 9, M = 19.1, SD = 4.94). The four-point difference in scores between the two groups is due in part to the solid circle containing four out of the seven high-school teacher participants, whereas the dotted circle does not include any of the high-school teachers and contains the only elementary teacher. High-school teachers generally have more training in content, and thus would be expected to have higher pre-PD content knowledge scores, especially in the disciplines they teach. No associations were discernible between content knowledge scores and teacher discipline (e.g., physical science, biological science, or earth science).

Pedagogical Content Knowledge Assessment

Table 4.7 presents the descriptive statistics for the pedagogical content knowledge assessment at different time points. The range of scores can be understood to roughly correspond to the interval values on the 5-point Likert-type scale of this instrument. For example, the mean pedagogical content knowledge score at the post-PD time point was 3.9, which is close to the response category “often” (4 on the interval scale), indicating the teachers thought they “often” implemented the inquiry activities in the item prompts. The range in scores is, at most, 1.4 points on the scale. At no time point did scores drop below 3.3 points. Pedagogical content knowledge scores at each time point are also illustrated in a histogram (Figure 4.7) and boxplot (Figure 4.8). At the pre-PD time point, scores were bimodal with the two modes towards the higher and lower end of the distribution. The post-PD time point scores were more normally distributed. At the follow-up time point, scores were evenly distributed with a clear mode at ~3.8 points.

Table 4.7

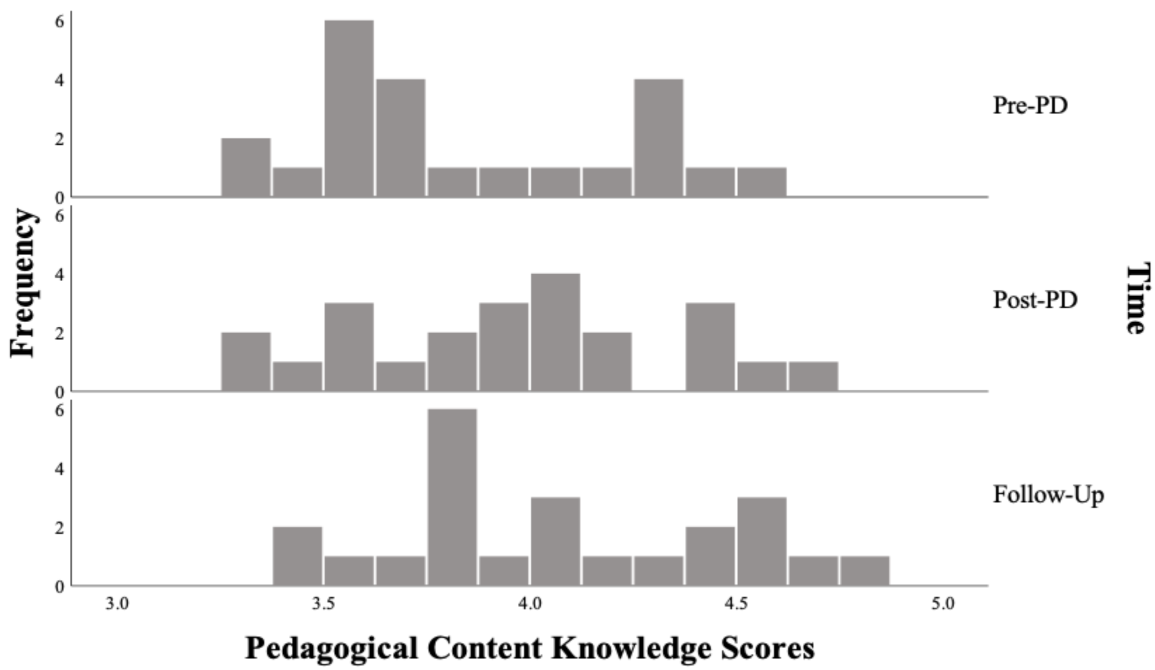
Descriptive Statistics of the Pedagogical Content Knowledge Assessment

Time	*M*	*SD*	se_m	Min	Max	Cronbach's α
Pre-PD	3.84	0.39	.08	3.3	4.6	.92
Post-PD	3.94	0.39	.08	3.3	4.6	.94
Follow-up	4.06	0.41	.09	3.4	4.8	.94

Note. *N* = 23.

Figure 4.7

Histogram of the Pedagogical Content Knowledge Assessment Scores at Different Time Points

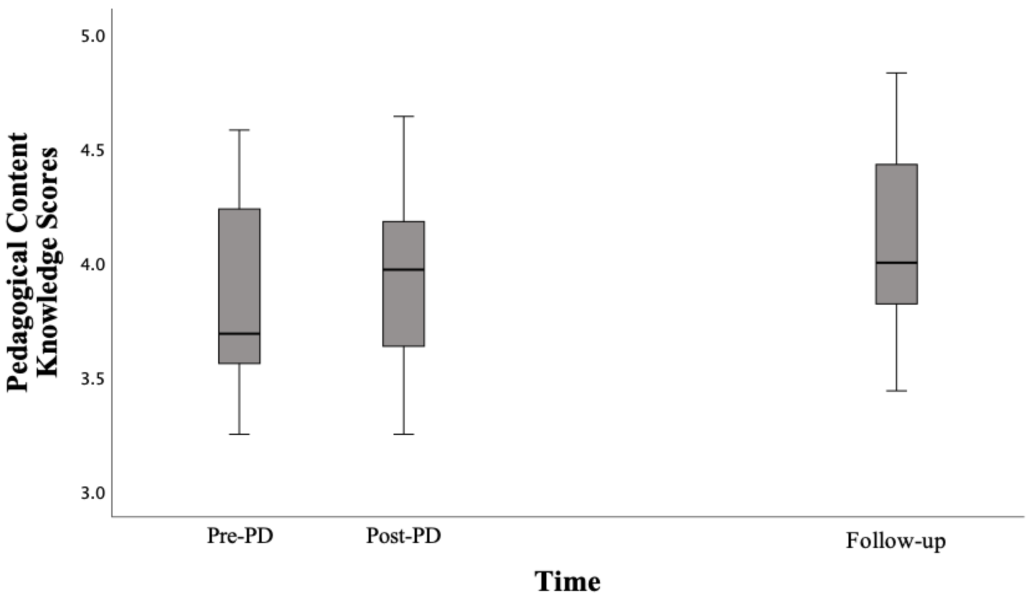


Note. *N* = 23.

The Shapiro-Wilks test of normality indicated one of the three difference scores examined for the pedagogical content knowledge assessment (post-PD to follow-up) deviated from normality ($W = .794$, $p < .001$; Appendix G). This set of difference scores was examined with a histogram (Appendix G). Non-normality was strong; two outliers affected the mean and caused the variance to expand. As the hypothesis of normality was rejected, Wilcoxon signed-rank tests, rather than paired-samples *t*-tests, were used to evaluate the difference between scores at each time point.

Figure 4.8

Boxplot of the Pedagogical Content Knowledge Assessment Scores at Different Time Points



Note. $N = 23$.

First, I examined the change in pedagogical content knowledge assessment scores pre- to post-PD for the sample of 23 teachers in order to understand if the results aligned with what was found in prior work (Duncan Seraphin, 2014). Row 1 of Table 4.8 indicates that the increase in scores from pre- to post-PD was not statistically significant. In the original PD evaluation, there was also no statistically significant difference in scores between these time points (Duncan Seraphin, 2014). This result demonstrates that the pedagogical content knowledge assessment scores between these periods in the subsample is similar to what was observed with all the 28 teachers in the PD. I then compared post-PD to follow-up scores to determine if there was a change in scores since the conclusion of the PD. Row 2 of Table 4.8 indicates that the increase in scores from the post-PD to the follow-up time point was statistically significant ($p < .05$). Finally, I compared the pre-PD and follow-up scores to determine if the follow-up scores were higher than the pre-PD baseline scores. Row 3 of Table 4.8 indicates that there was a statistically significant increase in scores from the pre-PD to the follow-up time point ($p < .01$). Statistical conclusions at these levels remained after applying the Benjamini-Hochberg adjustment. In summary, although there was no statistically significant change in pedagogical content knowledge assessment scores from pre- to post-PD, the teachers' scores were statistically higher at the follow-up time point than both the pre-PD and post-PD time points, although the measured scores differences were small.

Table 4.8

Wilcoxon Signed-Rank Test Results Between Different Time Points of the Pedagogical Content Knowledge Assessment

Time 1	Time 2	*z*	*p*	Statistical conclusion [a]	*r*
Pre-PD	Post-PD	0.86	.389	Not sig	0.13
Post-PD	Follow-up	2.52	.012	Sig*	0.37
Pre-PD	Follow-up	2.66	.008	Sig**	0.39

Note. $N = 23$.

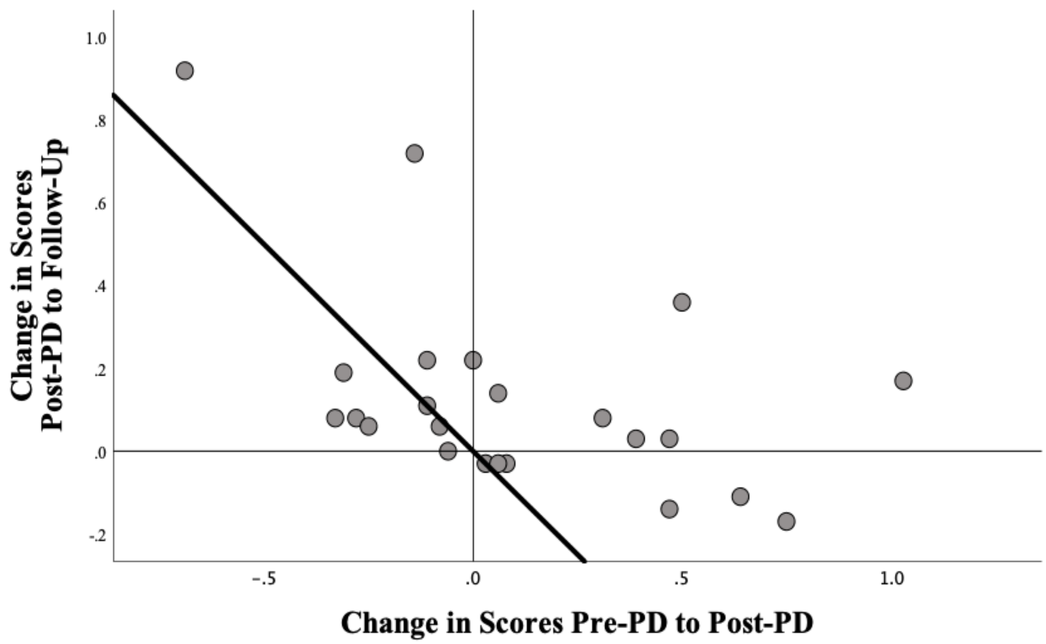
*$p < .05$. (two-tailed) **$p < .01$. (two-tailed)

[a] Statistical conclusion after applying Benjamini-Hochberg adjustment

Figure 4.9 shows a scatterplot of the relationship between the change in scores on the pedagogical content knowledge assessment from pre- to post-PD and the change in scores from the post-PD to the follow-up time point. About half of the teachers' scores increased from pre- to post-PD (12 of 23 teachers). All but one of the remaining teachers, who showed either no change or a decrease in scores from pre- to post-PD, showed a gain in pedagogical content knowledge from the post-PD to follow-up time point. Only six teachers' scores declined post-PD to follow-up, and only three of these lost more than about 0.05 points.

Figure 4.9

Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Pedagogical Content Knowledge Assessment



Note. $N = 23$. The line has a slope of negative one. The individuals falling near this line had pre- to post-PD gains approximately equivalent to their follow-up losses.

There was a significant negative correlation between the pre- to post-PD gains and the post-PD to follow-up losses, $r(22) = -.48$, $p = .022$ (two-tailed). This means that the more teachers gained pre- to post-PD, the more they lost at the follow-up. The line in Figure 4.9 has a slope of negative one and passes through the origin ($y = -x$). The individuals falling close to this line had pre- to post-PD gains approximately equivalent to their follow-up losses.

There was no significant correlation between experience and score changes on the pedagogical content knowledge assessment from pre- to post-PD, $r(22) = .07$, $p = .759$ (two-tailed); or post-PD to follow-up, $r(22) = -.11$, $p = .628$ (two-tailed).

I conducted independent-samples *t*-tests to evaluate the difference between the middle- and high-school teachers' pedagogical content knowledge scores (Table 4.9). The difference of 0.20 in pre-PD scores between these grade bands was not statistically significant. Similarly, the difference in magnitude between the middle- and high-school teachers' change scores pre- to post-PD (0.16), and change scores post-PD to follow-up (0.12) was not statistically significant. The was also no statistical significance between the scores of the middle-school teachers and the high-school teachers at the follow-up time point (difference of 0.08).

Table 4.9

Independent-samples t-Test Results by Grade Bands on the Pedagogical Content Knowledge Assessment

Scores	Middle-school teachers (n = 15)		High-school teachers (n = 7)		Est.	*SE*	$t(22)$	p	95% CI	Hedges' g
	M	*SD*	*M*	*SD*						
Pre-PD	3.75	0.36	3.95	0.44	0.20	0.17	1.13	.271	[-0.56, 0.17]	0.66
Change scores: Pre- to post-PD	0.16	0.45	0.00	0.31	0.16	0.19	0.84	.411	[-0.24, 0.56]	0.43
Change scores: Post-PD to follow-up	0.17	0.30	0.05	0.08	0.12	0.12	1.00	.328	[-0.13, 0.36]	0.37
Follow-up	4.01	0.43	4.01	0.41	0.08	0.19	0.39	.698	[-0.33, 0.48]	0.00

$^{*}p < .05$. (two-tailed)

Self-efficacy Assessment

Table 4.10 presents the descriptive statistics for the self-efficacy in science teaching assessment at different time points. The range of scores can be roughly interpreted as interval scale values corresponding to the 6-point Likert-type scale of this instrument. For example, the mean self-efficacy score at the pre-PD time point was 3.45, which is about half-way between the response category end points of the scale (1 = low ability, 6 = high ability), indicating that the teachers thought they had an "average" ability to implement the inquiry activities in the item prompts. Self-efficacy assessment scores at each time point are also illustrated in a histogram (Figure 4.10) and boxplot (Figure 4.11). From these depictions you can see there was a ceiling effect with this instrument. At the pre-PD time point, scores were fairly evenly distributed. The post-PD time point scores have a more normal distribution, with a slight skew towards the higher scores. At the follow-up time point scores were again more evenly distributed.

Table 4.10

Descriptive Statistics of the Self-efficacy Assessment

Time	*M*	*SD*	se_M	Min	Max	Cronbach's α
Pre-PD [a]	3.45	1.04	.22	1.9	5.8	.97
Post-PD	4.79	0.70	.15	3.3	5.9	.98
Follow-up	4.58	0.66	.14	3.5	6.0	.94

Note. *N* = 23.

[a] The pre-PD instrument was administered retrospectively, at the same time as the post-PD instrument.

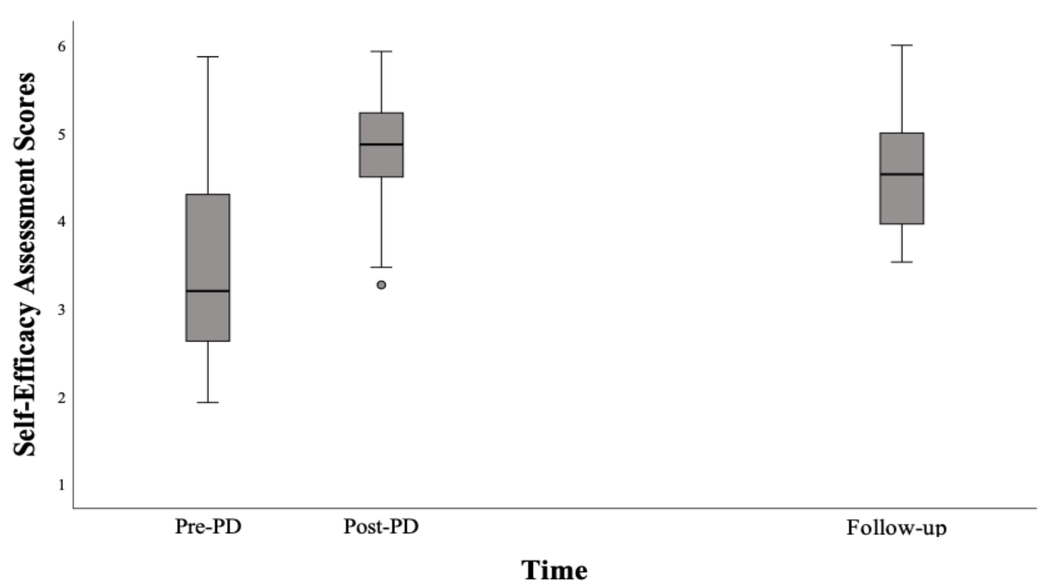
Figure 4.10

Histogram of the Self-efficacy Assessment Scores at Different Time Points

Note. $N = 23$.

Figure 4.11

Boxplot of the Self-efficacy Scores at Different Time Points



Note. $N = 23$.

The Shapiro-Wilks test of normality indicated two of the three difference scores examined for the self-efficacy assessment (pre-PD to post-PD and post-PD to follow-up) deviated from normality (Appendix G). This set of difference scores was examined with histograms (Appendix G). Non-normality was not severe for either of these differences. For the pre- to post-PD time point difference, the histogram showed that the distribution was bimodal, not due to outliers. The post-PD to follow-up time point difference histogram skewed to the right; it was non-normal, but also not due to outliers. For this second set of difference scores the Shapiro-Wilks test was barely significant ($W = .913$, $p = .048$). As violations of non-normality were not severe, I conducted paired-samples *t*-tests to evaluate the differences between scores at each time point (Table 4.11). However, these slight deviations from normality should be considered a limitation in interpreting results.

First, I explored the change in self-efficacy assessment scores pre- to post-PD for the sample of 23 teachers in order to understand if the results from this subsample aligned with what was found with the full sample of 28 teachers in the prior PD evaluation (Duncan Seraphin, 2014). Row 1 of Table 4.11 indicates that there was an increase in scores by 1.34 from pre- to post-PD that was statistically significant ($p < 0.01$). This result demonstrates that the change in self-efficacy assessment scores between these periods in this subsample is similar to what was observed in the prior evaluation (Duncan Seraphin, 2014). I then compared post-PD to follow-up scores to determine if there was a decrease in scores on the self-efficacy assessment since the PD. Row 2 of Table 4.11 indicates that the decrease in scores by 0.21 from the post-PD to the follow-up time point was statistically significant ($p < 0.05$). Finally, I compared the pre-PD and follow-up scores to determine if the follow-up scores were higher than the pre-PD baseline scores. Row 3 of Table 4.11 indicates that there was a statistically significant increase in scores by 1.14 from the pre-PD to the follow-up time point ($p < .01$). Statistical conclusions for these paired-samples *t*-tests at these levels remained after applying the Benjamini-Hochberg adjustment. In summary, although some gains pre- to post-PD were lost by the follow-up time point, the teachers still scored significantly higher at the follow-up than on the pre-PD assessment. While scores were significantly lower at the follow-up than the post-PD time point, the effect size was comparatively small. Thus, the loss in self-efficacy between the post-PD and the follow-up was minor compared to the gain in scores from pre-PD to post-PD.

Table 4.11

Paired-samples t-Test Results Between Different Time Points of the Self-efficacy Assessment

Time 1	Time 2	Estimate	*SE*	*t*(22)	*p*	Statistical conclusion [a]	95% CI	Hedge's *g*
Pre-PD	Post-PD	1.34	1.14	5.67	< .001	Sig**	[0.85, 1.84]	1.55
Post-PD	Follow-up	0.21	0.47	2.14	.022	Sig*	[-0.01, -0.41]	0.28
Pre-PD	Follow-up	1.14	1.01	5.41	< .001	Sig**	[0.70, 1.57]	1.35

Note. $N = 23$.

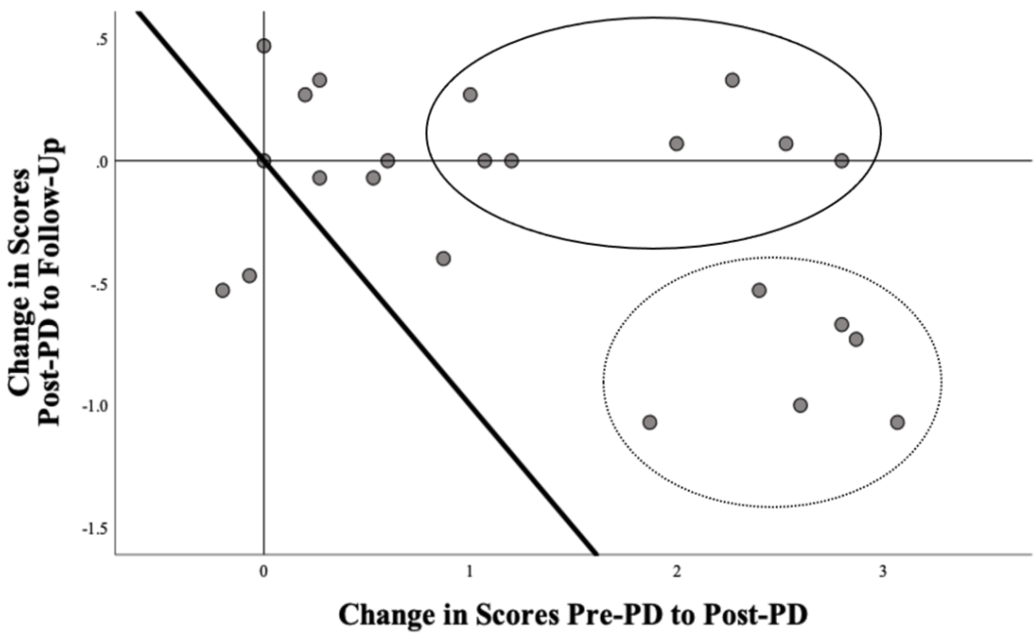
*$p < .05$. (one-tailed), **$p < .01$. (one-tailed)

[a] Statistical conclusion after applying Benjamini-Hochberg adjustment

Figure 4.12 shows a scatterplot of the relationship between the change in scores on the self-efficacy assessment from pre- to post-PD and the change in scores from post-PD to the follow-up time point. Only three of the teachers did not show gains from pre- to post-PD. Almost half of the teachers (10 of 23) did not lose any of their pre- to post-PD gains at the follow-up time point or gained in scores at the follow-up. There was a significant negative correlation between the pre- to post-PD gains and the post-PD to follow-up losses, $r(22) = -.47$, $p = .025$ (two-tailed). This means that the more the teachers gained pre- to post-PD, the more they lost at the follow-up. The line in Figure 4.12 has a slope of negative one (y = -x). The individuals falling close to this line had pre- to post-PD gains approximately equivalent to their follow-up losses.

Figure 4.12

Scatterplot Relationship Between the Change in Pre- to Post-PD Scores and Post-PD to Follow-up Scores on the Self-efficacy Assessment



Note. $N = 23$. The line has a slope of negative one. The individuals falling near this line had pre- to post-PD gains approximately equivalent to their follow-up losses. The individuals in the solid circle sustained their pre- to post-PD gains over time. The individuals in the dotted circle gained pre- to post-PD but had moderate losses post-PD to follow-up.

The individuals in the solid circle ($n = 7$, 30%) in Figure 4.12 sustained their pre- to post-PD gains over time. The individuals in the dotted circle ($n = 5$, 22%) gained pre- to post-PD, but had losses in self-efficacy post-PD to follow-up. The teachers in the solid circle had, on average, a couple more years of teaching experience than those in the dotted circle (solid circle $M = 12$ years, dotted circle $M = 9.8$ years). There was no clear difference in discipline or grade levels taught between the two groups; the solid circle includes two high-school teachers and five middle-school teachers; the dotted circle includes one high-school teacher and four middle-school teachers.

There was a significant correlation between the teachers' pre-PD self-efficacy scores and years teaching science, $r(22) = .48$, $p = .019$ (two-tailed). However, when an outlier teacher was removed who had 30 years of teaching experience and a gain of 2.8 points on the Likert-type scale from pre- to post-PD, this relationship was no longer significant, $r(22) = .41$, $p = .056$ (two-tailed).

I conducted independent-samples *t*-tests to evaluate the difference in magnitude between the middle- and high-school teachers' pre-PD self-efficacy assessment scores, their change scores pre- to post-PD, their change scores post-PD to follow-up, and at the follow-up time point (Table 4.12). The difference in pre-PD self-efficacy assessment scores of 0.23 between the middle-school teachers and the high-school teachers was not statistically significant. There was also no statistically significant difference between the magnitude of change scores pre- to post-PD (0.29) or post-PD to follow-up (0.09) between the middle-school and high-school teachers. Thus, it is not surprising that at the follow-up time point the two groups were also not statistically different (difference of 0.03).

Table 4.12

Independent-samples t-Test Results by Grade Bands on the Self-efficacy Assessment

Scores	Middle-school teachers (n = 15)		High-school teachers (n = 7)		Estimate	*SE*	t(20)	p	95% CI	Hedges' g
	M	*SD*	*M*	*SD*						
Pre-PD	3.27	0.95	3.49	0.88	0.23	0.42	0.54	.597	[4.27, 14.38]	0.21
Change scores: Pre- to post-PD	1.50	1.18	1.21	1.05	0.29	0.52	0.56	.583	[1.95, 8.97]	0.25
Change scores: Post-PD to follow-up	-0.23	0.45	-0.13	0.55	0.09	0.22	0.42	.677	[-6.74, -1.05]	0.20
Follow-up	4.54	0.63	4.57	0.77	0.03	0.31	0.09	.929	[-0.67, 0.62]	0.04

Effects of PD Participation After PD Under Study

At the start of the interview, the teachers were asked if they had participated in any significant PD since the PD under study had ended (dichotomous response—yes or no). I asked this question because prior research on the PD under study had shown that whether the teachers had participated in substantial PD was a significant factor in student learning (Duncan Seraphin et al., 2017). Although not specified, based on their answers the teachers interpreted "significant PD" to mean "significant science PD." Eight of the 21 teachers who answered this question (38%) indicated that they had participated in significant science PD and 13 teachers (62%) indicated that they had not. The term "significant" was not defined and thus was left up to the teachers to interpret. However, a few teachers indicated that they used the PD under study as a benchmark for "significant." For example, one of the teachers who responded "no" said "nothing as in-depth as the TSI PD." Three of the eight teachers who responded yes, they had participated in other significant PD, said they had attended PD focused on the NGSS, which were being rolled out in Hawaiʻi around the time of this follow-up study.

I conducted independent-samples *t*-tests to evaluate the difference between the teachers with and without significant PD experiences after the PD under study and their scores on the follow-up instruments. There were no statistically significant differences between the teachers who reported that they participated in additional substantial science PD and those who did not for any of variables as operationalized by the instruments (Table 4.13). However, it is important to note that the small sample size favors this null hypothesis. All of the estimates are higher for the additional PD group, which suggests that there may be some systematic variance in these variables explained by this additional-PD variable.

Table 4.13

Independent-samples t-Test Follow-Up Instrument Results Between the Teachers With and Without Significant PD Experiences After the PD Under Study

Instrument	No additional PD (n = 13)		Additional PD (n = 8)		Estimate	*SE*	$t(19)$	*p*	95% CI	Hedges' *g*
	M	*SD*	*M*	*SD*						
Inquiry Teaching Assessment	-0.02	1.42	0.12	1.01	0.14	0.58	0.24	.812	[-1.07, 1.35]	0.10
Content knowledge assessment	24.92	7.04	28.75	6.61	3.83	3.09	1.24	.231	[-2.65, 10.30]	0.54
Pedagogical content knowledge assessment	4.02	0.46	4.18	0.36	0.15	0.19	0.79	.440	[-0.25, 0.55]	0.45
Self-efficacy assessment	4.54	0.76	4.66	0.61	0.12	0.32	0.37	.719	[-0.55, 0.78]	0.27

Note: N = 21

Comparisons Between and Across Quantitative Assessments

To examine the credibility of the results from the quantitative assessments, I estimated the correlations between time points for the scores on each instrument using z-scores (Table 4.14). The comparisons between each of the time points of the content knowledge assessment were correlated the most tightly (*r* > .82 for all comparisons), followed by the pedagogical content knowledge assessment, the self-efficacy assessment, and then the Inquiry Teaching Assessment (*r* < .48 for all comparisons). The post- PD to follow-up time point scores were correlated most strongly for all of the instruments; the scores on the content knowledge assessment also had a high correlation between the pre-PD to follow-up time points. The pre-PD to post-PD score correlation was weakest for three of the four instruments, the exception being the Inquiry Teaching Assessment.

Table 4.14

Score Correlations Within Instruments Across Different Time Points

Instrument	Pre-PD	Post-PD	Follow-up
Inquiry Teaching Assessment			
Pre-PD	–		
Post-PD	.32	–	
Follow-up	.19	$.48^{*}$	–
Content knowledge assessment			
Pre-PD	–		
Post-PD	$.82^{**}$	–	
Follow-up	$.88^{**}$	$.88^{**}$	–
Pedagogical content knowledge assessment			
Pre-PD	–		
Post-PD	$.46^{*}$	–	
Follow-up	$.59^{**}$	$.80^{**}$	–
Self-efficacy assessment			
Pre-PD	–		
Post-PD	.19	–	
Follow-up	.36	$.77^{**}$	–

Note. $N = 23$. All scores in z-scores.

$^{*}p < .05$. (two-tailed) $^{**}p < .01$. (two-tailed)

I expected high, significant score correlations between all administrations of each instrument as they are measuring the same construct. This was true for all instruments and administrations except pre- to post-PD and pre-PD to follow-up score correlations for both the Inquiry Teaching Assessment and the self-efficacy assessment. Score correlations were particularly low between the pre- and post-PD time points for the self-efficacy assessment and between the pre-PD and follow-up times points for the Inquiry Teaching Assessment (both r = .19). Low correlations indicate a regression in scores to the mean or other measurement error problem.

Pearson correlations on z-scores were also estimated between instruments' scores at each time point (Table 4.15). I expected no or weak score correlations across instruments as the instruments are designed to measure different constructs. This was true for all instruments except the pedagogical content knowledge assessment and the self-efficacy assessment. The scores on these instruments were significantly positively correlated across all time points—$r(22) = .50, p < .01$ (pre-PD), $r(22) = .80, p < .01$ (post-PD), and $r(22) = .70, p < .01$ (follow-up). Positive relationships among scores on instruments indicates they are measuring related constructs and may be considered evidence of convergent validity, but it can also be indicative of commonalities in administration and test structure. Both the pedagogical content knowledge assessment and the self-efficacy assessment had Likert-type scale items. In addition, they were administered back-to-back on a single questionnaire. The pedagogical content knowledge assessment was first, the self-efficacy assessment was second. While undoubtedly some of the reason for the high correlation between the pedagogical content knowledge and self-efficacy scores is due to the similarity between the constructs, it is difficult to distinguish this from bias due to method effects, or other bias such as acquiescence on the part of the participants.

Table 4.15

Score Correlations Across Instruments and Different Time Points

Time point by instrument	Inquiry Teaching Assessment	Content knowledge assessment	Pedagogical content knowledge assessment	Self-efficacy assessment
Pre-PD assessments				
Inquiry teaching	–			
Content knowledge	.31	–		
Pedagogical content knowledge	.11	-.05	–	
Self-efficacy	-.03	.04	.58**	–
Post-PD assessments				
Inquiry teaching	–			
Content knowledge	-.10	–		
Pedagogical content knowledge	.19	-.06	–	
Self-efficacy	.12	-.11	.80**	–
Follow-up assessments				
Inquiry teaching	–			
Content knowledge	-.07	–		
Pedagogical content knowledge	.24	-.17	–	
Self-efficacy	.22	-.17	.70**	–

Note. $N = 23$. All scores in z-scores.

**$p < .001$. (two-tailed)

In Summary—Results of Research Questions 1 and IA

Table 4.16 is a summary of the quantitative results and interpretation of each data source by research question. Figure 4.13 illustrates how the constructs are related to each other graphically.

Table 4.16

Summary of Results and Interpretation of Data Sources by Research Question

Data sources	Research question	Results: Pre-PD	Results: Pre-PD to post-PD	Results: Post-PD to follow-up	Results: Pre-PD to follow-up	Results: Follow-up	Interpretation
Inquiry Teaching Assessment [b]	1	—	Increase $g = 0.99$	Decrease $g = 0.43$	Increase $g = 0.59$	—	Scores declined, but sig. higher than baseline
Content knowledge assessment	1	—	Increase $g = 1.34$	Decrease $g = 0.70$	Increase $g = 0.75$	—	Scores declined, but sig. higher than baseline
	1A [a]	MS [c] lower scores	MS [c] increased more	MS [c] decreased more	—	No difference	PD mitigated incoming MS/HS [c] content differences
Pedagogical content knowledge assessment [b]	1	—	No change $r = 0.13$	Increase $r = 0.37$	Increase $r = 0.39$	—	Scores rose over time, indicating maturation effect or enhanced self-perceived capacity to facilitate inquiry teaching strategies
Self-efficacy assessment [b]	1	—	Increase $g = 1.55$	Decrease $g = 0.28$	Increase $g = 1.35$	—	Scores declined, but sig. higher than baseline; may have played factor in follow-up confidence
All quantitative assessments	1/1A Credibility of results	Score correlations were particularly low between the pre- and post-PD time points for the self-efficacy assessment and between the pre-PD and follow-up times points for the Inquiry Teaching Assessment (both $r = .19$); this indicates a regression towards the mean or measurement error problem. The scores on the pedagogical content knowledge assessment and the self-efficacy assessment were significantly positively correlated across all time points $r(22) = .5 – .8$; indicating they are measuring related constructs and/or indicative of commonalities in administration and test structure. On all instruments, the more teachers' scores rose over the course of the PD, the more their scores declined by the follow-up $r(22) = -.61 – -.47$; which may indicate a regression towards the mean.					

