

A Realist Evaluation Approach to the Study of a K-12 Online Professional Development Course

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Stephanie Jean Teague
August 2019

Copyright © 2019 by Stephanie Jean Teague
All rights reserved.

ACKNOWLEDGEMENTS

First and foremost, I give thanks to God, who continually sustains me and gives me undeserved grace. His plans for me always exceed any I could conceive on my own. I am so grateful that He has taken me from a college freshman, eager to escape formal education, to a doctoral candidate with two master's degrees in education. I hope to continue seeking and glorifying Him as I look towards new adventures.

I also extend my wholehearted thanks to those who have made my time at the University of Tennessee so enjoyable, challenging, and rewarding—

To my dissertation advisor, Dr. Lisa Yamagata-Lynch. Through your guidance, knowledge, and genuine care for your students as individuals, you exemplify the qualities of leadership to which I aspire. Thank you for helping me navigate my time here at UT, for encouraging me to pursue research that I am passionate about, and for being an example of a servant-leader. Also, thank you for your infectious laugh and delightful animal stories!

To my dissertation committee members, Dr. Barbara Thayer-Bacon, Dr. Gary Skolits and Dr. Louis Rocconi. Thank you so much for lending me your expertise and providing invaluable feedback on my work. Special thanks to Dr. Barb for being such a caring, inspiring educator and mentor throughout my time here at UT. You truly made me feel at home here!

To my friend and classmate Erin Garty. From our first seminar class, you have taken me under your wing and become a true friend. Your thirst for knowledge, your work ethic, and your integrity have helped me to grow as a scholar. More importantly, your selflessness, encouragement, and heart for others (including our little furry and feathery friends) have helped me to grow as a person. I have loved the many talks and laughs we have shared together and look forward to many more!

To my fiancé, Jimmie Hostetter. Your abiding love and grace for me exemplify the spirit of Christ in my life and draw me closer to Him. I cannot thank you enough for the constant encouragement, support, and patience that you have provided throughout this process. Your cheerful spirit and servant's heart motivate me to love God, you, and others better. I can't wait to spend the rest of our lives serving, laughing, and growing together. I love you :)

To my parents, Richard and Cindy Teague. I will never be able to adequately articulate the gratitude I have for the countless sacrifices you have made, the fathomless love and grace you have shown me, and how you always put my, Christianne's, and David's needs before your own. You have shown us what it means to live as godly parents and I can only hope to follow your example. Thank you for supporting my pursuits in life, giving wise guidance and counsel, and being my best friends. I will forever enjoy our family adventures, inside jokes, and nerdy conversations. I love you and I am proud and honored to call you my parents.

To my sister and brother, Christianne and David Teague. Christianne, your humor always brings me joy and your sensitive spirit lets me know you are always there for me. I have enjoyed going through the graduate experience together (especially swapping bird memes) and am so proud of your accomplishments! David, thank you for asking thought-provoking questions that helped me strengthen the presentation of my ideas. You are going to be a transformative teacher. I love both of you!

And last, but never least, to my steadfast feathered companion and writing partner, Skittles. You have (literally) stayed by my side throughout this entire process, finding new ways to make me laugh and ensuring that I am never bored or lonely. You brighten each day and continue to show new facets of your fun-loving personality. I love you!

ABSTRACT

The primary purpose of this study is to determine the extent to which the desired outcomes for select K-12 teachers' participation in a specific arts-based online professional development (OPD) course were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. A secondary purpose of the study is to set a precedent for future studies of K-12 OPD courses, particularly those using a realist evaluation (RE) approach.

The RE approach develops, tests, and refines a set of program theories, or hypotheses, about what outcomes will be achieved, for whom, and in what circumstances (Pawson & Tilley, 1997).

In accordance with this approach, I began by developing initial program theories that hypothesized the extent to which the desired course outcomes were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. I then used multiple forms of data collection and analysis to test and refine these program theories during three iterative phases: Develop Initial Program Theories, Test and Revise Program Theories, and Refine and Finalize Program Theories. As RE approaches are retroductive, I analyzed the data as it was being collected. I gathered data from course completion rates, a realist review of the literature, interviews, course annotations, and responses to question and reflection prompts embedded in the course. Collectively, these data were used to determine the extent to which desired course outcomes were being achieved, for which teachers, in what types of contexts, and how. Ultimately, six program theories were finalized, from which five middle-range theories were derived. These findings identified relationships between the contextual factors, causal mechanisms, and outcomes from the *VTSB* course. They indicated that teachers engaged in, applied, and shared course content when the content aligned with their pedagogical beliefs, when they experienced pedagogical discontentment, and when they had strong interpersonal

relationships with school leadership. Findings also indicated that teachers' engagement in the course was constrained by mandated curricular requirements, time demands, an over-focus on accountability measures, and when course examples were not representative of their diverse students and instructional contexts.

Keywords: Online professional development, K-12 professional development, realist evaluation, program evaluation

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND GENERAL INFORMATION	1
Introduction.....	1
Background of Study	4
K-12 OPD Evaluative Research Approaches	13
Overview of Realist Evaluation	15
Purpose of Study	23
Research Questions.....	23
Significance of the Study	25
Assumptions and Limitations	26
Definitions of Terms	27
Organization of the Study	31
CHAPTER 2: LITERATURE REVIEW	32
Introduction.....	32
Methodology	32
Conceptual Analysis and Discussion	37
Conclusion	52
Realist Evaluation	53
Historical, Philosophical, and Theoretical Foundations	54
Strengths and Challenges of Realist Evaluation	70
Conclusion	88
CHAPTER 3: MATERIALS AND METHODS	89
Research Questions.....	89
RAMESES II Quality Standards for Realist Evaluation	90
Design of the Study.....	90
Institutional Review Board Compliance	92
Participants.....	92
Data Collection & Analysis	93
Phase I: Develop Initial Program Theories.....	94
Phase II: Test & Revise Program Theories.....	106
Phase III: Refine and Finalize Program Theories	115
Saturation	119
Trustworthiness.....	120
CHAPTER 4: RESULTS AND DISCUSSION	122
Phase One Overview	123
Review of the Literature	123
Watershed Communications and Exploration of VTSB Course Platform and Resources	132
Generating Initial Program Theories	133
Phase Two Overview	139
Review of the Literature	139
Interviews and Course Annotations	146

Revised Program Theories.....	179
Phase Three Overview	179
Review of the Literature	179
Finalized Program Theories.....	186
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS	189
Summary	189
Major Findings and Discussion	191
Middle-Range Theories (MRTs) and Implications.....	203
Future Research	210
Concluding Remarks.....	212
LIST OF REFERENCES.....	214
APPENDICES.....	237
Appendix A.....	238
Appendix B.....	241
Appendix C.....	243
Appendix D.....	244
Appendix E.....	249
Appendix F.....	251
Appendix G.....	253
Appendix H.....	254
Appendix I.....	257
Appendix J.....	259
Appendix K.....	261
Appendix L.....	263
Appendix M.....	267
Appendix N.....	269
Appendix O.....	270
Appendix P.....	271
VITA	273

LIST OF TABLES

Table 1. STARLITE Literature Search Components	35
Table 2. Definitions of OPD Evaluation Terminology	41
Table 3. STARLITE Literature Search Overview	55
Table 4. Timeline of TBE Theorists and Their Contributions	59
Table 5. Six Stages of a Realist Search	102
Table 6. Codebook Codes.....	116
Table 7. Contextual Factors That Affect PD Engagement	131
Table 8. Alignment between Contextual Factors That Affect PD Engagement and VTSB Course Content and Objectives	134
Table 9. Initial CMOCs and Program Theories.....	137
Table 10. DSMRI Components, Descriptions, and Elements	144
Table 11. Participants' Backgrounds and Level of Study Involvement	147
Table 12. Revised CMOCs and Program Theories	180
Table 13. Five Facets of Trust.....	183
Table 14. Finalized CMOCs and Program Theories	187

LIST OF FIGURES

Figure 1. Pawson & Tilley's Realist Evaluation Model.....	16
Figure 2. CMOC Example.....	20
Figure 3. Retroductive process	22
Figure 4. Digital literature review process	34
Figure 5. OPD evaluation terminology concept map.	40
Figure 6. Levels of Kirkpatrick's and Guskey's models in comparison to steps in OPD course development and implementation process	43
Figure 7. Black Box Evaluation Model.....	57
Figure 8. Chen's Black Box Solution: Process Evaluation Model.....	60
Figure 9. Weiss's Black Box Solution: Process Evaluation Model	62
Figure 10. Pawson & Tilley's Black Box Solution: Process Evaluation Model	64
Figure 11. Theoretical foundations typology	66
Figure 12. RAMESES II Quality Standards for Realist Evaluation.....	91
Figure 13. Three Phases of Study	95
Figure 14. Stratified social context.....	97
Figure 15. The DSMRI.....	142

ABBREVIATIONS

BCSD	Big City School District (pseudonym)
(C)	Content
CMOC	Context Mechanism Outcome Configuration
DSMRI	Dynamic Systems Role Model Initiative
ELL	English Language Learner
K-12	Kindergarten through twelfth grade
(M)	Mechanism
MRT	Middle-Range Theory
(O)	Outcome
OPD	Online Professional Development
PD	Professional Development
PS	Public School
PT	Program Theory
RAMESES	Realist and Meta-narrative Evidence Syntheses: Evolving Standards
RE	Realist Evaluation
SPED	Special Education
STEM	Science, Technology, Engineering, and Math
TBE	Theory-Based Evaluation
VICTORE	Volitions, Implementation, Contexts, Time, Outcomes, Rivalry, Emergence
VTs	Visual Thinking Strategies
<i>VTsB</i>	Visual Thinking Strategies Basics

CHAPTER 1: INTRODUCTION AND GENERAL INFORMATION

Introduction

In the thirty-five years since the Reagan Era report *A Nation at Risk: The Imperative for Educational Reform*, the United States has heightened its focus on accountability measures for K-12 public educators and students (Schroeder, Currin, & McCardle, 2016; Kim, 2017). Many schools and researchers assume a causal connection between teachers' participation in PD, their behavior and skill level, and student achievement (Muijs, Day, Harris, & Lindsay, 2004). Thus, school administrators perceive PD as a means to achieving higher scores on teacher evaluations and high-stakes tests and have increased PD requirements for teachers in K-12 schools (Owston, Wideman, Murphy, & Lupshenyuk, 2008). Consequently, this has increased the importance and relevance of K-12 teacher professional development (PD) and, by extension, the evaluation of teacher PD. As a result, schools are conducting more evaluations of PD in order to demonstrate its effects on teacher and student performance and justify its continued inclusion in school budgets (Hahs-Vaughn, Zygouris-Coe, & Fielder, 2007); however, according to the National Research Council of the US, the evaluation of K-12 PD remains an underexplored educational research endeavor (2010).

In particular, online forms of professional development (OPD) have become more prevalent during this time period, as, among other advantages, they offer widespread access, flexibility, and are cost-effective (Muijs & Lindsay, 2008; Rice & Hung, 2015). As a result, schools need to conduct evaluations of these OPD programs to gauge the extent to which they are achieving their desired outcomes. However, the evaluation of OPD is uniquely challenging. There are few evaluation methods and tools designed specifically for OPD and those "that have been used to evaluate face-to-face classes are not easily adapted to online environments" (Hahs-

Vaughn et al., 2007, p. 7). Consequently, very few evaluative studies of K-12 OPD programs have been conducted (Snyder, 2009; Muijs & Lindsay, 2008). An extensive search of the literature for K-12 OPD evaluations yielded studies that were either non-content specific or conducted in STEM fields; only one study was affiliated with a non-STEM field (Snyder, 2009). Thus, there is a need for the evaluation of K-12 OPD to be expanded and refined. In particular, there is a need for more evaluative studies in non-STEM subjects. This study uses a realist evaluation (RE) approach to help address these needs.

In this study, I examined an arts-based OPD course from Watershed Collaborative, a nonprofit organization that trains and equips K-12 teachers to cultivate student-centered classroom environments. The course, *VTSB*, is designed to help K-12 teachers learn how to facilitate inquiry-driven discussions and develop students' critical thinking skills by incorporating Visual Thinking Strategies (VTS) into their instruction (Watershed Collaborative, 2017). I sought to determine not only the extent to which participants were engaging in the course and applying course content, but also *how* and *why*, including why there were different outcomes for different participants (e.g., teachers who experienced pedagogical discontentment in different areas applied course content in different ways).

As part of my study, I extensively researched and compared several leading evaluation approaches. Based on my research questions and my ontological and epistemological positions, I concluded that the RE approach was most appropriate for this study. RE is founded on realist philosophy, which contends that a reality exists external to the researcher's perception of that reality (Van Belle et al., 2016). Thus, realist evaluators use multiple forms of data to develop, test, and refine theories hypothesizing how and why interventions have different outcomes for different individuals in different contexts (Pawson & Tilley, 1997). By conducting multiple

iterations of this empirical process, realist researchers endeavor to construct explanations of the relationships between contextual factors, causal mechanisms, and outcomes that are as close to reality as possible. Similarly, I used a RE approach in order to construct explanations for my research questions (which desired course outcomes were being achieved, for which teachers/subgroups of teachers, in what types of contexts, and how) that are as close to reality as possible.

This study is valuable for several entities, including Watershed Collaborative, future participants of the *VTSB* course, K-12 OPD course designers, researchers and evaluators of K-12 OPD, and researchers interested in exploring or adopting a RE approach. In particular, it sets a precedent for future studies of K-12 OPD courses, especially those using a RE approach. The final program theories serve as valuable contributions to a growing body of evidence that can inform future K-12 OPD REs (Timmins & Miller, 2007). At their roots, many program interventions consist of similar underlying mechanisms implemented in different contexts (Pawson & Tilley, 1997). Thus, elements of the final program theories in this study, in the form of middle-range theories, can be referenced by OPD courses operating in similar contexts or with similar intervention strategies (Pawson & Tilley, 1997; King et al., 2016; Stern, et al., 2012). Additionally, given the limited number of RE studies outside of the UK, this study also serves as an exemplar for other researchers interested in adopting a RE approach or comparing it to other evaluative research approaches.

In this chapter, I begin by providing background information relevant to my research topic and methodology, including Watershed Evaluation's OPD *VTSB* course, the development and evaluation of OPD, and realist evaluation. I then describe the purpose of my study, my research questions, and how this study will contribute to the development and evaluation of

future K-12 OPD courses. Next, I discuss the assumptions limitations and of this study. Finally, I provide definitions of terminology and an overview of this study's organization.

Background of Study

This section provides context for my study. I first review the background and organization of Watershed Collaborative's *VTSB* OPD course for K-12 educators, including a synopsis of VTS. Next, I review the definition of OPD and its growth over the past several decades, followed by a discussion on the research regarding effective OPD. Finally, I discuss the evolution of K-12 OPD evaluation and provide an overview of realist evaluation and its key concepts.

Watershed Collaborative's *VTB Basics* OPD Course

Watershed Collaborative is a nonprofit organization with a mission to "better equip classroom teachers for the complex task of enabling students to learn how to learn for success in school, work, and life in the 21st century" (Watershed Collaborative, 2017). They have identified several gaps in the current K-12 public education system that are not being adequately addressed:

- Engaging, efficient, and affordable professional development opportunities for teachers,
- Promoting collaborative professional development opportunities,
- Training teachers to be reflective practitioners,
- Promotion of student-centered pedagogy,
- Promotion of inquiry-driven classroom discussions, and
- Development of visual literacy skills and critical thinking skills.

To address these issues, Watershed offers services that train and equip K-12 teachers to cultivate student-centered classroom environments and to facilitate inquiry-driven discussions, including:

- Individual, school, and district-level packages for the *VTB Basics* course,
- Online and on-site coaching for individuals and teams,
- Option for organizations to customize the course for their instructional context and the specific needs of their learners, and
- Preparation to launch a Global Online Professional Learning Community to facilitate collaboration amongst VTS practitioners worldwide.

The *VTB* course is the main focus of this study. The course is designed as a self-paced, minimum ten-week long OPD course for K-12 teachers, schools, and school districts that seek to integrate student-centered, inquiry-based approaches into their curricula using methods that align with state, district, and school-specific curricular standards. Participating teachers learn how to implement VTS, facilitate learning, and engage in reflective practice through online instruction and job-embedded application. The course also facilitates opportunities for peer collaboration and provides access to a course facilitator and comprehensive course materials.

Visual Thinking Strategies (VTS). Visual literacy is defined as a “set of abilities that enables an individual to effectively find, interpret, evaluate, use, and create images and visual media” (ACRL, 2011). Consequently, visual literacy cultivates the development of critical and creative thinking skills (Baylen & Alba, 2015; Moeller et al, 2013). However, in K-12 education, visual literacy is not emphasized to the same extent as other forms of literacy, especially text-based literacy (Cappello & Walker, 2016). Despite this hyper-focus on textual literacy, K-12 students still perform poorly in tasks involving reading and writing (VTS, 2017).

The VTS curriculum was co-designed by Philip Yenawine, an experienced and renowned museum educator, and Abigail Housen, a cognitive psychologist, in order to improve students' textual literacy and critical thinking skills through the development of their visual literacy skills (Yenawine, 2013; VTS, 2017). Several longitudinal studies have demonstrated that implementing the VTS curriculum positively affects K-12 students' academic performance and development of critical thinking, communication, and social skills (Adams et al, 2006; DeSantis & Housen, 2007). In particular, the quality of students' writing increased after engaging in multiple VTS sessions (Cappello & Walker, 2016; Moeller et al, 2013; Grohe & Egan 2015).

As an instructional method, VTS consists of regular teacher-facilitated interactive sessions that foster engaging, open-ended discussions about select piece(s) of artwork. During VTS sessions, students observe and discuss specific pieces of artwork while the teacher poses a series of three simple, open-ended questions: What's going on in this picture? What do you see that makes you say that? What else can we find (Yenawine, 2013)? These questions encourage students to think critically by orally communicating their ideas and supporting them with evidence (O'Leary, 2010; Moeller et al, 2013). Thus, students make sense of these visual images based on their observations, analysis, and synthesis of information. Some K-12 teachers integrate VTS techniques into other subject areas. Instead of artwork, these teachers select poems, short stories, or word problems for their students to observe, analyze, and discuss collaboratively (Franco & Unrath, 2014; Yenawine, 2013).

Definition of Online Professional Development

Online professional development is a compound term consisting of two components: "online" and "professional development" (O-Brian, 2016, p. 40). Each component has a unique history that situates OPD in the appropriate context, which I address in this section.

Professional development. In the context of K-12 education, professional development is defined as “learning activities related to the profession of teaching that occur after initial certification (Fishman, 2016, p. 14). PD features several common characteristics, including: planned activities, exposing teachers to current pedagogical and subject-related research, an intention to advance teachers’ professional knowledge and skills, and an ultimate goal of bolstering student performance (Merchie, Tuytens, Devos, & Vanderlinde, 2016; Hahs-Vaughn et al., 2007). PD traditionally, and still commonly, occurs in face-to-face settings (Masters et al., 2010). However, online professional development, defined as “teacher learning experiences delivered partially or completely over the Internet” (Fishman, 2016, p. 427), has become increasingly popular with innovations in technology and the increased availability of the Internet (Holmes, Singer, & MacLeod, 2010). Nevertheless, whether face-to-face or online, the ultimate goal of PD is to improve both teacher performance and student achievement (Elliot, 2017).

Since the inception of the No Child Left Behind (NCLB) Act in 2001, there has been a demand for “highly qualified” or “high quality” teachers and “high quality” professional development. These official terms are defined by a series of standards listed in NCLB and the Every Student Succeeds Act (ESSA) (Dash, de Kramer, O’Dwyer, Masters, & Russell, 2012; Masters et al., 2010). Thus, the term “high quality” is not used arbitrarily in this study, but refers to the definition used in K-12 educational policy.

Online learning. Online learning provides a means for individuals to access instruction, even when separated geographically and temporally from an instructor(s). The history of online learning is marked by an evolving understanding of learning in online contexts as compared to learning in face-to-face contexts (National Research Council, 2007; Means, Toyama, Murphy, Bakia, & Jones, 2010). Joksimović et al. identify some of the benefits and disadvantages

associated with online learning in comparison to face-to-face settings. Benefits include "flexibility, alleviation of overcrowded classrooms, increased enrollment, reduced cost, and increased profit", while disadvantages include "the cost of training instructors, feelings of isolation, and technology gaps" (2015, p. 112).

Growth of OPD

Recent face-to-face and online PD programs are more learner-centered than traditional, face-to-face PD, which is known for its teacher-centered approach. In particular, traditional, face-to-face PD has been criticized for its limited accessibility, inflexibility, lack of extended collaborative opportunities, and lack of subject-specific topics (Owston et al., 2008). OPD addresses these criticisms and has become a prevalent form of PD as the result of several factors: The inception of new technologies, widespread theoretical shifts (Brown & Neal, 2013), a growing demand for "high quality" PD, and OPD's advantages over traditional PD (Surrette & Johnson, 2015).

Inception of new technologies. The technological evolution of OPD parallels that of distance education in general. OPD networks emerged in the 1980s, facilitating peer-to-peer learning and providing opportunities for educators to engage in PD courses in an online setting (Harasim, 2000). The public emergence of the World Wide Web in 1993 brought greater access to online and blended PD programs, which attracted teachers looking for an "anytime, anywhere" option" (Dash et al., 2012, p. 5) for PD. Teachers began enrolling in asynchronous online courses as part of OPD programs or as supplements to face-to-face PD sessions. OPD grew exponentially in subsequent decades, especially with the evolution of Web 2.0, which facilitates collaboration and the ability to create and share original content (Harasim, 2012).

The emergence of new technologies and high demand for flexible, widely accessible professional development have presented an increasing number of options for the design and delivery of OPD. In addition to self-paced and instructor-facilitated online courses, Brown & Neal emphasize the major roles of web conferences, webinars, and webcasts in twenty-first century OPD (2013). Web conferencing software affords online users the ability for synchronous communication that simulates the face-to-face interactions associated with traditional PD. Webinars (web-based seminars) also use conferencing software, but are less interactive than web conferencing. A webinar is “a [web-based] presentation, lecture, workshop, or seminar” (Brown & Neal, 2013, p. 186) on a particular topic, such as classroom management techniques or teaching fifth graders to find the area of an object. Lastly, webcasts use the Internet “to broadcast live or delayed audio and/or video transmissions” (Brown & Neal, 2013, p. 186).

Current OPD opportunities vary in content and structural elements, such as course duration, levels of interaction, and instructional approach (Roskos et al., 2007). For example, an OPD course can have single or multiple sessions and include synchronous or asynchronous activities; it can also be independent and self-paced, facilitated by an instructor, or centered around a community of practice (Bates et al., 2016; Elliot, 2017). Teachers can engage in synchronous activities, such as web conferencing, webinars, and online chat forums at the same time; alternatively, they can engage in asynchronous activities, such as discussion board postings, self-paced courses, at different times (Bates, Phalen, & Moran, 2016; Brown & Neal, 2013). Thus, single OPD sessions, such as webinars, webcasts, or stand-alone modules, are inherently either synchronous or asynchronous in nature, while OPD courses or programs have multiple sessions and can incorporate both types of activities.

Theoretical shifts. Behaviorism and constructivism constitute the primary learning theories that have been associated with OPD from its onset in the 1980s to the present, with popularity shifting from behaviorism to constructivism (Harasim, 2012). Behaviorism is predicated on the assumption that learning occurs only through observable changes in behavior. Thus, in a behaviorist model, students must “[demonstrate the] acquisition of knowledge or skills” (Harasim, 2012, p. 37) to achieve mastery of course topics. This model was prevalent in early forms of OPD, which were mostly self-directed and included lectures, text-based assignments, and traditional forms of assessment.

Alternatively, constructivism maintains that individuals create their own knowledge through interactions with their physical and social environments (Harasim, 2012). Many OPD programs are now designed using constructivist principles, which promotes the incorporation of collaboration, interactivity, and authentic, inquiry-based learning (Elliot, 2017; Singer, 2008). Thus, constructivist-based OPD courses and programs focus less on knowledge transmission and more on guided discovery.

Demand for “high quality” professional development. Many K-12 educational policy-makers and administrators maintain that there is a connection between PD, teacher quality, and student achievement; in this relationship, PD affects teacher quality (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), which affects student achievement (Masters, Kramer, O’Dwyer, Dash, & Russell, 2010). This theory is reflected in recent educational policies, including the 2001 NCLB Act and the 2015 ESSA, which emphasize the provision of “high-quality” PD as a means of producing high-quality teachers (Dash et al., 2012; Masters et al., 2010). Beginning with NCLB, “professional development has been adopted as a policy solution to improving the number of highly qualified teachers as well as helping all students to achieve high academic

standards” (Dash et al., 2012, p. 2). As a result, there has been an ongoing demand for high quality PD in K-12 schools. OPD has helped alleviate this demand by affording school districts and teachers the ability to access PD with widespread availability and temporal flexibility.

Advantages and challenges of online professional development. NCLB included a list of criteria for “high quality” PD that was used for evaluations of K-12 PD programs.

Traditional/face-to-face PD was found lacking in areas such as availability, differentiation, flexibility, cost-effectiveness, active and collaborative learning components, and subject-specific resources (National Research Council, 2007; Means, Toyama, Murphy, Bakia, & Jones, 2010). Teachers have also expressed their dissatisfaction with traditional modes of PD in their responses to multiple quantitative and qualitative research inquiries (Elliot, 2017). In contrast, OPD satisfies some of the needs that traditional, face-to-face PD does not (see Appendix B for comparison of OPD vs. traditional PD).

OPD has many advantages: it provides an “anywhere, anytime” availability, it connects teachers with peers and experts in their fields, it is ecological and cost-effective, and it promotes collaborative, inquiry-based learning. Additionally, teachers can select the OPD sessions or courses that address their immediate needs in terms of content, time commitment, and desired level of interactivity. However, OPD also presents challenges. First, school districts need to provide schools and teachers with equal access to high quality Internet and technological resources. They also need to ensure that teachers have at least a fundamental knowledge of the Internet and the hardware and software necessary for participation in OPD.

Research Regarding Effectiveness of Online Professional Development

Several empirical studies comparing face-to-face and online PD have been conducted over the past decade. The results of these studies suggest that there are no significant differences

between the effects of face-to-face and online PD on teacher knowledge and student performance (Russell, Carey, Kleinman, & Venable, 2009; Sankar & Sankar, 2010; Fishman et al., 2013).

However, while “online learning can be as good as, if not better than, in-person learning...educators must research the quality of an online opportunity and not make a decision based on cost or convenience alone” (Bates, Phalen, & Moran, 2016, p. 72). Unfortunately, identifying “high quality” or “effective” OPD may prove to be a difficult task.

There is no formal definition of “effective” PD, including OPD, on which scholars agree. Trends in the PD literature indicate a “growing consensus within the field regarding the central features of professional development that are effective in improving teacher practice” (Borko, Jacobs, and Koellner, 2010, p. 548), which is the ultimate goal of PD. However, there is limited empirical evidence to substantiate causal connections between these features of PD, teacher performance, and student achievement (National Research Council, 2007; Yoon et al., 2007). PD studies have tended to focus on changes in teachers’ content knowledge, attitudes, and satisfaction with PD, relying on self-reported data and case studies with non-generalizable conclusions (Borko et al., 2010; Russell, Kleinman, Carey, & Douglas, 2009; Guskey, 2009). In a widely-cited study by Yoon et al. (2007), only nine out of thirteen hundred studies analyzing the characteristics of effective PD met the What Works Clearinghouse’s (an initiative in the U.S. Department of Education’s Institute of Education Sciences) credible evidence standards. Consequently, as Guskey contends, there is a “knowledge gap” between “our beliefs about the characteristics of effective professional development and the evidence we have to validate those beliefs” (2009, p. 224). Thus, as many contemporary PD scholars advocate, more research is needed to distinguish which features of PD are most likely to achieve desired outcomes for teachers and students.

Moreover, the same PD intervention(s) may have varying effects from one school to the next, depending on a number of contextual factors. Thus, rather than attempting to determine universally applicable PD interventions, researchers should conduct studies that demonstrate how certain core characteristics manifest in different contexts. As Guskey asserts, “No professional development practice, strategy, approach, method, or activity works well under all conditions...success will come from finding the optimal mix of effective practices based on core elements that work well in a particular context or collection of contexts” (2009, p. 231).

K-12 OPD Evaluative Research Approaches

One of the primary challenges of evaluating K-12 OPD is that “little clarity exists on *how* [emphasis added] to evaluate professional development” (Merchie et al., 2016, p. 1). Consequently, researchers have adopted different approaches to designing and conducting K-12 OPD evaluations. In particular, these approaches include different underlying epistemologies, different methods of data collection and analysis, and different ways of addressing causal relationships, especially concerning outcomes and contextual factors (Snyder, 2009; Muijs & Lindsay, 2008). These differences are manifested in the long-standing debate between researchers who advocate for experimental or quasi-experimental approaches and those who advocate for theory-based approaches to evaluation (Crow, 2010). Thus far, researchers conducting K-12 OPD evaluations have primarily used non-experimental and quasi-experimental approaches to measure course outcomes and determine the extent to which they were achieved (see Chapter 2 and Appendix D). However, this study uses a RE approach, which belongs to the theory-based evaluation (TBE) family and emphasizes the relationships between contextual factors, causal mechanisms, and particular outcomes (Pawson & Tilley, 1997).

Evaluative research involving K-12 institutions is difficult to conduct, regardless of the approach. Researchers must comply with strict ethical standards and practical constraints, which can limit the scope of research questions, access to participants and information, and time allotted for implementing interventions and collecting data. Thus, designing and gaining approval for these studies can be a long and arduous process (Alibali & Nathan, 2010). For example, because studies with experimental or quasi-experimental approaches treat the program intervention as the unit of analysis, researchers create treatment and control groups to control for the effects of extraneous variables. This can present ethical and practical challenges in K-12 settings, as researchers must determine whether it is ethical to implement a treatment that unfairly advantages or disadvantages certain participants. Additionally, as random assignment is not often feasible in K-12 studies, researchers must work with groups that are as similar in characteristics and settings as possible. This increases the difficulty of controlling for all of the contextual variables that may affect the internal validity of the study. Nevertheless, experimental and quasi-experimental approaches can be efficient ways to study intervention outcomes, particularly through “summative evaluations and cost benefit analyses” (Hawkins, 2014, p. 47).

Proponents of a theory-based evaluation approach state that one of its major strengths is that, rather than evaluating “interventions in their entirety as successes or failures...there is a recognized need to uncover what aspects of an intervention have or have not been successful with which particular subgroups” (Mackenzie & Blamey, 2005, p. 151). Thus, they do not focus on controlling for extraneous factors, claiming that this attributes too much credit to the implementation of the intervention for generating outcomes without sufficiently considering the roles of other factors (Stame, 2004). In the context of K-12 OPD, TBEs are used to gather information about why and how course outcomes are achieved, which can be used to improve

OPD courses. The RE approach, which is used in this study, has become a unique and prominent member of the TBE family (Westthorp, 2014).