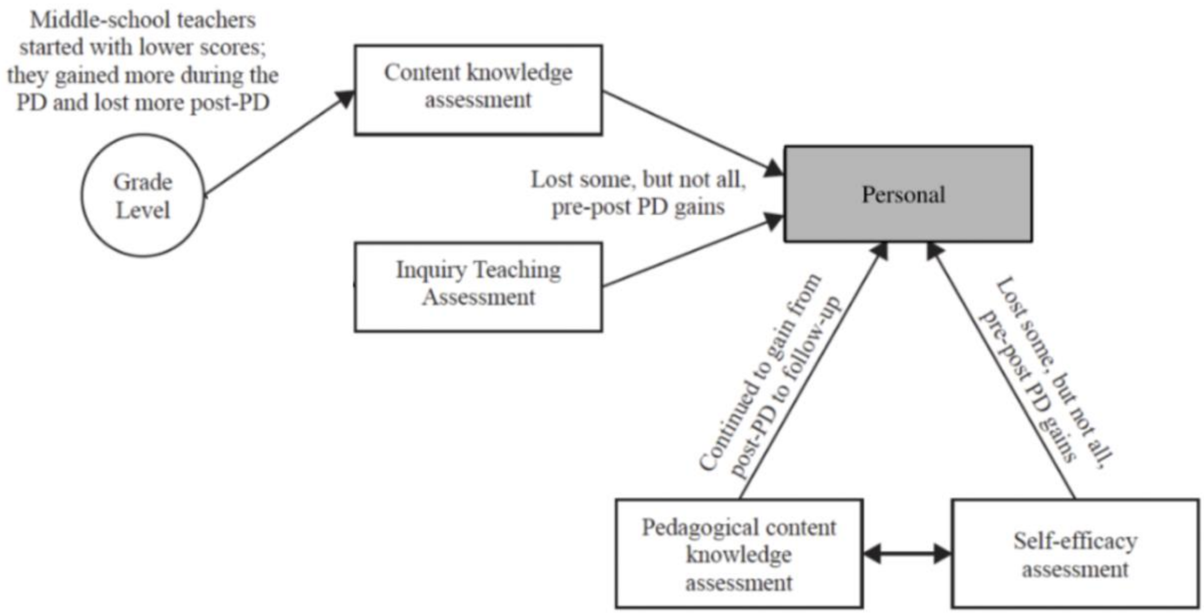
[a] All times points—no relationship between instrument scores and teacher experience

[b] All times points—no relationship between instrument scores and teacher experience or teacher grade level (research question 1A)

[c] MS = middle-school teachers, HS = high-school teachers

Figure 4.13

Synthesis of Quantitative Instrument Results



Note. The clear boxes denote the quantitative constructs measured. The grey box represents one of the corners of Bandura's Theory of Reciprocal Determination, the personal factor. The circle is a moderator factor. Arrows indicate relationships.

Research Question 1. The four quantitative instruments utilized in this study were designed to measure different aspects of the teachers' personal (cognitive) qualities. On three of the four instruments (Inquiry Teaching Assessment, content knowledge assessment, and self-efficacy assessment) scores increased significantly pre- to post-PD and regressed significantly toward the pre-PD baseline post-PD to follow-up. The declines were greater for the content knowledge assessment and the Inquiry Teaching Assessment than for the self-efficacy assessment. However, the teachers' follow-up scores on these instruments were still significantly higher than their pre-PD baselines (Hedges' $g = 0.5–0.8$, indicating a moderate statistical effect size). The teachers' scores on the pedagogical content knowledge assessment increased, but not significantly, pre- to post-PD and continued to increase post-PD to follow-up. At the follow-up time point, scores were significantly greater than at both the pre- and post-PD time points (g = 1.35; a very large effect). Thus, the teachers' self-efficacy, content knowledge, and pedagogical knowledge declined over time after the PD, but scores on the pedagogical content knowledge assessment increased over time. However, it is important to note that the self-efficacy assessment and pedagogical content knowledge assessment scores were highly correlated.

Research Question 1A. The grade level of the teachers was related to their content knowledge assessment scores. Additional PD, engaged in after the PD under study, did not significantly affected the teachers' quantitative assessment scores, although higher scores in the group that engaged in additional PD were found for each variable. Teacher experience was not significantly related to scores on the quantitative assessments. Thus, only one of the variables that was significant in the original PD evaluation was significant in this follow-up study.

Interview Qualitative Results—Research Question 2

Interviews were conducted with 22 out of 23 participants (96%) after completion of the quantitative instruments. This section describes the findings of the qualitative data collected during the teacher interviews. I arranged the interview results section roughly following the structure of Bandura's Theory of Reciprocal Determination. I first outline the results of an emergent environmental factor—that of the PD. I then describe findings aligned with the quantitative instruments, self-perceived cognitive (personal) changes in pedagogical and content knowledge. Enhancements and hindrances to PD implementation are organized by environmental, personal, and behavioral interactions. I then describe the teachers perceived behavioral (teaching) changes as a result of the PD. I end this section by summarizing the results of the second research question, including the use of a table and figure of the core findings. Throughout this section I have edited the teachers' quotes slightly for clarity and conciseness.

Environmental Factor

PD Embodied Adult Learning Principles. An unexpected theme that emerged from 82% of the teachers' interviews (18 of 22) was the importance of PDs embodying adult learning principles and being reflective of well-established core PD features (Table 4.17). Over two-thirds of the teachers (14 of 22, 64%) spoke of the PD as being high-quality—that the PD was "really good." The teachers asked the interviewer if the same facilitators were doing more PD they could participate in. Eight teachers (36%) said they appreciated the expertise of the PD facilitators, in terms of both their content and pedagogical expertise as well as their competence in planning and logistics. Approximately a quarter of the teachers liked the PD structure (e.g., year-long, with time for iterative implementation and reflection), that the PD was easy to incorporate in their practice, and how the PD built a professional learning community.

Only four teachers (of 22, 18%) mentioned PD enticements, such as the up to 12 PD credits the teachers were eligible for, the PD stipend of $1,200, and/or the provision of substitute teachers for four days over the course of the year to allow the teachers to attend PD sessions. This is in contrast to the more tangible benefits of supplies and curriculum, that the PD also provided and which a number of the teachers mentioned during the follow-up interview (13 of 22, 59%). Although the supplies and curriculum were more durable resources, and thus a longer-term benefit than short-term incentives like money or substitutes, the amount of PD credit available was substantial (the credit was optional and involved some additional requirements). The accumulation of PD credit enables teachers to move up in pay scale; 21 teachers (of 28 in the original PD, 75%) completed PD portfolios per PD module (many of the teachers completed

PD credit requirements for multiple modules). Having this outcome only mentioned by three teachers was surprising. Another surprising finding was that two of the experienced science teachers (with 20 and 30 years in the classroom) said the PD was non-threatening.

Table 4.17

Adult Learning Principles and Core PD Features

Adult learning principles and/or core PD feature	No. of teachers	Example quote(s) [a]
Quality PD	14 (64%)	The PD "was really good," the teachers said they would "love to do another course like this" and they "always recommend [the PD] to others." "I wanted to know if anything of that sort was going to be offered again in the future. Because it was like super good for my brain. Maybe a lot of work, but it was really good for me, and so, I would like to still be in the loop on any future offerings."
Expert PD facilitators	8 (36%)	"[The facilitators] do a great job with their planning and they really think about the teachers, which doesn't always happen." *Three teachers signed up for more PD by the same facilitators*. "When the OPIHI project came out [subsequent PD taught by same facilitators], I kind of was more excited about it, and inclined to do it, because I knew it would be done well, and I could feel confident and safe that I would be given the skills necessary to take my kids out to the intertidal."
"Good" PD structure	6 (27%)	"It forced me to look at, in detail, the teaching techniques and strategies in a little bit—in more detail. Breaking it down into parts, and that's where you, 'Ooooh! You know that, ooh ok! You know? That makes sense!'…TSI broke it down to each one of those parts… [and then] built it up a lot better." "[It was] wonderful to do it over a year's time...gives you plenty of time to digest it, experiment, and try different things while you adopted within your own, in the back of your mind, to put that in place." "It [the PD] only strengthened me, versus, made things more complex."
Development of teacher community	5 (23%)	"I feel like seeing what other teachers are doing is really powerful. So, the [online sessions], plus sitting at the table with somebody else when they went over a lesson that they had done was really very, very useful. Just like stepping outside of your own mind and seeing it from a different angle, was like 'oh, gosh, why didn't I think of that?' But, wait, ok, now I, now I know it, so now I'm going to incorporate that."

Adult learning principles and/or core PD feature	No. of teachers	Example quote(s) [a]
Easy to implement	5 (23%)	"It was set up in a way that was really easy and comfortable and made sense."
PD credit, stipend, and/or substitutes (to attend PD)	4 (18%)	"I got a lot of credit for it, yay! You know, there's nothing like another way to go up in the DOE system and maybe earn a little more money, thank you. I took my stipend and bought some new equipment, yay! That also was really good, because there's never any money ever, for anything!"
Engaging PD	3 (14%)	"it was really good, it was fun, hands-on"
Led to new opportunities	3 (14%)	"I ended up joining the OPIHI workshop [subsequent PD taught by same facilitators] experiences with my kids. And it kind of led to more things with my students, and for me as well, which was pretty amazing."
PD was non-threatening	2 (9%)	"I'm not as fearful to teach science, I think, because of being in the PD. The way it was taught was really relaxed so that it didn't feel threatening. So for sure I was able to absorb more…I think it's very non-threatening the way that you guys teach it, and in the order that you taught it in."

Note. $N = 22$.

[a] There is overlap in categories amongst quotes.

Two of the teachers, one of whom had become a district science resource teacher and another who had left the classroom to become an informal science educator, said that they learned about the pedagogy of PD by attending the PD under study. This experience helped them implement their own subsequent teacher PDs.

> A lot of PDs fail, they don't work because they don't have that before PD, during the PD, and then after PD. However, you guys did…I think that really helped me, professionally, grow because unfortunately a lot of the PDs that I go to, I don't implement, because it's just, 'it happened, it was so far away' or we get caught up in our own day-to-day things and there's no follow-through and there's no accountability because nobody else is checking on you, right? But with you guys it was quite different.

Personal Factors

Gained Pedagogical Knowledge. Two and a half years after the PD, 19 teachers (86%) said the PD had enhanced their pedagogical knowledge. This included all but one of the newer teachers (6 out of 7, 85%), and 13 out of 15 of the experienced teachers (87%). Approximately two-thirds of these teachers (12 out of 19, 63%) reported that their perceived enhanced pedagogical knowledge directly led to behavioral changes in their teaching practice. For example, the following teacher reported that understanding more about inquiry, and the inquiry circle of learning, resulted in a change in how she approached formatting lessons:

> So, you start with a hook, and then they create their own predictions, and then they share it out, and they adjust it through testing, and modeling, or applying, and then we come back to discussion. It's that circle of learning—I use it a lot.

While three teachers used vague terminology, describing the PD pedagogy as "just a good model for how to teach science," Table 4.18 lists the more specific pedagogical concepts the teachers said were significant. Half of the teachers (11 of 22, 50%) reported that one of their main takeaways from the PD was the concept that the scientific process is not linear, that the scientific method does not proceed in a lock-step fashion. The pedagogy the PD used, TSI, emphasized this in the way the TSI phases were represented, and we stressed this in how we modeled using the phases for planning and reflection in the PD. Although most of the teachers did not use the TSI terminology, the concept that the TSI inquiry cycle was designed to convey was retained. About a third of the teachers (8 of 22, 36%) were struck by the concept of metacognition and the importance of being reflective about their own teaching practices. A similar number of the teachers (7 of 22, 31%) said that learning there are many different ways to do science, conveyed in the PD through the use of the TSI modes of inquiry, was significant as they had previously been boxed in by the notion that "doing science" needed to involve experimentation and manipulation. Six teachers (27%) shared that they found the concept of the teacher-as-research-director in student-centered learning important.

> TSI made a lot of impact on me just in terms of how to engage students differently with any subject, and allows students to make mistakes and discover on their own with guidance from the teacher, as opposed to just direct instruction or differentiated instruction, like all those other different pedagogical approaches.

The teachers also reported that the purposeful questioning strategies covered in the PD, the idea of having students think like scientists, and the importance of initiation (a TSI phase), were all important concepts they learned in the PD that had become ingrained in their teaching philosophies. The teachers brought up all of these concepts spontaneously. They were asked what they thought was the most significant pedagogical method(s) or strategy(s) that they learned as a result of participating in the PD, but were not prompted to respond with any particular TSI term or concept.

Of the 19 teachers who said the PD enhanced their pedagogical understanding, one reported five of the pedagogical concepts listed in Table 4.18 as significant to their teaching practice. The rest of the teachers reported that one to three concepts were significant ($M = 1.9$, $SD = 1.23$).

Table 4.18

Pedagogical Concepts the Teachers Thought were Significant in the Follow-up Interview

Pedagogical concept	No. of teachers	Example quote(s) [a]
The scientific process is not linear (TSI phases; inquiry cycle)	11 (50%)	"I think the one important thing I learned was that in science we don't have a linear process of steps. That we go back and forth, and students can come up with new discoveries, or new predictions as they're working."
		"I know for me the main takeaway I had after TSI was realizing that there's not like one linear scientific method...So that has changed my own thinking of the scientific process. And therefore, that's what I relate to the students too."
The importance of metacognition/ being reflective (as a teacher)	8 (36%)	"[The PD gave me] a chance to look at my own teaching practices, how I view the student learner, how I want to enhance the student learners' experience within the classroom."
There are lots of different ways to do science (TSI modes)	7 (31%)	"The modes of science, that there are a lot of different ways to do science, not just like experiments, which is what a lot of people think."
		"Inquiry can also be looking stuff up in a book and researching, doing that kind of thing is also inquiry and not just hands-on experimenting all the time, which is an important part, but it's not the only part."
Teacher-as- research-director; student-centered learning	6 (27%)	"[TSI] allows me to put more of the work or the focus on the students and allow them to investigate in their own way, so with planning I try not to be so rigid in what I'm expecting of them, and let them kind of just investigate or explore on their own."
Questioning strategies	4 (18%)	"It's really important when our principals and administration go through and do a walk-through that they see that higher-level questioning happening via the teachers. However, it's not limited just to the teachers. It's also the students. They're observing and making sure that the students are asking those higher-level questions…[so] I implemented a lot. The primary reason is that—is the inquiry."
Having students think like a scientist (TSI scientific demeanors)	3 (14%)	"I really liked the part where there was an emphasis on thinking like a scientist. And so, I sort of make a point to emphasize that. That it doesn't matter what the topic is, if you're thinking like a scientist and you bring that to bear on any aspect of your—of much of your life, then you are a scientist."
Importance of initiation (one of the TSI phases)	3 (14%)	"I remember and today still apply the idea of trying to get them to—sort of beginning with a question. Like trying to lure them into thinking about the topic before any sort of lecture or anything is presented."

Note. $N = 22$.

[a] There is overlap in categories amongst quotes.

Gained Content Knowledge. Of the 22 teachers interviewed for this follow-up study, 14 (64%) said the PD enhanced their content knowledge, including 4 of the 7 high-school teachers (57%) and 10 of the 15 middle and elementary-school teachers (67%). Five of the 7 novice teachers (71%) and 9 of the 15 experienced teachers (60%) reported gains in content knowledge. Four of these teachers (29%) made explicit links between their perceived gains in content knowledge and behavioral changes in their classroom. There was no pattern in perceived content knowledge gained related to the discipline the teachers taught at the time of the PD.

Thirteen teachers shared that the content knowledge they gained in the PD was either in an area they had less prior knowledge or training in, and/or was directly related to content they had to cover in their classrooms. Six teachers said they lacked training in science; the PD served as an introduction to one or more science disciplines. For example, one teacher shared they had a history background, thus all of the content addressed in the PD was new to them. Another teacher shared that:

> [Ecology] was content that I didn't learn in high school. Even the science class that I took [at a university] to prepare me as a science teacher didn't teach any of those concepts. It taught microecology and macroecology, but not 'what does a scientist do with that information?' and then how do they come up with plots and transects and all this other stuff.

Four of the teachers said they had little training in a discipline they were teaching and the PD helped them deepen their knowledge. A middle-school teacher said she had a chemistry teacher who made her memorize the periodic table—which "didn't do much" for her understanding, "but I do actually need to teach the periodic table to 6th graders."

> [The PD] completely changed the way I was teaching chemistry 'cause I was doing such a bad job of it, right? I was doing things that didn't work for me when I was in high school and in middle school, 'cause I didn't have a better idea on how to do it. I think I got a lot of much better ideas about how to make that information more practical, more accessible, more logical.

A third group of the teachers shared that the content covered in the PD was directly related to what they were teaching and they felt the PD enhanced this knowledge. For example, a teacher that had a marine science undergraduate degree said:

> I think it was a good time for me to take the PD when I did because it was after my first year of teaching [marine science], and at that point was looking to kind of improve my teaching methods. I had gotten through the first year of teaching, but I wanted to be able to know the content better first of all, which helped a lot. TSI helped with that quite a bit.

Figure 4.14 shows the relationship between these three groups of teachers—those for whom the PD addressed content deficiencies due to their lack of training, those for whom the PD was directly relevant to their teaching practice, and those for whom both of these conditions were true.

Four of the teachers, including three non-marine science teachers, said the PD enhanced their understanding of how disciplines were integrated, and how different content could be related under the umbrella of, and taught through the lens of, aquatic science.

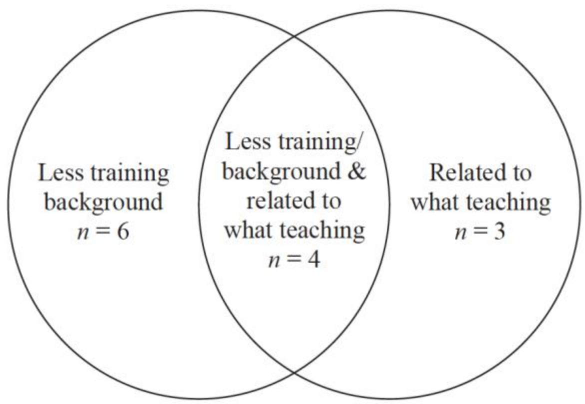
> I wanted to learn more, just plainly, about the ocean so I could teach it better. But I also liked the idea as the PD was going on, I was like 'Oh, I can actually weave—I can start from this point and have a lot of it mesh in.' I can kind of, not really, kill two birds with one stone, but I can address a lot more standards in this unit structured this way, coming

from the aquatic perspective, than I can from others. And that was a huge content shift — but I guess that's also more pedagogical too. In my head I was seeing more connections with the content.

When talking about specific content they thought was significant in the interview, two of the teachers mentioned the concept of one ocean—a foundational ocean literacy principle that was stressed throughout the PD. However, while the overarching theme of water in the PD tied content together, this focus could also limit implementation as most of the teachers did not teach marine science.

Figure 4.14

Venn Diagram of Relationship Between PD Content Knowledge and Teaching Relevance



Note. $N = 13$.

Erosion of Pedagogical and Content Knowledge. Hindrances to implementation relating to forgetting PD principles, terminology, and content emerged from the interviews of 11 of the teachers (out of 22, 50%). The most common hindrance (7 teachers, 32%) was forgetting the specifics of the pedagogical wording, apparent in the interviews due to stumbling over TSI terminology. For example, in reference to talking about one of the specific TSI inquiry phases, all of which start with the letter 'I', a teacher said "I don't remember which 'I' that is…" When asked if they were currently using the TSI pedagogical framework in their practice, one teacher said she was implementing some of the PD activities, and thus "I think I must be following, generally the TSI framework, but I couldn't know." She was honest about her lack of recall, but also knew the concept the pedagogical words were used to convey,

> I don't make specific reference, definitely not for the phases, 'cause that I have dropped completely… I can't even recall the phases actually, I mean I just remember what it looked like...like I know the purpose of the phases was that you could go through it fluidly, there was no one way to go about doing it, but—I don't know, something about it just didn't stick with me. Like, even the names. I think it's the concept of it that stuck with me, but not necessarily the specific phases and how to do the phase, and what each phase meant.

Several other teachers were also no longer explicitly using TSI terminology, but, as one teacher reported, the crux of the pedagogical concepts had become "part of [their] base teaching practice."

A slow forgetting of pedagogical concepts, not just terminology, over time was mentioned by six teachers (27%). While some of the teachers were a bit wistful about this loss and suggested additional PD sessions to jog their memories, others were more pragmatic, "[TSI] kind of fell to the wayside sometimes, just because there's so much stuff that you're always wading through. For me, I just tend to get overwhelmed with other stuff and then I get forgetful of things." Interestingly, only two of the teachers directly linked their forgetting of PD pedagogical concepts to a decline in their classroom implementation of PD material.

Two teachers (14%) who were actively using PD pedagogy and activities in their practices reported a loss of content knowledge since the PD, in part because they were not evenly implementing activities from all of the disciplines addressed by the PD.

> I would say I don't really use the sampling technique activities because it just doesn't feel like it's one of the 'big sciences,' like it doesn't fit into chemistry, biology, physics, it's ecology. And there's not a lot of stress on ecology in K–12 curriculum. So, unfortunately I don't really use those activities at all. Actually, when I was taking the content test again I completely had lost that knowledge, I was kind of just guessing because I haven't used it since TSI.

Interactions Between Environmental and Personal Factors

PD Affirmed Teaching Philosophy. The PD affirmed 11 of the teachers' beliefs (50%) about their teaching philosophy and practices. Ten of these teachers felt that the PD confirmed their pedagogical practices, the one elementary teacher in this study, who implemented all of the activities successfully with her 5th grade students, thought the PD confirmed her belief that young students "can absorb and learn and succeed…I didn't see why they couldn't do it, and they didn't either."

Fourteen of the 22 interviewed teachers (64%) were experienced, defined as having five or more years of classroom experience at the start of the PD. Seven of the 22 teachers (32%) were novices, defined as having three or less years of classroom teaching experience at the start of the PD. The experienced teachers were more likely to report that the PD affirmed their pedagogical practices (8 of 14 teachers, 57%), than the novice teachers (2 of 7 teachers, 29%). For example, a teacher who had been teaching for nine years at the start of the PD said:

> It [the PD] confirmed my belief that kids need to do science, they need to be scientists, and do the practices of science. And that inquiry does take a huge amount of time, and I'm not doing it wrong. 'Cause for awhile I thought 'Oh, I'm not doing something right because I'm taking so much time doing stuff.' And then when I took the class I realized, 'oh no, this is exactly what I've always wanted to do, and kind of been doing, but better now. It reinforced the things I was doing.

A teacher with ten years of experience reiterated this position. His response explains why, for some of the teachers, although they thought the PD was a positive learning experience they might not think it resulted in large teaching changes.

> I tried to do what it is TSI was advocating previously. And so, all it did was sort of formalize it for me in my mind. So, if I were to say on this 'not at all' to 'very much' scale how much did I change after the class, I wouldn't say that much, because I already agreed with their philosophy of education and tried to apply it as much as I could anyway.

However, most of the teachers thought that if even the PD confirmed existing practices, it also asked them to "take it up a notch." A teacher with 15 years of classroom experience said, "I normally would do that [teach via inquiry], now I do it more." Another teacher with nine years of experience said that, "the questions and answers that I used with students became more focused on modes of thinking, and that really put a little more shine on them." The teachers also reported that this confirmation led to behavioral changes like explicitly reminding students to embody the demeanors of scientists and being more purposeful in their lesson planning.

Three of the more experienced teachers who had 10, 15, and 20 years of experience said the PD was invigorating.

> I think TSI sparked something in me, turned me on a little bit for science and trying to share that out, and just to keep reigniting that a little bit, because sometimes you just need that little extra charge, a little extra gas.

Because the PD was aligned with many of the teachers' existing notions, one teacher said it had resulted in behavior change, "because it's something that reflects my own personal belief—my personal teaching philosophy—it's something that I'm more likely to put into practice, for years to come."

PD Increased Confidence. Seven out of 22 teachers (32%) said the PD increased their confidence. More of the new teachers at the start of the PD (4 of 7, 57%) reported gains in their confidence levels than the experienced teachers (3 of 15, 20%). Four teachers explicitly stated that their confidence stemmed from their gains in content and/or pedagogical knowledge,

> I think in the content there was so much for me for me to learn because I didn't have that background, so the whole chemistry module…I felt really weak in that area, but even so, being pushed to take those concepts back into the classroom and teach even the smallest part that I understood from that chemistry unit. I felt confident in getting that process across to my kids…knowing that I felt solid in just some real basics to take that back to my students I think was important.
>
> —teacher with 20 years of experience

Two teachers said their confidence grew from implementing the activities as students in the PD. Because the PD modeled the activities they were going to implement, the teachers said they were "not just floundering around, I know where I'm going," in large part because they understood how students would experience the activity.

> After doing the PD I was a lot more confident in what I was doing in a lot of the lessons, partly because it was the second time around, but also because I had a model of how to approach teaching, because in the workshops they did a good job of putting the teachers in the position of the students, so that we could kind of understand what the students are going through. So, that was really helpful. I kind of knew what kind of misconceptions the students might have, or how they might go about approaching the labs, because I had been in that position as a student-teacher.
>
> —Teacher with 2 years of experience

Four teachers explicitly reported that their gains in confidence led to changes in their teaching practice. Interestingly, two of the teachers spoke about the importance of trust and reputation in confidence,

> It makes me feel a little more solid in my practice that you guys know what the hell you're doing, so you made it right and I can just use that, and I can trust what I learned from you guys.
>
> —teacher with 20 years of experience

Another teacher with five years of experience shared that her confidence increased, "and the students know it, the students feel it."

Personal Preferences and Agency. Three teachers (14%) made the decision to revert to their pre-PD teaching style due to their personal preferences and teaching philosophies. One teacher with 18 years of experience said that while thinking of science as a process was an interesting pedagogical approach, implementing it was "wishful thinking;" as "doing science as scientists work [is] not realistic in the science classroom." She thought the PD was "idealistic." Another experienced teacher (15 years) was not keen on some aspects of the pedagogical framework. For example, he "could never quite get into the groove of the phase[s]" and could not "picture a teacher putting that in a classroom… it's like an extra layer of new jargon." While he "personally didn't see a whole lot of value in [the phases]," he did speak highly of other aspects of the pedagogical framework (e.g., the scientific demeanors). The third and final teacher who shared that their personal beliefs impeded implementation was a novice teacher with two years of experience who returned to her more traditional teaching style at the end of the PD.

> When I was doing the program, I did it according to how you guys gave it to us, but after that I did not...'Cause I prefer to do things more like how I teach, it's just more comfortable for me.

Curriculum Modifications for Students. Seven teachers, including three high-school and four middle-school teachers (32%), explicitly said that the activities and content were right at their students' level so they did minimal, if any, modifications to the activities for their students. Nine teachers (41%) said they adapted the lessons to align to their students' abilities. These teachers were making modifications based on their perception of their students' abilities and their beliefs in how their students would best navigate the PD material. Two of these teachers taught high school. One taught in a special education classroom. She explained that she simplified the language and modified some of the steps to make them more accessible to her students. The other high-school teacher was blunt, but vague, in sharing that he modified and customized everything, as he was "looking at where learners are, and how they're responding to different activities." Seven middle- and elementary-school teachers (of 14, 50%) were more specific in describing the modifications they made to the activities due to their perceived abilities of their students. Of these, both of the teachers who taught upper elementary school after the PD (100%) simplified the lessons, both of the 6th grade teachers (100%) reported modifications, one out of three (33%) of the 7th grade teachers made modifications, and one out of five (20%) of the 8th grade teachers modified the lessons. Thus, as the age and grade level of the students increased, fewer modifications were made due to the students' perceived abilities.

Most of the modifications were centered on simplifying the activities; the 5th grade teacher said, "I have to tone it down a little." This included, for her, front-loading more information "because they had less prior knowledge" and explaining questions to them that were beyond their reading abilities. However, she made a point to stress that:

> There wasn't anything that we did that they ever thought was too hard, or was too difficult to understand. I mean some of the vocabulary was a little challenging, but I would differentiate maybe by having it out on the desk for them, maybe more than you would in 9th grade—and I didn't expect them to memorize anything, but that didn't stop us, we just kept going.

However, one of the 6th grade teachers felt that her students came in "with just such limited experience or any kind of background knowledge of science" that even simplification was not feasible as she needed to "put other stuff on the side to get to more of the basic stuff." In contrast, a different 6th grade teacher, who worked with a lot of multilingual learners, reported using a number of the PD activities, particularly those designed to help build a foundation for understanding the scientific process. She described how she decreased the language load and adjusted the format of the student-facing materials to support her students, but did not change the steps of the activity itself. Most of the elementary- and middle-school teachers who reported making modifications due to their students' abilities took this adjustment in due course; they recognized that the activities were designed for older students and thus modifications were somewhat inevitable. However, even though two of the 7th grade teachers reported making no modifications to the activities and using at least some of them successfully in their classes, one thought the activities were “very high level, so I think it was really good for high school, middle school had a little bit of a hard time with the content.”

Only two of the middle-school teachers described modifications that I classified as clearly in conflict with the intended inquiry-based design of the PD activities. One of the teachers said she modified and shortened the activities so she would be teaching content as opposed to process. The other teacher said she taught her students *all* of the pertinent vocabulary before any of the activities, so they “understand what we’re trying to do before we do anything or we learn anything.”

Transfer of PD Principles Beyond PD “Focus” Class. During the PD, the teachers were asked to choose a “focus” class in which they would implement all of the PD activities and administer student surveys. The teachers were welcome to implement PD activities in additional classes as well, but these were optional. Eleven teachers (50%) reported applying PD principles outside of their original PD science “focus” class. Five teachers applied TSI concepts across science disciplines, for example:

> The lessons and the content that was in the TSI, it applies across a wide variety of the curriculum. When I was taking it [the PD] my biggest complaint was that, ‘oh, you know, you don’t directly deal with space, cause it [the PD] doesn’t. Like nowhere in there does it directly say ‘stars.’ But the content is applying what I teach in space. So, it gives it more meaning and context. Like when I talk about the electromagnetic spectrum and I do the Elodea lab with the Erlenmeyer flask…I’m doing that with my 7th graders this year to teach photosynthesis. But I’m also doing it with my 8th graders this year because

> they're measuring, how the plant has selected for those [certain electromagnetic] wavelengths. It connects in multiple ways.

Five teachers shared that TSI is applicable across disciplines. One teacher liked how we were intentional in the PD about pedagogical foci for each activity—what we were teaching and why we were teaching it. This teacher thought that the application of this intentionality helped her teaching practices, and "not just in science." Similarly, another teacher said, "There's a lot of concepts that I learned and now keep as part of my repertoire across subject areas besides just science." As this teacher described,

> I taught English and life science when I took the PD and it helped me in both of those areas because I transferred over some of those same things about what does a scientist look like and what's the difference between a law and a theory. I did that in my English class as well because I wanted them to be able to read things with a critical eye—and then I had Earth Science last year and I have no idea about Earth and space science but I did the whole startup again the same way I did the life science class with those activities, the beginning ones, of what does a scientist look like and what are the modes and what is, you know, the five I's on the chart.

Two teachers used the PD materials to develop new courses—a science camp in Africa and an oceanography course in a land-locked state.

Interactions Between Environmental and Behavioral Factors

Student Engagement and Learning. Of the 22 teachers who participated in the follow-up interview, 18 (82%) mentioned student engagement and/or student learning. Eleven teachers (50%) described evidence of student learning—in both content and scientific process understanding. "I could see them, you know, I could really see the light coming on for them," said one middle-school teacher, “it was the ability of the students to actually have fun doing it and that changed the amount that they understood and the amount that they wanted to participate.” Three teachers, including the special education teacher and a teacher with a large percentage of multilingual learners, liked how the PD activities made abstract concepts more concrete for their students. For example, the water properties activity helped one middle-school teachers’ students understand the difference between adhesion and cohesion, “I think it was a very abstract concept for them, but it brought it down to more concrete for them using the water and the penny. I think that was a huge ‘aha!’ for my students.” This teacher used the TSI phase diagram with her students to help them understand that science is not linear and that everyone approaches the scientific processes differently. She thought this helped her students,

> Especially with science fair, it helped them to say ‘oh, ok. You know what? I can go back and change my hypothesis. I can go back and make some more observations.’ ‘oh, I can change my procedure a little bit and test this instead of what I wanted to test.

Four teachers shared that this application of metacognition, as taught in the PD, enhanced their students’ critical thinking skills. One teacher shared, “it really did make them think about what are you doing at this time and how are you recording your results down—are you being careful about it?”

Student engagement was discussed by 15 teachers (68%). “Student investment that is intrinsic with the activity” was one teacher’s primary reason for using the PD activities, “the fact that students feel like they’re conducting real science, they’re answering meaningful questions, they’re a part of the thought process...it basically makes your job easier when you’re doing something that students are so invested in.” Another teacher used the PD activities to “get kids connected to the class faster… get engaged faster, not only with the class but with each other.”

One of the first things one of the teachers noticed was that the PD activities "started changing the attitudes of the students towards what the curriculum was." This teacher was one of the teachers observed for a concurrent study on the fidelity of PD implementation:

> It was wonderful that you could come down into my classroom [as part of Lawton's concurrent study (2017)] and kind of get to know the kids a little bit and get to see the difference from how they were, sitting there quiet and kind of shy and not really feeling like 'oh, we're very good at science' to when you came back and they're up and showing you things and excited and—for you to be able to see that change in them—it's really what the program was all about.

One teacher said her former students, who she would bump into around town, would "bring up the activities that we did in school that were related to what I learned [in the PD] and how much they learned and how fun it was." More specifically, four teachers said the PD activities made learning more relevant to their students. It was "really interesting helping students see the relevance in the information that they're learning instead of just thinking, 'oh this is just biology, and I have to pass it and move on.'"

PD Provided Resources for Implementation. As participants in the PD, the teachers were mandated to implement PD activities. To assist the teachers with this task, the PD provided a substantial amount of supplies to the teachers—enough for them to implement each of the PD activities with at least one class (in groups of 2–4 students, depending on the activity). Not only did the teachers report that having the supplies made it easier to start implementing the lessons immediately, they reported that having an adequate amount of supplies in their classrooms was pivotal in their practice. A middle-school teacher said this transformed her classroom.

> [The supplies] made it [my classroom] so different, it's not even the same in any realm, because it meant that instead of two or three students actually getting to do the experiments and see different ways that it could turn out, it meant that the whole class could. So, there wasn't anybody who didn't get a chance to see the results by doing them themselves….I think that it really changed everything—not just in helping me teach in a way that reached more students, but in their ability to say, "Ahh, I get it! Now it makes sense!"

These resources meant the PD was somewhat "easy" to implement. Six teachers (27%) said the PD was "practical" and "very easy to adapt" into their teaching. The PD passed one teacher's benchmark of "can I turn around and use this in my classroom? Yeah, I could. So it was really good." Another teacher said "we're always looking for PD classes, but not all of them are very valuable, or we can actually use in the classroom, and I think that [is why] this is one of the best PDs ever."

Sixteen teachers (of 22, 72%) talked about how having quality lessons in-hand enhanced their implementation. One teacher said the reason he continued to implement activities from the PD was that "the TSI activities were excellent….the way they were designed is excellent." Another particularly appreciated that the resources provided were more than disconnected lessons—"It's not a crutch, it's actually a curriculum." Six of the teachers specifically mentioned their continued use of the physical PD binders, which were comprehensive and included both teacher and student materials and information about the PD pedagogy.

The importance of supplies to continued classroom implementation was mentioned by 13 of the teachers (59%). For example, a high-school teacher said,

> Previously, before this, it [my practice] was more lecture-based. I didn't have the resources to come up with more hands-on, or lab activities. Especially since I teach the special education population, they really thrive from those experiences, and can understand concepts more when they do it hands-on. So, that was one of the main reasons I wanted to take this course, and then [why] I continue to implement what I've learned from it.

Two of the teachers said that the low cost of the supplies allowed them to keep implementing the activities (some of the supplies we provided were consumable). However, supplies were also cited as a limitation for seven teachers (32%). Two of these teachers said this was due to the size of their classes; four explained that while the supplies given to them in the PD allowed them to do the activities, even though the cost of replacing consumables was low, it was still significant.

> For the one [activity] with the LED bulbs, I know those all eventually burned out within like a year… And I never replaced those. Even getting Elodea [an algae]—I was trying a lot not to pay out-of-pocket so much.

Three of the teachers said it was the time required to prepare the supplies for an activity, rather than the supplies themselves, that hindered their implementation of some of the PD activities. The other resource limitation reported by the teachers was their classroom space (5 out of 22 teachers, 23%). For example, one teacher did not do any water activities with their students one year as they had been temporarily moved into the school library.

Importance of Class Alignment.

Content Alignment. More than half of the teachers (12, 55%) reported drawing and teaching mainly from the aquatic science discipline unit(s) from the PD that aligned to their class(es). For example, "with 6th grade, it's a lot of physical science, [so] the most that I used was lessons from the physical or chemical units rather than the biological or ecological content." Similarly, a 7th grade life science teacher said,

> Two of the modules, the chemical and physical, I couldn't find a way to tie those activities in real smoothly into life science. And then the ecological unit, I don't know why, but I always run out of time for my ecological—anything to do with ecosystems I always end up behind.

This makes intuitive sense, as one teacher said, "the topics that were applicable to what we were teaching I used a lot." However, it meant that activities that were not on topic were not implemented, and those that were only tangentially related became "extra," and thus over time were dropped from the curriculum. We knew this when designing the PD—having a variety of grade levels and disciplines in one course would likely result in long-term implementation of only the activities most relevant to their own disciplines. The goal of the PD was that the teachers would learn a variety of content, but only some of that would be directly applicable to their discipline. We hoped the teachers would come away with a greater understanding of cross-disciplinary connections and, as appropriate, make these connections explicit for their students. Only two of the teachers voiced frustration with going through a year-long PD where they only applied a portion of the material in their teaching.

> There were some [activities] that I couldn't ever use in my classroom. I mean I enjoyed doing it…but like the one about the fish—there's no way I could use that….I liked doing it, it was still fun. Just in the sense of me using it in the classroom, it just wasn't applicable.
>
> –8th grade earth science teacher

In contrast, the three marine science teachers who participated in the follow-up study said they used wide swaths of the PD material as it directly related to their content. Six additional teachers also reported using surprisingly large chunks of the content. Five of these were middle-school teachers, the sixth was the sole elementary teacher who taught 5th grade. She continued to implement activities from all of the aquatic science disciplines addressed in the PD.

> I used all of the content, all of the modules, and that I think is because I'm an elementary teacher. I'm not pigeon-holed into just teaching chemistry, right?…It wasn't like they [the students] were switching classes at the end of a period...I didn't have so much with time constraints.

Four of the teachers described minor modifications they made to the lessons to align better with their classroom content and objectives. For example, in the PD we modeled a sampling activity as a way to characterize the abundance of different organisms in an area. An 8th grade earth and space science teacher explained how he modified this activity to be aligned with stars and space,

> Instead of calling the different things organisms I might call them different frequencies of the electromagnetic spectrum that we're detecting, or different star colors, or different objects we see in space...But other than that, it's [the lessons]—they're pretty well laid out. So, I really don't have to change that much.

Science Standards. Between the end of the PD and the follow-up study interview, the NGSS were being rolled out in Hawai'i. This meant that the science teachers in the State were operating in a bit of a grey area, they were being encouraged to use the NGSS but their students were still being tested on older standards. Some schools encouraged implementation of the NGSS early in the rollout while others waited until the year the assessment reflected the new standards to switch. Thus, it is not surprising that only two of the teachers mentioned the NGSS specifically, a 7th grade life science teacher who felt the PD material was not aligned to NGSS, and a 6/7th grade teacher who thought that it was. Two of the middle-school teachers reported more generally that,

> After TSI there just seemed to be a whole lot of other things happening in education in general. I got bogged down by other state initiatives—and then TSI kind of just fell to the back burner.

However, two of the other middle-school teachers, as well as a high-school teacher, talked about how the PD helped them tie lessons and topics together for themselves and their students, and "complete the picture." One of the middle-school teachers liked how the PD lessons "incorporate so much stuff into this one nice, neat package that can be taken and can apply to multiple disciplines, and, within science, multiple standards." The other middle-school teacher spoke of how the PD's pedagogy helped him align the mandate to teach certain standards with his own teaching philosophy,

> My training as a scientist had me wanting students to investigate authentic problem solving, and documenting the problems they're solving, and trying to tell a story with the data. That's me as a scientist. When I stepped into teaching, teaching science as far as it was being communicated to me was all about the state science standards, and making sure students came away from the classroom understanding those standards. I would say TSI's value for me was finding a marriage between the two. Students can leave the classroom understanding the knowledge they're supposed to. But along the way they can also go through these extremely valuable thinking processes and actually exercise real science, and still leave the classroom knowing those standards that they're supposed to. I

think it helped me plan for that authentic thought process, instead of just planning for an accumulation of random science facts.

Pacing Guides, Testing, and Time. Ten of the teachers (45%), cited the overall lack of class time as a limitation to PD implementation. A marine science teacher said, “some of the really cool labs actually I don’t get to teach because of time-wise in the classroom.” Many of the teachers’ time limitations were due to school-mandated pacing guides.

> We have this curriculum map that we’re all supposed to follow, so at a certain time of whatever month or year it is, I’m supposed to be doing everything—covering this topic with this assessment, and I have to, I have to be locked into that.

This was particularly true for the teachers who taught classes where there was state-mandated science testing—such as in 8th grade (earth and space science, at the time of this study) and in high school biology. The lone high-school biology teacher in this follow-up study said that the end-of-course exams were stressful for teachers and that, due to this testing, “biology curriculum in Hawaiʻi is pretty narrow, it’s pretty straightforward. We don’t have a whole lot of room—wiggle room, on what we have to cover.”

Time and testing limitations led all but one of the 10 teachers who reported that a lack of class time limited implementation to pick and choose lessons to cover the content they needed to hit, or to modify lessons to fit within their time parameters. However, there was a distinct break in attitude between the teachers who were resigned to abiding by the constraints of the system (5 of the 10 teachers), and those who considered this adjustment normal and were rather proud of their adaptations and ability to bend the system (5 of the 10 teachers).

Recognition, Resistance, and the Importance of Community. Six teachers (27%) described supportive administration or being the recipients of other external validation. One of these received a prestigious science teaching award; she attributed her teaching success in part to her implementation of PD pedagogy and curriculum. She subsequently led a workshop on her application of PD pedagogy. Another teacher was also asked to lead a workshop based on her implementation of PD material. Two teachers attributed their promotion to department heads as a result of their participation in the PD. One teacher, who taught in a different state after the PD ended yet, at the time of the follow up study, was returning to the same school where she had taught during the PD, said that when she interviewed to come back, "the director asked me 'would you be willing to be teaching some of that.' She didn't know exactly what it was, but she remembered it." Another teacher said that he used one of the PD activities as a model lesson when applying for a job at a progressive school in another state, "they loved it, they hired me. So, I would say it [the PD] has a lot to do with my professional growth."

Implementing PD content led to productive conversations with administrators for two teachers that resulted in a more supportive environment. A teacher was asked by her administrator why she was implementing some activities, such as a beach excursion, in her physical science class. The teacher was able to explain how she was using non-traditional activities to make connections between the natural environment and the physical aspects of science. Another teacher, who was struggling with adapting her school's grading system to the more project-based learning espoused in the PD, convinced her administration to accept a more open grading system so students could get credit for "their ability to understand and participate and share what they learned with their peers" rather being boxed in with a traditional "right or wrong" points-system of grading. This teacher explained that the way administration at her school ended up seeing it, "the kids are learning and they're doing well, and the standardized testing was good," so they had no complaints.

Six teachers (of 22, 27%) expressed frustration with administration. While not overtly unsupportive of their use of PD pedagogy or curriculum, the teachers attributed some of their lack of use to administration decisions. Three teachers were frustrated with the lack of time they had for planning and implementation, in part due to "stupid meetings" that were a "waste of time." A high-school teacher with 10 years of experience said that this lack of time due to administrative burdens, conflated with his own frustration with the system, had resulted in less implementation of activities from this PD as well as other classroom innovations.

> I just don't have the time to devote to overcoming the amount of time that they [administration] take away from our teaching. If I were younger, more motivated, I could, you know. I could deal with that. I could overcome it. And I try to, but mostly it's the lack of time, because it's sucked up by administration burdening us with useless tasks that have nothing to do with education.

Four teachers had changed grades and disciplines, and sometimes schools, each year since the PD—although only two said that this was a hindrance to implementation. One teacher, who ironically had been promoted to department head, said that while he really "liked the idea of [the scientific process] not being linear" and using the TSI inquiry cycle and phases to teach this concept, that his students and peers were ingrained in the "linear ways" of "observation-ask a question-form a hypothesis."

> So, when I tried to change those words and go, well you're in this inquiry phase, and you're this da-da-da—that was a struggle for them [his students], just cognitively. They got it, but it was kind of just me, I was the only one doing it. When I asked them about it later, they're like, yeah, our [other] teachers didn't want us to know that.

However, this same teacher said he heavily promoted the PD material with his department, and ended up using the inquiry phases "sneakily" in his teaching—relaying their concepts but not using TSI terminology. He summarized his conflict between having agency in his classroom and his role as a teacher leader and department head as "when I veer, it doesn't look good." An additional three teachers who moved into supervisory or administration roles after the PD talked explicitly about promoting pedagogy they had learned in the PD with their teachers—without pushback.

Unexpectedly, five teachers (23%) said the PD itself fostered a supportive teacher learning community,

> I met some really good teachers through TSI. Some of the members who were from other schools, who I would never have had a reason to meet otherwise—I keep seeing them again at other PDs and at other places, and the conversations are always good, and the exchanges are always useful.

Another teacher reiterated how the PD "built our own capacity within. So, it didn't just stop in the classroom. That collaboration continued—it's a true testament what type of community you guys established." Three of these teachers wanted additional PD follow-ups, not to necessarily learn "new stuff," but to re-connect with peers and continue the conversation, "being able to talk with people—like 'how's it going?' 'are you doing this again this year?'…'do we need help with anything?' And then sharing what did we do."

Interactions Between Personal and Behavioral Factors

Importance of Practice. Ten teachers (45%) said that engaging in the curriculum activities as students in the workshop was key to their continued implementation of PD material. One teacher said, "because I had practice doing them, I knew how to do it. That was really helpful. Rather than make something up and figure it out on your own, I already had the experience of doing them." Due to this practice, the teachers also said they had a better understanding of how to modify the activities for their students if needed, how to mitigate any potential problems, and how to incorporate them into their instruction. Having experienced the PD lessons as students, and then having to implement the lessons in their classes during the PD, meant the activities had a longer life in the teachers' practices.

> I think the reason it has lived on in my practice is going through the activities myself as a teacher, but also kind of—putting on the shoes and the thinking cap of a student as we went through those activities—that was really helpful for me. And it helped me realize the intrinsic value of going through the thought process of real science. And, being able to take those tangible activities back to my classroom, practice them, and see the success, that's the reason I'm still using them today.

Two of the teachers reported that this in-person PD preparation directly enhanced their confidence, for example:

> Before I really used to depend on the teacher guides, the textbook one that had the answers in it. I used to have that always at hand in case—like when we went over questions, so that I could refer to it to make sure we were on the right track. But when I had done the activities, then I knew what they were supposed to get. Does that make sense...? 'Cause sometimes the [book] questions were a little bit vague. You know, it's like 'ok, this is what the book says you should have put down,' so I'd tell the students that. But I think after actually doing the activities all by myself or discussing them with the other teachers too it helped me to just be more sure of what I expected the students to get out of, and that usually would match what the book would say. And sometimes we

> would go even beyond that and have more information that the students could come up with that would also still answer the question correctly. Whereas before I didn't really know for sure 'oh is that an ok answer?' 'cause it's different than what's in the book.

Sharing PD Pedagogy and Curriculum. Seven teachers (32%) shared the PD pedagogy and/or curriculum with others. Five teachers shared with other teachers in their school. For example,

> I don't know if I'm supposed to tell you this, but I photocopy stuff out of that [PD] notebook and give it to other teachers because they walk by my room and they see my kids engaged.

Three teachers shared with other teachers outside of their school at workshops. One did a leadership symposium on middle school achievement in science,

> I used your [TSI] concepts, your pedagogy, your modes and your phases of inquiry in my classes, and so that was the very first thing I thought about that I wanted to share with the teachers that helped me to be a successful science teacher…Because I just think that it was a huge 'aha' for me, how science isn't linear, and I really wanted to share that.

The two teachers who were promoted into administrative positions also used knowledge they had gained from the PD in their new roles,

> One of the things I do is go and do walk-throughs and observations of teachers and then have conversations with those teachers about how things went, you know, reflective conversations on how they can improve. And so, I usually try, especially with science teachers—other teachers too—but especially science teachers, to try to talk to them about the power of teaching through inquiry and designing their labs and lessons that way.

Behavioral Factors—Changes in Teaching Practice

All of the teachers who were interviewed, but one, reported changes in their teaching behavior as a result of the PD (21 out of 22 teachers, 95%). Table 4.19 lists the teaching changes self-reported by the teachers at the follow-up interview. This table also includes the percent of the teachers who reported these pedagogical concepts as being significant for them (from Table 4.20). Thus, the table allows comparisons between self-reported effects of the PD on the teachers' thinking (in the right column of Table 4.19, personal factors) and changes to their teaching practice (in the left column of Table 4.19, behavioral factors). Self-reported behavioral changes were either (a) in line with pedagogy the teachers thought was significant, (b) implemented more often than expected based on the number of the teachers that reported them as personally significant, or (c) implemented less than expected. An example of a behavioral change that aligned with personal significance was the importance of metacognition. Seven out of 22 teachers (32%) shared that they had changed their practice to incorporate the explicit teaching of metacognition; eight out of 22 teachers (36%) had shared that this concept was significant to them. Other concepts that showed alignment between these behavioral and personal factors were that there are many different ways to do science (~30% of teachers), teacher-as-research-director (~30% of teachers), having student think like a scientist (~19% of teachers), and questioning strategies (~10% of teachers).

The teachers who reported applying the concepts of the teacher-as-research-director and metacognition with their students said they had changed their teaching practice to be more student-centered and student-driven, less lecture-based, and more collaborative—with more group work and peer-to-peer discussions. One teacher described this as having students "share what they're learning as they're learning it." Two teachers used the idea of metacognition, of "thinking about thinking," more generally with their students, but six teachers specifically taught their students this concept as a way for them to be more reflective. For example, one teacher used the concept of metacognition "to get kids to be more thoughtful, and more contemplative about—have I just settled? Or should I go back and rethink this step?" Four teachers used the TSI inquiry phase diagram with their students to demonstrate and reflect on the scientific process. One teacher said she used the diagrams during group activities to "talk about what they did as a group" and that "it was a good checkpoint" for her to see what they were doing and where they were as she circled around the room.

Thirteen teachers (59%) reported that the PD had changed the way they initiated activities or established norms in their classroom (Table 4.19). This is in contrast to just three teachers (14%) who said the concept of initiation was significant to them personally (Table 4.18). Initiation is one of the TSI phases, it represents the importance of not just "hooking" students, but getting them invested in the activity by making it relevant—of having them ask, and understand, "why?" Six of the teachers (27%) said initiation had changed the way they started activities. They reported "not giving everything away right at first" and approaching their lesson planning differently. "I go back to TSI and I ask questions like, 'what would be a good way to start a class, a new class?'" Three teachers adjusted their teaching to avoid front-loading as much information, and another reported learning the importance of accessing prior knowledge.

> A lot of times, I think as teachers, we just give our students the lab, say 'here you go, just do it.' Well there's no way you'll learn. Instead of, 'what do you think is going to happen first, give us your prediction, give us your hypothesis,' which helps them to understand more of the content. I think a lot of times when we just say just do it and tell us what you

learned, we really don't know what they learned because we don't know what they knew before.

Seven of the teachers (32%) talked about how they used the PD activities to initiate their class at the start of the academic year. The activities the teachers reported using were all, but one, practices of science activities the teachers used to set up classroom norms and habits. This included activities about what a scientist looks and acts like, scientific language (e.g., the difference between opinion, hypothesis, theory, and law), and an activity (the modes of learning) that explicates that science can be done in many different ways.

> I really liked the scientific modes of learning...I've used it [the activity] a lot, especially with 6th graders, because they really don't think scientifically. Like, do they learn any science at all in elementary school? I don't know. A lot of the kids I get in 6th grade seem to have had zero, zero input in scientific thinking before that. So, it's really powerful with them. It's one of the activities I do in August right away...And I have some posters up in the classrooms, so we're able to keep coming back to it as they tackle some new topic. Maybe they're getting ready for thinking about a science fair project. 'Ok, so, are you really using all your modes of thinking?' That works pretty well.

Two teachers who taught earth and space science started off their 8th grade classes with an activity that used floating and sinking soda cans to explore the concept of density. This was a concrete activity to introduce a concept that related to core ideas in earth science that the students would keep revisiting over the course of the year—like convection currents in the core of the Earth, in the ocean, and in the atmosphere.

Table 4.19

Follow-up Interview Self-reported Teaching Changes

Self-perceived teaching change	No. of teachers	Example quote(s) [a]	No. of teachers who said concept was sig. for them (Table 4.18)
The scientific process is not linear (TSI phases; inquiry cycle)	6 (27%)	"When I tried to make it, the 5 I's, they [the students] kind of freaked, but when I started to just say 'ohh. You're doing inquiry overall, you know, you're *investigating* here [emphasis added]'. And I started to use those [TSI] words...I put them in where they fit, and that seemed to work better. So, I liked that method/strategy. It actually helps me too, to realize that I needed to give my kids the chance to stop and revise their work, and revise the lab."	12 (55%)
The importance of metacognition/ being reflective (with students)	7 (32%)	"I do have students mapping their thought processes after the activity…Because there's something to be said about students being cognizant of their metacognition, and I think that's something that students aren't necessarily exposed to ever. And it's a new and interesting challenge for them." "I think when we make them [students] more aware, just like teachers, they caught the intent. It helped them to be more accountable."	8 (36%)
There are lots of different ways to do science (TSI modes)	6 (27%)	"[The modes] provide multiple ways of addressing a problem or looking at problems so they [the students] have the ability to explore the subject, rather than just memorize."	7 (32%)
Teacher-as- research-director; student-centered learning	7 (32%)	"The way that I allowed my students to explore and investigate labs, I kind of tried to take a step back and not instruct or influence them as much, and just let them investigate and have them come up with their own conclusions, or what they're learning from it." "I don't always have to script, you know, 'follow these directions and what do you come to' kind of thing…. but they [the students] were coming up with their own ideas and experimenting."	6 (27%)

Self-perceived teaching change	No. of teachers	Example quote(s) [a]	No. of teachers who said concept was sig. for them (Table 4.18)
Having students think like a scientist (scientific demeanors)	5 (23%)	"They [the students] put their science cap on, and they went out into the field, and they practiced science and then had to make a hypothesis and judgment and all those other things... They were things they couldn't necessarily Google the information for, you know?" "I think it [the PD] helped me plan for that authentic thought process, instead of just planning for an accumulation of random science facts."	3 (14%)
Questioning strategies	2 (10%)	"…the idea of the—the facilitating questioning that I felt I got such good training in this PD, it kind of has carried through in my teaching, and not just science but just asking better questions of my students that kind of takes them further into the content and into their thinking, explaining why, or—improving the classroom discourse. It certainly helped me as I've made lesson plans or even as I've used more scripted lesson plans, it's allowed me—it just made me start thinking about asking better questions and kind of guiding the students along in what I was teaching them…I think it's kind of broadened my facilitating abilities."	3 (14%)
Importance of initiation (one of the TSI phases)	13 (59%)	*Initiation/start of year (7 teachers):* "I use a lot of those activities that we did in the first module, sort of at the beginning of the year...to sort of lay the foundation of what does a scientist look like, and what are the practices of science and the laws of science and the theory. I...talk about science as being a process and the times you learn the most is if you fail and you have to go back and refigure and try again. So I use a lot of those base activities to set myself up for the rest of the year and lay out my expectation of the students in science." *Initiation of activities (6 teachers)*: "So I tend to try and let them [the students} discover a lot more on their own before just giving straight out information to them, so I think that since participating in the TSI PD, that that's changed kind of the way that I approach those activities."	3 (14%)

Note. N = 22 teachers.

Of the 12 teachers who found the concept that the scientific process was not linear personally significant, only half (6 of 12, 50%) talked explicitly about implementing this idea with their students. This may be because the concept of the scientific process being nonlinear was more important to the teachers' understanding and planning than a concept they thought needed to be explicated to students.

Of the 21 teachers who said the PD changed their teaching practice, two mentioned five of the teaching changes listed in Table 4.19, and one mentioned four changes in their practice. The rest of the teachers described one change to their practice (9 teachers, 42%), and 2–3 changes (9 teachers, 43%).

Almost half of the teachers (10 out of 22, 45%) said that one of the biggest changes to their teaching practice was in how they planned lessons. There was a lot of overlap in this category with the teachers who said the PD caused them to be more reflective and metacognitive—purposeful reflection caused the teachers to change their lesson plans. Some of the teachers reported that they were overall more organized and intentional with regard to student learning objectives and making clear to students "what we're teaching and why we're teaching it." One teacher reported that the PD made her "start thinking about asking better questions and kind of guiding the students along in what I was teaching them," while another said "weak lessons were really picked up by re-thinking the thinking behind how I was teaching it...overlying practices, modes of thinking elements started to be something I pulled into every lesson." Overall, the teachers reported that they were less rigid in what they expected of their students.

> [The PD] affected me as a teacher just in terms of how we do things through the year. It just opened my mind to different ways of teaching science, that it's not all about

> experimenting. Doesn't follow the scientific method. 'Cause that is what I was teaching the kids before TSI class. Like I would always be like...'I want to teach this but how do I make a lab out of it?' And how do I have to make the kids a hypothesis, and then they have to have control of the variables… whereas after TSI, I'd realized 'oh, there's other ways of doing science.' Like, it doesn't always have to have a hypothesis. It doesn't always have to have a controlled experiment. Maybe it just is collecting data. Maybe it's just analyzing data.

When talking about changes in their teaching practices, 10 (45%) of the teachers described changes that were modest. A few teachers addressed this subtle change in their teaching directly, for example one teacher said the PD "didn't really change it [my teaching] so much as added to it and enhanced some stuff." For other teachers, I inferred that there were modest changes to their teaching practice based on their descriptions of changes in their practice. For example, a teacher said, "I think [my teaching practice] has changed. Maybe not completely changed, but it's changed my approach.…I tend to try and let them [the students] discover a lot more on their own before just giving straight out information to them." Conversely, 11 teachers (50%) described substantial changes to their practice. Some of these teachers took a great deal of pride in accomplishments they attributed directly to the PD.

> I kind of wished they [the facilitators] could come to the class and see 5th graders doing the lessons and then sitting down and doing charts and thinking about their thinking. It really was kind of profound. *We're doing a lot of science* [emphasis added].

One teacher reported that her teaching practice did not change at all since the PD—not even in subtle ways. Of the seven high-school teachers, five described changes in their teaching practice as modest (72%), and two as substantial (29%). Of the 16 middle-school teachers, five described modest changes in their teaching practice as a result of the PD (21%) and nine described more substantial changes (56%). More of the middle-school teachers than the high-school teachers described substantive changes in their teaching practice as a result of the PD. There was no discernable difference in experience of the teachers between these two categories.

Although almost all of the teachers reported that the PD had changed their teaching practice, they also reported that their implementation of specific PD activities and practices had decreased over time. One teacher shared how this decline in implementation occurred.

> I didn't keep up with it as much as I had hoped…only because I think as time has gone by I've used less and less of it. And this year, I'd have to say I hardly—I didn't even use the modes activity, so, it's kind of dwindled over the years. I think like with a lot of other professional development, if it's not followed-up on, I think it doesn't stick. Or if it does, it's not easily applied or it doesn't fit super smoothly into what you're already doing. And I think that was a big hindrance.

Implementation of Specific PD Activities. Although the teachers were not asked nor prompted about specific PD activities, during the course of the follow-up interviews 95% of the teachers (21 out of 22) talked about implementing specific activities in their classrooms after the PD (Table 4.20). Out of 31 activities in the PD, the teachers mentioned 27 (87%) of them. Activities from the chemical aquatic science module were mentioned the most frequently (64% of teachers), followed by those in the physical aquatic science module (50%), practices of science activities (45%, interspersed throughout the PD modules), the biological aquatic science module (41%), and lastly the ecological aquatic science module (36%).

Table 4.20

Frequency of PD Activities Mentioned by the Teachers in the Follow-up Interview

Activity	Teachers who mentioned		Frequency rank of activity
	n	%	
Practices of science total	10	45	
Draw a scientist, scientific demeanors & practices of scientists	6	27	2
Modes of inquiry	5	23	5
Scientific language	3	14	8
Physical aquatic science total	11	50	
Density activity(s) Soda can (5 teachers) Density bags (3 teachers) Other (4 teachers)	8	36	1
Phases of the moon	3	14	8
Other (3 activities)	3	14	
Chemical aquatic science total	14	64	
Water properties	6	27	2
Electrolysis	5	23	5
Conductivity	3	14	8
Other (4 activities)	5	23	
Biological aquatic science total	9	41	
Microevolution	5	23	3
Fish printing	3	14	8
Other (3 activities)	3	14	
Ecological aquatic science total	8	36	
Transect/quadrat sampling	6	27	5
Other (2 activities)	3	14	

Note. N = 22 teachers.

In Summary—Results of Research Question 2

The teachers had a very positive perception of the PD and noted many of the features that made it high-quality according to the PD literature; they appreciated its extended duration, expert facilitation and modeling of activities, and the opportunities to implement and then reflect with colleagues. Almost all of the teachers reported that the PD had changed their teaching practice on some level; the changes described aligned to the TSI pedagogical framework. About half of the teachers reported transferring the skills and abilities they had gained in the PD to new situations. Most of the teachers reported gains in both content and pedagogical knowledge, but half of the teachers noted that this knowledge was eroding over time. All but six of the teachers said the PD either affirmed their teaching philosophy and/or enhanced their confidence in teaching science. The teachers modified activities in response to the grade level and perceived abilities of their students. The reasons the teachers continued to implement PD activities were (a) their perceptions of student engagement and learning, (b) because they had the supplies and curriculum, and (c) because they were familiar with the activities, having experienced them as students in the PD. The lack of class time was the most commonly cited reason for not implementing PD materials. A quantitative summary of these results, using themes derived from Bandura's Theory of Reciprocal Determination, is in Table 4.21. Figure 4.15 is a synthesis of the dominant findings from the interviews and emphasizes the relationships between the personal, behavioral, and environmental factors that affected the teachers' implementation of PD pedagogy and activities.

Table 4.21

Quantitative Summary of Interview Results Using Themes Derived from Bandura's Theory of Reciprocal Determination

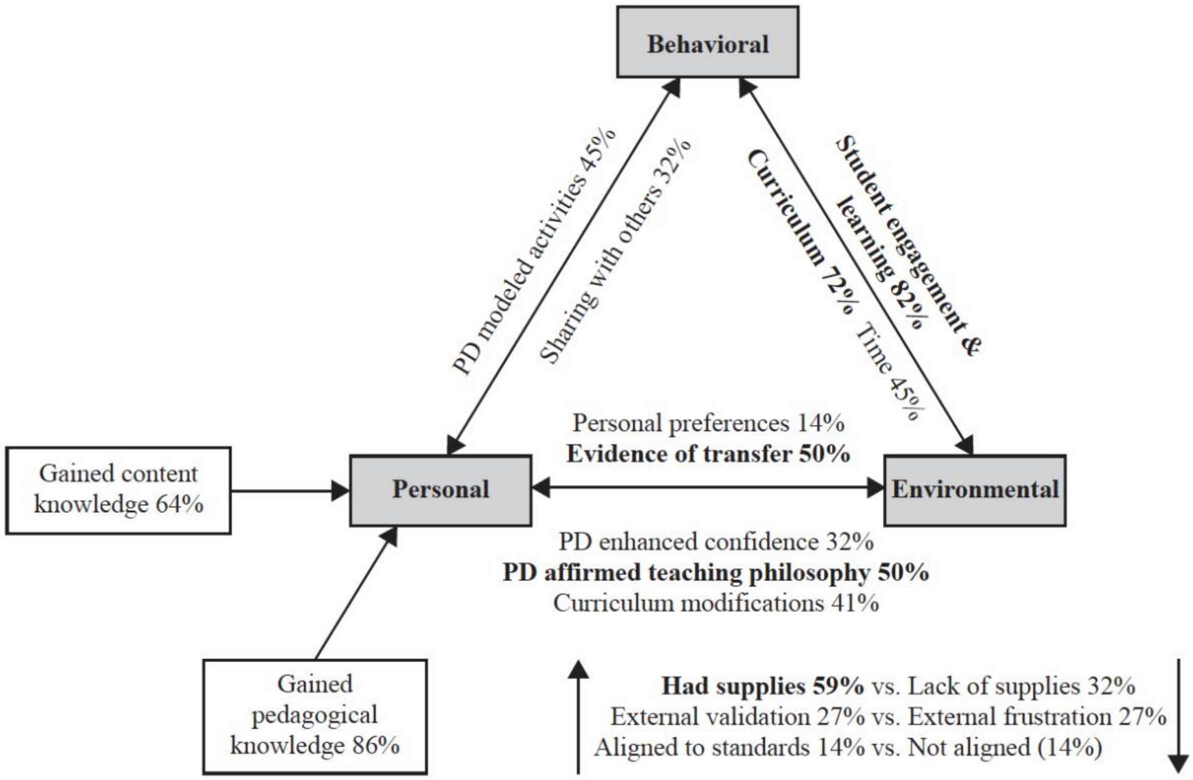
Major themes	Subthemes	Teachers who mentioned	
		n	%
Factors			
Environmental	Various indicators of quality PD [a]	18	82
Personal	Gained pedagogical knowledge [a]	19	86
	Gained content knowledge	14	64
	Forgetting PD principles, terminology, and/or content	11	50
Behavioral	Teaching behavior changed due to PD [a]	21	95
	Modest teaching changes	10	45
	Substantial teaching changes [b]	11	50
Interactions			
Personal – Behavioral	PD modeled activities, which enhanced confidence and therefore implementation	10	45
	Shared PD activities/pedagogy with others	7	32
Environmental – Personal	Evidence of transfer to new classes	11	50
	PD affirmed teaching philosophy	11	50
	Adapted the lessons to align to students' abilities	9	41
	PD enhanced confidence	7	32
	PD was invigoration	3	14
	Reverted to pre-PD teaching style	3	14
Environmental – Behavioral	Reported enhanced student engagement and/or learning	18	82
	Implemented activities due to quality curriculum	16	72
	Taught the activities from PD discipline module that aligned to their class(es)	12	55
	Lack of time as limiting factor	10	45
	PD fostered supportive teacher community	5	23
	Issues with classroom space as limiting factor	5	23
	Supplies: Had supplies	13	59
	Lack of supplies as limiting factor	7	32
	External validation: Recipients of	6	27
	Experienced frustration	6	27
	Standards: PD aligned to	3	14
	PD not aligned to	3	14

[a] These subthemes have their own results tables.

[b] More middle-school than high-school teachers described substantial changes in teaching practice

Figure 4.15

Synthesis of Interview Findings Overlaid on Bandura's Theory of Reciprocal Determination



Note. $N = 22$. The grey boxes represent the factors in Bandura's Theory of Reciprocal Determination. The clear boxes align to the quantitative constructs measured. Arrows indicate interactions. Bolded subthemes were reported by more than half of the teachers interviewed.

Mixed Methods Analysis—Research Question 3

This section describes the findings of the merging of the quantitative and qualitative data in a convergent mixed methods analysis. Recall that examination of the scatterplots of each quantitative assessment showed no clear patterns in teacher grouping, with the exception of the content knowledge assessment, where the separation of the teachers into two groups could be explained in part by the difference in scores between the middle- and high-school teachers. The difference between pre- and post-PD scores and post-PD and follow-up scores were significantly negatively correlated for all the instruments that showed pre- to post- PD gains. This means that, overall, the more teachers' scores rose over the course of the PD on an instrument, the more their scores had declined by the follow-up.

During the interviews participants were asked two questions about their application and implementation of PD practices. I first describe the findings of the item-level analyses and then share the estimate of the correlation between these item-level responses and scores on the quantitative instruments at the follow-up time point. I also present the result of independent *t*-tests examining the relationship between these self-reported behaviors and the teachers' grade levels.

I then further examined the dichotomous characterization of the teachers' descriptions of their extent of teaching change as a result of the PD (modest vs. substantial) from the interviews. I compared the two groups' quantitative instrument scores and their self-reported behaviors using independent-samples *t*-tests. Then I constructed two joint display tables, one of which is structured around the teachers' descriptions of the extent of their teaching change, the other of which is structured by the teachers' self-reports of their implementation of PD activities. I end this chapter with a graphical overview of my synthesis of the mixed methods results.