Overview of Realist Evaluation

Used most commonly in the field of program evaluation, RE has grown in popularity in recent years (Westthorp, 2014; Greenhalgh et al., 2017). Based on the principles of realist philosophy (see Chapter 2), REs seek to construct explanations of the relationships between contextual factors, causal mechanisms, and outcomes that align as closely to reality as possible. REs explore the premise that programs have different outcomes for different individuals in different contexts by answering several defining questions: What works, for whom, in what circumstances, and how (Pawson & Tilley, 1997)? To answer these questions, they develop, test, and refine program theories, which function as the unit of analysis, to determine the relationship between contextual factors, causal mechanisms, and outcomes (see Figure 1). This information can help stakeholders determine whether and how to revise or expand their programs. Furthermore, findings from REs of programs with similar interventions or contextual factors contribute to a shared knowledge base that can be used to inform future evaluative studies (Blamey & Mackenzie, 2007; King et al., 2016).

Pawson and Tilly state that RE begins with the generation of program theories, which hypothesize the relationship between contextual factors, mechanisms, and outcomes; they describe the various ways that a program's outcomes might (or might not) be reached (Van Belle et al., 2016). Once developed, the initial program theories are tested and refined through multiple methods of data collection and analysis, resulting in a finalized set of program theories. Based on these iterative cycles of program theory testing, revision, and refinement, elements of the final program theories can be transferred as middle-range theories to programs operating in similar

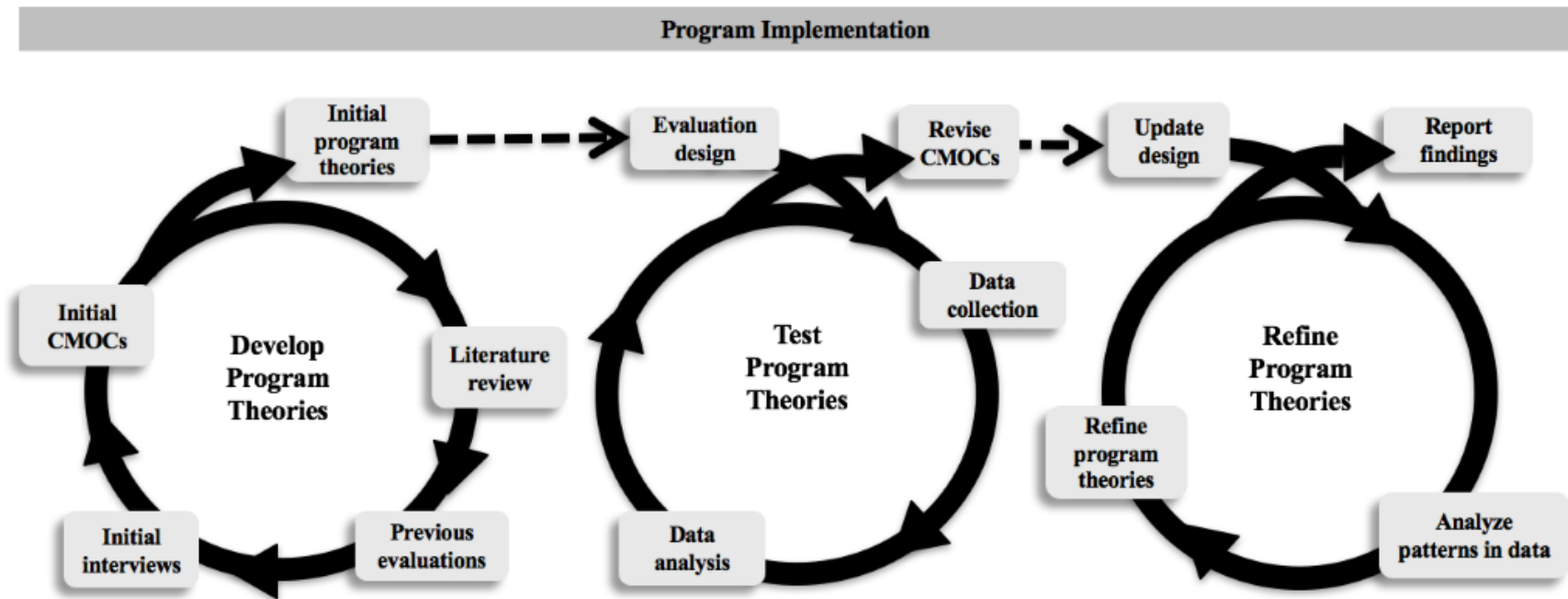


Figure 1. Pawson & Tilley's Realist Evaluation Model

contexts or with similar interventions. Middle-range theories are neither universal nor specific to a single context; thus, they can be referenced in subsequent evaluations that may have similar contextual factors and interventions (Pawson & Tilley, 1997; Astbury, 2013).

Realist Evaluation Key Concepts

As stated by Pawson and Tilley, “realist evaluation stresses four key linked concepts for explaining and understanding programs” (1997, p. 6), including mechanisms (M), context (C), outcomes (O), and context-mechanism-outcome configurations (CMOCs). The relationship between the four elements is commonly represented by the formula $C + M = O$. I further describe these concepts below and they are presented in a summary table in Appendix A.

Mechanism (M). Mechanisms are the powers, forces, interactions, and processes that bring about outcomes when activated by particular contextual factors. Pawson and Tilley define mechanisms as “what it is about programs and interventions that bring about any effects” (1997, p. 6). Ultimately, “identifying mechanisms involves the attempt to think through how a program actually changes behavior” (Pawson & Tilley, 2013, p. 5) through participants’ responses to program interventions and resources (Dalkin et al., 2015).

Pawson and Tilley align their usage of mechanisms with several key characteristics, including the facts that:

- Mechanisms should not be equated with program activities or variables,
- They are usually hidden,
- They are sensitive to variations in context, and
- They generate outcomes (Astbury & Leeuw, 2010; Dieleman, Wong, & Marchal, 2012; Van Belle et al., 2016; Pawson & Tilley, 1997).

These characteristics have important implications for realist evaluators. Because mechanisms are hidden, evaluators must uncover them through targeted data collection and analysis. Pawson and Tilley often use the example of a clock to describe this phenomenon; in order to determine how the clock works, one must look beyond its face and examine the gears and springs inside (1997, 2013). However, just because mechanisms are not always observable “in a direct, empirical sense...does not mean they are not real” (Astbury & Leeuw, 2010, p. 369). As Pawson and Tilley also point out, we believe in the existence of gravity, even though it is not tangible; instead, we see the effects of gravity, such as a ball dropping to the ground, and that is evidence of its reality (1997; Greenhalgh et al., 2017). Mechanisms are also sensitive to variations in context; thus, they may be activated in some contexts but not others. Pawson and Tilley relate this to the example of gunpowder, which “has the causal potential to explode, but whether it does so depends on it being in the right conditions” (2013, p. 6). Similarly, contextual factors have the potential to trigger mechanisms in different ways (Van Belle et al., 2016; Dalkin et al., 2015; Westhorp, 2014), so realist evaluators endeavor to identify relevant contextual factors when considering how mechanisms are activated. Finally, it is mechanisms, not program interventions, that generate outcomes. Individuals react to program interventions in different ways, bringing about certain outcomes based on the interaction between mechanisms and contextual factors (Van Belle et al., 2016). Those using a RE approach must identify the contextual factors that influence the activation of mechanisms and, by extension, the outcomes of particular interventions.

Context (C). In terms of realist evaluation, contexts include the existing social, cultural, political, historical, psychological, economic, and organizational factors in which programs are embedded (Greenhalgh et al., 2017; Van Belle et al., 2016). They exist in individual,

interpersonal, institutional, and infrastructural levels. Different contextual factors “interact and influence each other” (Greenhalgh et al., 2017, p. 1), as well as causal mechanisms, to bring about different outcomes. Realist evaluators are tasked with identifying relevant contextual factors in order to determine why mechanisms vary in different settings or circumstances (Van Belle et al., 2016; Tan & Harvey, 2016; Pawson & Tilley, 1997; Wong et al., 2016).

Even seasoned realist evaluators admit that it is sometimes challenging to distinguish between context and mechanisms (Jolly & Jolly, 2014; Pawson & Manzano-Santaella, 2012; Dalkin et al., 2015). In order to make this distinction more apparent, several researchers have revised the $C + M = O$ formula, defining the C and M elements more explicitly. These attempts at revision are detailed in the section of Chapter 2 that discusses the challenges of RE.

Outcomes (O). Pawson and Tilley define outcomes as “the intended and unintended consequences of programs, resulting from the activation of different mechanisms in different contexts” (1997, p. 8). Realist evaluators anticipate that programs will produce multiple outcomes, depending on the interactions between relevant contextual factors and causal mechanisms. Thus, when collecting data through interviews or focus groups, realist evaluators ask targeted questions about the achievement of outcomes for different individuals in different contexts and why these differences may exist (Manzano, 2016; Pawson & Manzano-Santaella, 2012; Greenhalgh et al., 2017). They analyze this data to create and refine program theories that explain how different outcomes result from the interactions between certain contextual factors and causal mechanisms (Tilley, 2000).

Context-mechanism-outcome configurations (CMOCs). CMOCs are the crux of a RE. Pawson and Tilley describe CMOCs as “propositions [that] bring together mechanism-variation and relevant context-variation to predict and explain [outcome] variation” (1997, p. 9). They



Figure 2. CMOOC Example

liken CMOCs to recipes, in which specific ingredients are measured, combined, and cooked at specific temperatures to create the desired product. Similarly, programs operate through the combination of certain mechanisms in certain contexts to generate desired outcomes (1997; see Figure 2 for an example).

Retroductive process. As Pawson and Tilley state, RE “is about theory testing and refinement” (1997, p. 9) through the use of retroduction, or abductive reasoning. The concept of retroduction was formally developed by Charles Sanders Peirce in the early 1900s as “the process of both generating hypotheses and selecting some for further pursuit” (CP 5.171, 6.468-477 in Kapitan, 1992, p. 1). It involves both deductive and inductive reasoning (see Figure 3). First, a hypothesis, or theory, is formed. Next, the researcher uses deductive reasoning to gather and analyze data to test the theory. Typically, data is gathered through multiple methods, including interviews, observation, document analysis, and quantitative information. The researcher then examines the data to see if it aligns with the initial theory; discrepancies are noted and inductive reasoning is then used to generate a new theory that better represents the data. Finally, deductive reasoning is again used to test the new theory. This process can iterate as long as the researcher continues to collect new data, which may depend on temporal and logistical constraints, as well as the practicality of continuing the study (Pietarinen & Bellucci, 2014; Gold, Walton, Cureton, & Anderson, 2011).

The process is the same in the context of a RE study, which begins with a “retroductive question about the causal powers of the policy, intervention or program, given the circumstances in which it is applied. How is it that Intervention X can produce outcomes $Y_{1..k}$ given conditions $Z_{1..k}$?” (Greenhalgh et al., 2017g, p. 2). Researchers develop an initial set of program theories, or hypotheses, about what outcomes will be achieved, for whom, and in what circumstances.

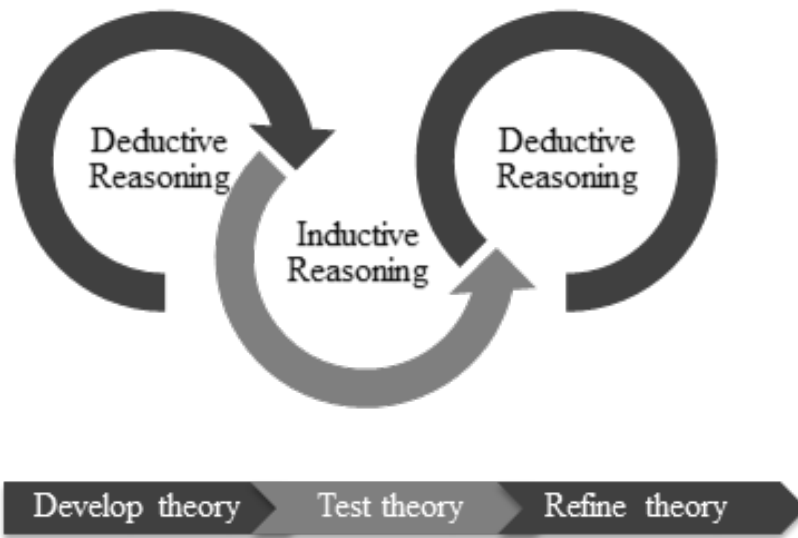


Figure 3. Retroductive Process

CMOCs are the basis of these program theories, and program theories are written in C-M-O format (Pawson & Tilley, 1997, 2013; Tilley, 2000; Westhorp, 2011; Greenhalgh, 2017).

Researchers then use multiple methods to collect and analyze data, which informs their refinement of the original program theories (Westhorp, 2011; Blamey & Mackenzie, 2007).

Ultimately, Pawson and Tilley hope that those using a RE approach will make greater efforts to share their tested program theories in order to create “reusable conceptual frameworks... [so that similar] evaluations then operate within a common set of program theories” (Astbury, 2013, p. 395).

Purpose of Study

The primary purpose of this study was to determine the extent to which the desired outcomes for select K-12 teachers’ participation in the *VTSB* OPD course were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. A secondary purpose of the study was to set a precedent for future studies of K-12 OPD courses, particularly those using a realist evaluation (RE) approach. The findings of this study identified relationship(s) between certain contextual factors, causal mechanisms, and outcomes that can be referenced in evaluative studies of other OPD programs operating in similar contexts or with similar intervention strategies (Pawson & Tilley, 1997; King et al., 2016).

Research Questions

In keeping with the primary RE questions, the study considered the extent to which the desired course outcomes were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how.

Guiding questions for the study include:

What works?

- Which of the desired outcomes (e.g., teachers develop the ability to adapt VTS for diverse groups of students, teachers implement VTS in the classroom) are achieved and to what extent?
- What are the unintended outcomes of the course?

For whom?

- To what extent do participants/sub-groups of participants (e.g., teachers in different locations or who completed different stages of the course) benefit from this course, as indicated by achieving desired or unintended outcomes?

In what circumstances?

- What contextual factors influence the achievement of desired outcomes and to what extent (e.g., level of administrative support, relevance to instructional goals)?

How?

- What types of participant reactions/responses do course resources (e.g., course interface, structure) and activities (e.g., course annotations, reflection, applying content) elicit? To what extent do these responses relate to the achievement of desired outcomes?

These guiding questions cannot be answered separately, as they are interdependent; the answer to “what works” depends on the answers the other questions. Therefore, the answers to these questions are presented in Chapter 5 as program theories representing configurations of contextual factors, causal mechanisms, and outcomes.

Significance of the Study

In this era of accountability, K-12 schools are expected to meet specific performance standards based on teacher evaluations and student achievement. Many schools and researchers assume a causal connection between PD, teachers' behavior and skill level, and student achievement (Muijs, Day, Harris, & Lindsay, 2004). Thus, they perceive PD as a means to higher scores on teacher evaluations and high-stakes tests, and have increased PD requirements in K-12 schools (Owston, Wideman, Murphy, & Lupshenyuk, 2008). Many teachers have gravitated towards OPD programs, as they are "convenient and flexible, and...make teaching and learning possible at any time and in any place" (Rakap, 2014, p. 222). As a result, more evaluative studies of OPD programs are needed to ensure that they are effectively achieving their intended outcomes. However, few OPD evaluative studies have been conducted in K-12 settings and those that have are mostly in STEM related fields (Snyder, 2009; Muijs & Lindsay, 2008).

This study provides information regarding Watershed Collaborative's K-12 OPD course, *VTSB*, which trains K-12 teachers to use artwork to incorporate student-centered and inquiry-driven activities into their classrooms. Using a RE approach provided a systematic way to determine not only the extent to which participants are applying course content, but also *how* and *why*, including why there are different outcomes for different participants.