Self-reported Levels of PD Behaviors

During the interviews the teachers were asked to verbally answer one Likert-type scale question and one forced-choice question with ordered categorical responses about the extent to which they had applied what they learned in the PD to their teaching practice and how much they had implemented PD activities. For the question on the teachers' application of PD practices to their teaching, the response scale used by the teachers included over five distinct points (the teachers were offered a 7-point Likert-type scale), so I conducted my analysis assuming that the scores were on an interval (rather than an ordinal-level) scale (Rhemtulla et al., 2012). The mean score was 5.5, which is much closer to the response category seven ("very much" applied what had learned in the PD) than response category one (applied PD learning to practice "not at all"). The teacher's responses ranged from 2.5 to 7; the interviewer accepted teachers' responses that were between two scale points. Table 4.22 presents the descriptive statistics derived from these questions and Figure 4.16 represents the teachers' PD application scores as a histogram.

Table 4.22

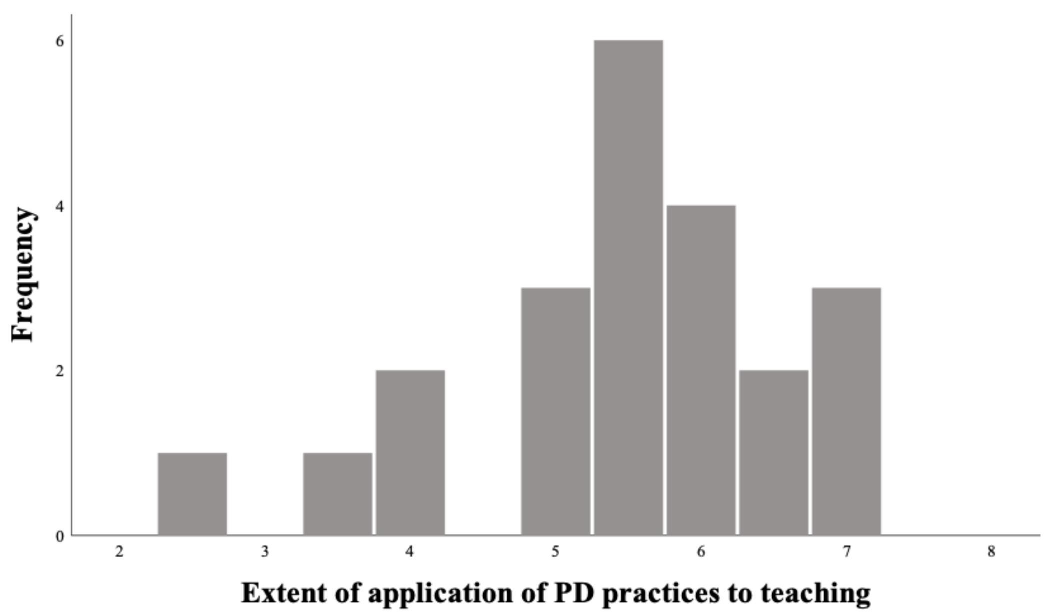
Descriptive Statistics of the Likert-Scale Interview Questions

Interview question	$N_{teacher}$ [a]	*M*	*SD*	se_M	Min	Max
Applied PD practices to teaching [b]	22	5.45	1.15	.25	2.5	7
Implemented PD activities [c]	21	2.09	0.93	.20	1	3

[a] One teacher did not answer how much they implemented activities. [b] Response categories ranged from *1* (not at all) to *7* (very much). [c] Response categories ranged from *0* (none) to *3* (a lot).

Figure 4.16

Histogram of Extent of Application of PD Practices to Teaching

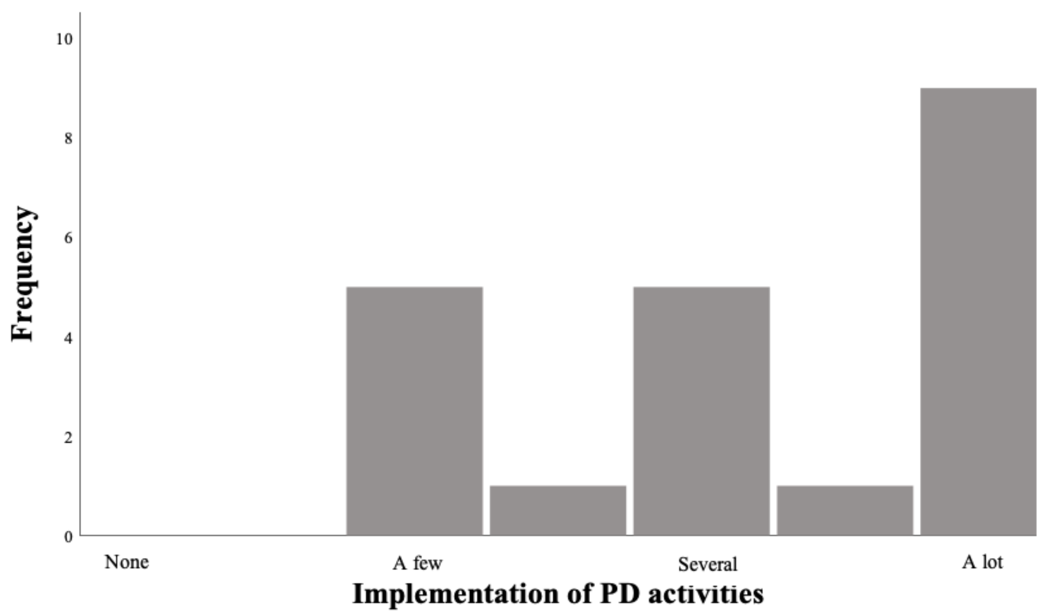


Note. *N* = 22. Response categories ranged from *1* (not at all) to *7* (very much).

There were four response choices for the teachers to choose from when they were asked how much they implemented PD activities: “none,” “a few,” “several,” and “a lot.” These categories were coded on a scale of 0–3, with zero representing no implementation of PD activities and three representing “a lot” of implementation of PD activities. Descriptive statistics for this question are in Table 4.23. One teacher did not answer this question (21 of 22 did, 95%). The mean score was 2.2, which corresponds to the category indicating the teachers had implemented “several” PD activities. Figure 4.17 illustrates these responses in a histogram.

Figure 4.17

Histogram of Implementation of PD Activities



Note. $N = 21$

The teachers’ self-perceived application of PD practices to their teaching followed a positively skewed normal distribution, which shows that of the 21 teachers who said they implemented PD activities 43% (9 teachers) said they implemented them “a lot.” The relationship (using z-scores) between the teacher’s self-perceived application of PD practices to their teaching practice and their implementation of PD activities is, unsurprisingly, statistically significant, $r(21) = .79$, $p < .001$ (one-tailed).

Pearson correlations were calculated between instruments' z-scores at the follow-up time point and the teachers' self-reported application of PD practices and implementation of PD activities (Table 4.23) to explore if there were any relationships between instrument scores and self-reported behaviors. Scores on the pedagogical content knowledge assessment and self-efficacy assessment—but not the Inquiry Teaching Assessment or content knowledge assessment—were significantly positively correlated with the teacher's self-reported implementation of PD activities. A statistically significant correlation was also found between follow-up scores on the pedagogical content assessment and the teachers' self-reported application of PD practices. Thus, the higher the teachers' self-efficacy and pedagogical content knowledge, the more likely they were to report that they implemented PD activities. Higher pedagogical content knowledge scores were also linked to more application of PD practices. Content knowledge and inquiry-teaching knowledge scores, as measured by PD assessments, were not significantly correlated with self-reported application or implementation of PD material.

Independent-samples *t*-tests were conducted to explore the relationship between the teachers' self-reported behaviors and the grade band of the teachers. The high-school teachers reported that they implemented more PD activities (0.43 standardized units) than the elementary- and middle-school teachers, but this difference was not statistically significant (see Table 4.24). Similarly, although the high-school teachers reported applying PD principles to their teaching more than the elementary and middle-school teachers by a magnitude of 0.78 standardized units, the two groups were not significantly different. Thus, the teachers' grade bands were independent of their self-reported behavior.

Table 4.23

Score correlations Between Instruments at the Follow-up Time Point and the Teachers' Application of PD Practices and Implementation of PD Activities

Instrument	Application of PD practices ($N = 22$)	Implementation of PD activities ($N = 21$)
Inquiry Teaching Assessment	.03	.27
Content knowledge assessment	.19	.29
Pedagogical content knowledge assessment	$.43^{*}$	$.54^{**}$
Self-efficacy assessment	.33	$.56^{**}$

Note. All correlations conducted on z-scores.

$^{*}p < .05$. (one-tailed) $^{**}p < .01$. (one-tailed)

Table 4.24

Independent-samples t-Test Results Comparing Grade Level with Self-Reported Implementation

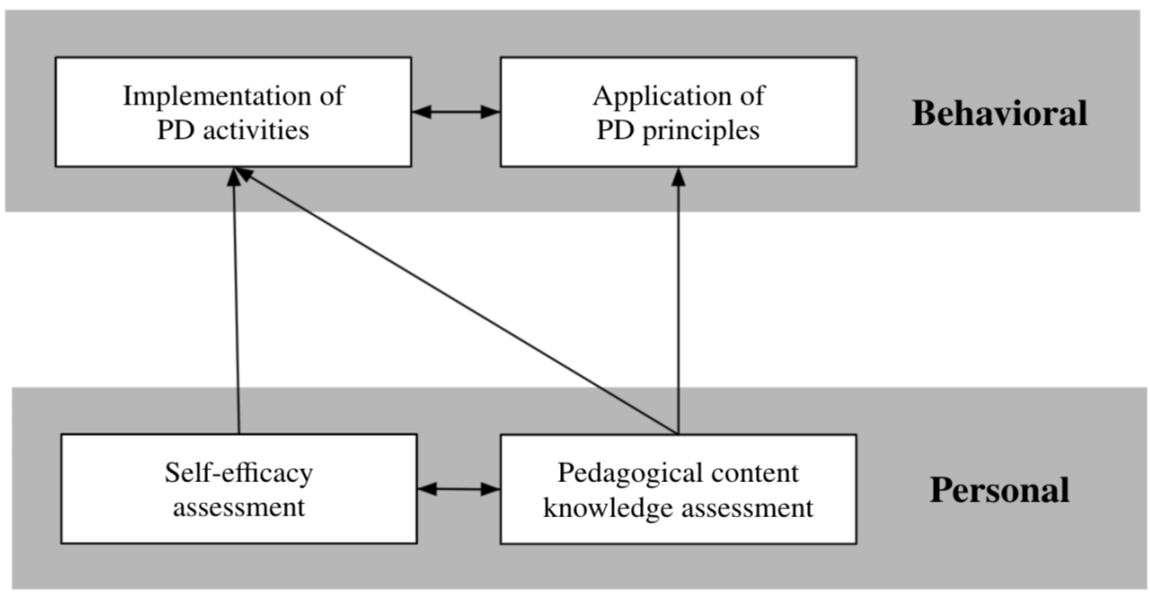
Self-reported behavior	Middle-school teachers ($n = 14$)		High-school teachers ($n = 7$)		Est.	*SE*	$t(19)$	*p*	95% CI	Hedges *g*
	M	*SD*	*M*	*SD*						
Implementation of PD activities	-0.14	1.11	0.29	0.73	0.43	0.46	0.93	.365	[-1.41, 0.54]	0.41
	Middle-school teachers ($n = 15$)		High-school teachers ($n = 7$)		Est.	*SE*	$t(20)$	*p*	95% CI	Hedges *g*
	M	*SD*	*M*	*SD*						
Application of PD principles	-0.25	1.09	0.53	0.46	0.78	0.43	1.80	.087	[-1.69, 0.12]	0.79

Note. $N = 22$ for application of PD principles and 21 for implementation of PD activities. All scores in z-scores. All p-values are two-tailed.

In summary, the teachers' self-reported levels of PD implementation were correlated with scores on the self-efficacy and pedagogical content knowledge assessments; self-perceived application of PD principles to their teaching practice was correlated with just the pedagogical content knowledge assessment. Self-reports of implementation and application were highly correlated. Figure 4.18 is a synthesis of these results.

Figure 4.18

Synthesis of Self-Reported Behavior Relationships with Personal Constructs



Note. The grey boxes represent two of the corners of Bandura's Theory of Reciprocal Determination, the personal factors measured by the quantitative instruments and the behavioral factors from the interviews. Arrows indicate significant relationships.

Teaching Change Relationships

Independent-samples *t*-tests were conducted (using z-scores) in order to explore if the extent the teachers described the changes in their teaching practice (modest or substantial) in the interviews was associated with the quantitative instrument follow-up time point scores or self-reported behaviors in the interviews (Table 4.25). Ten teachers described modest changes in their practice in the interviews, 11 teachers described more substantial changes. One teacher did not provide enough information in their interview to make this distinction. Statistical conclusions were determined by applying the Benjamini-Hochberg adjustment. Table 4.25 indicates a difference of 0.39 between scores of the teachers who described substantial and modest teaching changes at the follow-up time point on the Inquiry Teaching Assessment which was statistically significant ($p < 0.01$). None of the scores at the follow-up time point on the other instruments (content knowledge assessment, pedagogical content knowledge assessment, or self-efficacy assessment) nor the self-reported behaviors (implementation of PD activities or application of PD principles) were significantly associated with the extent of the teachers' behavioral changes.

Mixed Method Joint Display Tables

The two categories of teaching representing the extent of changes in teaching practice described in the interviews, modest and substantial, were compared to self-reported environmental factor themes (Table 4.26). The teachers who described more substantial gains in their teaching practice all reported gains in confidence and/or affirmation of their teaching philosophy as a result of the PD. Most of these teachers also reported evidence of student engagement and/or learning as a result of implementation of PD activities or pedagogy. The teachers who described more modest changes to their practice also reported gains in student learning gains and self-confidence, but to a lesser extent. The teachers' quotes corroborate these thematic categories.

Table 4.25

Independent-samples t-Test Results of Instruments, Self-Reported Knowledge and Behaviors, and Extent of Teaching Change

Instrument	Modest teaching change (n = 10)		Substantial teaching change (n = 11)		Est.	*SE*	$t(20)$	*p*	Stat. con. [a]	95% CI	Hedges *g*
	M	*SD*	*M*	*SD*							
Inquiry Teaching Assessment	-0.59	0.79	0.55	0.97	1.14	0.39	2.93	.005	Sig.**	[0.33, 1.96]	1.23
Content knowledge assessment	0.45	0.88	-0.31	1.01	0.76	0.42	1.82	.042	Not Sig.	[-1.63, 0.11]	0.77
Pedagogical content knowledge assessment	-0.37	0.90	0.52	0.90	0.89	0.39	2.26	.018	Not Sig.	[0.07, 1.71]	0.95
Self-efficacy assessment	-0.33	1.19	0.29	0.83	0.62	0.45	1.38	.092	Not Sig.	[-0.32, 1.55]	0.59
Implementation of PD activities [b]	-0.23	1.03	0.37	0.86	0.61	0.42	1.43	.085	Not Sig.	[-0.28, 1.49]	0.61
Application of PD principles	-0.05	1.06	0.28	0.59	0.32	0.37	0.88	.196	Not Sig.	[0.45, 1.10]	0.37

Note. N = 21. All scores in z-scores. All instrument scores are from the follow-up time point.

$^{*}p < .05$. (one-tailed) $^{**}p < .01$. (one-tailed)

Add in Self-reported behaviors above Implementation of PD activities

[a] Statistical conclusion after applying Benjamini-Hochberg adjustment [b] N = 20 for implementation of PD activities

Table 4.26

Influence of Environmental Factors on Changes in Teaching Practices

Changes in teaching practice	% of teachers reporting gains in:		Congruent quotes
	confidence and/or affirmation	Student engagement and/or learning	
Substantial (*n* = 11, 52%)	100%	91%	"It gave me a lot more knowledge and confidence in being able to teach. And also, the way that I allowed my students to explore and investigate. I kind of tried to take a step back and not instruct or influence them as much, and just let them investigate and have them come up with their own conclusions, or what they're learning from it. So I would say, it definitely had a big change on my teaching practice." "I grew as a teacher tremendously. I think I grew in my confidence as an educator. TSI kind of revolutionized the way I was thinking about education, and education really should be centered on the student, and the student's thoughts, and especially science. Students should be allowed to ask and answer their own questions…. I feel like it [the PD} has really lived on in my teaching."
Modest (*n* = 10, 48%)	50%	70%	"[The PD] kind of affected things, again, not that I was like consciously writing it out or anything, but if you looked back, like there were subtle differences. Instead of trying to be more teacher-led 'cause of time constraints, it was a little bit more student-led than perhaps I would have done." "I think one thing that the PD did for me is that it just opened my eyes to teaching science a different way where there's just not in a script, more open-ended, allowing children to kind of come up with their own experiments. Not that I always do this, I find this hard to do sometimes, but it just kind of opened my eyes up to the way that kids learn and can build meaning and engage in their learning, and just that they can. And that I don't always have to script, 'follow these directions and what do you come to' kind of thing. But to get them at a higher critical thinking level, which I again, I have trouble—it's hard. And I still find it difficult to allow them to do that."

Note. N = 21. All scores in z-scores. All instrument scores are from the follow-up time point.

In the interview, the teachers were provided with four responses to choose from to categorize their implementation of PD activities. None of the teachers indicated "not at all." Two teachers choose responses in between two categories; these teachers were excluded from this analysis. The remaining teachers indicated they implemented "a few," "several," or "a lot" of the PD activities. These self-reported implementation categories were compared to environmental factor themes from the interviews (Table 4.27). The teachers who reported implementing more PD activities were more likely to attribute gains in their confidence and their students' learning to the PD. The teachers' quotes corroborate these thematic categories at the lower "a few" and upper "a lot" implementation categories. However, some of the responses from the teachers in the middle "several" category indicated they were implementing a large number of activities. As the categories were not well-defined, they appear to have been interpreted differently by different teachers; the teachers may have chosen a scale response based on personality characteristics (Harrison, 2020).

In Summary—Results of Research Question 3

Figure 4.19 is a synthesis of the mixed methods analyses. In this figure, self-reported implementation of PD activities and application of PD pedagogy are represented together as they were significantly correlated. Pedagogical content knowledge and self-efficacy are also represented together. These measures were combined as at the follow-up time point scores from these assessments were significantly correlated. Self-reported interview themes of gains of confidence and/or affirmation of teaching philosophy, as well as reports of enhanced student engagement and/or learning, are depicted together as they varied similarly when compared to implementation measures (Tables 4.26 and 4.27).

Table 4.27

Influence of Environmental Factors on Self-Reported Frequency of Implementation of PD Activities

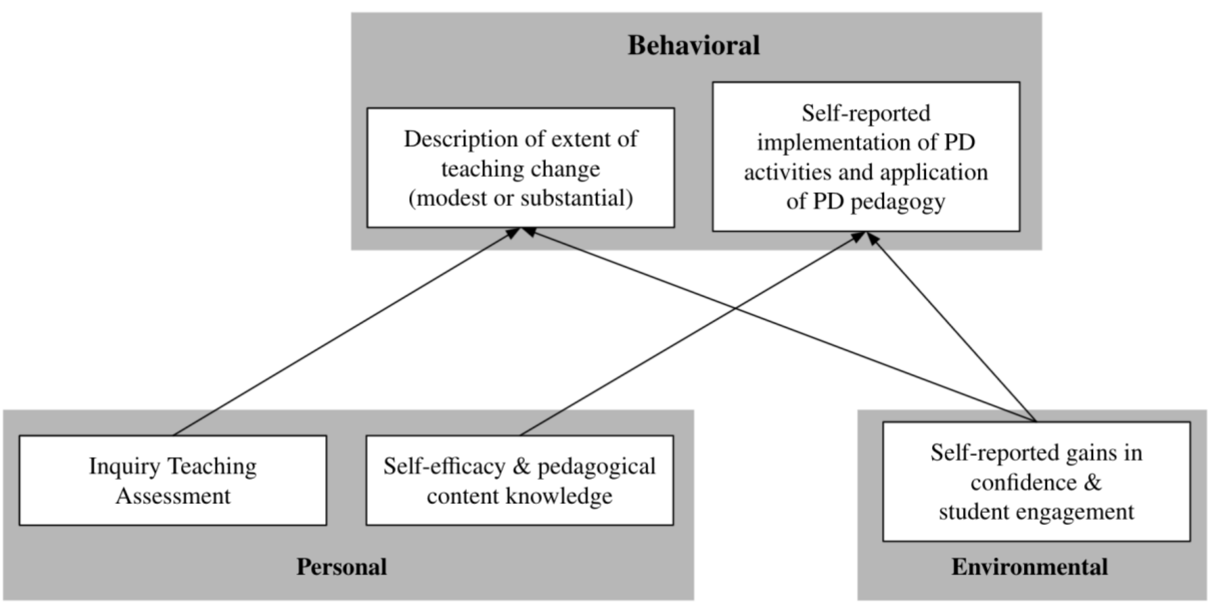
Frequency of implementation of PD activities	% of teachers	% of teachers reporting gains in		Congruent and discrepant quotes
		Confidence and/or affirmation	Student engagement and/or learning	
A lot	47	100	100	"I refer back to all the lessons—the things that we've done and I literally use it all—whatever I can from, the PD." "It's really important when our principals and administration go through and do a walk-through that they see those higher-level questioning happening via the teachers. However, it's not limited just to the teachers. It's also the students. They're observing and making sure that the students are asking those higher-level questions. So the inquiry component was—I did a lot. I implemented a lot. The primary reason is the inquiry."
Several	26	80	80	*Congruent:* "The TSI focused quite a bit on water. And water in a general chemistry class, a lot of it can be used, but a lot of times I have to move on from that….And so I've cut back in order just to cover other content that doesn't necessarily deal with water specifically." *Discrepant:* "I would say [I use] actually a majority of the activities that we did during the PD course….because I teach marine science."
A few	26	20	40	"I'm not too sure why I haven't used all of them [the activities] 'cause thinking back, I think a lot of them could kind of fit. Part of it was supplies….And I think maybe just more my comfort of being able to set up enough equipment for all the students. "The one [activity] that I used every year that I kind of continue to use is the bacteria. Actually that really is the only one that I've continued to use 'cause that one fits in very, very easily, its like a super good connection with what I was already teaching."

Note. $N = 21$. All scores in z-scores. All instrument scores are from the follow-up time point.

Figure 4.19 reveals that the environmental factors of self-reported levels of student engagement and confidence gains are related to both measures of behavioral change (Likert-type scale self-assessments and interview descriptions of teaching changes). Scores on the Inquiry Teaching Assessment at the follow-up time point are significantly related to descriptions of teaching change, but not self-reported behaviors. Pedagogical content knowledge and self-efficacy are related to self-reported Likert-scale behaviors but not descriptions of teaching change. It is important to note that the two measures of behavioral change are not significantly related to each other, and that of the four constructs of personal attributes earmarked for positive gains as outcomes for the PD (self-efficacy, pedagogical content knowledge, pedagogical knowledge, and content knowledge) only three are related to changes in behavior as coded in this study. Content knowledge was not related to any measure of behavior.

Figure 4.19

Overview of Mixed Methods Results



Note. Arrows indicate significant relationships between instruments' scores and/or interview themes.

CHAPTER 5

DISCUSSION

If a sustained increase in student learning is the ultimate goal of teacher PD, at the classroom level this equates to a sustained change in teachers' behaviors. However, there is a lack of research investigating the long-term effects of PD on teachers' understandings, beliefs, and praxis. The purpose of this study was to determine the durability of pre-to-post PD gains on measured constructs as well as the factors that enhanced or hindered 23 teachers' classroom behaviors tied to PD practices 2.5 years after the end of a year-long inquiry-based PD. Earlier research on the effects of this PD found significant gains in the teachers' aquatic science content knowledge, pedagogical knowledge, and self-efficacy on post-PD measures. This study's results show that scores on these instruments declined significantly beyond the post-PD time point, although they all remained significantly higher than baseline measures. The decline in scores over time may support the assertion made in Duncan Seraphin et al. (2017) that gains during this PD were due to the PD, and not a result of maturation effects, because when the PD support was removed there was a decline in knowledge. Scores on the pedagogical content knowledge assessment rose pre- to post-PD, but the gain was not significant; at the follow-up time point scores were significantly higher than at both the pre- and post-PD time points. In interviews, the teachers said the PD increased their content and pedagogical knowledge. The teachers indicated that they perceived changes in their teaching behavior as a result of the PD; half of the teachers described substantial changes to their practice. The interviews provided a more complete picture of how, and the extent to which, the PD became integrated into each teacher's practice.

The goal of this study was not simply to document whether the teachers recalled knowledge conveyed by the PD, but whether it lived on in their practice. It was also an attempt to understand the personal and environmental challenges teachers endured and supports they encountered if they continued to implement PD-based approaches in their classrooms. Results of this study provide useful information for PD researchers, funders, and providers who are interested in the longitudinal effects of PD; and suggest that quantitative and qualitative measures should be used in tandem to understand these effects. Moreover, the results yielded important details about extrapolating behavior changes from pre-post PD gains.

In this chapter I present conclusions and interpretations for each of the research questions. I begin by interpreting the scores of the four quantitative instruments (research question 1), including information gleaned from the interviews that aligned to these constructs where appropriate. I then address the variables found in this study to influence instrument scores and support these findings with interview results (research question 1A). I follow this with a discussion of the qualitative results as aligned to Bandura’s Theory of Reciprocal Determination—the environmental and personal factors that affected behavior as well as self-described behavioral changes the teachers ascribed to the PD (research question 2). I finish the discussion by interpreting the mixed methods results (research question 3). I conclude this chapter with comments about the study’s limitations, implications for PD providers, and suggestions for future research.

Quantitative Instrument Constructs—Research Question 1

To answer the first research question, I compared the teachers’ scores on each of the four quantitative instruments at three time points (pre-PD, post-PD, and follow-up). Results for each of the instruments are discussed in more detail below, although it was difficult to situate the findings in the literature as there are few long-term PD studies.

Within-instrument time-point score correlations, when the pre-PD time point was included, were particularly low for the self-efficacy assessment and the Inquiry Teaching Assessment, indicating measurement problems. For the self-efficacy assessment, this may be due in part to the administration of both the pre- and post-PD instruments at the post-PD time point. The Inquiry Teaching Assessment was a constructed-response instrument that required much more time to complete than the other assessments. Changes in motivation, self-efficacy, and social desirability over the course of the PD may have affected responses at later time points, especially post-PD. Indeed, all post-PD scores across instruments may have been artificially high as the teachers had just devoted a year of their professional life to the PD and knew the facilitators well. At the follow-up time point, these subject-expectancy effects and social desirability biases may have been mitigated. The self-efficacy and pedagogical content knowledge assessments were significantly positively correlated across all time points, indicating that they measured related constructs. However, due to the small sample size of this study, interpretations of instruments' score correlations need to be done with caution.

Understanding of Inquiry-based Teaching

Two and a half years after the PD, there was a moderate loss in the teachers' understanding of inquiry-based teaching, an effect size of Hedges' $g = 0.43$. However, the teachers still had considerably more knowledge of inquiry-based teaching than prior to the PD ($g = 0.59$). This is in contrast to the sustained gains found a year after a science teacher educator PD (aka, for faculty). In that study, St. Clair et al. (2020) found measured knowledge of inquiry remained steady a year after the PD ended. Although scores declined post-PD on the Inquiry Teaching Assessment, most (86%) of the teachers reported in their interviews that the PD had enhanced their pedagogical knowledge. The magnitude of this self-perceived gain in understanding could not be determined, however.

Throughout this dissertation I have referred to the Inquiry Teaching Assessment as a measurement of the teachers' understanding of inquiry-based teaching. However, because the instrument prompts the teachers to respond to short science-teaching vignettes, it is not capturing declarative knowledge of inquiry-based teaching as much as measuring the teachers' pedagogical content knowledge of inquiry—the teachers' understanding of what inquiry-based science teaching looks like in the classroom. This distinction is important as previous researchers have not found evidence linking scientific inquiry knowledge to classroom practice (Bartos & Lederman, 2014; Lakin & Wallace, 2015). Duncan Seraphin et al. (2017) argued that this more procedural measure of pedagogical content knowledge of inquiry "is likely to be less susceptible to self-report biases and better targeted to classroom practice than measures of scientific inquiry knowledge" (p. 6). In support of this assertion, this study found that the teachers' scores on the Inquiry Teaching Assessment were related to their description of the degree of their teaching change as a result of the PD.

Content Knowledge

In the interviews, approximately two-thirds of the teachers said the PD enhanced their content knowledge, but they had differing views about their retention since the program ended. The teachers' aquatic science content knowledge decreased after the PD (a moderate effect size; $g = 0.70$), but scores were still significantly higher than pre-PD levels ($g = 0.75$). The erosion of content knowledge scores over time was expected. The content knowledge assessment was narrow and purposefully tailored to the PD. Furthermore, few teachers taught all of the aquatic science disciplines covered, and thus were unlikely to continue to engage with all or even the majority of the PD content. When knowledge is not actively used, decays in gains are particularly pronounced (Kim et al., 2013). Thus, it was surprising that teachers from different grades and disciplines retained substantial gains over their baseline content scores 2.5 years later. However, recall that this assessment likely did not meet the assumption of unidimensionality, and the items chosen for inclusion showed growth over the course of the PD, but the cause of this growth is uncertain. Thus, inferences should be interpreted with caution.

Other programs have examined PD gains in content knowledge longitudinally, but results have varied widely. For example, two years after a 3-year science PD for K–2 teachers, Sandholtz and colleagues (2016) found that content knowledge remained stable. In 2012, Heller and colleagues found that fourth-grade teacher science content knowledge declined from post-PD to the follow-up 1-year later, but compared to pre-PD measures gains were very large (g = 1.14–1.96, N = 19–30). In contrast, Clary et al. (2018) found that content knowledge was retained in only one of three science content areas measured 1–2 years after an extended inquiry-based middle-school PD (geoscience, d = 1.16); content knowledge in other disciplines (chemistry and physics) showed no significant difference compared to pre-PD levels. (N = 16–23). In an extreme example of content knowledge loss, Liu and Phelps assembled a large dataset (N = 3,340 K–12 teachers) encompassing multiple math PD programs, they found that average math content knowledge score gains were lost in just 37 days (2020).

The focus on water as an overarching theme for the PD enhanced relevance to local teachers but was also limiting. Only a few of the teachers taught marine science and could easily continue to use all of the activities. However, only a couple of the teachers mentioned being bothered by learning content and activities that were not applicable to their practice. The multidisciplinary content the teachers learned, and retained, may have helped them make deeper cross-disciplinary connections with their students.

Pedagogical Content Knowledge

While there was no statistically significant change in pedagogical content knowledge assessment scores from pre- to post-PD, at the follow-up time point the teachers' scores were statistically greater than at both the pre-PD and post-PD time points ($g = 0.74$ and 0.49 respectively). This was the only instrument on which the teachers' scores increased over time. One probable cause of this increase is due to the maturation effect—novice teachers have generally been shown to have less pedagogical content knowledge (Lee et al., 2007; Magnusson et al., 1999; although see Rowan et al., 2001). After 2.5 years of additional classroom experience post-PD, none of the teachers could truly be considered novices at the follow-up time point. Alternatively, the growth reflected by the instruments' scores may be due to seeds planted during the PD. Scarlett (2008), who developed the pedagogical content knowledge scale, wrote, "it is not expected that [a] PD will have instant impact upon the teachers and their classes but lay the foundation for future growth" (p. 86).

The pedagogical content knowledge scale utilized in this study emphasized "pedagogy" items over "context" and "content" items, and thus emphasized classroom management strategies more than other aspects of pedagogical content knowledge. Collaboration, discussion, and facilitation strategies, and accompanying teacher and student behaviors, were explicitly modeled in the PD. During the PD, the teachers may have only implemented the required PD lessons, which included these strategies, and may not have changed aspects of their overall teaching practice. In the 2.5 years between the PD and the follow-up, the teachers had more time to use and become comfortable with these strategies and transfer them to their general practice. Thus, this instrument may be accurately reflective of, and capturing, changes in classroom management strategies—but not other aspects of pedagogical content knowledge. In the future, using an instrument with a more balanced measure of pedagogical content knowledge, or weighing the questions prior to scoring so each cluster of items is considered equally, may address this asymmetry. The pedagogical content knowledge assessment and Inquiry Teaching Assessment are likely measuring different aspects of pedagogical content knowledge. This would explain why they have a lower-than-expected correlation—they are not different methods of measuring the same construct but are measuring different constructs.

Reviews of PD programs show that those designed to teach teachers to use specific instructional processes, including classroom management strategies like cooperative learning, are very effective (Baye et al., 2019; Pellegrini et al., 2021). Introducing teachers to high-leverage strategies is likely to lead to student learning gains in the classroom when implemented (Windschitl et al., 2012), as they are well-defined and easy to learn. These strategies are likely to be maintained if they address problems of practice, such as issues with student engagement and motivation (Barron & Darling-Hammond, 2008). Thus, even if the pedagogical content knowledge measurement used in this study was unbalanced, emphasizing classroom management strategy gains over other aspects of pedagogical content knowledge, gains indicate the PD enhanced the teachers' self-perceived capacity to facilitate strategies associated with inquiry-based teaching.

Literature on PD's long-term effects on pedagogical content knowledge are mixed, and it is unclear how long gains are sustained post-PD. Rozenszajn and Yarden (2014) followed-up with three biology teachers one year after a PD and found that teachers had retained most of their pedagogical content knowledge gains. However, two years after a PD, Kyriakides and colleagues

(2017) attributed issues with implementation to gaps in content and pedagogical content knowledge. Although not a follow-up study, over the course of a longer-duration PD, Copur-Gencturk and Papakonstantinou (2016) found that changes in practice associated with student interactions were easy to make but difficult to sustain. Thus, it is difficult to predict if the sustained gains in pedagogical content knowledge will continue to be maintained.

Self-efficacy

Two and a half years after the PD, there was a limited loss in the teachers' self-efficacy in teaching science (g = 0.28), but the teachers had retained most of their gains compared to pre-PD levels (g = 1.35). This effect size is substantially greater than the 0.06 effect size that Sandholtz et al. (2016) found two years post-PD. Herrington and colleagues found that over the course of a 2.5-year-long PD, self-efficacy scores decreased on one instrument and increased or remained unchanged on another—while interview data and classroom observations indicated an increase in the teachers' knowledge of and ability to use inquiry-based practices (2016). A follow-up one to two years post-PD showed a stabilization of self-efficacy scores on the instrument that had shown a decline during the PD (Herrington et al., 2016). Thus, long-term stability of self-efficacy appears to vary widely and may depend on the instruments used to measure it.

The self-efficacy instrument utilized in this study was designed to measure aspects of pedagogy taught in the PD. Due to the narrow focus of this construct, the smaller effect size of the decrease in scores from the post-PD to the follow-up time point compared to the instrument measuring inquiry-based teaching was surprising, but not unheard of. Sandholtz & Ringstaff (2016) found substantial gains in confidence in teaching science attributed to a PD ending two years prior. As teachers' perceptions and beliefs are a significant predictor of behavioral change (Opfer & Pedder, 2011), the finding that this PD resulted in sustained self-efficacy gains is notable. Self-efficacy is strongly related to confidence, and the social and emotional states of teachers affect their self-efficacy (Shahzad & Naureen, 2017). Confidence has been shown to be important in implementing inquiry-based teaching approaches (Fitzgerald et al., 2019).

In the interviews, half of the teachers said the PD affirmed their teaching philosophy, a third said the PD increased their confidence, and 14% said that the PD was professionally invigorating. The nature of the TSI PD pedagogy, where the teachers are positioned as the research directors in their classroom (a significant pedagogical concept mentioned by 27% of the teachers), may have increased the teachers' self-efficacy in their facilitation skills. A couple of the teachers explicitly mentioned that the PD enhanced their questioning skills, perhaps due to their observation of well-modeled facilitation skills by the PD practitioners (Bandura, 1977).

Relationship Between Pedagogical Content Knowledge and Self-efficacy

The scores on the pedagogical content knowledge and self-efficacy assessments were highly correlated. Due to this correlation, the continued increase in the scores of the teachers on the pedagogical content knowledge assessment over time may be a mediating factor in the comparatively smaller decline in self-efficacy scores post-PD to follow-up compared to the decline in scores noted on the other instruments. The strong relationship between the instruments' scores indicates they are measuring the same, or very similar, constructs; or that similar methodological effects and biases are driving the correlation. High pre-PD scores on both of these instruments meant there was not a lot of room to grow, a ceiling effect. Both of these instruments used agreement scales, and thus may have invoked acquiescence and made the teachers susceptible to subject-expectancy effects (Dillman et al., 2014; Harrison, 2020). Subject-expectancy effects might have occurred because the teachers expected to gain knowledge and skills over the course of the PD, thus they unconsciously reported their expected results as opposed to their actual practice.

Effect of Grade Level, Experience, and Subsequent PD Participation—Research Question 1A

Most of the variables that significantly affected the teachers' scores on instruments in the original PD evaluation did not affect scores on the instruments in this follow-up study. However, interview findings lend a more nuanced view of these results. The teachers who participated in additional significant PD(s) had higher scores on all of the instruments at the follow-up time point than the teachers who did not report participating in additional PD after the PD under study, although this difference was not statistically significant. The teachers' years of experience in the classroom also did not have any detectable effect on their scores on any of the instruments in this study nor in their self-reports of content or pedagogical knowledge gained from the PD. However, the more experienced teachers reported that the PD affirmed their teaching philosophy, whereas more novice teachers reported that the PD enhanced their confidence. This difference by experience lends support to research that has found that more experienced teachers have greater pedagogical content knowledge (Lee et al., 2007; Yang et al., 2020b) and more stable self-efficacy (Ross, 1994). Three of the more experienced teachers said the PD was invigorating. As more experienced teachers have been shown to have less innovative energy over time (Hargreaves & Goodson, 2006), high-quality PD may serve as a platform to revitalize these teachers.

Grade level did affect PD outcomes—the middle- and high-school teachers had different initial levels of, and gains in, content knowledge as measured by the quantitative assessment. But at the follow-up time point scores were similar between the grade levels. This follow-up study supports previous research that this PD appears to have minimized some of the content knowledge disparities that existed between the middle- and high-school teachers (Duncan Seraphin et al., 2017). This is substantiated by the teacher interviews. Although the middle-school and high-school teachers reported similar gains in pedagogical and content knowledge at the follow-up interview, more of the middle-school teachers than the high-school teachers described substantive changes in their teaching practice as a result of the PD. Previous research indicated this PD also minimized gaps between grade levels in pedagogical knowledge (Duncan Seraphin et al., 2017), but this study did not find differences among middle- and high-school teachers on the instruments measuring this construct.

Due to perceived student abilities, the middle-school teachers reported making more modifications to the PD activities they implemented in their classes than the high-school teachers. The middle-school teachers' ways of thinking about the PD broke into two camps: those who thought the PD was a good fit for their students, and those who felt the activities were too advanced. In this PD the activities were originally developed for high-school students. Thus, if the activities in a PD are misaligned to the participating teachers' grade levels, they may benefit from support in appropriately modifying activities.

Interview Findings—Research Question 2

To answer the second research question, I coded the interviews using Bandura's Theory of Reciprocal Determination, detailed below by the interactions between personal, environmental, and behavioral factors. By taking part in this study, the teachers reflected on their experiences. A teacher who participated in the verbal cognitive interviews said it was like a "walk down memory lane," and that she planned to revisit some of the PD activities and pedagogy materials as a result of her recollections during the interview. Reflection is an important part of learning (Zeichner & Liston, 2014), and thus of effective teacher PD. Participation in the study itself may have served as a type of follow-up experience.

Interactions Between Personal and Behavioral Factors

This study had a high response rate (23 of 25 eligible teachers, 82%). While many teachers are largely dissatisfied with the PD experiences presented to them (Darling-Hammond et al., 2009; Penuel et al.; 2007), interview findings indicated the teachers viewed this PD extremely favorably; over two-thirds described it as being of high-quality. While teachers may not like most PD offerings, this was an opt-in PD. The response rate indicates the teachers continued to be invested on some level, likely in part due to the small cohorts and rapport developed with the facilitators. Individuals are more likely to adopt modeled behavior if the model has admired status (Bandura, 1972). Participant enthusiasm is important to maintaining PD gains,

> ...because ultimately the long-term success of any reform depends on teacher buy-in. Educators may comply in the short term with mandates and reforms that are handed down from above, but real change demands sustained professional learning that teachers want to engage in. (Penuel et al., 2020, p. 41)

The features of the PD that the teachers remembered and appreciated are reflective of qualities of effective PD well-established in the literature—it was of an extended duration, had expert facilitators, and allowed time for experimentation, reflection, and sharing (e.g., CRDG, 2012; Desimone, 2009; Loucks-Horsley et al., 2010; Timperley et al., 2007; Yoon et al., 2007; Zeichner & Liston, 2014). While the importance of these features is debatable (Hill et al., 2013), the teachers in this PD were attentive to these components. This PD required the teachers to implement modeled PD activities, providing them with an opportunity to experience the activities within their own contexts, develop confidence, and assess student response. These are key variables in Guskey's PD theory of change (2002b). The teachers acknowledged the importance of experiencing the activities as learners before implementing the activities in their classrooms, and attributed this modeling to enhanced classroom enaction. Teachers are likely to persist in adopting new strategies when they expect that they can succeed or when they experience success (Abrami et al., 2004; Gorozidis & Papaioannou, 2014).

A number of the teachers shared PD concepts with their colleagues in both formal and informal settings. Some of the teachers also reported that the PD led to educational leadership roles. This indicates the PD helped the teachers identify as leaders and for some of them to feel confident and competent enough to lead workshops for their colleagues. It is critical that PD prepares educators to assume leadership roles (Shaw et al., 2018), as educational ownership and strategy sharing are critical components of bringing reforms to scale (Coburn, 2003; Datnow & Stringfield, 2000).

Interactions Between Environmental and Personal Factors

While the teachers opted-into the PD (indicating internal agency), their ability to translate innovations from the PD into their classrooms is related to their self-efficacy as well as their belief in their sense of control over their behavior (external agency; Bandura, 1989). Three of the teachers explicitly shared that they reverted to their pre-PD teaching style after the PD. This may have been due to personal preferences—they made a judgement that the PD was not useful to them. When new practices do not align with preferences, teachers tend to abandon them (Larrivee, 2000). Alternatively, these teachers may have felt they did not have sufficient autonomy with respect to their curriculum and pacing guides or they needed additional support that was no longer available through the PD. The experiences of these three teachers suggests that the other teachers who continued to implement PD pedagogy and curriculum were doing so of their own accord and that they valued and saw utility in what the PD had taught them.

Surprisingly, two of the experienced science teachers said the PD "didn't feel threatening," which suggests that they had previously attended intimidating PD. Teachers learn more effectively when they feel secure and are in a supportive, collegial environment, especially when being pushed beyond their respective comfort zones (Darling-Hammond et al., 2020; Opfer & Pedder, 2011; Podolsky et al., 2019). It is worth noting that half of the teachers reported applying inquiry-based approaches learned in the PD to other classes after the PD ended. Desimone and colleagues (2002) called this a "spillover" effect. The prevalence of this effect in this study may be due to the number of teachers who changed schools and transitioned across grades and disciplines since the PD ended. The teachers were able to transfer some of the PD practices to their new contexts.

Modification and customization of PD activities was probably encouraged by the PD facilitators too soon. For example, the teachers were encouraged to post and share their activity modifications to the online learning community early in the project, when the focus should have probably been on fidelity of implementation (Fixsen et al., 2005, Loucks-Horsley et al., 2010). Fixsen and colleagues (2005) argued that adaptation of programs should not occur until an intervention is accepted practice. PD literature supports this "fidelity first" philosophy (e.g., Loucks-Horsley et al., 2010) as, in general, fidelity correlates with better program outcomes (Hill & Erickson, 2019). Asking teachers to implement a program with high fidelity maintains its integrity, while adapting it risks altering program principles. This is because adaptations may later be characterized as either innovations, customizations that lead to desirable project modifications, or "program drift" (Fixsen et al., 2005). However, the curriculum utilized in the

PD was written for a high-school audience, and there were a number of middle-school teachers in the PD who could not have been expected to implement the activities without modifications. And while a number of the teachers shared how they adapted lessons to their context, I suspect even more were doing so but, as one teacher said, because “the modifications are fairly natural ones to be making in class anyways,” they were not significant enough to bring up in the interview.

This tension between fidelity and adaptation is especially pronounced in educational programs (Protheroe, 2009). For example, a recent survey found that most science teachers agreed that students should be given definitions for new vocabulary at the beginning of instruction and that laboratory activities should be used to reinforce ideas that students have already learned (Smith, 2020). If teachers altered PD activities in this fashion, the activities would no longer follow inquiry-based teaching principles. However, there is usually a need to adapt to local needs and circumstances to enhance program fit. Effective instruction entails teachers being able to use a variety of instructional strategies so they can modify their practice according to the needs of the situation (Darling-Hammond et al., 2020). As adaptations are inevitable, “the rule rather than the exception” (Castro et al., 2004, p. 44), leaning in and providing educators with strategies for adapting materials can enhance their agency over reform and improve science teaching and learning (Shaw et al., 2018). Fishman and Krajcik (2003) claimed that sustainability is fostered primarily by adaptation. Ten years after a PD, four teachers were using the PD strategies but had all adapted them considerably (Shaharabani & Tal, 2017). As innovations are adapted and altered, it may appear over time that PD practices have declined as implementation levels may not be easily detected—even if internalization is a long-term goal of the intervention. Some of the teachers in this study reported that they had integrated PD strategies in their practice, one teacher said what she had learned in the PD was now part of her "professional bag of tricks." It short, over time it becomes more difficult to trace the origins of now-routine practices to any specific PD.

Interactions Between Environmental and Behavioral Factors

With the exception of one teacher, all of the PD participants continued to use inquiry-based teaching strategies they learned in the PD and to implement PD activities. However, similar to other follow-up studies (e.g., Sandholtz & Ringstaff, 2016), the teachers reported that over time, their use of the PD curriculum and pedagogical language had declined. The teachers reported they were most likely to continue to implement the activities most relevant to their classrooms. Thus, the frequency of use of the chemical aquatic science activities was surprising as only one teacher during the PD, and two teachers at the time of the follow-up, were teaching chemistry. These activities may have been used more than others as they were able to support the teachers in their weakest subject. As the teachers in general chose the activities a la carte, rather than following the carefully constructed learning progression of activities modeled in the PD, this may mean they considered aspects of the PD in isolation, and not related to their practice as a whole. Small changes are unlikely to sufficiently change behavior to result in improved student outcomes (Pellegrini et al., 2021).

Most of the teachers reported enhanced student engagement and/or student learning as a result of the PD. If teachers perceive that their students are benefiting from an innovation, they are more likely to continue to do it (Fullan, 1987; Guskey, 2002b; Klingner et al., 1999), which can in turn shift their beliefs (Deglau & O'Sullivan, 2006).

The teachers reported that the PD curriculum and supplies were important to their implementation; aligning with previous research (e.g., Buczynski & Hansen, 2010; Loucks-

Horsley et al., 2010). The curriculum and supply resources allowed the teachers to begin implementation immediately. For some of the teachers, as these supplies ran out, so did implementation. Only two of the teachers said that an activity they tried did not work well for them. The activities, including the supplies, teacher guides, and student materials were extensively tested prior to the PD. This fits with research that shows that innovations are more likely to be successfully implemented when their technical quality and accuracy are already proven (Fullan 1987).

The teachers valued the community that was developed over the course of the PD and appreciated the opportunities to debrief and problem solve with their colleagues. However, only a couple of the teachers taught at the same schools, so there was little opportunity for extending the learning through organically established informal school networks, a capacity-building mechanism that has been shown to be important in sustaining implementation of innovative practices (Lynch et al., 2019). Although an online learning community was designed to sustain the PD community, the teachers did not utilize it after the PD ended. A sense of isolation may have contributed to the decline in implementation over time, as teacher networks influence their abilities to sustain innovative practices (Coburn et al., 2012; Garet et al., 2001; Opfer & Pedder, 2011).

As other researchers have found, the teachers in this study reported being subjected to, and overwhelmed by, competing initiatives that impeded implementation of PD practices (American Federation of Teachers, 2017; Blumenfeld et al., 2000; Klingner et al., 1999; Murrill et al., 2013; Padua, 2018; Pomerantz & Condie, 2017). Instructional pacing guides, testing constraints, and new standards tend to inhibit teacher change (Hill & Chin, 2018). Having to attend to educational initiatives, as well as non-instructional tasks, meant that time was cited as the main limiting factor to implementation. The activities that the teachers reported keeping in their repertoire were those that were perceived as relevant to the standards. This fits with research that teachers are generally unwilling to invest in changes to their practice if the changes will not provide return on high-stakes assessments or other evaluation criteria (Klingner et al., 1999). The practices and activities that were kept by the teachers complemented what they were already doing. “Great ideas” that require extra work will often not be sustained.

TSI and the NGSS. The NGSS, one of the larger initiatives in recent history in science education, were rolled out in Hawaiʻi between the end of the PD and the follow-up study. Some of the decline in the use of the PD curriculum and TSI pedagogical language can be attributed to the new language introduced by the NGSS. However, some of the concepts and ideas established in the PD were more durable than others. The aspects of TSI that were retained, such as developing interest in and focusing an inquiry, metacognition and reflection, and teacher-as-facilitator, may indicate these are aspects of the nature of science that NGSS, and most NGSS-focused PD, does not explicate as well as TSI.

The durability of TSI concepts indicates that the TSI pedagogical framework is complementary to the NGSS. The NGSS establish what students should learn and presents a vision of what students should do while engaged in scientific investigations. However, the complexity of the NGSS can make the standards challenging for teachers to understand and implement (Friedrichsen & Barnett, 2018; Fulmer et al., 2018). Pedagogical frameworks like TSI may serve as a bridge, or foothold, to understanding the process of science, helping teachers see the variety of ways to teach in an inquiry-based manner to achieve the goals of the NGSS.

Lastly, the pedagogical approach used in the PD, TSI, provided a common framework and vocabulary for the PD facilitators and the teachers to use when teaching and talking about the scientific process. Although years later many of the teachers did not recall the specifics of TSI (e.g., the terminology), it was clear they remembered the underlying key concepts. Thus, the TSI framework was useful in supporting teachers' understanding of the scientific process. While NGSS provides language to scaffold this understanding as well, the teaching of the cyclical, iterative, multi-modal nature of science may benefit from pedagogical frameworks that emphasize these aspects of inquiry.

Mixed Methods—Research Question 3

To determine how the personal constructs measured in this study interacted with the teachers' environment to lead to sustained teacher behavioral change, I conducted a mixed methods analysis. Exploratory scatterplots of each instruments' change scores, initially designed to elucidate patterns in teachers' scores across instruments, indicated a regression to the mean for all constructs, a potential validity threat and reminder that scores should be interpreted with caution (Shadish et al., 2002).

Recall that self-reported levels of PD implementation and application of PD principles were Likert-type scale questions the teachers were asked during the interviews. The teachers were categorized as having made modest or substantial teaching changes as a result of the PD based on information gleaned from their interviews. Self-reported levels of PD implementation were significantly positively correlated with scores on the self-efficacy and pedagogical content knowledge assessments; self-reported application of PD principles was significantly positively correlated solely with the pedagogical content knowledge assessment. As scores on the item measuring implementation of PD activities was significantly correlated with scores on the item measuring application of PD pedagogy, but not with the extent of self-reported teaching practice, these self-reported behavior questions may have prompted the teachers to reflect on the amount of PD activities they implement, whereas their descriptions in the interviews of their teaching practice may have better captured changes associated with PD pedagogy. Thus, implementation of PD activities may not be related to the extent of pedagogical changes in teaching practice. After the initial implementation of PD activities, long-term changes in teaching practice may not be tied to activity implementation.

Higher Inquiry Teaching Assessment scores at the follow-up time point were associated with descriptions of more substantial teaching change. Both more substantial teaching change and higher frequencies of implementation of PD activities were associated with gains in confidence and/or affirmation of teaching philosophy as a result of PD participation as well as reports of student engagement and/or learning. Higher levels of classroom implementation were associated with gains in confidence, perceived student engagement, pedagogical knowledge, self-efficacy, and pedagogical content knowledge—but not content knowledge. Previous literature has demonstrated that just enhancing teachers' content knowledge does not equate to improved practices (Kennedy, 2016), but may be a moderating factor. In this study, the PD content often did not align to the teachers' disciplines. Therefore, the teachers may not have been further exposed to large portions of the content covered since the end of the PD. But pedagogy is content-agnostic, and thus all of the teachers of each grade level and discipline could continue to utilize PD strategies if they chose. Although pedagogy cannot be taught without content for context, content may not be as important when assessing long-term PD effects.

Other studies have found that self-reported practices were not related to knowledge (Heller et al., 2012, Lakin & Wallace, 2015), self-efficacy (Herrington et al., 2016), or pedagogical content knowledge (Lakin & Wallace, 2015). Although Herrington and colleagues studied a longer duration PD, and was not a long-term follow-up, they attributed this disconnect to teachers' more sophisticated understanding of inquiry after the PD, which they hypothesized lead the teachers to be more critical of their implementation. Conversely, Lakin and Wallace (2015), examining another longer-duration PD, suspected that teachers inflated their reported use of inquiry-based strategies based on the researchers' measurements of the teachers' knowledge of inquiry practices and pedagogical content knowledge. Thus, self-reported practice yields disparate results in the long-term PD literature, as in other domains, and its accuracy may depend on PD-level variables.

Limitations

There are a number of limitations to this study. It is of a single project, with a self-selected sample, relied on self-report, and did not include a comparison group. As with all quantitative findings derived from small sample sizes, this poses a threat to both internal and external validity, limits causal interpretability, and limits the degree to which findings can be generalized (e.g., Shadish et al., 2002). This study should be considered exploratory and its results considered tentative rather than prescriptive.

The Inquiry Teaching Assessment had arguably the most rigorous development process of the assessments utilized in this study. However, there were some issues with this instrument. Although the teachers were scored on their written responses to different teaching scenarios, they still had to choose one of the multiple-choice options to respond to; in the open-ended responses, the teachers then had to justify their choices. Some of the teachers did not like the options. It was also difficult for some of the teachers to imagine themselves as the teachers in the scenarios. For example, while coding the responses to one of the scenarios the coders noted "This question seems very different in tone than the other questions. We believe that is because it is a fourth-grade question and most participants have no elementary experience."

Although scores on all quantitative instruments were validated through an iterative process of administration during the development of the PD, this study inherited instrument design issues. Most of this study's instruments were developed for this project. Effect sizes on researcher-developed measures are, on average, larger than those produced independently (Hill et al., 2020; Wolf, 2021). However, Lynch and colleagues (2019) argued that while standardized assessments have face validity, "they are not especially sensitive to instructional improvement efforts, due to differences in the skills measured by the tests versus those targeted in the intervention" (p. 285). To fully capture the effects of an intervention they suggested using assessments that are closely tied to the outcomes of the intervention as well as those that are more standardized (Lynch et al., 2019).

The qualitative findings in this study would be strengthened by having additional researchers analyze the data to establish interrater reliability for the coding scheme and therefore minimize bias (Corbin & Strauss, 2014). The number of significant pedagogical concepts and changes in practice described by each of the teachers may have been a function of how loquacious they were in their interviews, with longer responses offering more opportunities for the teachers to share. Also, for some of the teacher interviews it was difficult to differentiate between what the teachers were currently doing as opposed to what they did during or immediately after the PD. The teachers who moved positions, schools, and grade levels may have been especially prone to discussing no longer extant practices.

In both the original PD and this study there was little focus on school context. Although the teachers brought up administration support, or lack thereof, in their interviews, as well as other environmental limitations and affordances, none of the interview questions asked the teachers to reflect specifically on this aspect. Given the importance of school context to PD implementation (e.g., Banilower et al., 2007; Opfer & Pedder, 2011; Yang et al., 2020b), this oversight means that these results are more focused on the teacher level as opposed to the system level.

Suggestions for Future Research

The costs of PD, and the effectiveness of PD, vary widely (Odden et al., 2002; Shear & Penuel, 2010). Thus, PD efforts have been criticized for being a poor investment on return (Hill et al., 2013; Jacob & McGovern, 2015; Whitehead, 2010). However, PD is the primary scalable mechanism of teacher learning. Thus, the solution is not to defund PD but rather fund effective PD. Examining the long-term effects of PD programs should thus become a higher priority and included in initial proposals for large-scale PD interventions rather than done on an ad hoc basis.

Many questions remain and many questions were generated by this study's findings. For example, what aspects of the PD design were more influential in contributing to long-term change? How can these components be improved? In order to answer these questions, researchers need to standardize their terminology so long-term PD effects can be better understood. Currently in the literature researchers use derivatives of the terms sustain, transfer, long-term, longitudinal, and maintenance to refer to the act of following up with teachers after the PD has ended. It is also difficult to situate this study in relation to prior research as the methodologies of PD studies, including those on long-term PD effects, vary widely. Details on the number of teachers followed, follow-up timeline, long-term goals of the PD, outcomes examined, specifics of the measures used to determine effects, and effect sizes of quantitative measures are often vague or buried. This makes researching the long-term effects of PD, as well as designing new studies to explore these effects, difficult.

Understanding teachers' motivations for participating in PD is particularly important as teachers who are successful at fulfilling their initial motivations are more likely to say the PD was valuable and implement PD practices in their classrooms (Kennedy, 2016). This study did not examine teacher motivations for opting-into the PD, thus it is difficult to know if buy-in and long-term effects are related to initial motivations. PD facilitators should prioritize asking teachers about their goals and motivations prior to the PD to further research in this area. Thinking about voluntary PD as free-choice learning may further our understanding of teachers' motivations, as this field has long understood that learning is related to why people choose an educational experience (e.g., explorers vs. hobbyists; Falk, 2006).

A preliminary analysis of individual teachers in this study supports research that different teachers can have very disparate responses to the same experiences (Desimone & Garet, 2015; Kennedy, 2016; Minor et al., 2016; Van Eekelen et al., 2006). The teachers' instrument scores did not always align to their effusiveness, or lack thereof, in the interviews and their descriptions of their practice. Different types of professional learning may be needed for teachers at different stages of their professional trajectories (e.g., Franke et al., 2001; Short & Hirsh, 2020). Further, the desire for customization and personalization of PD as a means to enhance its effectiveness has only increased over time (Starkey et al., 2009). However, both tailoring PD and supporting teachers in choosing specific areas for enhancement aligned to their needs, interests, and abilities requires more intimate knowledge of what works and for whom. In order to gain more insight into how different types of PD serve some teachers better than others, more detailed information is needed on teacher characteristics and other differentiating factors that may help explain the variability in teacher outcomes. This will require both looking at more case studies of individual teachers as well as larger studies with the statistical power to tease out population differences.

Any one PD, however long, is a transient event in a teacher's professional trajectory. Reforms do not take place in isolation (Hill & Erickson, 2019), thus it is difficult to isolate the effects of any one initiative from other factors. Few innovations reach the stage at which they become a routine part of teachers' practice (Hargreaves & Goodson, 2006). However, conceding this is defeatist. Understanding how any particular PD contributes to teachers' practice can help elucidate why some efforts succeed and others fail. Only having one follow-up time point means this study did not have enough information to speculate on the rate of decay, or gain, of measured constructs. Purposefully investigating the progress of teacher change during and after PD participation can help answer this question. This study found that follow-up support would be beneficial to maintaining high-levels of PD usage after the conclusion of the PD. Furthermore, although teachers are adaptive experts (Muijs et al., 2014), and as the educational landscape shifts, teachers need support in determining how PD innovations can be modified to align with ongoing changing circumstances. Teachers want a menu of options for follow-up support (Sandholtz & Ringstaff, 2020), but more research needs to be done on the timing and structure, and thus cost, of this support. However, it is important to note that not all teachers need this type of considerable support (Shaharabani & Tal, 2017).

Lastly, the PD field needs to confront two questions that have loomed over this study—what is long-term success? What does "success" look like? As the goal of PD is enhanced student learning, long-term success should be defined by long-term student success. However, measuring long-term student outcomes is difficult due to numerous factors, including constantly changing student assessments and learning objectives (Desimone et al., 2002; Garet et al., 2001; Kennedy, 2016). Focusing at the teacher-level, is success long-term implementation of PD activities? Is it evidence of meaningful integration of PD strategies and concepts into teaching practice? If the goal is for teachers to make PD-informed instructional choices, how might we measure that on a self-report instrument? How do the different stakeholders (e.g., teacher participants, PD facilitators) perceive program success? It is unrealistic to expect a program to be used forever, so what does "long-term" mean—five years? Criteria or benchmarks indicative of long-term success need to be defined during program conception, so that the longer-term effectiveness of PD can be evaluated more systematically. Long-term (post-PD) success may, and likely should, be defined differently than short-term success. However, if there is no or little evidence of gains in the skills or knowledge targeted pre- to post-PD, it is unlikely that gains will solely happen post-PD (Rose & Church, 1998); in these cases, effort and funding may be better spent elsewhere.

Implications

This study underscores the need to provide longer-term support to sustain PD gains. The declines in scores over time indicate that while one year is generally considered a long-term PD, it may not be long enough for teachers to become comfortable enough with new practices that their understanding and application of inquiry-based teaching is maintained at the same levels without support. Thus, in order to maximize the longevity of PD outcomes, facilitators need to plan to support teachers at some level after the PD ends.

This study also has implications for interpreting PD results. Evidence of PD effects based on pre- to post-PD measures, measures over a limited timeline in a supportive environment, may not be indicative of long-term sustainability of constructs or of classroom teaching changes. Program success (short-term pre-post gains) need to be differentiated from long-term success criteria. As relayed by most of the teachers in the interviews, the PD under study was successful. As indicated by the scores on the quantitative instruments, long-term effects are more muddled.

Self-described classroom changes were not clearly related to instrument score outcomes, and almost all teachers reported teaching changes that were not captured in their scores on the instruments. While about half of the teachers reported more subtle changes, more modest changes should not be automatically categorized as a poor PD outcome. Perhaps innovations had become ingrained or routine and were no longer recognized as significant, or the teachers were already teaching in an inquiry-based manner and the PD enhanced their instruction but did not significantly shift it. This study adds to the literature that gains in knowledge do not necessarily equate to instructional changes, particularly over long time frames. Employing both quantitative and qualitative measures capture a more complete picture of the long-term effects of programs. This study indicates that scores on content knowledge instruments may not be indicative of long-term behavior changes and that continued implementation of PD activities may not equate to changes in pedagogical practice, just as changes in pedagogical practice may not be reflected in reported implementation of PD activities.

Conclusion

PD program stakeholders value interventions that have enduring effects beyond the project. While pre-post measures are important, only by following up with teachers after projects end can we measure their effectiveness. However, studies examining the long-term effects of PD initiatives on teacher knowledge and instructional practices are rare. This study addressed this research need. Instruments were administered to teachers 2.5 years after they completed an inquiry-based science PD to determine its persistence in their current teaching practice and the factors that enhanced and/or hindered PD implementation.

The teachers reported that the PD enhanced their confidence, inquiry and content knowledge, and attributed modest to substantial inquiry-based changes in their teaching practice to the PD. The teachers' explicit use of PD pedagogy had declined due to competing initiatives and a lack of time, but there was evidence that inquiry-based teaching principles had become embedded in the teachers' practices. The results of both quantitative and qualitative measures in this study indicate that high-quality, well-resourced PD leads to long-term self-perceived meaningful changes in teacher practice years after the PD ended. However, declines in pedagogical and content knowledge, as measured on instrument scores, and activity implementation as reported in interviews, indicates that the measurable effects of high-quality PD fade over time. Results from this study may inform how researchers think about, measure, and interpret success in their own programs.

APPENDIX A

Instructions for Rating the Inquiry Teaching Assessment Responses

The Assessment

The purpose of this assessment is to find out how well teachers understand what teaching science as inquiry looks like in the classroom. Teachers (N = 23) in the 2012–2013 TSIA Oʻahu and Kauai cohorts were given the Inquiry Teaching Assessment (ITA); which is an adapted version of the Pedagogy of Science Teaching Test (Schuster & Coburn, 2010). In our version of this test, there are seven teaching scenarios, each followed by a four-option multiple-choice (MC) question that asks the teacher to select which choice best provides an inquiry-based approach to teaching the lesson and to meeting the lesson objective. Each MC question is followed by a prompt that asks teachers to explain their selected answer and to show what they know about teaching science as inquiry. Thus, there are seven items that teachers have responded to, each with a multiple-choice question and a constructed-response question. In effect, the teaching scenario and MC question serve as a vignette prompt for teachers to write their responses to.

The set of responses (which are in this document after the practice set) includes those from the original post-PD administration of the ITA from seven teachers. The original responses (2013) are embedded in the new responses (2016) to examine if there has been any drift in rater's scoring over time. The seven teachers were chosen from a stratified random sample of high, middle, and low scoring participants (two from each of the "high" and "low" categories, three form the "middle" category). The teachers' responses have been assigned a random number so you cannot identify the teacher nor determine which are original and which are recent based on their position in this response set.

Here are the instructions on the first page of the ITA:

> We would like to know about your understanding of what inquiry-based science looks like in the classroom. For each question, read the description of the teaching situation, then select the option that you believe best matches an inquiry-based approach to teaching that lesson objective. ***Be sure to read all possible answers before making your selection***. After completing your selection for the item, provide your explanation for selecting that choice. Please keep in mind the following:
>
> - Focus your writing on why you picked your choice, *you do not need to explain why you did not choose the other options*.
> - Explain your choice in the context of the given classroom scenario. We understand some of the scenarios do not provide many contextual details. Please do the best you can and refrain from referencing your own students and classroom, stick to the scenario!
> - Please provide as through an explanation for your choices as you can (avoid using explanations such as *see my previous response*).
>
> Your explanations will be used to assess your understanding of inquiry in the context of each teaching situation.

Rating

Your task is to rate each response on the scale of understanding of what teaching science as inquiry looks like in the classroom. The scale is operationalized by the rubric. Attached to this document is the rubric, a practice set of responses, and the full set of responses. The ITA is a separate document.

Before you get started rating the full set of responses, please read these instructions and practice rating responses in the practice set. Then you and the other rater will share your ratings and discuss any discrepancies. This is to make sure you are in agreement about how to translate the rubric into your ratings. Discrepancies greater than two points' difference are especially important to discuss. After you are closer in agreement on the practice items you can rate the full set of responses. When you make ratings on the full set of items, do so independently. In other words, do not work with anyone else when rating responses.

Before you rate the responses, carefully read the rubric and the ITA item (including the teaching scenario, the possible MC choices, and item's constructed-response instruction prompt).

There are two dimensions you have to deal with: 1) the teacher's understanding of what inquiry looks like in the classroom (ranging from misconceptions to deep and comprehensive understanding), and 2) the relevance of the information in their response (how much they give us to work with in rating their response). Our main goal, of course, is to measure this first dimension.

For this first dimension, focus on the indicators that that suggest the teacher understands what teaching science as inquiry looks like. Do not focus too much on the TSI terminology. One teacher might, for example, use the term "curiosity mode", while another teacher might provide a description that shows they understand the curiosity mode even though they might not use the TSI terminology. Treat both of these equally when possible, and even consider the possibility that some responses might show that teachers know the terms but do not really understand what TSI looks like in practice.

To address the second dimension, the rubric includes statements such as "There is some evidence that…" or "There is strong evidence that…". Give a score of zero when there is not enough information in the response to make a judgment.

Rubric-based scores range from -1 to +3, with half points, which is a 9-point scale. Zero is the point where we would either expect the response to be too short or too vague to provide adequate information for us to judge. Zero is also the point where there is a balance between the teachers' understanding of inquiry and their misconceptions about what inquiry looks like. Negative scores should be given when there is evidence that the teacher has some misconceptions and when there is little or no evidence that the teacher understands what inquiry looks like.

It is best to rate each item's set of responses in one sitting before rating the next item's responses (that is, read and rate all the responses to Item 1 before rating Item 2, and so forth). It is best to break it up into chunks in order to avoid fatigue (which can affect your rating). That is, ***rate one item or even half of the set of an item's responses at a time***. Each item may take around two hours to rate. ***You can rate responses with an item in any order you like*** (that is, you do not have to rate the first response first—you can start in the middle, the end, and so forth, to avoid an ordering effect in your rating).

TSIA Follow-up Inquiry Teaching Assessment

TSI
aquatic

Inquiry Teaching Assessment

We would like to know about your understanding of what inquiry-based science looks like in the classroom. For each question, read the description of the teaching situation, then select the option that you believe best matches an inquiry-based approach to teaching that lesson objective. ***Be sure to read all possible answers before making your selection*** . After completing your selection for the item, provide your explanation for selecting that choice. Please keep in mind the following:

- Focus your writing on why you picked your choice, *you do not need to explain why you did not chose the other options* .
- Explain your choice in the context of the given classroom scenario. We understand some of the scenarios do not provide many contextual details. Please do the best you can and refrain from referencing your own students and classroom, stick to the scenario!
- Please provide as thorough an explanation for your choices as you can (avoid using explanations such as *see my previous response*).