This study is valuable for several entities, including Watershed Collaborative, future participants of the *VTSB* course, K-12 OPD course designers, researchers and evaluators of K-12 OPD, and researchers interested in exploring or adopting a RE approach. In particular, it sets a precedent for future studies of K-12 OPD courses, especially those using a RE approach. The final program theories serve as valuable contributions to a growing body of evidence that can inform future K-12 OPD REs (Timmins & Miller, 2007). At their roots, many program

interventions consist of similar underlying mechanisms implemented in different contexts (Pawson & Tilley, 1997). Thus, elements of the final program theories in this study can be referenced by OPD courses operating in similar contexts or with similar intervention strategies (Pawson & Tilley, 1997; King et al., 2016; Stern, et al., 2012). Additionally, given the limited number of RE studies outside of the UK, this study also serves as an exemplar for other researchers interested in adopting a RE approach or comparing it to other evaluative research approaches.

Assumptions and Limitations

Assumptions

There are several underlying assumptions for this study. First, concerning the topic of the study, K-12 OPD is assumed to be an expanding, yet under-researched area of study that has potential benefits for K-12 OPD course designers, educators, and students. Second, the theoretical framework for the study, which is based on realist philosophy, involves both ontological and epistemological assumptions. According to realist ontology, reality exists separately from an individual's perception of that reality. Each person experiences this reality in unique ways, as influenced by biological, psychological, social, and cultural contextual factors. However, these subjective experiences of the world do not constitute multiple realities; rather, they are varying perceptions of a single reality. Thus, while participants may engage with the same intervention, it is assumed that various contextual factors influence their experiences of, and responses to, the intervention in unique ways. According to realist epistemology, individuals construct their knowledge of the world based on their experiences and perceptions of reality. Realist researchers assume that exploring these differing perspectives brings them closer to a representation of reality that aligns with actual reality. Consequently, my goal as a researcher is

to construct answers to my research questions that are as close to reality as possible by using multiple methods of data collection and analysis. Third, in relation to the study's methodology, it is assumed that the study's multiple methods of data collection and analysis complement each other in the process of testing and refining program theories. Furthermore, the use of multiple methods, such as the involvement of participants in refining program theories, enhances the study's credibility. Finally, in relation to the validity of the study, it is assumed that participants provided honest survey and interview responses.

Delimitations

The delimitation of this study relates to the population of interest, which was purposively selected. Data was only collected from the leadership at Watershed Collaborative and consenting participants enrolled in the *VTSB* OPD course in the fall of 2018 and the spring of 2019.

Limitations

There are a couple of limitations for this study. First, the sample size for interviews and course data was limited based on the number of consenting participants. Second, practical limitations included time, resources, and accessibility to participants. Thus, data collection and analysis had to be completed within a specific time frame, allowing for a limited number of cycles of theory testing and refinement and access to short-term outcomes. Finally, lack of access to participating teachers' classrooms did not allow for observations of their applications of course concepts.

Definitions of Terms

Middle-Range Theories (MRT)

Theories that can be transferred to programs operating in similar contexts or with similar interventions, as they are neither universal nor specific to a single context.

Online Professional Development (OPD)

Online professional development (OPD) is defined as “teacher learning experiences delivered partially or completely over the Internet” (Fishman, 2016, p. 427). OPD has become increasingly popular with innovations in technology and the increased availability of the Internet (Holmes, Singer, & MacLeod, 2010).

Professional Development (PD)

In the context of K-12 education, professional development (PD) is defined as “learning activities related to the profession of teaching that occur after initial certification (Fishman, 2016, p. 14). PD traditionally, and still commonly, occurs in face-to-face settings (Masters et al., 2010), but is rapidly moving to online settings (Harasim, 2000).

Program Theories (PT)

In the context of realist evaluation, the evaluator gathers information from existing program documentation and evaluation reports, reports of similar evaluations, relevant literature, and interviews with program stakeholders to help craft theories that “[comprise] one or more C-M-O configurations” (RAMESES Project Team, 2017, p. 4; Pawson & Tilley, 1997). These theories describe the specifics of how a particular program operates, as well as the (assumed) causal relationship between its design and intended outcomes; they inform the evaluation design and evaluators will continue to refine them throughout the evaluation process.

Realist Evaluation (RE)

Realist evaluation is a member of the theory-based evaluation family (Westhorp, 2014). Realist evaluators assert that programs do not affect individuals in exactly the same way; rather, the same program will have different outcomes for different individuals in different contexts.

Thus, the defining questions of RE are: What works, for whom, in what circumstances, and how (Pawson & Tilley, 1997)?

Theory-Based Evaluation (TBE)

This study adopts Coryn et al.'s definition of TBE, as "any evaluation strategy or approach that explicitly integrates and uses stakeholder, social science, some combination of, or other types of theories in conceptualizing, designing, conducting, interpreting, and applying an evaluation" (2010, p. 201).

Visual Thinking Strategies (VTS)

The Visual Thinking Strategies (VTS) curriculum was co-designed by Philip Yenawine, a long-term museum educator, and Abigail Housen, a cognitive psychologist (Yenawine, 2013; VTS, 2017). Yenawine defines Visual Thinking Strategies as "the use of art to teach visual literacy, thinking, and communication skills" (Yenawine, 2013). As an instructional method, VTS consists of regular teacher-facilitated interactive sessions that foster engaging, open-ended discussions about select piece(s) of artwork.

Realist Evaluation Key Concepts

As stated by Pawson and Tilley, "realist evaluation stresses four key linked concepts for explaining and understanding programs" (1997, p. 6), including mechanisms (M), context (C), outcomes (O), and context-mechanism-outcome configurations (CMOCs). The relationship between the four elements is commonly represented by the formula $C + M = O$.

Context (C)

In terms of realist evaluation, contexts include the existing social, cultural, political, historical, psychological, economic, and organizational factors in which programs are embedded (Greenhalgh et al., 2017; Van Belle et al., 2016). Different contextual factors "interact and

influence each other” (Greenhalgh et al., 2017, p. 1), as well as causal mechanisms, to bring about different outcomes. They exist in individual, interpersonal, institutional, and infrastructural levels.

Mechanism (M)

Mechanisms are the powers, forces, interactions, and processes that bring about outcomes when activated by particular contextual factors. Pawson and Tilley define mechanisms as “what it is about programs and interventions that bring about any effects” (1997, p. 6). Ultimately, “identifying mechanisms involves the attempt to think through how a program actually changes behavior” (Pawson & Tilley, 2013, p. 5) through participants’ responses to program interventions and resources (Dalkin et al., 2015).

Outcomes (O)

Pawson and Tilley define outcomes as “the intended and unintended consequences of programs, resulting from the activation of different mechanisms in different contexts” (1997, p. 8). Realist evaluators anticipate that programs will produce multiple outcomes, depending on the interactions between relevant contextual factors and causal mechanisms (Pawson & Tilley, 1997; Astbury, 2013; Greenhalgh et al., 2017).

Context-Mechanism-Outcomes Configurations (CMOCs)

CMOCs are the crux of a RE. Pawson and Tilley describe CMOCs as “propositions [that] bring together mechanism-variation and relevant context-variation to predict and explain [outcome] variation” (1997, p. 9). They liken CMOCs to recipes, in which specific ingredients are measured, combined, and cooked at specific temperatures to create the desired product (1997). CMOCs are often presented in a table format.

Organization of the Study

This study has five chapters. In addition to providing relevant background information, the first chapter includes the statement of the problem, research questions, the purpose and significance of the study, and definitions of key terminology. The second chapter includes a review of the literature on several topics, including the history and current state of: K-12 OPD, K-12 OPD evaluations, and realist evaluation. The third chapter describes the study's methodology. The fourth chapter contains my data analysis and findings. The fifth and final chapter presents a summary of my findings, middle-range theories and their implications, and recommendations for future studies in this area of research.

CHAPTER 2: LITERATURE REVIEW

Introduction

The purpose of this literature review is twofold: First, it establishes the historical, philosophical, theoretical, and methodological contexts for my study. Second, it critically evaluates my decision to adopt a RE approach for this study. Thus, in the first section of the review, I provide a comprehensive overview, synthesis, and critique of key concepts from K-12 OPD evaluative studies from 2005-present. Specifically, I focus on these studies' use of evaluation terminology, models, purposes, designs, and data collection methods. I also identify several gaps in the literature and discuss how they are addressed by this study. In the second section of the review, I first provide an overview of RE's philosophical and theoretical foundations. I then identify the primary strengths and challenges associated with a realist evaluative approach and discuss how my study leverages these strengths and responds to these challenges.

Methodology

The methodology for this section of the literature review consists of several steps. The workflow was informed by Lubke, Britt, Paulus, & Atkins, (2017) (See Figure 4):

1. Establish inclusionary/exclusionary criteria and keywords for potential sources.
2. Conduct keyword, subject, Internet, and backwards citation searches to find sources, including peer reviewed articles, official evaluation reports, and evaluation standards from various databases and Internet searches.
3. Apply inclusionary/exclusionary criteria to filter sources.
4. Compile sources in Paperpile, a web-based reference management software.

5. Annotate sources in PDF Expert and continue to eliminate sources that do not meet criteria.
6. Upload and code sources using Nvivo, a qualitative analysis software.
7. Create matrices displaying coding frequencies.
8. Use coding matrices and additional Nvivo queries to conduct a comprehensive conceptual analysis of source materials.
9. Interpret and report findings.

STARLITE Literature Search Overview

As outlined in Table 1, I used Booth's STARLITE mnemonic "to convey the essential elements for reporting [my literature search]" (2006, p. 421). Using the following search strategies, I first obtained 13 evaluation studies, 4 evaluation reports, 7 frameworks/models, 10 theoretical/descriptive sources, 2 critiques, and 2 literature reviews. After applying the inclusionary and exclusionary criteria, the number of sources in each category was reduced to, 7, 6, 4, 2, 2, and 2, respectively. See Appendix C and D for a visual breakdown of the final sources and their citation information.

Annotation and Coding

I read and annotated all of my sources before importing them to Nvivo for coding. As I chose not to start with a predetermined set of codes, I created codes as I re-read through the material. In total, I created fifty-one codes, which I organized under six parent codes, which acted as concepts for my conceptual analysis: Evaluation Design, Evaluation Levels, Models/Frameworks, Professional Development, Research, and Definitions/Terminology (See Appendix E). I also created a memo for each source, which helped me to track how my research

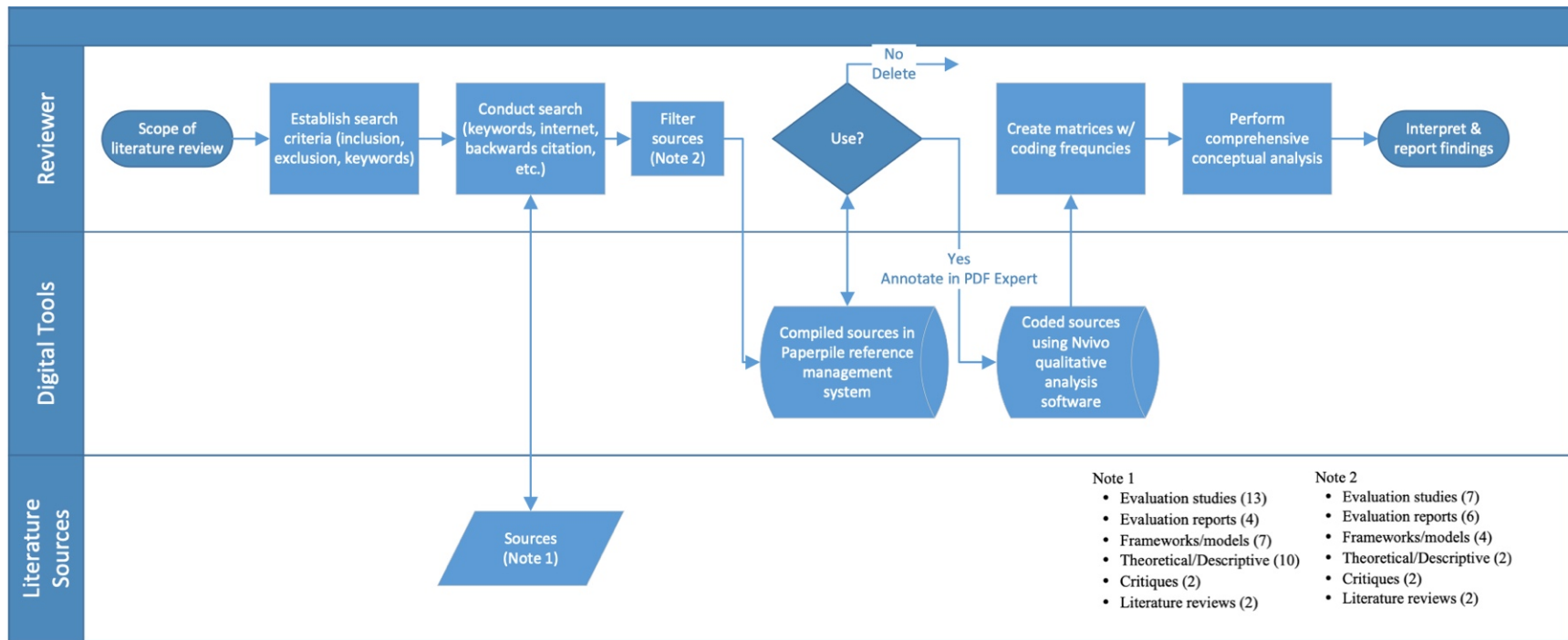


Figure 4. Digital Literature Review Process

Table 1

STARLITE Literature Search Components

Elements	Explanatory Notes
Sampling Strategy	• Purposive
Types of Studies/Literature (22 sources)	• Evaluation studies (7) • Evaluation reports (6) • Frameworks/models (4) • Theoretical/Descriptive (2) • Critiques (2) • Literature reviews (2)
Approaches	• Backward citation searches • Keyword searches • Subject searches • Internet searches
Range of Years	• 2005-present
Limits	• English
Inclusions and Exclusions	• Inclusions • Sources that are peer reviewed, disseminated by a professional organization, or supplied by a government entity, including national/state/local public school documents • Studies and reports must involve K-12 PD or OPD • Studies and reports must involve non-experimental, quasi-experimental, experimental, or theory-based designs • Frameworks, models, and critiques of frameworks/models must involve PD or OPD evaluation • Literature reviews must focus on the evaluation of PD or OPD; they may focus on general evaluation principles if referred to in at least two sources involving PD or OPD • Exclusions • Studies not conducted in the United States • Studies or reports involving OPD for pre-service teacher education • Theses/Dissertations
Terms Used	• See Appendix F for key terms and combinations of terms

Table 1. Continued

Elements	Explanatory Notes
Electronic Sources	• Databases • Academic OneSearch, ERIC (Education Resources Information Center), Professional Development Collection from EBSCO • Journals • Professional Development in Education, Journal of In-Service Education, Educational Research and Evaluation, Studies in Educational Evaluation, American Journal of Evaluation

practices and reflections evolved over the course of my analysis. Additionally, the memos helped me to avoid unintentional plagiarism when writing this literature review.

Conceptual Analysis and Discussion

Terminology and Definitions

I first define the key terms and concepts that are frequently referenced in this review, which are derived from the twenty-three sources used for the review. The definitions and usage of many of these terms vary among sources, with authors sometimes using different terms for the same concept. Thus, to establish consistency and expose any personal bias, I will present the definitions that I adopted for the purposes of this review. First, I adopted Hahs-Vaughn et al.'s definition of OPD, which states that “online professional development is multidimensional. It encompasses: a) an online, web-based format (e.g. self-paced, custom-built courses, or blended learning); b) professional development; and most likely c) specific objectives tailored to and created for the respective online professional development course” (2007, p. 5). As OPD evaluation is comprised of two primary components, “OPD” and “evaluation”, I will address terminology from each of these components separately before discussing OPD evaluation itself.