Your explanations will be used to assess your understanding of inquiry in the context of each teaching situation.

This assessment will take about 45 minutes to complete. You need to submit all of your answers in one sitting.

If you have any questions, please do not hesitate to call (808-956-4920) or send me an email (tsistudy@hawaii.edu).

(This instrument is adapted from the Western Michigan University's Pedagogy of Science Teaching Test.)

TSIA Follow-up Inquiry Teaching Assessment

Question 1 out of 7

Lesson on Thermometers and How They Work

Objective:
Students develop an understanding of the basic structure and mechanisms of thermometers.

* 1a. Four different 5th grade teachers each use a different approach. Whose approach below do you believe best provides an inquiry-based approach to teaching the lesson objective?

- A. Ms. Ash says, "Today you will make a mystery device, see how it behaves, and then tell me what you think it is for." Each group has a bottle, a cork, a straw, and colored water, which they put together as she shows them. Students then explore what happens if they hold the bottle in their hands or in hot or cold water, etc. Students are then to explain what their product can be used for, what to call it, and how it works.
- B. Ms. Brown says, "You will be discovering something for yourselves today." Each group has a bottle, a cork, a straw, and colored water. She does not tell students how to assemble the materials or what they are for, but asks them to figure this out for themselves and to try anything they wish. Students are then to explain what their product can be used for, what to call it, and how it works.
- C. Ms. Connolly writes the lesson title "Thermometers" on the board, draws a thermometer diagram and labels the parts. She then carefully explains how it works. She leads discussion based on student questions. Then groups try out various thermometers at their benches; for example, in hot and cold water.
- D. Ms. Dole asks the class, "What do you already know about thermometers?" She lists responses on the board. Working from some of these, she draws a thermometer on the board and explains how a thermometer works. Then groups try out various thermometers at their benches; for example, in hot and cold water.

* 1b. Please explain why you selected this teacher's approach for the lesson. In your explanation, show what you know about teaching science as inquiry.

TSIA Follow-up Inquiry Teaching Assessment

Question 2 out of 7

Lesson on Finding the Density of a Mystery Substance

Background:
Mr. Cobb's 8th-grade students have learned the concept of density in class by measuring the mass and volume of solid geometric objects. Students have also learned how to measure the volume of solid geometric objects using the liquid displacement method.

Task:
Mr. Cobb gives students an application experiment where they have to apply their knowledge of density. He provides a mystery element, which is insoluble in water and is in granular form. The students' challenge is to devise a method of finding the density of this substance, record the data, calculate the density, and suggest what the mystery element might be. (They will have to use a water displacement method to measure volume since there are air spaces between granules.)

* 2a. Thinking about teaching science as inquiry, which one of the following approaches would you suggest that Mr. Cobb use for this activity?

- A. Have students first propose a method they intend to use. Students then discuss this with the teacher, get feedback, revise their plans if necessary, and then go ahead with their experiment.
- B. Leave students to their own devices as much as possible. The teacher asks students to work out a method on their own and to decide what measurements to take and how. Students then write up their method, experiment, and results in their own way.
- C. Provide students with lab worksheets that have headings for the main steps in the procedure, blank tables for students to enter experimental data, and suitable spaces for calculations, results, and conclusions.
- D. Provide students with a lab instruction sheet which outlines the experimental method. Students follow this procedure and record data in their own lab notebooks. They then calculate density and write up their results in their own way.

* 2b. Please explain why you selected this approach for Mr. Cobb's class. In your explanation, show what you know about teaching science as inquiry.

Handling Student Misconceptions in Finding the Density of a Mystery Substance

Background:
Mr. Cobb's 8th-grade students have learned the concept of density in class by measuring the mass and volume of solid geometric objects. Students have also learned how to measure the volume of solid geometric objects using the liquid displacement method.

Task:
Mr. Cobb gives students an application experiment where they have to apply their knowledge of density. He provides a mystery element, which is insoluble in water and is in granular form. The students' challenge is to devise a method of finding the density of this substance, record the data, calculate the density, and suggest what the mystery element might be. (They will have to use a water displacement method to measure volume since there are air spaces between granules.)

* 3a. One group of students decides to measure the volume of their granular sample by pouring the sample dry into a measuring cylinder. This will give an inaccurate value for the actual volume of granules because of air spaces. How do you think Mr. Cobb should address this?

- A. Tell them immediately that this method will give the wrong volume because of air spaces in the sample, and that they should use the water displacement method instead.
- B. Before they go any further, ask them to think about their volume measurement, and prompt the idea of air spaces between granules if necessary. Once they recognize the problem, ask them to think of another method and then to continue.
- C. Let them go through with the experiment using their method, allowing them to calculate an inaccurate density value and to suggest a possible element. If students do not notice the error, bring their attention to the anomalous result and ask them to think of an alternative method. Then, have them re-do the experiment.
- D. Let them go through with the experiment using their method, allowing them to calculate an inaccurate density value and to suggest a possible element. But, do not ask students to re-do the experiment correctly; rather, have them attribute their anomalous result to "experimental error".

*3b. Please explain why you think Mr. Cobb should take this approach. In your explanation, show what you know about teaching science as inquiry.

TSIA Follow-up Inquiry Teaching Assessment

Question 4 out of 7

Lesson on Air as a Material Substance

Objective:
Ms. Harvey wants her 4th grade students to realize that air is a substance (is matter) with certain properties that are evidence for that.

Background:
She leads into the topic by saying: “Here are some questions of interest about air. Is it some sort of substance, or not? Air is invisible, so is there really something there? And how could we be sure if there was something there or not?”

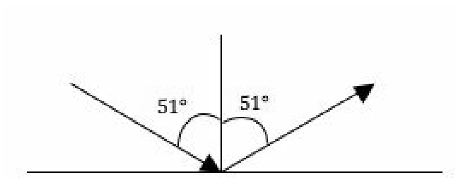
* 4a. After this opener to the lesson, and thinking about teaching science as inquiry, how do you think Ms. Harvey should continue? (Focus on overall approach rather than minor points of detail.)

- A. Tell the students that, yes, air is a substance. Say that although air is not very dense, there is something there that can be felt, exerts pressure, and occupies space. Then demonstrate these properties by showing how air can be felt (as wind from a fan), and by blowing up a balloon (showing the pressure and space aspects).
- B. Tell the students that, yes, air is a substance. Say that although air is not very dense, there is something there that can be felt, exerts pressure, and occupies space. Then let each student experience feeling air (as wind from a fan), and have them each blow up a balloon (to see the pressure and space aspects).
- C. Ask the students, “How can we tell if air is a substance or not and what could we do to test this?” Take students’ suggestions, then guide groups toward feeling wind from a fan and blowing up balloons themselves in order to see evidence for these air properties. To close the lesson, groups report back to the class on the various things they did and what they found.
- D. Ask student groups to think up ways to test if air is a substance. Then let them go ahead and try these as they wish, using any equipment they request that is available. Ms. Harvey should not prescribe, suggest, or interfere, but be available for individual help or discussion. To close the lesson, groups report back to the class on the various things they did and what they found.

* 4b. Please explain why you selected this approach for Ms. Harvey. In your explanation, show what you know about teaching science as inquiry.

Planning the First Steps in a Lesson on the Reflection of Light

Objective:
By the end of the lesson, Ms. Baker wants her 8th grade students to know how a light beam reflects when it strikes a mirror, and to be able to state and apply the law of reflection. Formally, this is “When a ray of light strikes a mirror surface, it reflects such that the angle of reflection is equal to the angle of incidence.”



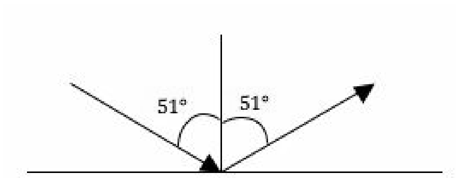
* 5a. Ms. Baker has to decide on an approach for starting the lesson. Thinking about teaching science as inquiry, which one of the following first steps do you think she should use?

- ○ A. Start by writing the law of reflection of light on the board, illustrating it with a diagram, and explaining it. With a light beam, a mirror, a protractor, and paper at each lab bench, students then model the illustration.
- ○ B. Using a light beam and mirror, start by demonstrating what happens when a light beam strikes a mirror. With a light beam, a mirror, a protractor, and paper at each lab bench, students then model the demonstration.
- ○ C. Start by having the students try out what happens to a light beam when it strikes a mirror. Provide each lab bench with a light beam, a mirror, a protractor, and paper. Encourage students to come up with specific questions about light behavior.
- ○ D. Start by having students find out what they can about light behavior. Provide each lab bench with a light beam, a mirror, a protractor, and paper. Do not prescribe a specific goal for students.

* 5b. Please explain why you selected this approach for Ms. Baker's class. In your explanation, show what you know about teaching science as inquiry.

Planning the Second Part of the Lesson on the Reflection of Light

Objective:
By the end of the lesson, Ms. Baker wants her 8th grade students to know how a light beam reflects when it strikes a mirror, and to be able to state and apply the law of reflection. Formally, this is "When a ray of light strikes a mirror surface, it reflects such that the angle of reflection is equal to the angle of incidence."



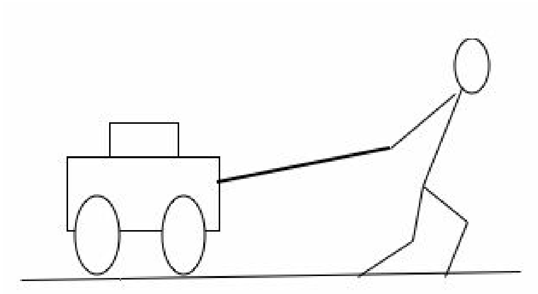
* 6a. Student groups, at their lab benches, have already used the light beam, a mirror, a protractor, and paper to have hands-on experience with the reflection of light. Now, Ms. Baker has to decide what she will do next. To get students to think like scientists and to reach the lesson objective, which activity below do you think she should use?

- A. Draw several different paths of light on an illustration on the board and ask students to explain why each one is the correct or incorrect path. Ms. Baker should then finish the lesson by restating to the class the formal law of reflection.
- B. Using the formal law of reflection, written on the board and illustrated with a ray diagram, Ms. Baker should ask student groups to verify whether their findings at their benches matched the diagram. If there are findings that do not match the law, she should have the class discuss the groups' methods and reasons for the discrepancy.
- C. Ask students to present their groups' hypotheses and findings to the class. The class discusses the findings and together formulates a law of reflection. Ms. Baker should then state the formal law of reflection and have students compare their law to the formal law.
- D. Ask students to present their groups' observations and discoveries to the class. The teacher then combines the groups' observations and discoveries into a large table, which the class then discusses. Ms. Baker should then finish the lesson by stating the formal law of reflection.

* 6b. Please explain why you think Ms. Baker should take this approach. In your explanation, show what you know about teaching science as inquiry.

Lesson on Force and Motion (Newton’s Second Law)

Objective:
The instructional goal is for students to learn that a net force will cause an object to change (speed up or slow down) its motion. This is Newton’s Second Law.



* 7a. Four 5th-grade teachers all have the same teaching equipment available, including a loaded wagon to which a pulling force can be applied. But, they have different approaches to teaching this topic and they design their lessons differently. Of the following approaches, which approach do you believe is the best way to teach students to think like scientists?

- A. Mr. Adams writes up a clear statement of Newton’s Second Law, and explains it carefully. He then demonstrates the law by pulling a loaded wagon in front of the class as students observe the motion and record what they see. The teacher asks students to restate the law while he pulls the wagon again.
- B. Mr. Brandt writes up a clear statement of Newton’s Second Law, and explains it carefully. He then has groups of students verify the law by pulling a loaded wagon, following his instructions and lab worksheet.
- C. Mr. Campos raises the question, “What kind of motion results if there is constant force on an object?” He guides groups of students to explore this themselves by pulling a loaded wagon and observing what happens. From the evidence, they then propose a possible law. The teacher then introduces Newton’s Second Law.
- D. Mr. Doyle raises the question of whether there is any relationship between force and motion. Groups of students are free to explore this in any way they wish, using any suitable and safe equipment in the lab. He does not interfere but helps with any equipment that students request. Groups report back to the class with findings. The teacher then asks students what they know about Newton’s Second Law.

* 7b. Please explain why you selected this teacher's approach for the lesson. In your explanation, show what you know about teaching science as inquiry.

TSIA Follow-up Classroom Instruction Questionnaire

Classroom Instruction Questionnaire

The purpose of this questionnaire is to learn about your science teaching practices and your beliefs about your ability to implement inquiry-based teaching. The results will help us to understand the long-term effects of the TSI-A project.

This questionnaire will take about 10 minutes to complete. Please carefully read each question. ***You will need to submit all of your answers in one sitting.***

All the best,
Lisa

TSIA Follow-up Classroom Instruction Questionnaire

Classroom Instruction Questionnaire

Part 1.

Instructions for Items 1–36

Please reflect on the activities that you and your students do in your science classroom. For each item, select the answer that best reflects your classroom.

	Never	Rarely	Sometimes	Often	Always
1. Students in my science classes talk with each other about how to solve problems.	○	○	○	○	○
2. Students in my science classes learn how science can be part of their daily lives.	○	○	○	○	○
3. Students show respect for the ideas of others in my science classes.	○	○	○	○	○
4. Students in my classes share knowledge in the classroom science laboratory.	○	○	○	○	○
5. Students identify sources of error in experimental data in my classes.	○	○	○	○	○
6. Students propose alternative explanations to phenomena in my science classes.	○	○	○	○	○
7. My science students like to keep their data to themselves.	○	○	○	○	○
8. Students in my science classes work best individually.	○	○	○	○	○
9. Students in my science classes use mathematics to support convincing explanations.	○	○	○	○	○
10. Students in my science classes are able to interpret graphs.	○	○	○	○	○
11. Students in my science classes offer to explain their ideas to one another.	○	○	○	○	○
12. A grade is sufficient feedback for my science students.	○	○	○	○	○
13. Students in my science classes explain their ideas to each other.	○	○	○	○	○
14. Students in my science classes discuss the connections between classroom learning and their daily lives.	○	○	○	○	○
15. I provide science students with needed feedback to improve their work.	○	○	○	○	○
16. My primary assessment of students focuses on completion of work.	○	○	○	○	○
17. I use multiple science assessment types (e.g., quizzes, projects, reports, or lab practicals).	○	○	○	○	○

	Never	Rarely	Sometimes	Often	Always
18. Part of my job is to supply materials, tools, and resources for my science students.	○	○	○	○	○
19. New data from science investigations can change students' ideas.	○	○	○	○	○
20. My students feel safe in expressing themselves in science discussions.	○	○	○	○	○
21. My science students learn best when they work alone.	○	○	○	○	○
22. My science classroom supports collaboration among students.	○	○	○	○	○
23. My science classroom is arranged for interaction among students.	○	○	○	○	○
24. It is more important to teach for content than to teach for broad conceptual understanding.	○	○	○	○	○
25. In my science classroom, students must support statements with evidence.	○	○	○	○	○
26. In my science classes, learning science content is more important than learning science processes.	○	○	○	○	○
27. If students ask me if their data are correct, I answer yes or no.	○	○	○	○	○
28. Teachers should use a variety of assessments in the science classroom.	○	○	○	○	○
29. Students in my science classes have opportunities to talk to each other about their classroom work.	○	○	○	○	○
30. I provide feedback to my science students.	○	○	○	○	○
31. I provide examples of how science concepts apply to daily life.	○	○	○	○	○
32. I ask students to evaluate their own data or conclusions.	○	○	○	○	○
33. I like to keep science student desks in straight rows.	○	○	○	○	○
34. I ask students in my science classes why they think their answers are correct.	○	○	○	○	○
35. I change my science teaching based on assessment results.	○	○	○	○	○
36. I ask students if they agree or disagree with data presented by their peers.	○	○	○	○	○

Classroom Instruction Questionnaire

Part 2.

Instructions for Items 37-51

Please select the number that best estimates your ability as a science teacher to implement the activity.

	Low ability 1	2	3	4	5	High ability 6
37. My ability to include various modes of inquiry in my science classroom.	○	○	○	○	○	○
38. My ability to provide students with an activity, demonstration, or other experience which will address a problem or question that will be the focus of inquiry.	○	○	○	○	○	○
39. My ability to direct students to identify a problem or to pose a question that will become the intended focus of inquiry.	○	○	○	○	○	○
40. My ability to direct students to consider or explore possible approaches, procedures, and outcomes that will address a scientific problem or question.	○	○	○	○	○	○
41. My ability to direct students to develop a hypothesis that will address a problem or question.	○	○	○	○	○	○
42. My ability to direct students to develop, design, or identify a procedure that will address a problem or question.	○	○	○	○	○	○
43. My ability to direct students to invent or identify a means of collecting data.	○	○	○	○	○	○
44. My ability to direct students to implement the procedures and to revise them when needed.	○	○	○	○	○	○
45. My ability to direct students to assess the quality of the procedures.	○	○	○	○	○	○
46. My ability to direct students to test a hypothesis.	○	○	○	○	○	○
47. My ability to direct students to evaluate the quality of their plan, procedure, and results.	○	○	○	○	○	○
48. My ability to direct students to evaluate whether the data collection occurred as they had predicted.	○	○	○	○	○	○
49. My ability to direct students to discuss whether their procedures were appropriate for addressing the problem or question.	○	○	○	○	○	○
50. My ability to direct students to become aware of their own thinking processes.	○	○	○	○	○	○
51. My ability to direct students to think like scientists.	○	○	○	○	○	○

TSI-A Follow-up Study: Teacher Content Assessment

TSI
aquatic

Teacher Content Assessment

This is an assessment addressing some of the content covered in the TSI-A project. Your responses will help us understand the project's effect on teachers' content knowledge. In the original project everyone's scores increased from pre- to post-project on the content assessments. The purpose of this assessment is not to evaluate your knowledge of this content, rather, it is to examine the longevity of the participant content knowledge of the concepts covered in the professional development.

This assessment will take about 40 minutes to complete. You will need to submit all of your answers in one sitting.

Directions:
Please answer each of the questions to the best of your ability (work on your own, without consulting any materials or other people). For each question, *select the single* ***best*** *answer*. Be sure to read the entire question and all possible answers.

Please email me if you have any questions!
Lisa

tsistudy@hawaii.edu

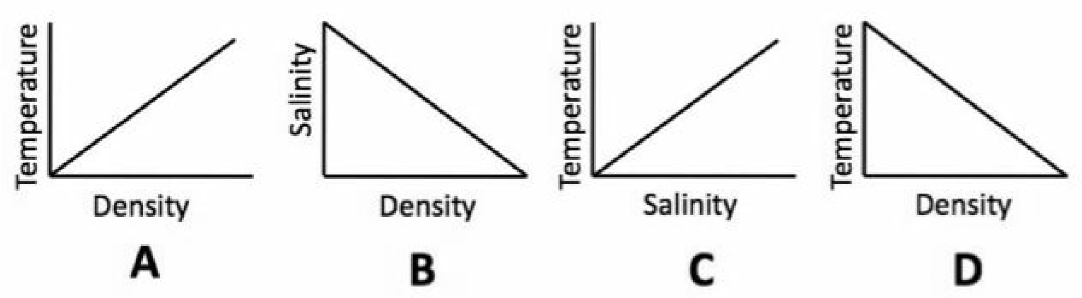
* 1. Which of the following best describes the distribution of saltwater on Earth?

- There is one ocean.
- There are four oceans.
- There are five oceans.
- There are seven oceans.

* 2. Which of the following is the most dense?

- Cold salty water
- Warm salty water
- Cold fresh water
- Warm fresh water

Use this figure to answer Question 3.



* 3. Which graph shows the correct relationship in liquid water?

- Graph A
- Graph B
- Graph C
- Graph D

* 4. The warm Amazon River flows into the warm Atlantic ocean basin. What do you expect will happen?

- The river and ocean water will mix evenly.
- The river water will be repelled by the ocean water, and flow back up the river.
- The river water will sink in the ocean.
- The river water will float on the ocean water.

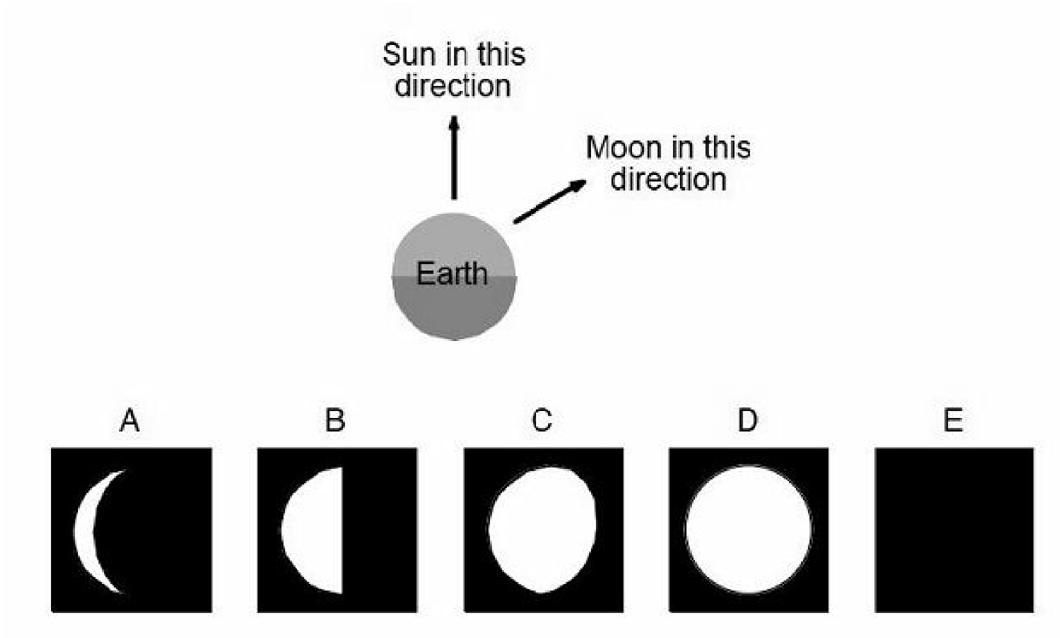
* 5. When an object is subsurface floating (neutrally buoyant):

- the gravitational force is greater than the buoyant force.
- the buoyant force is greater than the gravitational force.
- the gravitational force and the buoyant force are equal.

* 6. Imagine that the moon suddenly disappeared. What would happen to the tides in Earth's oceans?

- The tides would be the same as before the moon disappeared.
- There would only be smaller tides caused by the sun.
- There would be a permanent low tide everywhere.
- There would be no tides.

Use the following figure to answer Question 7.



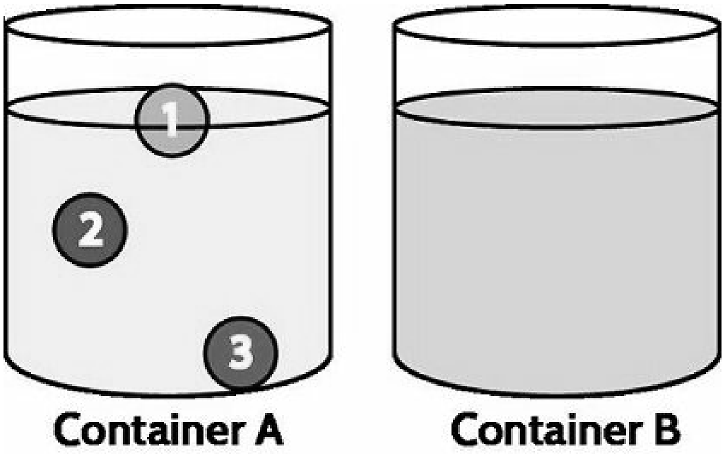
* 7. If you could look down from space at Earth from far above its north pole, the sun and moon would be in the directions shown by the arrows in the picture above. What would the moon look like to a person on Earth facing the moon? (*Select from among the boxes above labeled A through E.*)

- A
- B
- C
- D
- E

* 8. The Moon ________ on its axis. (*Select the best choice to fill in the blank.*)

- rotates once a day
- rotates once a month
- rotates once a year
- does not rotate

Use the following figure to answer Question 9.



* 9. Look at the figure above. In Container A, there are three different objects with the same size and shape, and each is either floating, subsurface floating, or sinking. Which of these three objects would you use to see if the liquid in container B is more or less dense than the liquid in container A?

- ◯ Object #1
- ◯ Object #2
- ◯ Object #3

* 10. Water can be separated into its elemental components of hydrogen and oxygen by:

- ◯ boiling, condensing, and then freezing.
- ◯ digestion by ultraviolet sensitive bacteria.
- ◯ passing an electric current through liquid water.
- ◯ heating to high temperature (e.g., in volcanic eruptions).

* 11. Chemically separating water into its elemental components is what kind of reaction?

- ◯ Single replacement
- ◯ Double replacement
- ◯ Synthesis
- ◯ Decomposition

* 12. Which of the following forces best explains the surface tension of water?

- ◯ Ionic bonds
- ◯ Non-polar covalent bonds
- ◯ Strong forces
- ◯ Hydrogen bonds

* 13. After you take a shower, your hair remains wet for some time. The property of water that helps explain its wetting ability is:

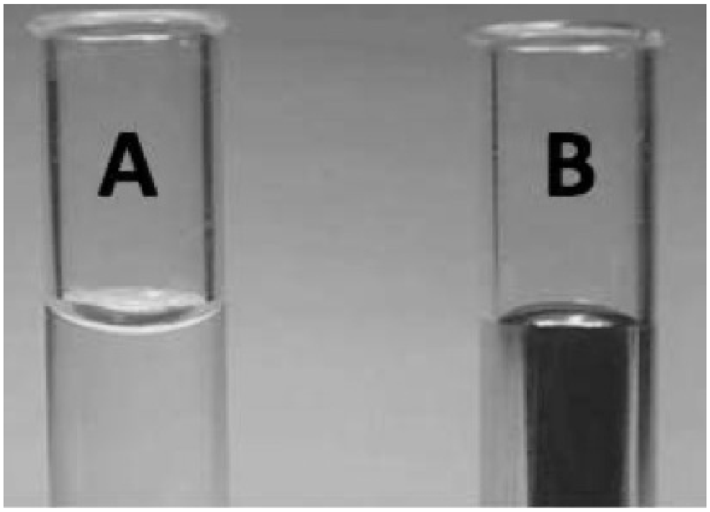
- ◯ adhesion.
- ◯ cohesion.
- ◯ surface tension.
- ◯ holding capacity.

* 14. Hydrogen bonds are ____________ caused by the ____________ sharing of electrons in the O-H bond.

(Please select the pair of terms that correctly fills in the blank spaces.)

- ◯ bonds within the water molecule, equal
- ◯ intermolecular forces between water molecules, equal
- ◯ bonds within the water molecule, unequal
- ◯ intermolecular forces between water molecules, unequal

Use the figure below to answer Question 15. Tube A shows how water appears in a glass tube. Tube B shows how the same volume of mercury appears in a glass tube. Why do water and mercury take on different shapes in a glass tube?



* 15. The ____________ forces between molecules of mercury are stronger relative to the ____________ forces between the molecules of mercury and those of the container.

(Please select the pair of terms that correctly fills in the blank spaces.)

- ◯ cohesive, cohesive
- ◯ cohesive, adhesive
- ◯ adhesive, adhesive
- ◯ adhesive, cohesive

* 16. Water is a good solvent for salts because it:

- is a polar compound.
- is a very cohesive liquid.
- dissolves nonpolar substances easily.
- has a high surface tension.

* 17. Although water is often called the "universal solvent" for its ability to dissolve a large range of matter, water is not an effective solvent of oil. This is because:

- water is denser than oil.
- oil is a nonpolar compound.
- oil is more cohesive than water.
- oil is made largely of carbon.

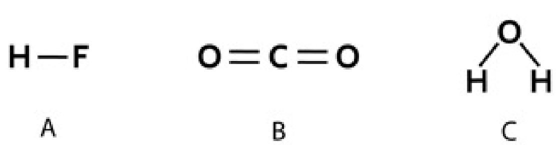
* 18. Which of the following substances will most increase the conductivity of pure water?

- Sugar
- Alcohol
- Starch
- Baking soda

* 19. Which of the following is the best example of a chemical change?

- Osmosis of water
- Electrolysis of water
- Distillation of water
- Dissolution of salt in water

Use the figure below to answer Question 20.



* 20. Which of the compounds above is/are very likely to be polar?

- A and B
- A and C
- B and C
- only C

* 21. Which of the following is the most accurate statement about scientific laws?

- They are general descriptions explaining events that occur under a broad range of conditions.
- They can change based on the rulings of the scientific authority.
- They can incorporate facts, theories, and tested hypotheses.
- They mathematically predict the behavior of natural phenomena.

* 22. Which of the following represents a *scientifically correct* use of the word *theory*?

- The scientist proved her theory about ocean pH in an experiment.
- It is a well-known theory that shaving your hair makes it grow back thicker.
- Gene theory describes how hereditary information is transmitted.
- Newton had a theory that force is proportional to acceleration or F = ma.

* 23. Why is the law of gravity not considered a theory?

- There is not enough experimental evidence to elevate it to a theory.
- One person alone, Newton, cannot establish a theory.
- It is concrete and well-supported by evidence.
- It predicts or describes events rather than explaining them.

* 24. In what statement is the bolded word used *correctly*?

- My **hypothesis** is that if I hear creaking sounds at night, my house is haunted.
- In my **opinion**, if I make both vanilla and chocolate cupcakes, more people will eat the chocolate cupcakes.
- My **theory** is that it was the salad at the party that caused everyone to get sick.
- Newton's **law** of inertia causes you to keep moving forward a bit if the car you are riding in stops moving.

* 25. Of the following statements, which is most correct in explaining why evolution by natural selection is considered a theory?

◯ It is the best-supported explanation of the world based on evidence.

◯ It incorporates a number of scientific laws.

◯ It is the prevailing hypothesis among scientists.

◯ It has been tested and observed to the point that it no longer needs testing.

* 26. The role of the fish operculum is to:

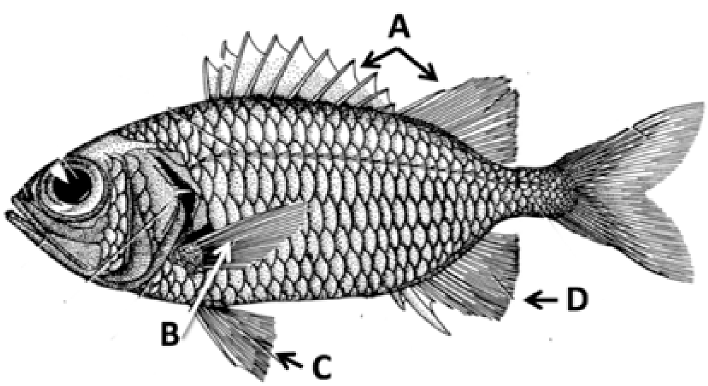
◯ protect the gills.

◯ help the fish detect water movement.

◯ help the fish maintain neutral buoyancy.

◯ help filter-feeding fish consume plankton.

Use the picture directly below to answer question #27.



* 27. Which of the fin(s) in the figure directly above is/are *not* positioned farther forward (towards the fish's head) in the more evolutionarily advanced fishes?

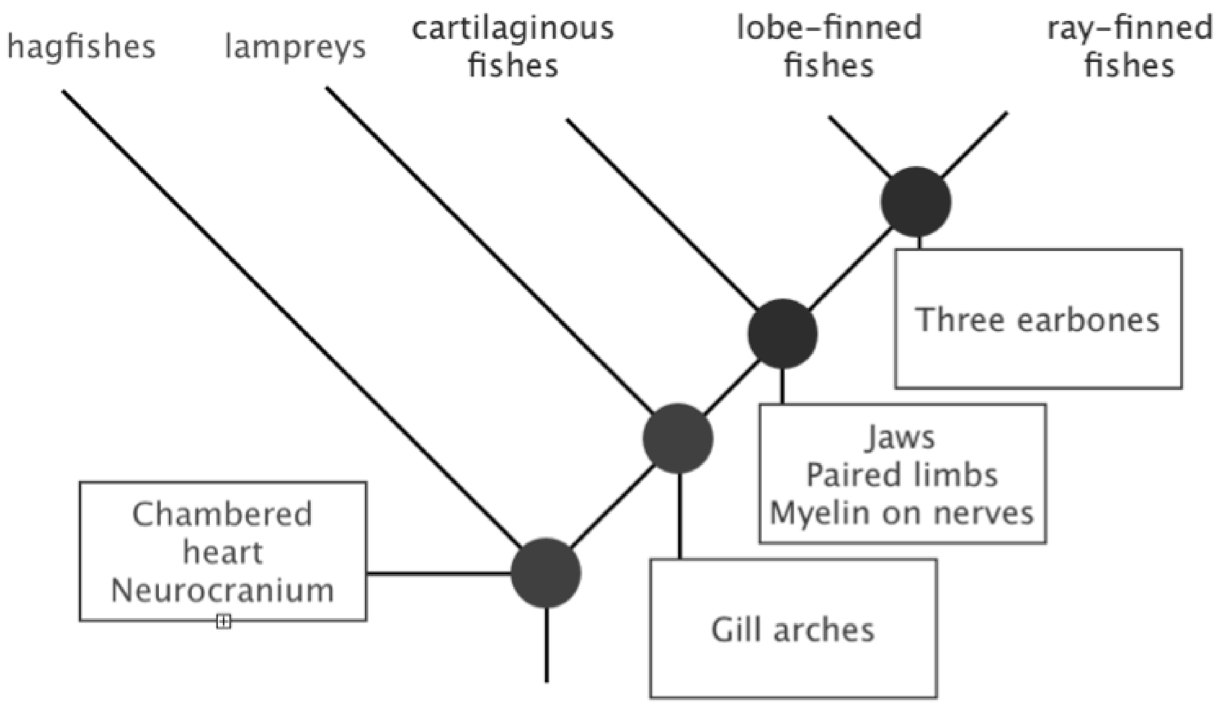
◯ A

◯ B

◯ C

◯ D

Use the figure directly below to answer question #28.



* 28. A phylogenetic tree (such as the one directly above) shows:

- ◯ cartilaginous fish are central in the evolution of fish.
- ◯ common ancestry.
- ◯ kingdom classification.
- ◯ shared ecological niches.

Use the figure directly below to answer question #29.

Lake Malawi Cichlids

* 29. Lake Malawi is a lake that has many different species of cichlids (see picture directly above). This is an example of:

- changing adaptations.
- population interbreeding.
- species radiation.
- gene regulation.

* 30. Variation in populations arises through:

- replication.
- mutation.
- environmental pressure.
- geologic change.

* 31. Microevolution is:

- change within a population.
- change within an individual.
- the evolution of a new species.
- the evolution of small changes.

* 32. Even when you use a reliable random sampling method and have a good sample size, your data might not represent that of the whole population. For this reason, scientists generally:

- prefer to conduct a census where all individuals are represented.
- disregard samples that are unrepresentative due to random variation.
- take multiple samples and calculate the probability of accuracy.
- modify sampling techniques for each data point collected to ensure representative sampling.

* 33. Which of the following is true of systematic sampling?

- It is not a valid way to survey an area when random sampling is possible.
- It characterizes an environment but does not allow us to generalize beyond the specific sample.
- It is an appropriate way of characterizing an environment, depending on the survey goal.
- It is used to collect information about every item you are interested in studying in an area.

* 34. Clancy wants to describe the fish population in a pond. Which of the following is the most important issue in designing his sampling procedure?

- Taking as many fish samples as possible.
- Making sure the study design allows him to answer his study questions.
- Making sure he measures the number of fish with a least three different sampling strategies.
- Making sure sampling locations within the pond are randomly chosen.

* 35. Jordan does an experiment on his students. He places a handful of carrots in the middle of the room and the entire bag of chocolate kisses in the corner of the room in identical glass bowls. At the end of the day, there are no carrots left but there are lots of chocolate kisses remaining, so Jordan thinks his students like carrots better than chocolate. What is the *biggest* mistake in Jordan's experiment?

- Jordan did not repeat the procedure multiple times.
- Jordan does not have a control variable for this experiment.
- Jordan did not standardize enough variables to be able to compare preferences.
- Jordan chose variables that are not comparable—chocolate and vegetables.

* 36. Which of the following is the most appropriate strategy for sampling high mobile organisms like fish?

- transect point intercept
- quadrat point intercept
- quadrat percent cover
- timed visual estimate

* 37. Which of the following is the most appropriate sampling strategy for quickly studying the substrate in a large area (such as the deep ocean floor) that is dominated by a few species?

- transect point intercept
- quadrat point intercept
- quadrat percent cover
- timed visual estimate

* 38. The transect point intercept sampling method would work best in:

- a complex high relief area, like a coral reef.
- an area with many active mobile organisms, like a forest canopy.
- a large uniform area, like a field or sandy beach.
- a location with many small cryptic organisms, like a tidepool.

* 39. A kelp forest has been altered through predation by a population of sea urchins. There are still many species, but the numbers of individuals of most species were depleted. Despite predation, this area still has high species:

- ◯ biodiversity.
- ◯ abundance.
- ◯ evenness.
- ◯ richness.

* 40. Which of the following fish tanks has the highest species evenness?

Tank A: 3 brittle stars, 2 sea cucumbers, 15 urchins, 1 cichlid, 6 oysters, 64 snails, 40 mussels
Tank B: 6 shrimp, 5 flatworms, 4 crabs, 5 sea slugs, 6 wrasse
Tank C: 964 blennies, 321 gobies
Tank D: 4 dolphins, 1 great white shark, 1 humpback whale

- ◯ Tank A
- ◯ Tank B
- ◯ Tank C
- ◯ Tank D

* 41. An oceanographer needs to go out in a boat to retrieve an important temperature-and-salinity data logger that is attached to an underwater stake on a coral reef. When the oceanographer's Global Positioning System (GPS) indicates that she is at the location of the stake, she anchors the boat and jumps in the water to collect the data logger. However, she can't find the stake. The other GPS units belonging to her colleagues on the boat also indicate that they are at the correct location. After an extensive search, the oceanographer finds the stake 50 meters (m) from the boat. Assuming the stake was at its assigned location, the GPS units in this scenario were:

- ◯ accurate but not precise.
- ◯ precise but not accurate.
- ◯ accurate and precise.
- ◯ neither accurate nor precise.

APPENDIX E

Content Knowledge Assessment Blueprint

Module	Item No.	Content	*p*-value change	*p*-value pre	pbcor	Item-total cor
1	1	one ocean	0.46	0.43	0.66	0.63
1	9	density	0.14	0.82	0.35	0.32
1	10	density	0.14	0.54	0.48	0.42
1	12	density	0.21	0.75	0.32	0.29
1	19	density	0.14	0.82	0.43	0.41
1	21	moon/tides	0.18	0.68	0.64	0.61
1	22	moon/tides	0.32	0.36	0.50	0.45
1	24	moon/tides	0.32	0.46	0.19	0.13
1	30	density	0.21	0.71	0.46	0.43
2	3	electrolysis	0.18	0.79	-0.12	-0.14
2	5	electrolysis	0.25	0.68	0.25	0.21
2	10	water properties	0.32	0.46	0.74	0.71
2	11	water properties	0.36	0.54	0.22	0.18
2	13	water properties	0.14	0.39	0.54	0.49
2	15	water properties	0.14	0.61	0.27	0.21
2	18	solubility/conductivity	0.14	0.54	0.77	0.73
2	19	solubility/conductivity	0.18	0.57	0.75	0.72
2	21	solubility/conductivity	0.46	0.54	n/a	n/a
2	28	electrolysis	0.43	0.50	0.21	0.17
2	34	water properties	0.14	0.43	0.73	0.69
3	4	sci lang	0.36	0.25	0.31	0.25
3	5	sci lang	0.25	0.57	0.33	0.28
3	8	sci lang	0.29	0.29	0.53	0.47
3	9	sci lang	0.25	0.21	0.20	0.13
3	11	sci lang	0.14	0.71	0.37	0.32
3	18	fish	0.36	0.54	0.43	0.39
3	19	fish	0.14	0.07	0.24	0.18
3	21	fish	0.18	0.68	0.55	0.51
3	28	evolution	0.18	0.57	0.73	0.69
3	29	evolution	0.18	0.71	0.48	0.45
3	33	evolution	0.29	0.54	0.29	0.24
4	2	sampling	0.18	0.82	n/a	n/a
4	3	sampling	0.25	0.43	0.64	0.59
4	5	sampling	0.18	0.43	0.23	0.16
4	6	sampling	0.14	0.68	0.14	0.08
4	8	abundance	0.43	0.36	0.34	0.28
4	10	abundance	0.14	0.25	0.51	0.45
4	13	abundance	0.32	0.32	0.59	0.54
4	17	abundance	0.29	0.14	0.35	0.28
4	18	abundance	0.18	0.71	0.57	0.54
4	20	sampling	0.14	0.32	0.48	0.42

Note. Item numbers from original questionnaires, correlations to final test (of 41 items, not original 141 items)

APPENDIX F

Teaching Science as Inquiry
Aquatic Professional Development
Follow-up Interview
Spring 2016

1

Logistics

- What subject(s) and grade(s) have you taught since you participated in the TSI PD (ended spring 2013)? At what school(s)?
- Briefly describe any significant professional development experiences you have had since the end of the TSI PD.
- Please share your contact information for us (for stipend purposes).
- Please share a contact number I can reach you at in case something happens to the internet connection during the interview.

2

2

Introduction

- Confidential *(your name will not be shared in any presentations or publications based on these results)*
- Not an evaluation of you
- Please be forthright and honest

3

3

Practical Significance

- What do you think is the most significant ***content*** you learned as a result of participating in the TSI PD?
- What do you think is the most significant ***pedagogical method or strategy*** that you learned as a result of participating in the TSI PD?

4

4

Application

- To what extent have you applied what you learned in the PD to your teaching practice?

Not at all	1	2	3	4	5	6	7	Very much

Please explain your rating.

- What, if anything, has enhanced or hindered the degree to which you can apply what you learned in the PD?

TSI aquatic

5

5

Implementation

- How much would you say you have implemented TSI-A activities? (none, a few, several, a lot)
- What are your primary reasons for implementing TSI-A activities?
- If you modify or customize TSI-A activities when you implement them, please explain how and why.
- What, if anything, has enhanced or hampered the degree to which you have used the activities?

TSI aquatic

6

6

Teaching Practice

Do you think your teaching practice changed as a result of participating in the TSI PD?

- If you believe it has changed, please describe how.
- If it has not changed, why do you think that is?
- If you think the PD did not change your teaching practice, do you believe it confirmed existing practices, and if so, what?

TSI aquatic

7

7

Professional Trajectory

How do you think the TSI PD fit into your overall learning and professional growth?

What could we have offered that would have helped sustain the implementation of TSI-A in your teaching?

TSI aquatic

8

8

Parting Comments

Do you have any additional thoughts or comments?

Thank you for you time!

TSI aquatic

9

9

APPENDIX G

Examinations of Non-Normality

Table G.1

Shapiro-Wilks Test of Normality for Quantitative Instruments

Difference scores between time points	$W(23)$	p-value
Inquiry Teaching Assessment		
Pre-PD to post-PD	.964	.547
Post-PD to follow-up	.979	.885
Pre-PD to follow-up	.931	.115
Content assessment		
Pre-PD to post-PD	.980	.907
Post-PD to follow-up	.952	.321
Pre-PD to follow-up	.950	.291
Pedagogical content knowledge assessment		
Pre-PD to post-PD	.965	.565
Post-PD to follow-up	.794	$< .001^{**}$
Pre-PD to follow-up	.935	.138
Self-efficacy assessment		
Pre-PD to post-PD	.896	$.021^{*}$
Post-PD to follow-up	.913	$.048^{*}$
Pre-PD to follow-up	.956	.393

$^{*}p < .05$. (two-tailed), $^{**}p < .01$. (two-tailed)

Figure G.1

Histogram of the Pedagogical Content Knowledge Assessment Difference Scores Between the Post-PD and Follow-up Time Points

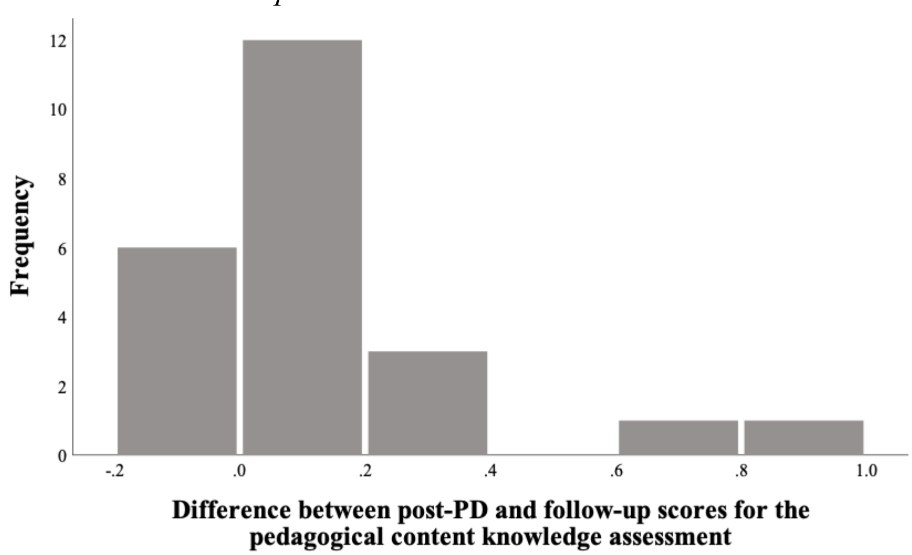


Figure G.2

Histogram of the Self-Efficacy Assessment Difference Scores Between the Pre-PD and Post-PD Time Points

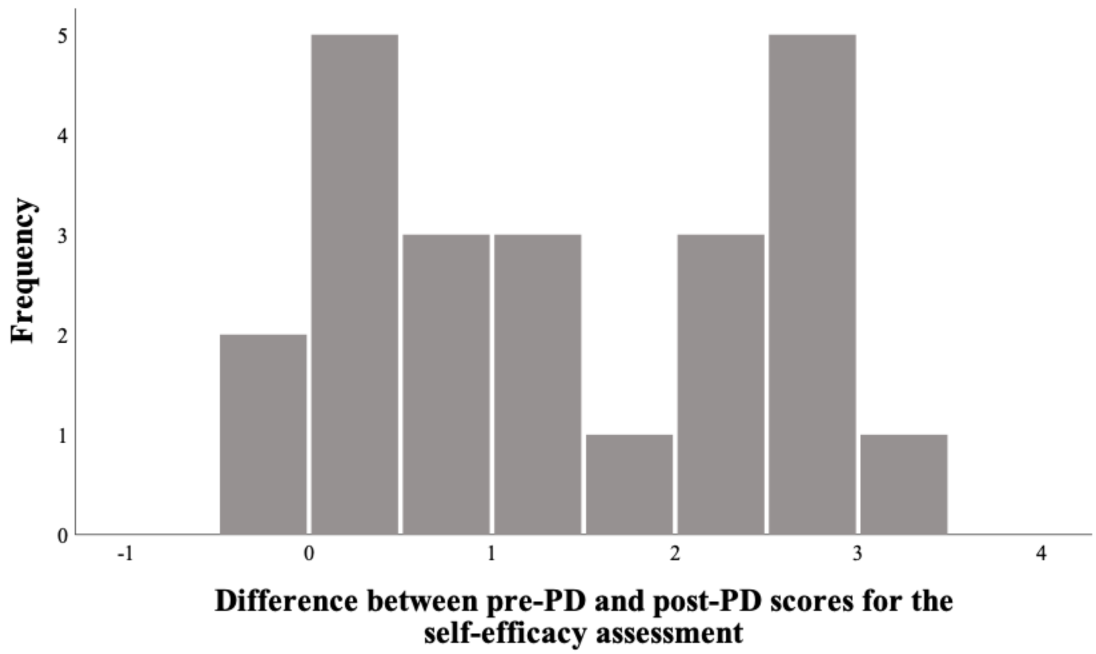
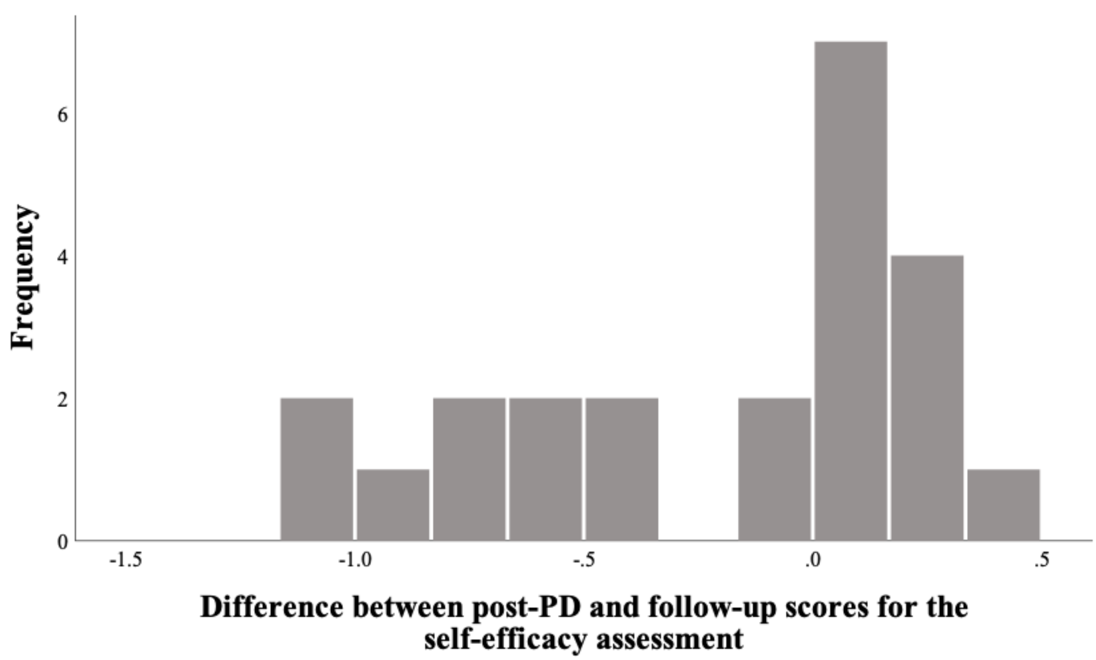


Figure G.3

Histogram of the Self-Efficacy Assessment Difference Scores Between the Post-PD and Follow-up Time Points



REFERENCES

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, *82*(4), 417–436. https://doi.org/10.1002/(SICI)1098-237X(199807)82:4<417::AID-SCE1>3.0.CO;2-E

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A.,Niaz, M., Treagust, D., & Tuan, H. L. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*(3), 397–419. https://doi.org/10.1002 /sce.10118

Abrami, P. C., Poulsen, C., & Chambers, B. (2004). Teacher motivation to implement an educational innovation: Factors differentiating users and non-users of cooperative learning. *Educational Psychology*, *24*(2), 201–216. https://doi.org/10.1080/ 0144341032000160146

Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review*, *14*(4), 374-390. https://doi.org/10.1177/0193841x9001400403

Airasian, P. W., & Walsh, M. E. (1997). Constructivist cautions. *Phi Delta Kappan*, *78*(6), 444–449. https://www.proquest.com/docview/218492309?pqorigsite=gscholar &fromopenview=true

Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development program. *Journal of Research in Science Teaching*, *44*(5), 653–680. https://doi.org/10.1002/tea.20159

Akiba, M., & Liang, G. (2016). Effects of teacher professional learning activities on student achievement growth. *The Journal of Educational Research*, *109*(1), 99–110. https://doi.org/10.1080/00220671.2014.924470

Akuma, F. V., & Callaghan, R. (2019). Characterizing extrinsic challenges linked to the design and implementation of inquiry-based practical work. *Research in Science Education*, *49*(6), 1677–1706. https://doi.org/10.1007/s11165-017-9671-x

Alexander, D., Heaviside, S., & Farris, E. (1999). *Status of Education Reform in Public Elementary and Secondary Schools: Teachers' Perspectives*. Statistical analysis report. National Center for Education Statistics. https://nces.ed.gov/pubs99/1999045.pdf

Allen, C. D., & Penuel, W. R. (2015). Studying teachers' sensemaking to investigate teachers' responses to professional development focused on new standards. *Journal of Teacher Education*, *66*(2), 136–149. https://doi.org/10.1177/0022487114560646

Alozie, N. M., Moje, E. B., & Krajcik, J. S. (2010). An analysis of the supports and constraints for scientific discussion in high school project-based science. *Science Education*, *94*(3), 395–427. https://doi.org/10.1002/sce.20365

American Association for the Advancement of Science (AAAS). (1990). *Science for all Americans*. Oxford University Press.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. Oxford University Press.

American Federation of Teachers. (2017). *2017 Educator Quality of Work Life Survey*. American Federation of Teachers. https://www.aft.org/sites/default/files/2017_eqwl_survey_web.pdf

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/bf02293814

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education*, *29*, 1–12. https://doi.org/10.1016/j.tate.2012.08.001

Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2012). *Effects of Curriculum and Teacher Professional Development on the Language Proficiency of Elementary English Language Learner Students in the Central Region* (NCEE 2012–4013). Mid-continent Research for Education and Learning. https://files.eric.ed.gov/fulltext/ED530839.pdf

Armbruster, P., Patel, M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE—Life Science Education*, *8*(3), 203–213. https://doi.org/10.1187/cbe.09-03-0025

Asterhan, C., & Lefstein, A. (2020). *Evidence-based Design Principles for Effective Professional Development: A Critical Appraisal of the Evidence*. ICLS 2020 Proceedings, 2046-2052. https://repository.isls.org/bitstream/1/6491/1/2046-2052.pdf

Aulls, M. W., & Shore, B. M. (2008). *Inquiry in Education: Vol. 1. The Conceptual Foundations for Research as a Curricular Imperative*. Taylor & Francis Group; Lawrence Erlbaum Associates.

Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, *27*(1), 10–20. https://doi.org/10.1016/j.tate.2010.08.007

Ayala, C. C. (2005, October 26-29). *Development and Validation of an Inquiry Science Student Achievement* [Paper presentation]. American Evaluation Association Annual Conference, Toronto, Ontario, Canada.

Bandura, A. (1972). *Modeling theory: Some traditions, trends, and disputes*. In R.D. Parke (Ed.), Recent trends in social learning theory (pp. 35–61). Academic Press. https://doi.org/10.1016/B978-0-12-545050-8.50008-0

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, *33*(4), 344–358. https://doi.org/10.1037/0003-066X.33.4.344

Bandura, A. (1983a). Self-efficacy determinants of anticipated fears and calamities. *Journal of Personality and Social Psychology*, *45*(2), 464–469. https://doi.org/10.1037/0022-3514.45.2.464

Bandura, A. (1983b). Temporal dynamics and decomposition of reciprocal determinism: A reply to Phillips and Orton. *Psychological Review*, 90(2), 166–170. Ihttps://doi.org/10.1037/0033-295X.90.2.166

Bandura, A., & National Institute of Mental Health. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175

Bandura, A. (Ed.). (1995). *Self-efficacy in Changing Societies*. Cambridge university press.

Bandura, A. (1997). The anatomy of stages of change. *American Journal of Health Promotion*, *12*(1), 8–10. https://doi.org/10.4278/0890-1171-12.1.8

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the national science foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching*, *44*(3), 375–395. https://doi.org/10.1002/tea.20145

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education* (ED541798). ERIC. https://files.eric.ed.gov/fulltext/ED541798.pdf

Barron, B., & Darling-Hammond, L. (2008). Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning. In L. Darling-Hammond, B. Barron, P. D. Pearson, A. H. Schoenfeld, E. K. Stage, T. D. Zimmerman, G. N. Cervetti, & J. Tilsonn (Eds.), *Powerful Learning: What we Know About Teaching for Understanding*. Jossey-Bass. https://files.eric.ed.gov/fulltext/ED539399.pdf

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*(9), 1150–1184. https://doi.org/10.1002/tea.21168

Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2019). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly*, *54*(2), 133–166. https://doi.org/10.1002/rrq.229

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. https://doi.org/10.1007/BF03173676

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, *71*(2), 287-311. https://doi.org/10.1093/poq/nfm006

Belden, Russonello, & Stewart; & American Viewpoint. (1999, October). Communicating about oceans: Results of a national survey. *The Ocean Project*. https://theoceanproject.org/wp-content/uploads/2017/08/Communicating-About-Oceans_Results-of-National-Survey-1999.pdf

Bell, B., & Gilbert, J. K. (1996). *Teacher Development: A Model from Science Education*. Psychology Press.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* (Methodological), *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bierman, K. L., Sanford DeRousie, R. M., Heinrichs, B., Domitrovich, C. E., Greenberg, M. T., & Gill, S. (2013). Sustaining high-quality teaching and evidence-based curricula: Follow-up assessment of teachers in the REDI project. *Early Education & Development*, *24*(8), 1194–1213. https://doi.org/10.1080/10409289.2013.755457

Blank, R. K., & de las Alas, N. (2009). *The Effects of Teacher Professional Development on Gains in Student Achievement: How Meta analysis Provides Scientific Evidence Useful to Education Leaders* (ED544700). ERIC. https://files.eric.ed.gov/fulltext/ED544700.pdf

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist*, *35*(3), 149–164. https://doi.org/10.1207/S15326985EP3503_2

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Bos, J. M., Sanchez, R. C., Tseng, F., Rayyes, N., Ortiz, L., & Sinicrope, C. (2012). *Evaluation of Quality Teaching for English Learners* (*QTEL) Professional Development* (NCEE 2012–4005). National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences; U.S. Department of Education. https://files.eric.ed.gov /fulltext/ED529843.pdf

Boubonari, T., Markos, A., & Kevrekidis, T. (2013). Greek pre-service teachers' knowledge, attitudes, and environmental behavior toward marine pollution. *The Journal of Environmental Education*, *44*(4), 232–251. https://doi.org/10.1080/00958964 .2013.785381

Brandon, P. R., Taum, A. K. H., Young, D. B., Speitel, T. W., Pottenger III, F. M., Nguyen, T. T., & Gray, 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*. Curriculum Research & Development Group, University of Hawai‘i at Mānoa. https://manoa.hawaii.edu/crdg/wp-content/uploads/SCUP_final_report.pdf

Bredo, E. (1997). *The Social Construction of Learning*. In: G. D. Phye (Ed.) Handbook of academic learning: Construction of knowledge (pp. 3–46). Academic Press.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Weinsert, & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Lawrence Erlbaum Associates.

Bruner, J. S. (1966). *Toward a Theory of Instruction* (Vol. 59). Harvard University Press.

Buczynski, S., & Hansen, C. B. (2010). Impact of professional development on teacher practice:

Uncovering connections. *Teaching and Teacher Education*, *26*(3), 599–607. https://doi.org/10.1016/j.tate.2009.09.006

Bulunuz, N., & Jarrett, O. S. (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 and Technology Education*, *6*(2), 85–99. https://doi.org/10.12973/ejmste/75230

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*. BSCS. https://media.bscs.org/bscsmw/5es/bscs_5e_full_report.pdf

Cairns, D., & Areepattamannil, S. (2019). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education*, *49*(1), 1–23. https://doi.org/10.1007/s11165-017-9639-x

Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics*, *103*(4), 177–185. https://doi.org/10.1111/j.1949-8594.2003.tb18116.x

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*(3), 291–318. https://doi.org/10.1007/s10972-012-9275-2

Carley, S. (n.d.). *Why ocean literacy?*. Ocean Literacy. http://oceanliteracy.wp2.coexploration .org/?page_id=62

Castro, F. G., Barrera, M., & Martinez, C. R. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, *5*(1), 41–45. https://doi.org/10.1023/B:PREV.0000013980.12412.cd

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Clary, R. M., Elder, A., Dunne, J., Saebo, S., Beard, D., Wax, C., & Tucker, D. L. (2018). Beyond the professional development academy: Teachers' retention of discipline-specific science content knowledge throughout a 3-year mathematics and science partnership. *School Science and Mathematics*, *118*, 75–83. https://doi.org/10.1111/ssm.12267

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, *38*(2), 136–163. https://doi.org/10.2307/30034954

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education*, *119*(1), 137–182. https://doi.org/10.1086/667699

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, *44*(4), 263–272. https://doi.org/10.1177/0022487193044004004

Cohen, D. H., & Spillane, J. P. (1993). Policy and practice: The relations between governance and instruction. In S.H. Fuhrman (Ed.), *Designing Coherent Education Policy: Improving the System* (pp. 35–95). Jossey-Bass publishers.

Cohen, D. K., & Hill, H. C. (1998). *Instructional Policy and Classroom Performance: The Mathematics Reform in California*. Consortium for Policy Research in Education.

Cohen, D. K., & Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. Yale University Press.

Copur-Gencturk, Y., & Papakonstantinou, A. (2016). Sustainable changes in teacher practices: A longitudinal analysis of the classroom practices of high school mathematics teachers. *Journal of Mathematics Teacher Education*, *19*(6), 575–594. https://doi.org/10.1007/s10857-015-9310-2

Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage publications.

Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching*, *44*(4), 613–642. https://doi.org/10.1002/tea.20157

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage publications.

Curriculum Research & Development Group (CRDG). (2001). *Foundational Approaches in Science Teaching (FAST): Description and Overview*. Curriculum Research & Development Group, University of Hawaiʻi at Mānoa.

Curriculum Research & Development Group (CRDG). (2012). *Designing professional development for educators*. Curriculum Research & Development Group, University of Hawaiʻi at Mānoa.

Curriculum Research & Development Group (CRDG). (2020). *CRDG Professional Development Change Mode*l. Curriculum Research & Development Group, University of Hawaiʻi at Mānoa.

Darling-Hammond, L., Chung-Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the U.S. and Abroad*. National Staff Development Council; The School Redesign Network at Stanford University. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.180.6634&rep=rep1&type=pdf

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development* (ED606743). ERIC. https://files.eric.ed.gov/fulltext/ED606743.pdf

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, *5*(1–2), 183–204. https://doi.org/10.1080/10824669.2000.9671386

Deglau, D., & O'Sullivan, M. (2006). The effects of a long-term professional development program on the beliefs and practices of experienced teachers. *Journal of Teaching in Physical Education*, *25*(4), 379–396. https://doi.org/10.1123/jtpe.25.4.379

DeMonte, J. (2013). *High-Quality Professional Development for Teachers: Supporting Teacher Training to Improve Student Learning* (ED561095). ERIC. https://files.eric.ed.gov/fulltext/ED561095.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189x08331140

Desimone, L. M., & Garet, M. S. (2015). Best practices in teacher's professional development in the United States. *Psychology, Society and Education*, *7*(3), 252–263. https://doi.org/10.25115/psye.v7i3.515

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112. https://doi.org/10.3102/01623737024002081

Desimone, L. M., & Stuckey, D. (2014). Sustaining teacher professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of Professional Development in Education: Successful Models and Practices, PreK–12* (pp. 467–482).

Devlin-Scherer, W., Spinelli, A. M., Giammatteo, D., Johnson, C., Mayo-Molina, S., McGinley, P., Michalski, C., Schmidek, S., Tomaiuolo, L., & Zisk, L. (1998). *Action Research in Professional Development Schools: Effects on Student Learning* (ED418077). ERIC. https://files.eric.ed.gov/fulltext/ED418077.pdf

Dewey, J. (1938). *Experience and Education*. Kappa Delta Pi.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-mode Surveys: The Tailored Design Method* (4th ed.). John Wiley & Sons.

Drits-Esser, D., Gess-Newsome, J., & Stark, L. A. (2017). Examining the sustainability of teacher learning following a year-long science professional development programme for inservice primary school teachers. *Professional Development in Education*, *43*(3), 375–396. https://doi.org/10.1080/19415257.2016.1179664

Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950–963. https://doi.org/10.1016/s0003-9993(03)00035-2

Duncan, R. M. (1995). Piaget and Vygotsky revisited: Dialogue or assimilation?. *Developmental Review*, *15*(4), 458–472. https://doi.org/10.1006/drev.1995.1019

Duncan Seraphin, K. (Ed.). (2014). *Accessible Professional Development for Teaching Aquatic Science Inquiry: Final report*. Curriculum Research & Development Group, University of Hawaiʻi at Mānoa. https://manoa.hawaii.edu/crdg/research-development/research-programs/teaching-science-as-inquiry/teaching-science-as-inquiry-aquatic/

Duncan Seraphin, K., & Baumgartner, E. (2010). Students as scientists: guidelines for teaching science through disciplinary inquiry. In R. E. Yager (Ed.), *Exemplary Science for Resolving Societal Challenges* (pp. 33–50). NSTA Press.

Duncan Seraphin, K., Harrison, G. M., Philippoff, J., Brandon, P. R., Nguyen, T. T. T., Lawton, B. E., & Vallin, L. M. (2017). Teaching aquatic science as inquiry through professional development: Teacher characteristics and student outcomes. *Journal of Research in Science Teaching*, *54*(9), 1219–1245. https://doi.org/10.1002/tea.21403

Duncan Seraphin, K., Philippoff, J., & Kaupp, L. J. (2014). *Introduction to Teaching Science as Inquiry* [PowerPoint slides].

Duncan Seraphin, K., Philippoff, J., Kaupp, L., & Vallin, L. M. (2012). Metacognition as means to increase the effectiveness of inquiry-based science education. *Science Education International*, *23*(4), 366–382. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1091.5776&rep=rep1&type=pdf

Duncan Seraphin, K., Philippoff, J., Parisky, A., Degnan, K., & Warren, D. P. (2013). 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*, *22*(3), 235–251. https://doi.org/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*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Eylon, B. S., & Bagno, E. (2006). Design model for professional development of teachers: Designing lessons with physics education research. Physical review special topics-*Physics Education Research*, *2*(2), 020106. https://doi.org/10.1103/PhysRevSTPER.2.020106

Falk, J. H. (2006). An identity-centered approach to understanding museum learning. Curator: *The Museum Journal*, *49*(2), 151–166. https://doi.org/10.1111/j.2151-6952.2006.tb00209.x

Fauville, G., McHugh, P., Domegan, C., Mäkitalo, Å., Møller, L. F., Papathanassiou, M., Chicote, C. A., Lincoln, S., Batista, V., Copejans, E., Crouch, F., & Gotensparre, S. (2018). Using collective intelligence to identify barriers to teaching 12–19 year olds about the ocean in Europe. *Marine Policy*, *91*, 85–96. https://doi.org/10.1016/j.marpol.2018.01.034

Fishman, B. J., & Krajcik. (2003). What does it mean to create sustainable science curriculum innovations? A commentary. *Science Education*, *87*(4), 564–573. https://doi-org.eres.library.manoa.hawaii.edu/10.1002/sce.10088

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, *19*(6), 643–658. https://doi.org/10.1016/s0742-051x(03)00059-3

Fitzgerald, M., Danaia, L., & McKinnon, D. H. (2019). Barriers inhibiting inquiry-based science teaching and potential solutions: perceptions of positively inclined early adopters. *Research in Science Education*, *49*(2), 543–566. https://doi.org/10.1007/s11165-017-9623-5

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. University of South Florida; Louis de la Parte Florida; The National Implementation Research Network. https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (231–235). Lawrence Erlbaum.

Ford, J. K. (1994). Defining transfer of learning: The meaning is in the answers. *Adult Learning*, *5*(4), 22–30. https://doi.org/10.1177/104515959400500412

Franke, M. L., Carpenter, T., Fennema, E., Ansell, E., & Behrend, J. (1998). Understanding teachers' self-sustaining, generative change in the context of professional development. *Teaching and Teacher Education*, *14*(1), 67–80.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, *38*(3), 653–689. https://doi.org/10.3102/00028312038003653

Friedrichsen, P. J., & Barnett, E. (2018). Negotiating the meaning of Next Generation Science Standards in a secondary biology teacher professional learning community. *Journal of Research in Science Teaching*, *55*(7), 999–1025. https://doi.org/10.1002/tea.21472

Fullan, M. (1987). Implementing the implementation plan. In M. Wideen, I. Andrews, & I. H. Andrews (Eds.), *Staff development for School Improvement: A Focus on the Teacher* (pp.213–222). Falmer Press.