Online Professional Development

As stated by Hahs-Vaughn et al., “professional development can be defined in a variety of ways and can take many forms” (2007, p. 5). Although I encountered multiple definitions of PD, they share several defining characteristics that informed my perception of PD for this review: PD includes planned activities, exposes teachers to current pedagogical and subject-related research, is intended to advance teachers’ professional knowledge and skills, and has an ultimate goal of bolstering student performance (Merchie et al., 2016; Hahs-Vaughn et al., 2007; Muijs et al., 2004).

PD can be delivered face-to-face, online, or through a combination of both (a blended format). While recent face-to-face PD programs have been more student-centered, traditional face-to-face PD is known for its teacher-centered approach. In comparison to OPD, traditional, face-to-face PD is criticized for its limited accessibility, inflexibility, lack of extended collaborative opportunities, and lack of subject-specific topics (Owston et al., 2008).

Alternatively, OPD is widely available, affords teachers the ability to access subject-specific PD at various times and locations, and facilitates collaboration, communication, and resource-sharing among teachers (Rice & Hung, 2015).

Evaluation

Since its inception, the field of evaluation has struggled to establish consistent definitions and usage of its specialized terminology (Reio, Rocco, Smith, & Change, 2017). To prevent ambiguity, I will identify and define the key terms that I used for coding and conceptual analysis: evaluation framework, evaluation model, evaluation type, evaluation design (see Table 1 and Figure 4). While frequently cited, some of these terms are used interchangeably, even by the same authors. For example, King (2014), Merchie et al., (2016), Muijs & Lindsay (2008), Owston, Sinclair, & Wideman (2008), and Hahs-Vaughn, Zygoris-Coe, & Fielder (2007) all use the terms “model” and “framework” interchangeably. In comparison, Coldwell & Simkins (2011), Bradley (2011), and Foss Hansen (2005), distinguish between the two terms. Other pairs of terms that are used interchangeably include: approach and framework, type and approach, and model and approach.

Because the same terms were often used in different ways, it was challenging to create codes and code descriptions that applied to all sources equivalently. Thus, for the purposes of this review, I coded all of the sources’ terminology according to the definitions in Table 2 and

Figure 5. For example, although King (2014) uses the terms model and framework interchangeably, I assigned both terms to the code for models, as their usage aligns with the definition of model in Table 1. This strategy afforded me the ability to generate an accurate representation of the frequencies of concepts and ideas, rather than of words with inconsistent meanings. As a result, the usage of terminology in my conceptual analysis also mirrors the definitions in Table 2 and Figure 5.

Evaluation Models

Evaluation models “provide a graphical or textual overview of the steps involved in evaluation” (Ogle, 2002, p. 11). All twenty-two of the sources for this review incorporate evaluation models in some capacity, whether describing them, comparing and critiquing them, and/or applying them in a study. Eleven sources explicitly discuss level models, which Coldwell and Simkins define as “a family of evaluation approaches that share the characteristic of tracing the efforts of training and development interventions through a series of ‘levels’, each of which more closely approaches the ‘ultimate’ intentions or outcomes of the evaluation” (2011, p. 145). Thus, each level uses different methods to collect and analyze data in an effort to improve the quality of the program. Two of the most prominent level models include Kirkpatrick’s four-level evaluation model and Guskey’s professional development evaluation model (see Figure 6). Both of these, whether independently or conjoined with other models, have commonly been used in the evaluation of PD, including OPD (Hahs-Vaughn et al., 2007; King, 2014; Merchie et al., 2016). Of the ten OPD evaluation studies and reports, eight incorporated level models into their designs.

Kirkpatrick’s model and OPD: Description and application. Kirkpatrick’s model, derived from a series of papers he published in 1959, is “the most well-known and utilized model

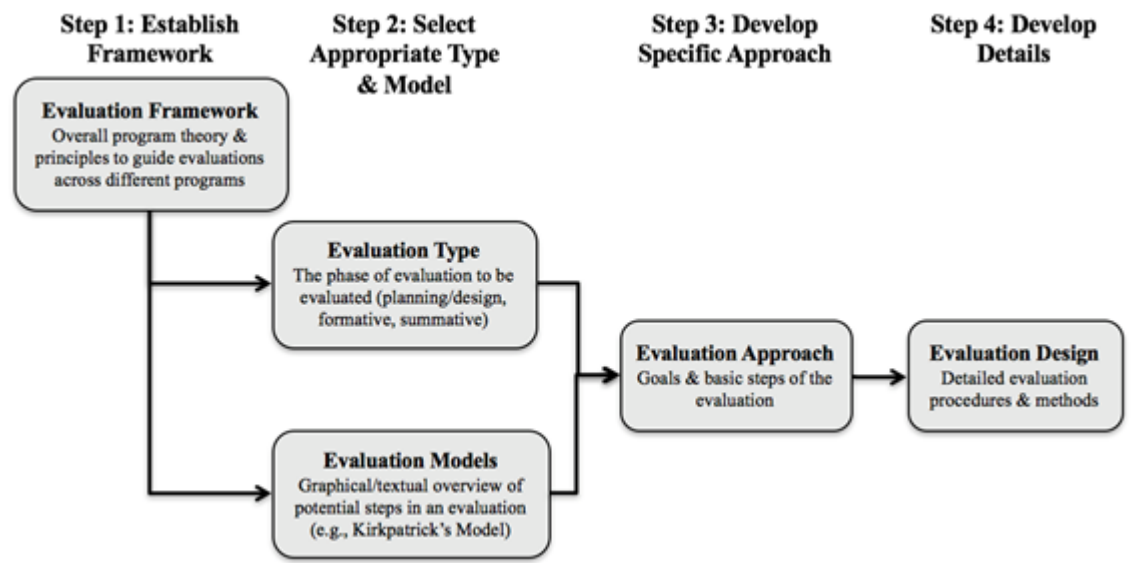


Figure 5. OPD Evaluation Terminology Concept Map.

Table 2

Definitions of OPD Evaluation Terminology

Term	Definition	Examples
Evaluation approach	Coherent set of ideas about what an evaluation should accomplish and how it should be carried out; the goals and basic steps of an evaluation.	• <i>Goal-Free approach</i> • <i>Objectives-Based approach</i> • <i>Theory-Based approach</i>
Evaluation framework	Provides an overall framework for evaluations across different programs or different evaluations of a single program (e.g. process evaluation; impact evaluation); can include guidance on data sources and management processes. It sometimes includes an overall program theory and principles to guide the planning, management and conduct of evaluations.	• <i>Desimone (2009)</i> • <i>Merchie et al. (2016)</i>
Evaluation model	Provides a graphical or textual overview of the steps involved in evaluation.	• <i>Kirkpatrick's 4-Level Evaluation Model</i> • <i>Guskey's PD Evaluation Model</i> • <i>Logic Model</i>
Evaluation type	The phase of evaluation to be evaluated; the type selected depends on the purpose of the evaluation.	• <i>Planning/Design</i> • <i>Formative</i> • <i>Summative</i>
Evaluation design	Provide the method by which the evaluation will be conducted, including detailed evaluation procedures.	• <i>Qualitative</i> • <i>Quantitative</i> • <i>Mixed-Methods</i>

Note. The definitions for evaluation model, type, and approach are from Ogle (2002), the definition for evaluation approach is from Stecher & Davis (1987), and the definition for evaluation framework is from Betterevaluation (2017). Examples provided by the author of the dissertation

for evaluating training programs” (Reio et al., 2017, p. 35). The model consists of four levels, which I will describe in the context of an OPD program (see Figure 6). The first level of the model gauges how teachers perceived different aspects of an OPD program, including “the training content, materials, instructors, facilities, [and] delivery methods” (Reio et al., 2017, p. 36). Teachers may be interviewed or asked to fill out surveys to provide qualitative and quantitative data that can be analyzed to improve the quality of the program’s design and implementation.

The second level of the model measures how teachers’ knowledge, skills, and attitudes change after participating in an OPD program. These changes are often measured by comparing pre and post-test scores or examining teachers’ reflections (Hahs-Vaughn et al., 2007), which can indicate ways to improve the program’s content and instructional strategies. Level three examines how extensively teachers transfer their new knowledge and skills from an OPD program to their classroom. Through observations, interviews, and/or focus groups, evaluators can determine the extent to which an OPD program influences teachers’ behavior outside of the PD environment. Finally, the fourth level of the model measures changes in student achievement resulting from teachers’ participation in the OPD program.

Kirkpatrick’s model and OPD: Critiques. Despite its ubiquity in the realm of training evaluation, Kirkpatrick’s model has been thoroughly critiqued. Reio et al. (2017) and Coldwell & Simkins (2007) challenge three of the model’s assumptions that were first identified by Alliger and Janak (1994). The first assumption is “that the levels are arranged in ascending order and the model is hierarchical in nature” (Reio et al., 2017, p. 37). The authors claim that this is troublesome, as it could make OPD evaluators feel justified in devoting their efforts to the higher levels of the model, while neglecting the lower levels. The second assumption “is that the four

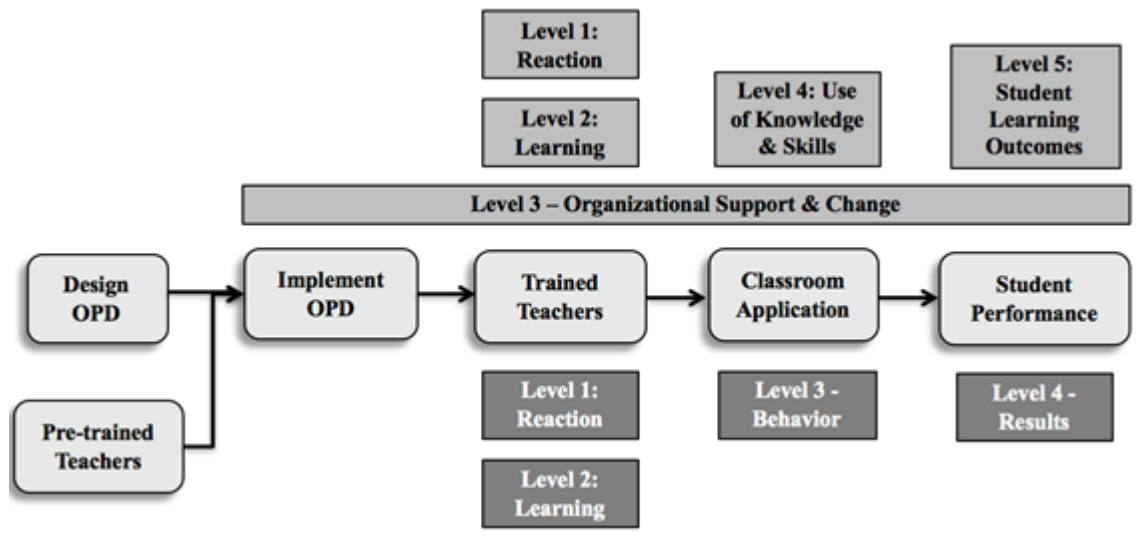


Figure 6. Levels of Kirkpatrick's Model (Bottom) and Levels of Guskey's model (Top) in Comparison to Steps in OPD Course Development and Implementation Process (Middle). Similar to Hahs-Vaughn et al.'s Hybrid Evaluation Model (2007).

levels of the evaluation are causally linked” (Reio et al., 2017, p. 37). In response, the authors state that OPD evaluators and researchers might unfoundedly attribute changes in higher levels to successes in preceding levels. The third and final assumption is “that the four levels are positively intercorrelated” (Reio et al., 2017, p. 37). Under this assumption, OPD evaluators might assume that desirable results in one level will inevitably produce desirable results in the following level. Thus, if participant reactions are satisfactory, the rest of the evaluation levels will be satisfactory, even if supporting evidence is lacking (Coldwell & Simkins., 2007).

Guskey’s model: Description and application. While Kirkpatrick’s model is the most well known in the context of training program evaluation, Guskey’s is the most well known in the context of educational PD evaluation (Merchie et al., 2016, p. 3). Although Guskey adapted his model from Kirkpatrick’s, he asserts that Kirkpatrick’s model has “limited use in education because of explanatory power. While helpful in addressing a broad range of ‘what’ questions, many find it lacking when it comes to explaining ‘why’” (2014, p. 1225). Thus, Guskey “sought to add explanatory power for evaluators” (Hahs-Vaughn et al., 2007, p. 10) in his five level model.

The first, second, and fourth levels of Guskey’s model correlate to the first, second, and third levels of Kirkpatrick’s model, respectively (Hahs-Vaughn et al., 2007; King, 2014). Thus, they gauge participants’ reactions, learning, and transfer of learning to the classroom environment (see Figure 6). Guskey’s third level is a new addition and the fifth level is modified from Kirkpatrick’s fourth level. The third level, labeled organizational support and change, “focuses on the organization as well as the support the organization provides for individuals to implement what was learned once they return to their respective positions within the organization” (Hahs-Vaughn et al., 2007, p. 10). Guskey contends that this level in particular sets

his model apart from Kirkpatrick's, as it examines factors that may influence the other levels.

For example, if teachers do not implement PD content in their classrooms, the fault may not lie at the individual level but at "the building and district levels" (Guskey, 2014, p. 1227).

Evaluators can gather information for this level by examining school records and policies, administering questionnaires, and conducting interviews with administrative personnel. This data is intended to provide the explanatory power that Kirkpatrick's model lacks. The fifth level focuses on student learning outcomes. Guskey stipulates that "providing acceptable evidence for judging the effects of professional learning activities will almost always require multiple sources of evidence" (2014, p. 1228). He also calls for data to be collected from randomized comparison studies if possible. School administrators can examine this data to determine the value of an OPD program and whether or not to continue funding it.

Guskey's model: Critiques. Interestingly, Guskey appears to embrace the assumptions that were critiqued in Kirkpatrick's model, stating that "the five levels in this model are hierarchically arranged, from simple to more complex. With each succeeding level, the process of gathering evaluation information requires more time and resources. And because each level builds on those that come before, success at one level is usually necessary for success at higher levels" (2014, p. 1225). However, King asserts, "level 3 of Guskey's model, Organization Support and Change...is not a consequence of level 2, Participants' Learning" (2014, p. 94). Furthermore, researchers contend that Guskey's third level is not empirically supported as a factor of "'high quality' and 'effective' PD as advocated by other researchers" (King, 2014, p. 95). Merchie et al. extend this criticism by stating that "an important element lacking in Guskey's five-stage evaluation model is the evaluation of a [PD program's] effective features"

(2016, p. 5). Finally, while Guskey's third level does provide some information regarding the influence of external factors, it is by no means comprehensive.

New evaluation models. As a result of the shortcomings of both Kirkpatrick's and Guskey's models, the authors of five sources generated their own PD evaluation models (Merchie et al., 2016; King, 2014; Bradley, 2011; Coldwell & Simkins, 2011; Hahs-Vaughn et al., 2007). Of these five new models, three are intended to evaluate PD programs in general, while Bradley's and Hahs-Vaughn et al.'s are designed specifically to evaluate OPD programs. Hahs-Vaughn's model is a synthesis of several models, including a logic model, Guskey's model, Kirkpatrick's model, and the Sloan Consortium's five pillars of online education. This "hybrid" model is Hahs-Vaughn's response to the fact that "there is no single evaluation model that 'fits' when evaluating online professional development" (2007, p. 7).

Evaluation Purposes

All of the sources in this review describe either process, outcome, or impact evaluations of K-12 OPD courses. Process evaluations provide information about the implementation of a program and assess the extent to which it is being implemented as intended; they help stakeholders determine why certain outcomes were or were not achieved (Royse, Thyer, & Padgett, 2010). For example, Snyder conducted a process evaluation to "identify to what extent teachers who had completed the [reading intervention] program...felt they were effectively prepared to teach reading. [Another] key goal of [the] evaluation was to identify areas of the program that could be improved or revised to better serve the needs of teachers" (2009, p. 147). Similarly, Lebec and Luft examined "the manner in which course participants learned and how the online environment influenced this process" (2007, p. 554).