Fulmer, G. W., Tanas, J., & Weiss, K. A. (2018). The challenges of alignment for the Next Generation Science Standards. *Journal of Research in Science Teaching*, *55*(7), 1076–1100. https://doi.org/10.1002/tea.21481

Gaikhorst, L., Beishuizen, J. J., Zijlstra, B. J., & Volman, M. L. (2017). The sustainability of a teacher professional development programme for beginning urban teachers. *Cambridge Journal of Education*, *47*(1), 135–154. https://doi.org/10.1080/0305764X.2015.1125449

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*

(NCEE 2008-4030). National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences; U.S. Department of Education

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development* (NCEE 2016-4010). National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences; United States Department of Education. https://ies.ed.gov/ncee/pubs/20164010/pdf/20164010.pdf

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation* (NCEE 2011-4024). National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences; U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED519922.pdf

Gawlik, M. A., Kearney, C. P., Addonizio, M. F., & LaPlante-Sosnowsky, F. (2012). Teacher quality in Michigan: A school-level analysis of the Detroit metropolitan region. *Education and Urban Society*, *44*(4), 412–447. https://doi.org/10.1177/0013124510392568

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome, & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (pp. 3–17). Springer. https://doi.org/10.1007/0-306-47217-1_1

Gess-Newsome, J., & Lederman, N. G. (1993). Preservice biology teachers' knowledge structures as a function of professional teacher education: A year-long assessment. *Science Education*, *77*(1), 25–45.

Givvin, K. B., & Santagata, R. (2011). Toward a common language for discussing the features of effective professional development: The case of a US mathematics program. *Professional Development in Education*, *37*(3), 439–451. https://doi.org/10.1080/19415257.2010.527711

Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long?. *Economics of Education Review*, *29*(3), 432–439. https://doi.org/10.1016/j.econedurev.2009.10.002

Goodale, T., & Sakas, C. (2019). Evaluating differences in outcomes and participant perspectives in marine science professional development conducted by informal educators compared to specialized university faculty. In G. Fauville, D. L. Payne, M. E. Marrero, A. Lantz-Andersson, & F. Crouch (Eds.), *Exemplary Practices in Marine Science Education* (pp. 109–121). Springer. https://doi.org/10.1007/978-3-319-90778-9_7

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomized controlled trial of quality teaching rounds. *Teaching and Teacher Education*, *68*, 99–113. https://doi.org/10.1016/j.tate.2017.08.007

Gorozidis, G., & Papaioannou, A. G. (2014). Teachers' motivation to participate in training and to implement innovations. *Teaching and Teacher Education*, *39*, 1–11. https://doi.org/10.1016/j.tate.2013.12.001

Gulamhussein, A. (2013). *Teaching the Teachers: Effective Professional Development in an Era of High Stakes Accountability*. Center for public education, *1*, 1–47.

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, *15*(5), 5–12. https://doi.org/10.3102/0013189X015005005

Guskey, T. R. (2000). *Evaluating Professional Development*. Corwin press.

Guskey, T. R. (2002a). Does it make a difference? Evaluating professional development. *Educational Leadership*, *59*(6), 45–51. http://www.ascd.org/publications/educational-leadership/mar02/vol59/num06/Does-It-Make-a-Difference¢-Evaluating-Professional-Development.aspx

Guskey, T. R. (2002b). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development?. *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Halawa, S., Hsu, Y. S., Zhang, W. X., Kuo, Y. R., & Wu, J. Y. (2020). Features and trends of teaching strategies for scientific practices from a review of 2008–2017 articles. *International Journal of Science Education*, *42*(7), 1183–1206. https://doi.org/10.1080/09500693.2020.1752415

Handler, A., & Duncan, K. (2006). Hammerhead shark research immersion program: Experiential learning leads to lasting educational benefits. *Journal of Science Education and Technology*, *15*(1), 9–16. https://doi.org/10.1007/s10956-006-0352-1

Hansen, W. B., & McNeal Jr, R. B. (1999). Drug education practice: Results of an observational study. *Health education research*, *14*(1), 85–97. https://doi.org/10.1093/her/14.1.85

Hargreaves, A., & Goodson, I. (2006). Educational change over time? The sustainability and nonsustainability of three decades of secondary school change and continuity. *Educational Administration Quarterly*, *42*(1), 3–41. https://doi.org/10.1177/0013161X05277975

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Harrison, G. M. (2020). Validity arguments with survey scales [Accepted chapter]. In U. Luhanga & A. G. Harbaugh (Eds.), *Basic Elements of Survey Research in Education: Addressing the Problems Your Advisor Never Told You About*. Information Age Publishing.

Havice, W., Havice, P., Waugaman, C., & Walker, K. (2018). Evaluating the effectiveness of integrative STEM education: Teacher and administrator professional development. *Journal of Technology Education*, *29*(2), 73–90. https://doi.org/10.21061/jte.v29i2.a.5

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*(3), 333–362. https://doi.org/10.1002/tea.21004

Herrington, D. G., Yezierski, E. J., & Bancroft, S. F. (2016). Tool trouble: Challenges with using self-report data to evaluate long-term chemistry teacher professional development.

Journal of Research in Science Teaching, *53*(7), 1055–1081. https://doi.org/10.1002/tea.21323

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, *42*(9), 476–487. https://doi.org/10.3102/0013189X13512674

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, *55*(5), 1076–1112. https://doi.org/10.3102/0002831218769614

Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, *48*(9), 590–598. https://doi.org/10.3102/0013189X19891436

Hill, H. C., Lynch, K., Gonzalez, K. E., & Pollard, C. (2020). Professional development that improves STEM outcomes. *Phi Delta Kappan*, *101*(5), 50–56. https://doi.org/10.1177/0031721720903829

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark(2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Hoffman, M., & Barstow, D. (2007). *Revolutionizing Earth System Science Education for the 21st Century: Report and Recommendations from a 50-State Analysis of Earth Science Education Standards* (ED497331). ERIC. https://files.eric.ed.gov/fulltext/ED497331.pdf

Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. I. Honig (Ed.), *New Directions in Education Policy Implementation*: *Confronting Complexity* (1–24). State University of New York Press.

Jacob, A., & McGovern, K. (2015). The mirage: Confronting the hard truth about our quest for teacher development. *The New Teacher Project*. https://tntp.org/assets/documents/TNTP-Mirage_2015.pdf

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, *22*(3–4), 255-278. https://doi.org/10.1080/00461520.1987.9653052

Johnson, C. C., Kahle, J. B., & Fargo, J. D. (2007). A study of the effect of sustained, whole-school professional development on student achievement in science. *Journal of Research in Science Teaching*, *44*(6), 775–786. https://doi.org/10.1002/tea.20149

Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, *14*(1), 22–37. https://doi.org/10.1080/1464536X.2011.573008

King, A. R., Jr., & Brownell, J. A. (1966). *The Curriculum and the Disciplines of Knowledge: A Theory of Curriculum Practice*. John Wiley & Sons, Inc.

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, *17*(2), 79–103. https://doi.org/10.1002/(sici)1520-6793(200002)17:2<79::aid-mar2>3.0.co;2-0

Kini, T., & Podolsky, A. (2016). *Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research* (ED606426). ERIC. https://files.eric.ed.gov/fulltext/ED606426.pdf

Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, *37*(3), 248–255. https://doi.org/10.1177/00222194040370031001

Klingner, J. K., Vaughn, S., Tejero Hughes, M., & Arguelles, M. E. (1999). Sustaining research-based practices in reading: A 3-year follow-up. *Remedial and Special Education*, *20*(5), 263–287. https://doi.org/10.1177/074193259902000502

Knapp, N. F., & Peterson, P. L. (1995). Teachers' interpretations of "CGI" after four years: Meanings and practices. *Journal for Research in Mathematics Education*, *26*(1), 40–65. https://doi.org/10.5951/jresematheduc.26.1.0040

Knowles, M. S. (1978). *Andragogy: Adult Learning Theory in Perspective*. Community college review, *5*(3), 9–20.

Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*(1), 51–67. https://doi.org/10.1177/0022487114549599

Kolbe, T., & Jorgenson, S. (2018). Meeting instructional standards for middle-level science: Which teachers are most prepared?. *The Elementary School Journal*, *118*(4), 549–577. https://doi.org/10.1086/697540

Kowalski, S. M., Taylor, J. A., Askinas, K. M., Wang, Q., Zhang, Q., Maddix, W. P., & Tipton, E. (2020). Examining factors contributing to variation in effect size estimates of teacher outcomes from studies of science teacher professional development. *Journal of Research on Educational Effectiveness*, *13*(3), 430–458. https://doi.org/10.1080/19345747.2020.1726538

Kranjc Horvat, A., Wiener, J., Schmeling, S., & Borowski, A. (2021). Learning goals of professional development programs at science research institutions: A Delphi study with different stakeholder groups. *Journal of science teacher education*, 1–23. https://doi.org/10.1080/1046560X.2021.1905330

Kyriakides, L., Christoforidou, M., Panayiotou, A., & Creemers, B. P. M. (2017). The impact of a three-year teacher professional development course on quality of teaching: Strengths and limitations of the dynamic approach. *European Journal of Teacher Education*, *40*(4), 465–486. https://doi.org/10.1080/02619768.2017.1349093

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*(2), 139–162. https://doi.org/10.1007/s10972-014-9412-1

Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching*, *48*(5), 534–551. https://doi.org/10.1002/tea.20404

Lambert, J. (2001). *A Quantitative and Qualitative Analysis of the Impact of High School Marine Science Curricula and Instructional Strategies on Science Literacy of Students* (Publication No. 3016049) [Doctoral dissertation, Florida State University]. ProQuest Dissertations and Theses Global.

Lambert, J. (2006). High school marine science and scientific literacy: The promise of an integrated science course. *International Journal of Science Education*, *28*(6), 633–654. https://doi.org/10.1080/09500690500339795

Larkin, K., Wouters, N., Caetano, A.-T., & McDonough, N. (Eds.). (2015). *EurOcean 2014: Connecting Science, Policy and People*. Publications Office of the European Union: Luxembourg. https://op.europa.eu/en/publication-detail/-/publication/292840a4-73fa-11e5-86db-01aa75ed71a1

Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice*, *1*(3), 293–307. https://doi.org/10.1080/713693162

Lawton, B. E. (2005, April 11-15). *The Differential Effects of Two Versions of Middle-school Inquiry-based Science Program Professional Development Institutes on Teachers' Self-Efficacy as Inquiry-based Science Teachers* [Paper presentation]. American Educational Research Association Annual Conference, Montreal, Quebec, Canada.

Lawton, B. E. (2017). *Evidence about the convergent and discriminant validity of the components of fidelity of implementation in a science-teacher professional development projec*t [Doctoral Thesis, University of Hawaiʻi at Mānoa]. ScholarSpace at University of Hawaiʻi at Mānoa. https://scholarspace.manoa.hawaii.edu/bitstream/10125/62369/2017-12-phd-lawton.pdf

Lee, O., Hart, J. E., Cuevas, P., & Enders, C. (2004). Professional development in inquiry-based science for elementary teachers of diverse student groups. *Journal of Research in Science Teaching*, *41*(10), 1021–1043. https://doi.org/10.1002/tea.20037

Lee, O., Llosa, L., Jiang, F., Haas, A., O'Connor, C., & Van Booven, C. D. (2016). Elementary teachers' science knowledge and instructional practices: Impact of an intervention focused on English language learners. *Journal of Research in Science Teaching*, *53*(4), 579–597. https://doi.org/10.1002/tea.21314

Lee, O., Luykx, A., Buxton, C., & Shaver, A. (2007). The challenge of altering elementary school teachers' beliefs and practices regarding linguistic and cultural diversity in science instruction. *Journal of Research in Science Teaching*, *44*(9), 1269–1291. https://doi.org/10.1002/tea.20198

Linacre, J. M. (2010). Predicting responses from Rasch measures. *Journal of Applied Measurement*, *11*(1), 1. https://www.winsteps.com/a/Linacre-Predicting.pdf

Lin, A. (2020). *Making Sense of Teachers' Initial Steps with the NGSS* [Doctoral Thesis, University of Hawai'i at Mānoa]. ScholarSpace at University of Hawai'i at Mānoa. https://scholarspace.manoa.hawaii.edu/bitstream/10125/73332/Lin_hawii_0085A_10842.pdf

Linn, G. B., Gill, P., Sherman, R., Vaughn, V., & Mixon, J. (2010). Evaluating the long-term impact of professional development. *Professional Development in Education*, *36*(4), 679–682. https://doi.org/10.1080/19415250903109288

Liu, S., & Phelps, G. (2020). Does teacher learning last? Understanding how much teachers retain their knowledge after professional development. *Journal of Teacher Education*, *71*(5), 537–550. https://doi.org/10.1177/0022487119886290

Lotter, C., Harwood, W. S., & Bonner, J. J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, *44*(9), 1318–1347. https://doi.org/10.1002/tea.20191

Loucks-Horsley, S., Stiles, K., & Hewson, P. (1996). Principles of effective professional development for mathematics and science education: A synthesis of standards. *National Institute for Science Education Brief*, *1*(1), 1–6. https://files.eric.ed.gov/fulltext/ED409201.pdf

Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2010). *Designing Professional Development for Teachers of Science and Mathematics* (3rd ed.). Corwin press.

Lovell, S., Hamer, M., Slotow, R., & Herbert, D. (2009). An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation*, *18*(12), 3295–3307. https://doi.org/10.1007/s10531-009-9642-2

Luckie, D. B., Aubry, J. R., Marengo, B. J., Rivkin, A. M., Foos, L. A., & Maleszewski, J. J. (2012). Less teaching, more learning: 10-yr study supports increasing student learning through less coverage and more inquiry. *Advances in Physiology Education*, *36*(4), 325–335. https://doi.org/10.1152/advan.00017.2012

Luft, J. A., & Hewson, P. W. (2014). Research on teacher professional development programs in science. In S. K. Abell & N. Lederman (Eds.), *Handbook of Research on Science Education* (Vol. II, pp. 903–924). Routledge. https://doi.org/10.4324/9780203097267-56

Lydon, S., & King, C. (2009). Can a single, short continuing professional development workshop cause change in the classroom?. *Professional Development in Education*, *35*(1), 63–82. https://doi.org/10.1080/13674580802264746

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260–293. https://doi.org/10.3102 /0162373719849044

Maeng, J. L., Whitworth, B. A., Bell, R. L., & Sterling, D. R. (2020). The effect of professional development on elementary science teachers' understanding, confidence, and classroom implementation of reform-based science instruction. *Science Education*, *104*(2), 326–353. https://doi.org/10.1002/sce.21562

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education* (pp. 95–132). Springer, Dordrecht. https://doi.org/10.1007/0-306-47217-1_4

Marshall, J. C., Smart, J. B., & Alston, D. M. (2017). Inquiry-based instruction: A possible solution to improving student learning of both science concepts and scientific practices. *International Journal of Science and Mathematics Education*, *15*(5), 777–796. https://doi.org/10.1007/s10763-016-9718-x

McFadden, J. (2019). Transitions in the perpetual beta of the NGSS: One science teacher's beliefs and attempts for instructional change. *Journal of science Teacher Education*, *30*(3), 229–258. https://doi.org/10.1080/1046560X.2018.1559559

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*(2), 233–268. https://doi.org/10.1002/sce.20294

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons.

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*(4), 474–496. https://doi.org/10.1002/tea.20347

Minor, E. C., Desimone, L., Lee, J. C., & Hochberg, E. D. (2016). Insights on how to shape teacher learning policy: The role of teacher content knowledge in explaining differential effects of professional development. *Education Policy Analysis Archives*, *24*(61), 1–34. https://doi.org/10.14507/epaa.24.2365

Monroe, B. L., & Brunsell, E. (2019). A metacognitive Approach to the next generation science and engineering practices. *Science Scope*, *43*(1), 51–55. https://www.jstor.org/stable /26899050

Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art–teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, *25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Murray, D. W., Rabiner, D. L., Kuhn, L., Pan, Y., & Sabet, R. F. (2018). Investigating teacher and student effects of the Incredible Years Classroom Management Program in early elementary school. *Journal of School Psychology*, *67*, 119–133. https://doi.org/10.1016 /j.jsp.2017.10.004

Murrill, L. D., Thomas, T. G., & Reynolds, T. L. (2013). Place of learning, place of practice: Elements that affect the transfer of teachers' professional development to students' learning in classrooms. *The Teacher Educators' Journal*, *20*, 39–56.

National Research Council (NRC). (1996). *National Science Education Standards*. National Academies Press. https://www.nap.edu/4962

National Research Council (NRC). (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. National Academies Press. https://www.nap.edu/9596

National Research Council (NRC). (2001). Knowing what students know: The science and design of educational assessment. *National Academies Press*. https://www.nap.edu/10019

National Research Council (NRC). (2007). Taking science to school: Learning and teaching science in grades K-8. *National Academies Press*. https://www.nap.edu/11625

National Research Council (NRC). (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press. https://www.nap.edu/13165

Next Generation Science Standards (NGSS) Lead States. (2013). *Next Generation Science Standards: For States, by States*. National Academies Press. https://www.nap.edu/18290

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. *Educational Evaluation and Policy Analysis*, *26*(3), 237–257. https://doi.org/10.3102 /01623737026003237

Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance*, *28*(1), 51–74. https://www.jstor.org/stable/40704157

Onwuegbuzie, A. J. (2002). *A Conceptual Framework for Assessing Legitimation in Qualitative Research* (ED471659). ERIC. https://files.eric.ed.gov/fulltext/ED471659.pdf

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron?. *Quality & Quantity*, *41*(2), 233–249. https://doi.org/10.1007/s11135-006-9000-3

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*(3), 376–407. https://doi.org/10.3102/0034654311413609

Padua, J. F. M. (2018). *The Journey After: A Phenomenological Examination of Teachers' Transfer of Learning From a Two-year Professional Development Program* (Publication

No. 10992863) [Doctoral dissertation, University of Hawaiʻi at Mānoa]. ProQuest Dissertations and Theses Global.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* Sage publications.

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). *Effective Programs in Elementary Mathematics: A Meta-Analysis. AERA open*, *7*(1), 1–29. https://doi.org/10.1177/2332858420986211

Penner-Williams, J., Diaz, E. I., & Worthen, D. G. (2019). Sustainability of teacher growth from professional development in culturally and linguistically responsive instructional practices. *Teaching and Teacher Education*, *86*, 102891. https://doi.org/10.1016/j.tate.2019.102891

Penuel, W. R., Bell, P., & Neill, T. (2020). Creating a system of professional learning that meets teachers' needs. *Phi Delta Kappan*, *101*(8), 37–41. https://doi.org/10.1177/0031721720923520

Penuel, W. R., & Fishman, B. J. (2012). Large-scale science education intervention research we can use. *Journal of Research in Science Teaching*, *49*(3), 281–304. https://doi.org/10.1002/tea.21001

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation.

American Educational Research Journal, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221

Plankis, B., & Marrero, M. E. (2010). Recent ocean literacy research in United States public schools: Results and implications. *International Electronic Journal of Environmental Education*, *1*(1). https://dergipark.org.tr/en/pub/iejeegreen/issue/7903/104036

Podolsky, A., Kini, T., Darling-Hammond, L., & Bishop, J. (2019). Strategies for attracting and retaining educators: What does the evidence say?. *Education Policy Analysis Archives*, *27*(38). http://dx.doi.org/10.14507/epaa.27.3722

Pomerantz, F., & Condie, C. (2017). Building bridges from pre-service experiences to elementary classroom literacy teaching: Challenges and opportunities. *Teaching and Teacher Education*, *64*, 211–221. https://doi.org/10.1016/j.tate.2017.02.011

Porter, A. C., Garet, M. S., Desimone, L., Yoon, K. S., & Birman, B. F. (2000). *Does Professional Development Change Teaching Practice? Results From a Three-Year Study* (ED455227). ERIC. https://files.eric.ed.gov/fulltext/ED455227.pdf

Pottenger, F. M. (2007, January). *Inquiry and Disciplinary Natural Science Teaching* [Paper presentation]. Hawai‘i International Conference on Education Annual Conference, Honolulu, Hawai‘i, United States.

Pottenger, F. M., Baumgartner, E., & Brennan, C. A. (2007, September). *Teaching Science as Inquiry (TSI): Teacher Professional Development* [Paper presentation]. Hawai‘i Charter School Administrative Office Breaking Barriers to Learning Professional Development Conference, Honolulu, Hawai‘i, United States.

Pottenger, F. M., & Berg, K. (2006, June). *Inservice Training in Science Inquiry* [Paper presentation]. Pacific Circle Consortium Conference Annual Conference, Mexico City, Mexico.

Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation*, *21*(3), 341–349. https://doi.org/10.1177/109821400002100305

Prawat, R. S. (1992). Teachers' beliefs about teaching and learning: A constructivist perspective. American journal of education, *100*(3), 354–395. https://doi.org/10.1086/444021

Protheroe, N. (2009). Fidelity of implementation. *Principal's Research Review*, *4*(6), 1-7.

Pulliam, J. D., & Van Patten, J. J. (2007). *History of Education in America* (9th ed.). Pearson.

Rankin, L. (2000). Lessons learned: Addressing common misconceptions about inquiry. In *Foundations: A Monograph for Professionals in Science, Mathematics, and Technology Education: Vol. 2. Inquiry: Thoughts, Views, and Strategies for the K–5 Classroom* (pp. 33–37). Division of Elementary, Secondary, and Informal Education; National Science Foundation. https://www.nsf.gov/pubs/2000/nsf99148/pdf/nsf99148.pdf

Reiser, B. J. (2013, September). *What Professional Development Strategies are Needed for Successful Implementation of the Next Generation Science Standards*. Invitational Research Symposium on Science Assessment; The Center for K–12 Assessment & Performance Management at Educational Testing Service. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.1266&rep=rep1&type=pdf

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246. https://doi.org/10.3102/00346543075002211

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Richardson, V. (2003). Constructivist pedagogy. *Teachers College Record*, *105*(9), 1623-1640. https://www.sfu.ca/~jcnesbit/EDUC220/ThinkPaper/Richardson2003.pdf

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, *28*(3), 559–586. https://doi.org/10.3102/00028312028003559

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Rose, D. J., & Church, R. J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education*, *8*(1), 5–35. https://doi.org/10.1023/A:1022860606825

Ross, J. A. (1994). The impact of an inservice to promote cooperative learning on the stability of teacher efficacy. *Teaching and Teacher Education*, *10*(4), 381–394. https://doi.org/10.1016/0742-051X(94)90020-5

Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *The Journal of Educational Research*, *101*(1), 50–60. https://doi.org/10.3200/joer.101.1.50-60

Rowan, B., Schilling, S. G., Ball, D. L., Miller, R., Atkins-Burnett, S., Camburn, E., Harrison, D., & Phelps, G. (2001, October). *Measuring Teachers' Pedagogical Content Knowledge* *in Surveys: An Exploratory Study.* https://citeseerx.ist.psu.edu/viewdoc/download?doi =10.1.1.591.1047&rep=rep1&type=pdf

Rozenszajn, R., & Yarden, A. (2014). Expansion of biology teachers' pedagogical content knowledge (PCK) during a long-term professional development program. *Research in Science Education*, *44*(1), 189–213. https://doi.org/10.1007/s11165-013-9378-6

Rudolph, J. L. (2014). Dewey's "science as method" a century later: Reviving science education for civic ends. *American Educational Research Journal*, *51*(6), 1056–1083. https://doi.org/10.3102/0002831214554277

St. Clair, T. L., Wheeler, L. B., Maeng, J. L., & Bell, R. L. (2020). Mixed-methods analysis of science teacher educator professional development: a five-year longitudinal study. *Professional Development in Education*, 1–24. https://doi.org/10.1080 /19415257.2020.1787191

Sancar, R., Atal, D., & Deryakulu, D. (2021). A new framework for teachers' professional development. *Teaching and Teacher Education*, *101*, 103305. https://doi.org/10.1016 /j.tate.2021.103305

Sandholtz, J. H., & Ringstaff, C. (2016). The influence of contextual factors on the sustainability of professional development outcomes. *Journal of Science Teacher Education*, *27*(2), 205–226. https://doi.org/10.1007/s10972-016-9451-x

Sandholtz, J. H., & Ringstaff, C. (2020). Offering modest supports to extend professional development outcomes and enhance elementary science teaching. *Professional Development in Education*, 1–16. https://doi.org/10.1080/19415257.2020.1725594

Sandholtz, J. H., Ringstaff, C., & Matlen, B. (2016). Temporary fix or lasting solution? Investigating the longitudinal impact of teacher professional development on K–2 science instruction. *The Elementary School Journal*, *117*(2), 192–215. https://doi.org/10.1086/688920

Scarlett, T. (2008). *An Exploratory Study of the Impact of Two Versions of Inquiry-Based Science Program Professional Development on Teachers' Perceptions of their Pedagogical Content Knowledge* (Publication No. 304603723) [Doctoral dissertation, University of Hawaiʻi at Mānoa]. ProQuest Dissertations and Theses Global.

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know?. *Journal of Research on Educational Effectiveness*, *2*(3), 209–249. https://doi.org/10.1080/19345740802641527

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1), 111–139. https://doi.org/10.1007/s11165-005-3917-8

Schunk, D. H. (2012). *Learning Theories: An Educational Perspective* (6th. ed.). Pearson.

Schuster, D., Cobern, W., Applgate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2006, October 19-21). 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). Drury University. http://web.archive.org/web/20170406030708/http://openwatermedia.com/downloads/STEM(for-posting).pdf

Settlage, J. (2007). Demythologizing science teacher education: Conquering the false ideal of open inquiry. *Journal of Science Teacher Education*, *18*(4), 461–467. https://doi.org/10.1007/s10972-007-9060-9

Settlage, J., Southerland, S. A., Smith, L. K., & Ceglie, R. (2009). Constructing a doubt-free teaching self: Self-efficacy, teacher identity, and science instruction within diverse settings. *Journal of Research in Science Teaching*, *46*(1), 102–125.

Seung, E., Park, S., & Jung, 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*(4), 507–529. https://doi.org/10.1007/s11165-013-9390-x

Sewell, A. (2002). Constructivism and student misconceptions: why every teacher needs to know about them. *Australian Science Teachers Journal*, *48*(4), 24–28. https://www.proquest.com/docview/194480833

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs*: For generalized causal inference. Houghton Mifflin.

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*(1), 29–34. https://doi.org/10.19030/jier.v11i1.9096

Shaharabani, Y. F., & Tal, T. (2017). Teachers' practice a decade after an extensive professional development program in science education. *Research in Science Education*, *47*(5), 1031–1053. https://doi.org/10.1007/s11165-016-9539-5

Shahzad, K., & Naureen, S. (2017). Impact of teacher self-efficacy on secondary school students' academic achievement. *Journal of Education and Educational Development*, *4*(1), 48–72. http://dx.doi.org/10.22555/joeed.v4i1.1050

Shaw, S., Moulding, B., Mackey, C., Templeton, S., Penuel, W., Van Horne, K. (2018). *Science Professional Learning Standards Rubric 2018*. Council of State Science Supervisors. https://indd.adobe.com/view/b5e3a2fd-72d2-4f06-a7a3-28ee2f95fb1d

Shear, L., & Penuel, W. R. (2010). Rock-solid support. *The Learning Professional*, *31*(5), 48. https://www.proquest.com/docview/928957666/fulltext/8A81D61993204FB2PQ/

Short J. & Hirsh, S. (2020). *The Elements: Transforming Teaching Through Curriculum-Based Professional Learning*. Carnegie Corporation of New York. https://www.carnegie.org /topics/topic-articles/professional-learning-educators/elements-transforming-teaching-through-curriculum-based-professional-learning/

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10 .1.1.923.4203&rep=rep1&type=pdf

Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: a critical review. *School Effectiveness and School Improvement*, *32*(1), 47–63. https://doi.org/10.1080/09243453.2020.1772841

Slavin, R. E. (2018). *Educational Psychology: Theory and Practice* (12th ed.). Pearson.

Slavin, R. E. (2019, December 19). *Classroom Management: What Works in Professional Development*. Robert Slavin's Blog. https://robertslavinsblog.wordpress .com/tag/classroom-management/

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, *51*(7), 870–901. https://doi.org/10.1002/tea.21139

Smith, P. S. (2020). What does a national survey tell us about progress toward the vision of the NGSS?. *Journal of Science Teacher Education*, *31*(6), 601–609. https://doi.org/10.1080/1046560X.2020.1786261

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*(3), 169–199. https://doi.org/10.3102/0162373707306025

Smithson, J. L., & Porter, A. C. (1994, October). *Measuring Classroom Practice: Lessons Learned from the Efforts to Describe the Enacted Curriculum—The Reform Up Close Study* (CPRE Report #31). Consortium for Policy Research in Education. https://curriculumanalysis.org/Reference/MCPcpreReport.pdf

Southerland, S. A., Granger, E. M., Hughes, R., Enderle, P., Ke, F., Roseler, K., Saka, Y., & Tekkumru-Kisa, M. (2016). *Essential Aspects of Science Teacher Professional Development: Making Research Participation Instructionally Effective*. AERA open, *2*(4), 1–16. https://doi.org/10.1177/2332858416674200

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431. https://doi.org/10.3102/00346543072003387

Sprague, J. (2005). *Feminist Methodologies for Critical Researchers: Bridging Differences*. Rowman & Littlefield.

Starkey, L., Yates, A., Meyer, L. H., Hall, C., Taylor, M., Stevens, S., & Toia, R. (2009). Professional development design: Embedding educational reform in New Zealand.

Teaching and Teacher Education, *25*(1), 181–189. https://doi.org/10.1016/j.tate.2008.08.007

Steel, B. S. (2006). Ocean and coastal literacy in the United States. *Sea Technology*, *47*(7), 45–48.

Steel, B. S., Smith, C., Opsommer, L., Curiel, S., & Warner-Steel, R. (2005). Public ocean literacy in the United States. *Ocean & Coastal Management*, *48*(2), 97–114. https://doi.org/10.1016/j.ocecoaman.2005.01.002

Strang, C. (2008). Education for ocean literacy and sustainability: Learning from elders, listening to youth. *Current: The Journal of Marine Education*, *24*(3), 6–10. http://marestage.lawrencehallofscience.org/sites/mare.lawrencehallofscience.org/files/images/mare_education_for_ocean_literacy_and_sustainability.pdf

Strang, C., DeCharon, A., & Schoedinger, S. (2007). Can you be science literate without being ocean literate?. *Current: The Journal of Marine Education*, *23*(1), 7–9. https://www.researchgate.net/profile/S-Schoedinger/publication/237286927_Can_you_Be_Science_Literate_Without_Being_Ocean_Literate/links/55342b4f0cf27acb0def9056/Can-you-Be-Science-Literate-Without-Being-Ocean-Literate.pdf

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy*, *14*(3), 331–356. https://doi.org/10.1177/0895904800014003001

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*(9), 963–980. https://doi.org/10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0

Sutton, M., & Levinson, B. A. U. (Eds.). (2001). *Policy as Practice: Toward a Comparative Sociocultural Analysis of Educational Policy*. Ablex Publishing.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration (BES)*. New Zealand Ministry of Education. https://www.educationcounts.govt.nz/__data/assets/pdf_file/0017/16901/TPLandDBESentireWeb.pdf

Van Eekelen, I. M., Vermunt, J. D., & Boshuizen, H. P. A. (2006). Exploring teachers' will to learn. *Teaching and Teacher Education*, *22*(4), 408–423. https://doi.org/10.1016/j.tate.2005.12.001

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

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*(1), 63–89. https://doi.org/10.1007/s10972-006-9031-6

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional Development in the United States: Trends and Challenges*. National Staff Development Council. https://edpolicy.stanford.edu/sites/default/files/publications/professional-development-united-states-trends-and-challenges.pdf

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). *Professional Learning in the Learning Profession: A StatusRreport on Teacher Development in the United States and Abroad*. National Staff Development Council;

School Redesign Network at Stanford University. https://edpolicy.stanford.edu/sites/default/files/publications/professional-learning-learning-profession-status-report-teacher-development-us-and-abroad.pdf

What Works Clearinghouse. (2020). *What Works Clearinghouse Standards Handbook: Version 4.1*. Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Whitehead, D. (2010). The year after: Sustaining the effects of literacy professional development in New Zealand secondary schools. *Language and Education*, *24*(2), 133–149. https://doi.org/10.1080/09500780903350273

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*(3), 276–301. https://doi.org/10.1002/tea.20329

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, *72*(2), 131–175. https://doi.org/10.3102/00346543072002131

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903. https://doi.org/10.1002/sce.21027

Wolf, R. (2021). *Average Differences in Effect Sizes by Outcome Measure Type* (ED610568). ERIC. https://files.eric.ed.gov/fulltext/ED610568.pdf

Wolf, S., & Peele, M. E. (2019). Examining sustained impacts of two teacher professional development programs on professional well-being and classroom practices. *Teaching and Teacher Education*, *86*, 102873. https://doi.org/10.1016/j.tate.2019.07.003

Wulff, A., & Johannesson, K. (2019). Bring the ocean to the classroom—Introducing experimental studies to teachers with fair or no science knowledge. In G. Fauville, D. L.Payne, M. E. Marrero, A. Lantz-Andersson, & F. Crouch (Eds.), *Exemplary Practices in Marine Science Education* (pp. 363–376). Springer. https://doi.org/10.1007/978-3-319-90778-9_20

Yang, R., Porter, A. C., Massey, C. M., Merlino, J. F., & Desimone, L. M. (2020a). Curriculum-based teacher professional development in middle school science: A comparison of training focused on cognitive science principles versus content knowledge. *Journal of Research in Science Teaching*, *57*(4), 536–566. https://doi.org/10.1002/tea.21605

Yang, Y., Liu, X., & Gardella Jr, J. A. (2018). Effects of professional development on teacher pedagogical content knowledge, inquiry teaching practices, and student understanding of interdisciplinary science. *Journal of Science Teacher Education*, *29*(4), 263–282. https://doi.org/10.1080/1046560X.2018.1439262

Yang, Y., Liu, X., & Gardella Jr, J. A. (2020b). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, *57*(7), 1028–1057. https://doi.org/10.1002/tea.21620

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the Evidence on how Teacher Professional Development Affects Student Achievement*. Institute of Education Sciences; Regional Educational Laboratory: Southwest; U. S.

Department of Education; National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=70

Young, D. B. (n.d.). *Impact on Education: Essential Elements of Successful Change Implementation*. Curriculum Research & Development Group, University of Hawai‘i at Mānoa.

Zeichner, K. M. & Liston, D. P. (2014). *Reflective Teaching: An Introduction* (2nd ed.). Routledge.

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, *32*(4), 437–463. https://doi.org/10.1023/A:1022411822951

Zimmerman, B., & Schunk, D. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 349–367). Lawrence Erlbaum.