Outcome evaluations provide information about a program's results or effects and the extent to which it achieved its objectives; they help stakeholders determine whether to improve, continue, or discontinue the program (Royse, Thyer, & Padgett, 2010). For instance, Kellogg, Corn, & Booth state that "the purpose of [their] evaluation is to provide detailed information about the extent to which the online professional development components of the RttT application have been carried out" (2012, p. 12). Maxfield, Huynh, & Mueller also conducted an outcome evaluation, which was part of a grant requirement; beyond this, their results were used to "help guide [Minneapolis Public Schools] in making important programmatic decisions to improve professional development and enhance technology integration district-wide" (2007, p. 7).

Finally, impact evaluations look at larger scale effects of a program and, often through rigorous empirical design, attempt to establish causal connections between program interventions and outcomes (Posavac, 2011). Jaciw et al. conducted an impact evaluation using "a mixture of causal and correlational analyses, [telling them] about the causal impact of the program" (2016, p. 7) on teachers' classroom practices and students' general reading literacy. Sherman, Byers, & Rapp also conducted an impact evaluation to determine whether "the professional development experience would have a positive impact on the professional practice of the participants" (2008, p. 23). Stakeholders from both of these evaluations used the results to inform their decisions about future iterations of the programs.

Evaluation Designs

The evaluation designs were driven by the evaluation purposes. Thus, the process, outcome, and process/outcome evaluations adopted either non-experimental or quasi-experimental designs, and the impact evaluations used either experimental or quasi-experimental

designs. Interestingly, many of the other sources expressed a desire for more controlled experimental designs in the evaluation of OPD, especially regarding the impact of OPD on teacher performance and student achievement (Rakap et al., 2014; Sherman et al., 2008; Owston et al., 2008). However, experimental studies are challenging to carry out in K-12 OPD settings for logistical and financial reasons, including the fact that many teachers are pre-enrolled in OPD courses prior to an evaluation. For this reason, Maxfield et al., stated that “it wasn’t possible to have a comparison group (of teachers who did not participate in the program)” (2007, p. 8). Furthermore, some schools and teachers are unwilling to participate in a control group, in which they receive no OPD interventions while other teachers do (Owston et al., 2008).

Data collection methods. Specific data collection methods differed according to evaluation designs; however, all of the evaluations used a mix of quantitative and qualitative methods, including survey questionnaires, pre/post assessments, interviews, document analysis, and observations. Several evaluations indicated that multiple data sources were used in order to triangulate data, which would increase the validity of their findings. For example, Maxfield et al. state that, while their evaluation did not have a comparison group, they used “multiple data collection methods to triangulate the results” (Maxfield, 2007, p. 9). Similarly, Lebec and Luft “[considered] the conclusions possible from the triangulation of all forms of data” (2007, p. 565). Snyder et al. also used triangulation to “[enhance] the reliability, validity, and usefulness of [their] findings” (2009, p. 147).

Limitations. Multiple sources acknowledged several common limitations of their evaluation designs and data collection methods. Several evaluations only obtained small sample sizes, which “limited the degree to which conclusions could be made” (Lebec & Luft, 2009, p. 561; Rakap, 2014; Owston et al., 2008). Furthermore, attrition rates were high in many instances.

Another reported limitation is the amount of self-reported data, which may not be consistently reliable (Rakap, 2014; Maxfield, 2007). Finally, several sources mentioned that the short duration of the evaluation limited their ability to draw conclusions regarding teachers' implementation of skills in the classroom (Lebec & Luft, 2009; Rakap, 2014; Maxfield, 2007).

Implications of Common Results

It is beyond the scope of this review to report the results of the evaluation studies in detail. However, some commonalities will be noted, as they have implications for future OPD evaluations. First, teachers in several studies expressed that, while they appreciated the convenience of OPD, they still had a desire for face-to-face interactions with their peers and the instructor (Sherman et al., 2007; Snyder, 2009; Rakap, 2014). Thus, they were more satisfied with OPD that incorporated face-to-face components, whether through synchronous online communication or in-person. Teachers also indicated that they preferred OPD courses that were structured and provided a strong teaching presence. Owston et al. stipulate that this is because teachers "were generally not used to self-directed professional development as it was normally something 'delivered' to them" (2008, p. 1058).

Additionally, multiple sources reported that attendance was lower and attrition rates were higher than anticipated. Owston et al. infer that teachers may have been unable to attend due to time constraints; thus, "more research needs to be done to find ways of increasing teacher participation" (2008, p. 209) in OPD. Teachers in Jaciw et al.'s empirical study implied that they might have been more engaged if "they were implementing [new skills] alongside other teachers from their school or district" (2016, p. 57). Teachers in Maxfield's report indicated that they would have been more involved if there had been more mandated activities, if there were greater consequences for non-engagement, and if there had been more technological support available"

(2007, p. 16). Finally, Lebec & Lufy tie teachers' struggles with technology to self-efficacy theory, in which "the aggravation felt by these individuals translated into a lack of engagement" (2007, p. 566).

Gaps in Literature

After reviewing the literature on OPD evaluation, I noted four major gaps that I address in this study:

1. The majority of evaluations encountered for this review use non-experimental or quasi-experimental designs, which do not focus on or control for "extraneous" contextual factors. Some of the evaluators advocate for more theory-based evaluation approaches, which can help in analyzing causal relationships between interventions and outcomes (Bradley, 2011; Coldwell & Simkins, 2011). For example, Jaciw et al. studied how certain contextual factors, such as teachers' levels of self-confidence or students' levels of engagement, affected teacher and student outcomes following teachers' completion of the iRAISE OPD course (Jaciw et al., 2016). Identifying these factors and their effects can help OPD developers and facilitators to revise their courses and tailor them for different clients with different needs.
2. Second, there is no mention of cultural competence in any of the twenty-three sources. According to the American Evaluation Association (AEA), culture affects all evaluation, all phases of evaluation, and the ways in which evaluation is conceptualized; "evaluations cannot be culture free" (p. 3, 2011). Thus, evaluators need to maintain an awareness of their own ethnocentrism and how it might affect their perceptions of situations and interactions with those from other cultures. Additionally, evaluators should strive to gain "specific knowledge of the people and place in which [an] evaluation is being conducted"

(p. 3, 2011). Exercising cultural competence when evaluating online environments is easy to neglect, as stakeholders can seem anonymous. This makes it all the more important for evaluators to ensure that they are making an effort to understand and work to accommodate diverse cultural backgrounds and perspectives.

3. Third, all of the OPD programs in this review are for STEM related fields. There were no evaluations conducted for OPD programs in ELA, foreign languages, art, or music. These are areas for future OPD developers and evaluators to consider.
4. Lastly, more research is needed regarding the ratio(s) of self-directed, asynchronous activity to synchronous instructor-facilitated interactions. Teachers indicate a preference for a blend of asynchronous and synchronous OPD activities, but there is limited evidence to compare how differing amounts of these activities affect OPD courses' outcomes (Sherman et al., 2007; Snyder, 2009; Rakap, 2014).

Addressing the Gaps

1. For this study, I adopted a realist evaluation approach, which is a member of the family of theory-based evaluation approaches. RE is unique in its underlying realist philosophy and emphasis on the importance of context (Westhorp, 2014). It uses multiple methods of data collection and analysis to examine how particular contextual factors and causal mechanisms affect program outcomes. Thus, using a RE approach helped to identify the contextual characteristics and conditions under which certain causal mechanisms were triggered during participants' completion of the *VTS Basics* OPD course, resulting in various outcomes.
2. I made it a priority to exhibit cultural competence throughout my study by striving to maintain an awareness of my own cultural perspectives and biases. Additionally, RE is

inherently concerned with contextual factors such as “culture, class, gender, religion, economic systems... [and how they] can have real effects on whether and how programs work” (Greenhalgh et al., 2017, p. 2).

3. The K-12 OPD course that I studied, *VTS Basics*, is intended for K-12 teachers of any grade level and subject area. By participating in the course, teachers learn to foster their students’ critical thinking, communication, and social skills by using works of art to engender inquiry-based discussions.
4. The *VTS Basics* course features a “3-4-1” shared enrollment, in which each enrollee may designate two collaborative partners who also engage in course activities and have complementary access to course resources. While the coursework is primarily asynchronous, partners are encouraged to observe each other’s live VTS implementation sessions and provide constructive feedback. Additionally, there is time allotted for participants to engage in synchronous chat sessions with the course facilitator. This provides opportunities to study the synchronous and asynchronous, as well as independent and collaborative, aspects of the course.

Conclusion

A review of the literature indicates that few evaluative studies of OPD have been conducted in K-12 settings (Snyder, 2009; Muijs & Lindsay, 2008). Of these, the majority are in STEM-related fields and use approaches that focus on outcomes; thus, according to TBE proponents, these studies do not sufficiently account for the influence of moderating and mediating factors on these outcomes (Astbury & Leeuw, 2010; Blamey & Mackenzie, 2007). Consequently, there is a need for studies of K-12 OPD in non-STEM-related fields that account

for these factors. In the following section, I describe the RE approach and discuss what it is appropriate for this study, which examines an arts-based K-12 OPD course.

Realist Evaluation

Having explored the gaps in the OPD evaluation literature, it is important to evaluate and justify my decision to adopt a RE approach for this study as opposed to other approaches. As Crandall, Caellegh, and Steinecke state, quality literature reviews should explore the different “views that exist in the literature base, that is, conflicting, consensus, or controversial opinions” (2001, p. 926). To explain why I selected a RE approach, I begin with an exploration of its historical, philosophical, and theoretical foundations. I next examine RE from a critical perspective and discuss its strengths and challenges in relation to other approaches. Through this discussion, I demonstrate why the RE approach is appropriate for this study and describe how I will address potential challenges and limitations.

Methodology

The methodology for this section of the literature review consists of several steps:

1. Establish inclusionary/exclusionary criteria and keywords for potential sources.
2. Conduct keyword, subject, Internet, and backwards citation searches to find sources, including peer reviewed articles, official evaluation reports, and evaluation standards from various databases and Internet searches.
3. Apply inclusionary/exclusionary criteria to filter sources.
4. Compile sources in Paperpile, a web-based reference management software.
5. Annotate sources in PDF Expert and continue to eliminate sources that do not meet criteria.
6. Upload and code sources using Nvivo, a qualitative analysis software.

7. Create matrices displaying coding frequencies.
8. Use coding matrices and additional Nvivo queries to conduct a comprehensive conceptual analysis of source materials.
9. Interpret and report findings.

STARLITE search overview. Using the following search strategies, I first obtained 36 critiques, 103 descriptive/theoretical sources, 84 studies/reports, and four standards/protocols. After applying the inclusionary and exclusionary criteria, the number of sources in each category was reduced to 24, 30, eight, and four, respectively (see Table 3). See Appendix F for a visual breakdown of the final sources, their genres, and their citation information.

Annotation and Coding

I annotated my sources and imported them to Nvivo for coding. I created codes while re-reading the sources in Nvivo. In total, I created forty-five codes, which I organized under six parent codes: Non-Realist Evaluation Approaches, Methodology, Programs, Realist Evaluation, Theory-Based/Theory-Driven Evaluation, and Research (see Appendix E). I created memos for some of the sources, which helped me to track how my research practices and reflections evolved over the course of my analysis.

Historical, Philosophical, and Theoretical Foundations

As Kenneth Hammond states, “every methodology implies a theory, expressed or not” (1980, as cited in McGaghie, Bordage, Crandall, & Pangaro, 2001, p. 930). Thus, to establish a holistic understanding of the RE approach, I must establish its historical, philosophical, and theoretical foundations. I first address the historical foundations of TBE, followed by its philosophical and theoretical foundations.

Table 3

STARLITE Literature Search Overview

Elements	Explanatory Notes
Sampling Strategy	• Purposive
Types of Studies/Literature (59 sources)	• Critiques (24) • Standards (4) • Descriptive/theoretical (30) • Studies/reports (8)
Approaches	• Backward citation searches • Keyword searches • Subject searches • Internet searches
Range of Years	• Did not specify
Limits	• English
Inclusions and Exclusions	• Inclusions • Sources must involve realist evaluation • Sources must be peer reviewed or cited in peer reviewed source(s) • Studies/reports must be cited in source from another type of literature • Exclusions • Theses/Dissertations
Terms Used	• See Appendix G for key terms and combinations of terms
Electronic Sources	• Databases • Academic OneSearch • Journals • British Journal of Criminology, Evaluation, American Journal of Evaluation, BMJ Open

Historical Foundations of RE: Theory-Based Evaluation

TBEs first appeared as a response to growing discontent with methods-based evaluation approaches, which had dominated the field of evaluation since the War on Poverty (Stame, 2004). As the name implies, methods-based approaches focus on “developing a methodology for verifying the internal validity (causality) and external validity (generalization) of programs” (Stame, 2004, p. 59), preferably through the use of randomized controlled trials (RCTs) or quasi-experimental designs. Consequently, evaluation results were used to determine whether a program “worked” or not, based on changes in the independent variables. These approaches were expedient and provided definitive results, which was attractive to many program managers. However, some evaluators began to criticize their rigid methodologies, lack of emphasis on contextual factors and failure to address the “black box” problem (see Figure 7), which “[refers] to the practice of viewing social programs primarily in terms of effects, with little attention paid to how those effects are produced” (Astbury & Leeuw, 2010, p. 364; Blamey & Mackenzie, 2007).

In response, these evaluators started putting forth “white (or clear) box” evaluation approaches, which became synonymous with TBEs (Astbury & Leeuw, 2010; Salter & Kothari, 2014). TBEs have both a conceptual and an empirical component, as they not only “explicate a program theory or model... [but also] seek to investigate how programs cause intended or observed outcomes” (Coryn et al., 2010, p. 203). Thus, unlike the traditional black box evaluation approaches, TBE approaches examine the contextual factors and causal mechanisms associated with program outcomes. Since the 1980s, there have been several widely adopted TBE approaches (see Table 4), including theory-driven evaluation, theory-based evaluation, and realist evaluation (Stame, 2004). While all members of the TBE family, each of these approaches



Figure 7. Black Box Evaluation Model

perceives the black box problem differently, which affects their priorities and methodology (Stame, 2004).

Chen's theory-driven evaluation. Chen's approach views the black box as empty and devoid of theory (see Figure 8). He advocates a "stakeholder-oriented approach" (Coryn et al., 2010, p. 203), in which the role of a theory-driven evaluator is to collaborate with stakeholders to devise and evaluate a program theory that is grounded in social science. The program theory, defined as "a systematic configuration of stakeholders' prescriptive assumptions and descriptive assumptions underlying programs, whether explicit or implicit" (Chen, 2012, p. 18), is method-neutral and serves as a framework for the evaluation process. Chen divides the program theory into two components: the change model (descriptive, or causal, assumptions) and the action model (prescriptive, or normative, assumptions).

Change models describe the relationship between a program's goals, intervention(s), causal mechanisms, and desired outcomes. Action models present the program's plan for achieving the outcomes specified in the change model; they answer the questions of who will be doing what, at which places, and during which times (Chen, 2012). Thus, the change model represents the overarching, abstract theory of why a program will work, while the action model indicates how the program will be carried out on a practical level. By distinguishing between these two components, evaluators can better pinpoint program failures; either there was a flaw in the logic of the action model, or the change model was not plausible (Chen, 2012). This is important "because if a program is based on a faulty theory, then it will not bring about desired changes, irrespective of how well it is implemented" (Astbury & Leeuw, 2010, p. 364).

Table 4

Timeline of TBE Theorists and Their Contributions

	1980	→	1985	→	1990	→	1995	→	2000
Chen & Rossi	1980: First to discuss TBE				1989: Proposes theory-driven as solution to the black box problem				
Chen					1990: Redefines theory-driven evaluation by distinguishing between change & action theory; advocates a stakeholder-oriented approach as a solution to the black box problem				
Weiss					1995: Advances the concept of theories of change as solution to the black box problem; distinguishes between implementation and programmatic theories				
Pawson & Tilley					1997: Publish <i>Realistic Evaluation</i> , which proposes a form of TBE grounded in realist theory; advocate a generative theory of causation as solution to the black box problem				

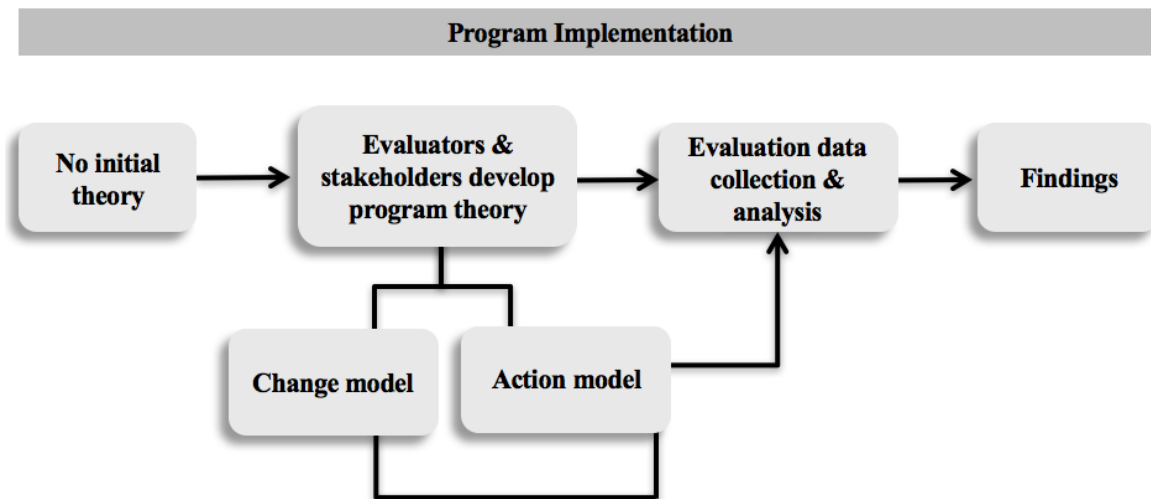


Figure 8. Chen's Black Box Solution: Process Evaluation Model

Weiss's theory-based evaluation. In contrast to Chen's notion of an empty black box, Weiss asserts that the box is filled with many different theories, or theories of change, that "are based on explicit or implicit theories about how and why the program will work" (1995, p. 66). She contends that evaluators should help program stakeholders identify the program's underlying theories of change, including any inherent assumptions (see Figure 9). Then, evaluators can select the appropriate data collection and analysis methods in order to "track the unfolding of the assumptions and...examine the extent to which [the] program theories hold" (1995, p. 67). Ultimately, this TBE approach affords the ability to determine which theor(ies) of change align with the way the program operates in reality.

Weiss divides a theory of change into two components, consisting of implementation and programmatic theories. Implementation theory tests the theoretical assumption "that if the program is conducted as planned...the desired results will be forthcoming" and programmatic theory "deals with the mechanisms that intervene between the delivery of program service and the occurrence of outcomes of interest" (Weiss, 1997b, p. 46). These are similar to Chen and Rossi's action and theory models, respectively. However, Weiss believed that evaluators were helping to expose pre-existing theories, while Chen and Rossi believed that they were helping to create program theories that had yet to be established (Stame, 2004). Nevertheless, they both emphasized the importance of identifying a program's causal mechanisms in order to determine why and how a program did or did not achieve its desired outcomes.

Pawson and Tilley's realist evaluation. Following Chen, Rossi, and Weiss, Pawson and Tilley introduced RE in their 1997 book *Realistic Evaluation* (see Table 4). They took yet another approach to addressing the black box problem that is premised on a generative theory of causality (see Figure 10). According to Pawson and Tilley, experimental and quasi-experimental

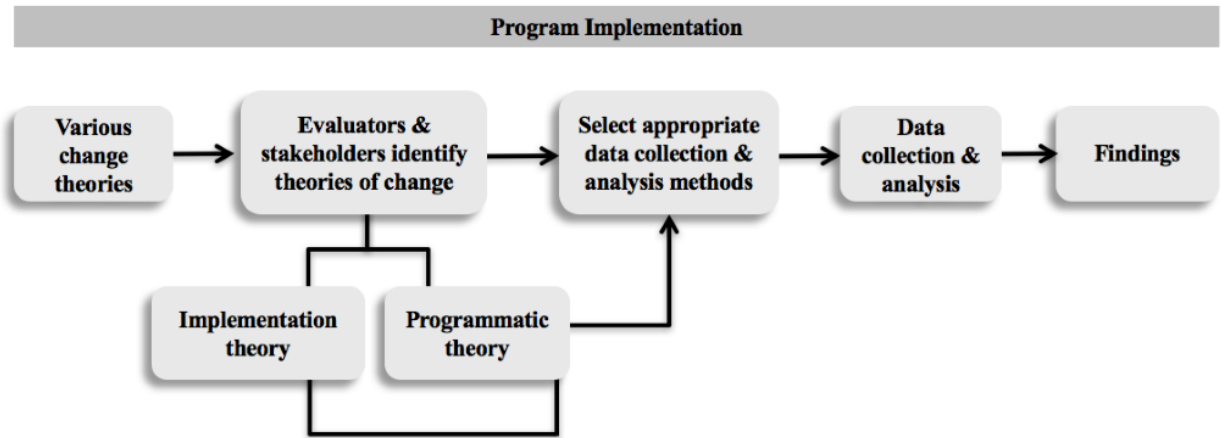


Figure 9. Weiss's Black Box Solution: Process Evaluation Model

evaluations are based on a successionist theory of causality, which considers only the observable changes before and after a program's implementation, and attributes these changes to the program intervention (similarly to an independent variable) (1997). The authors claim that, alternatively, a generative theory of causation assumes that changes occur due to the activation of underlying mechanisms, which they describe as "[accounts] of the makeup, behavior, and interrelationships of those processes that are responsible for [program outcomes]" (2013, p. 4). Thus, it is the individuals who make up the program, along with their actions and reactions, who fill the elusive black box (Stame, 2004).

TBE Critics

It should be noted that TBE has been widely critiqued (Scriven, 1998, 1999; Stufflebeam, 2002; Patton, 1989, Rogers, 2000), raising some valid concerns and difficult questions for TBE proponents. In particular, Michael Scriven defends black-box evaluation and argues that evaluators do not need to know how or why a program works in order to explain the outcomes of an intervention (1999, p. 521). He distinguishes between internal and external theories; internal theories address how outputs are produced, while external theories address how outputs produce outcomes. According to Scriven, "it is often quite possible to understand the external theory of a program...without knowing anything about how the program produces the outputs, i.e., about the internal theory of the program" (1998, p. 60).

Chen responds to Scriven's arguments by claiming that they are mostly based on material products, which should be evaluated differently than social programs (1994). He emphasizes the

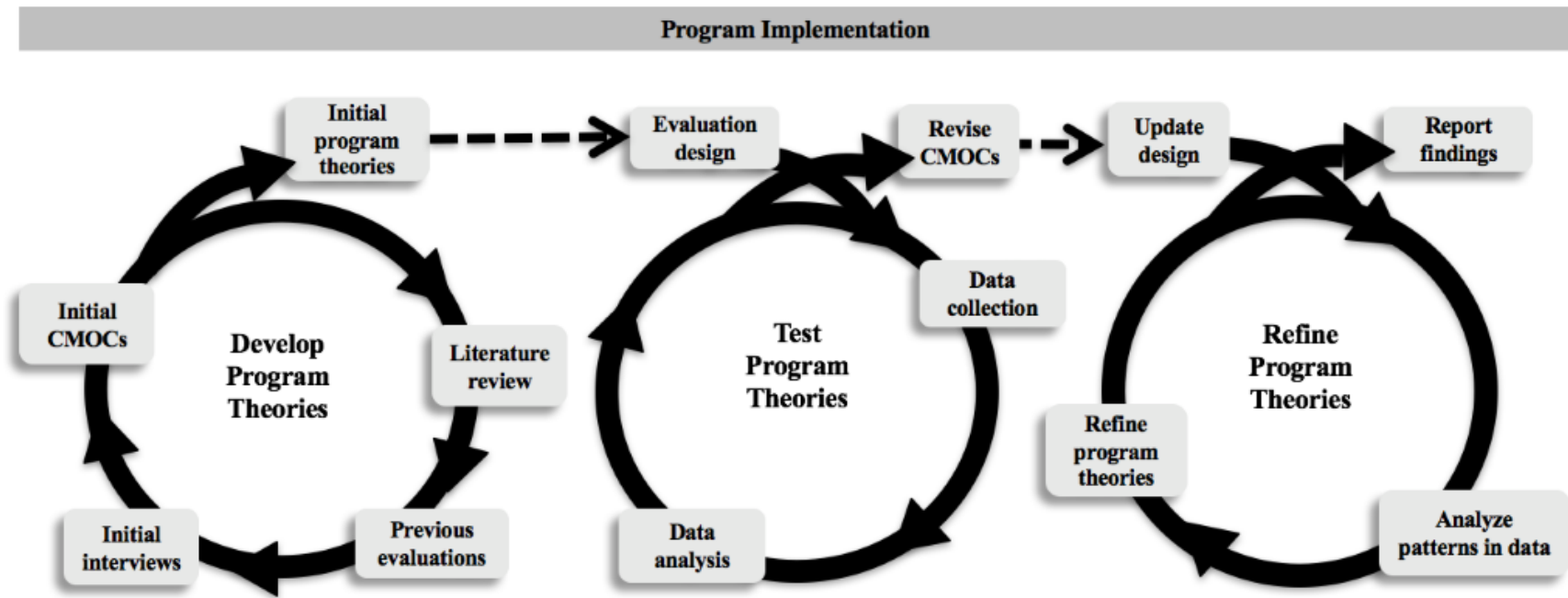


Figure 10. Pawson & Tilley's Black Box Solution: Process Evaluation Model

importance of determining how programs work, as the same outcomes can be achieved by a variety of processes. For example, the desired outcomes of a program might be achieved through processes that are efficient but unethical; if these processes are identified they can be revised in future iterations of the program. However, Scriven maintains that the evaluator's job is to determine the extent to which the intervention succeeded or failed; if the client desires a deeper explanation of why or how the intervention succeeded or failed, they should consult with subject matter experts (1998).

Patton is “concerned that [TBE] is so academically-driven that evaluators will be forced to fit the program into the model rather than constructing a model that is true to the conceptualization of the people involved directly with the program” (1989, p. 377). Similarly, Stufflebeam worries that TBE may be counter-productive, as evaluators may “incur the conflict of interest associated with having to evaluate the theory they developed” (2002, p. 39). He, along with Rogers (2000), also questions whether the theories can be sufficiently tested, considering TBE's high demands on time and financial resources. Weiss acknowledges and addresses these critiques in her description of twelve primary concerns for TBE. Most of her concerns relate to the resource-intensive and intellectually challenging nature of TBE. Ultimately, she argues that, “while some of [these challenges] may be intrinsic” (1997a, p. 521), others should decrease over time as TBE continues to be refined.

Philosophical and Theoretical Foundations of RE

The philosophical and theoretical foundations of RE are separated into four categories (see Figure 10): philosophy, formal theory, evaluation theory, and program theory, as depicted in Figure 11 (Westhorp et al., 2011; Greenhalgh et al., 2017).

Philosophy. As its name implies, RE is based on a realist philosophy, which claims a

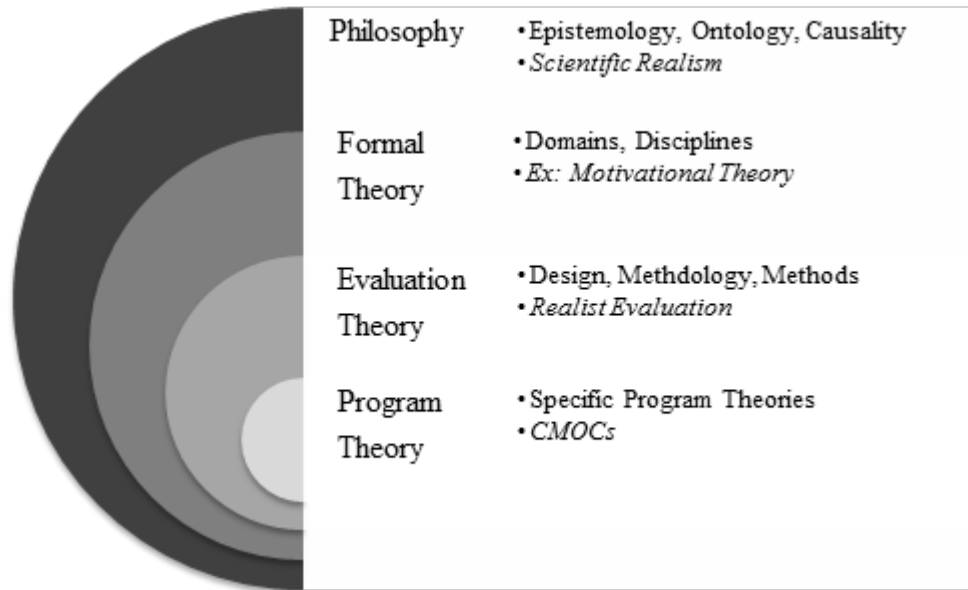


Figure 11. Theoretical Foundations Typology

unique ontology and epistemology. Realism is often described as a positionality “between” positivism and constructivism (Wong et al., 2015; Westhorp et al., 2011; Greenhalgh et al., 2017; Vrijhoef, 2017). Thus, a realist ontology asserts that there is a reality, but that it “exists separate from and independent of one’s perception of it” (Salter & Kothari, 2014, p. 2). Furthermore, something is considered to be “real” if it has real effects on something else, implying that there is both a material and social reality. This has important ramifications for realist evaluators, as “social institutions and constructs (culture, class, gender, religion, economic systems...) can have real effects on whether and how programs work” (Greenhalgh et al., 2017, p. 2). In a given evaluation, realist evaluators strive to determine which of these contextual factors affect the program’s outcomes and in what ways, forming the bases of the evaluation’s program theories. Realists recognize that individuals interpret reality in different ways, resulting in the epistemological tenet that final truth or knowledge is not attainable. However, realist evaluators believe that, by continuing to increase their knowledge base, they can come closer to an understanding of the truth over time (Van Belle et al., 2016; Astbury & Leeuw, 2010; Westhorp et al., 2011). Consequently, they encourage the use of multiple data collection and analysis methods when conducting program evaluations. They also encourage the use of realist synthesis to identify links between program theories from evaluations of programs with similar interventions (Greenhalgh et al., 2017). Essentially, evaluations should “be iterative and knowledge should be treated as cumulative, across policies and programs” (Westhorp et al., 2011, p. 4).

Substantive/formal theory. Substantive theories are those “that operate in different domains or disciplines” (Greenhalgh et al., 2017, p. 2), such as learning or motivational theories in the field of psychology. In the context of RE, these theories can help evaluators to create

program theories, determine influential contextual features and causal mechanisms, make sense of evaluation results, and make connections between different sets of evaluation results (Westthorp et al., 2017). For example, Sorinola et al. (2015) found that motivational theories helped them to understand why and how participants became more engaged in some professional development activities than in others. This helped them to articulate the dynamics between contextual factors and underlying causal mechanisms that produced certain outcomes.

Evaluation theory. Evaluation theories are “theories about evaluation itself” (Greenhalgh et al, 2017, p. 1); they relate to different aspects of evaluation, including its purpose, methodology, and evaluator roles, amongst others. Unsurprisingly, realist evaluators subscribe to a realist evaluation theory, which is an evaluation theory grounded in the tenets of realism, and which “[sets] out to develop, support, refute, or refine aspects of realist program theory (Wong et al., 2016, p. 7). RE theory serves as a guiding framework for evaluators as they design and implement the various steps of a program evaluation (Westthorp, 2011; Van Belle et al., 2016). It involves a cycle of hypothesizing and testing various program theories to determine the relationships between a program’s contextual factors, causal mechanisms, and outcomes” (Pawson & Tilley, 1997).

Program theory. Program theory is the most specific theoretical category; it describes the specifics of how a particular program operates, as well as the (assumed) causal relationship between its design and intended outcomes. Evaluators often express program theories through models and verbal descriptions. This can be challenging, as there are dozens of models and descriptions that have been used by realist evaluators in the past. Moreover, they can “have different purposes, contain different categories, imply different data needs, and justify different

sorts of evaluative decisions” (Greenhalgh et al., 2017, p. 2). Evaluators’ philosophies and evaluation theories inform their creation of program theories.

Realist evaluators have a unique perspective on programs that affects their approach to crafting program theories. Pawson and Tilley describe programs as “theories incarnate” (1997, p. 3), which are embedded in social systems, active, and function as open systems. First, they contend that all programs begin as an individual’s or group’s conceptual ideas, or theories, about how a particular intervention will bring about certain desired results. Programs are created to “test” these theories and identify which outcomes are achieved, for which individuals, in which circumstances, and why (Astbury, 2013; Van Belle et al., 2016; Salter & Kothari, 2014; Wong et al., 2016). Second, realist evaluators also assume that programs are embedded in social systems and relationships that are constantly interacting with one another; this affects programs’ operations and outcomes (Pawson & Tilley, 1997, 2013; Vriethoef, 2017). For example, a school tutoring program is embedded within the larger system of the school, which is comprised of different individuals with different motivations and priorities. All of these elements interact with one another to affect the success of the tutoring program. Third, realist evaluators identify programs as “active” in nature, meaning that they require individuals to actively respond and make choices in order to function. Pawson & Tilley contrast active programs with experimental trials, in which evaluators focus on the effects of an independent variable on a designated treatment group. In these situations, “human volition is seen as a contaminator” (1997, p. 5) and is not considered as contributing to outcomes. Finally, realist evaluators perceive programs as open systems, which also affects programs’ operations and outcomes. In a given program, individuals and resources can come and go, there can be expansion or consolidation, and there can be political influences that create tension (Pawson & Tilley, 1997). For example, if the

aforementioned school tutoring program was placed under a new supervisor, the program might change its procedures, resulting in increased or decreased student performance. Realist evaluators need to keep these four aspects of programs in mind when conducting evaluations, as they all contribute to programs' processes and outcomes.

Strengths and Challenges of Realist Evaluation

In this section, I analyze five purported strengths of RE within the context of a debate dubbed the “Paradigm Wars” (Pawson and Tilley, 1998; Crow, 2010, 2011). Over the course of this debate, Pawson and Tilley critique the use of the quasi-experimental (or OXO) research model, advocating instead for the adoption of a realist approach. The debate unfolded over two series of exchanges, the first with Trevor Bennett in the *Journal of Criminology*, and the second with David Farrington in *Evaluation*. Bennett and Farrington responded to Pawson and Tilley with defenses for the OXO model as well as criticisms of the realist approach. While the two evaluative research studies take place in the field of criminology, the concepts and arguments discussed are applicable to the social sciences in general, including education (de Souza, 2013).

This debate provides an excellent “real-world” context in which to examine the support for and criticisms against the purported strengths of RE, as both “sides” present their arguments in relation to specific research studies. I first provide a summary of the two studies, including their purpose, methodology, and findings. Next, for each of the five strengths, I provide an overview of the strength, followed by Pawson and Tilley’s supporting arguments and Bennett or Farrington’s responses.

Summary of the Studies

Bennett. In their article, *What Works in Evaluation Research?* (1994), Pawson and Tilley compare the quasi-experimental and realist approaches to evaluation research in the context of

Bennett's study exploring the effects of a community-based policing program on the level of community members' fear of crime (Bennett, 1991). The purpose of this year-long program was to reduce levels of fear in the community by establishing an increased sense of police presence. During this time, the local police force made itself more visible by initiating a greater number of cordial interactions with community members. Officers in each location were required to document their engagements with at least one individual from every residency.

Bennett selected the two participating communities according to specific criteria, including their pre-existing levels of fear and their desire to be a part of the study. In accordance with the quasi-experimental research design, he then formed similarly-comprised control and treatment groups in each location and applied the same treatment to the treatment group in each community. Each groups' level of fear was measured prior to and following the treatment using interviews and surveys. This data was compared at the conclusion of the study, with differences being attributed to the presence or absence of treatment. Ultimately the study found no significant effects of the policing program on fear of crime; however, there were positive effects in other areas. Bennett concludes that the program intervention was responsible for these effects, as evaluated by the use of the quasi-experimental method, whose "statistical analysis was based on multivariate techniques [that] help eliminate the influence of extraneous variables" (Bennett, 1991, p. 13).

Farrington. In a subsequent exchange with David Farrington, Pawson and Tilley (1998a, 1998b) critique Farrington's plan to evaluate the Communities that Care (CTC) program, a "community crime prevention program" (Farrington, 1997, p. 157) that originated in the United States and that he argues should be replicated in the United Kingdom (UK). The purpose of the CTC program was to combat major risk factors for youth crime (e.g., substance abuse, teenage

pregnancy, domestic violence) using targeted approaches developed for participating communities (1997). In each community, the program would first help community leaders to establish a Community Board comprised of individuals from the local public and private sectors. With the program's assistance, the Board would then work to identify and implement preventative strategies for the risk factors most relevant to their particular community.

Farrington proposes that the program should be evaluated using a quasi-experimental approach, with three pairs of experimental and control communities. Ideally, according to Farrington, the communities would be randomly selected; however, due to financial and practical constraints, he suggested that they be selected using pre-determined criteria. He also recommends that, to maintain internal validity, each experimental and control community pair should be as similar as possible. In this way, differences in program outcomes would be attributable to the treatment and not to other extraneous factors. In each community, outcomes would be measured before, during, and after the implementation of risk prevention strategies. Consequently, the "effectiveness" of the program would be indicated by the amount of the difference between the initial and final outcome measures.

RE Strength #1: Specific Questions and Causal Attribution

Overview. Pawson and Tilley frequently contrast the quasi-experimental evaluative research model with the realist approach to evaluation, criticizing the former for, among other things, the oversimplification of causal attribution and the failure to satisfactorily explain mixed findings (1994, 1996, 1997, 1998a, 1998b; Pawson, 2000, 2002, 2003, 2004, 2006, 2013, 2015, 2016; Tilley, 2000). As previously established, realist evaluators begin from the premise that “nothing works everywhere for everyone” (Westhorp, 2014, p. 4). Thus, they assert that they do not only ask if something works, as in quasi-experimental designs; rather, they ask what works,

for whom, in what circumstances, and how, as based on a generative perception of causation (Pawson & Tilley, 1997, 2004, 2013). From this perspective, REs provide stakeholders with a fine-grained understanding of how their program operates by identifying the interactions between underlying causal mechanisms and specific contextual factors. Stakeholders can then use this information to make targeted, effective decisions about program revisions or expansion (Westthorp, 2014; Greenhalgh et al., 2017; Tilley, 2000).

Pawson and Tilley's arguments. While Pawson and Tilley commend Bennett for the methodological rigor of his study, they take issue with his quasi-experimental approach (1994, 1996). Moreover, they express a position against quasi-experimentalism in general, calling for “an end to the domination of the quasi-experimental (or OXO) model of evaluation” (1994, p. 292), which they consider more suitable for the evaluation of commercial products than of social programs. They claim that a primary reason for the “failure” of the widely adopted OXO approach is its successionist perception of causation, which causes researchers to focus on whether or not a program “works” without first establishing an understanding of the relationships between various contextual factors and outcomes.

For example, they contend that Bennett over-simplifies the complex nature of social interactions in his study by attempting to control “all explanatory factors, save for the influence of the program itself” (Bennett, 1994, p. 293). Pawson and Tilley consider such an effort to be futile and even “absurd” (1994, p. 298), stating that it is neither possible nor beneficial to ignore contextual factors that might affect the implementation of the intervention, such as the “character of [police] contact” (1994, p. 297) or “character of the community” (1994, p. 299). Thus, they contend that Bennett does not sufficiently address *how* or *why* the program outcomes were or were not achieved (1994, 1996).

Pawson and Tilley instead advocate for a realist approach to evaluation research, which "starts with a theory of what makes programs work and a theory of the circumstances in which such ideas are likely to be efficacious" (1994, p. 292). The authors hold a generative perception of causation, in which they seek to identify the underlying mechanisms that result in particular outcomes in particular contexts; in other words, they seek to determine not only what works, but what works, for whom, in which circumstances, and why (1994, 1996; Pawson, 2007). Pawson and Tilley suggest that a realist approach would have improved Bennett's study by creating and testing "a series of implicit hypotheses about how police presence may be persuasive in changing local patterns of thought and deed on crime" (1994, p. 297). Such an approach would have helped to identify the specific contextual factors affecting when, where, how, and why program mechanisms were activated. For instance, Bennett concluded that, based on their pre and post-treatment survey and interview responses, community members' levels of fear did not differ significantly between control and treatment groups (Bennett, 1991). Pawson and Tilley claim that, by using a realist approach, Bennett could have determined why there was no apparent difference in residents' responses, which might have been due to a variety of reasons. A subsequent series of studies would further indicate which contextual factors activated which mechanisms to produce various outcomes, ultimately allowing for more generalizable conclusions (1994, 1996).

Bennett's response. Bennett disputes the claim that quasi-experimentalists hold a perception of causation in which conclusions are drawn based solely on the associations between program initiatives and outcomes, with no need for "further information or reasoning about why [they] are linked" (Pawson and Tilley, 1994, p. 293). Rather, he argues, quasi-experimentalists agree that, in order to make valid claims about causality, researchers must maintain an awareness

of factors that might influence the relationship between a treatment and the ensuing outcomes, including characteristics of the treatment locations and study participants. He also agrees that critiques and recommendations for improvement to the OXO research design should be pursued, as it still poses issues that “cannot be wholly corrected by statistical analysis” (1996, p. 572), such as “non-equivalence between experimental and control groups (and between pre-test and post-test surveys)” (1996, p. 572).

Bennett further contends that Pawson and Tilley make a category error by contrasting realism with quasi-experimentalism, stating that the former is a philosophical orientation and the latter is a methodology, which is itself informed by the researcher's philosophical orientation. Consequently, he claims that the authors' critiques against quasi-experimentalism are based on “a philosophy of their own creation” (1996, p. 568). Personally, Bennett does not consider realism to be incompatible with quasi-experimental methodology and believes it may have utility for evaluation research. However, he does not think that Pawson and Tilley made a convincing case for a realist approach in particular, concluding that its tenets do not seem to be as exclusive as the authors claim and their arguments against other approaches are based on inaccurate assumptions.

Strength #2: Central Role of Theory/Theory-Driven

Overview. Theory-driven evaluations, such as RE, have both a conceptual and an empirical component, as they not only “explicate a program theory or model... [but also] seek to investigate how programs cause intended or observed outcomes” (Coryn et al., 2010, p. 203). From the realist perspective, the program constitutes a theory, or series of theories, to be tested and refined over the course of the study. By testing program theories, realist evaluators can determine which underlying causal mechanisms and contextual factors are associated with

program outcomes (Salter & Kothari, 2014; Westhorp et al., 2011; Pawson & Tilley, 1997, 2004). Furthermore, because REs are theory-driven, they are method-neutral; researchers can use multiple research methods and sources of data, which provides more flexibility and enhances validity (Pawson & Tilley, 1997, 2004; Hewitt et al., 2012; Wong et al., 2012).

Pawson and Tilley's arguments. The authors address quasi-experimentalism's "deficient and defective conception of the program which is built into [its] methodology" (1994, p. 297). From this perspective, the program is simply the sum of its parts, including inputs, outputs, the treatment, and outcomes. To illustrate this point, they cite examples from Bennett's study that indicate a narrow focus on the proper implementation of the intervention and measurement of outcomes, to the exclusion of contextual factors that may have influenced participants' behaviors (1994, 1996). Pawson and Tilley present the realist alternative by listing several potential theories that might have explained why and how the intervention was affecting residents' attitudes and behaviors, based on the triggering of certain mechanisms in the presence of particular contextual circumstances. They question why Bennett did not explore these theories and sought to answer the binary question of whether or not the program "worked." Ultimately, they attribute this to the method-driven nature of quasi-experimental approach, as opposed to the theory-driven nature of the realist approach (1994, 1996).

Bennett's response. Bennett disagrees that his study's OXO approach reduces the program treatment to an inherently uniform "series of mechanical operations" (1996, p. 570) that does not sufficiently address causal mechanisms. He states that, "testing some of [his previously published] theories [of fear or crime] was the main point of the research" (1996, p. 571) and that the program itself, as well as the survey questions, were designed based on these theories. Thus, differences in the implementation of the intervention would be reflected in participants'

responses to these questions. Bennett ultimately concedes that his paper could have incorporated more overt discussion on these theories, but that this does not diminish their integral role in the study. He adds that his study's "failure" to explore theories regarding causal mechanisms was to budgetary constraints, as opposed to his epistemological stance on causality. The iterative process of testing and refining these theories, as in RE, would have been too expensive and time consuming for this study.

Strength #3: Handling Complexity

Overview. While realist evaluators take the complexity of programs into account, they do not attempt to evaluate them in a comprehensive manner. Programs "have long implementation chains with different decisions being made by different [entities] along the way" (Westthorp, 2014, p. 6); each point on the chain includes a unique interplay of contextual factors and mechanisms, resulting in various outcomes. This results in a list of potential program theories that is "infinitely and disputatiously long" (Pawson, 2004, p. 102), requiring evaluators to narrow their focus to particular program areas and prioritize which theories they will test. Thus, realist evaluators must select certain parts of an implementation chain to be the areas of focus for an evaluation. Then, they create a list of potential program theories, and iteratively test and refine those that they find most relevant to their areas of focus (Pawson, 2004).

Because "programs are [perceived as] complex interventions introduced into complex social systems" (King et al., 2016; p. 22), the contextual factors associated with a given program are numerous and volatile. To this end, Pawson (2013) created an acronym to help realist evaluators identify the primary elements of program complexity: VICTORE (Volitions, Implementation, Contexts, Time, Outcomes, Rivalry, Emergence). Evaluators can use the VICTORE acronym as a checklist to help inform their evaluation designs and ensure that they

are addressing all of the factors that potentially contribute to the program's complexity (Pawson, 2013; King et al., 2016; Astbury, 2013).

Pawson and Tilley's arguments. Pawson and Tilley understand the CTC program to be complex and emergent, deeming it "a flexible, iterative, and evolving construction" (1998a, p. 74; 1998b) that cannot be replicated with the preciseness that Farrington claims. While they acknowledge that CTC guidelines for selecting and implementing interventions are detailed and specific, they maintain that outcomes will differ based on contextual factors, and not on solely the presence or absence of the intervention. Thus, they disagree with Farrington that communities are variables whose effects can be controlled for with a quasi-experimental design. For these reasons, they believe that a realist, theory-based approach to evaluation is more appropriate than a quasi-experimental, methods-based approach in this situation (1998a, 1998b).

Pawson and Tilley also state that the CTC program operates through multiple mechanisms in a stratified reality, resulting in "incomparable vertical and horizontal complexity" (1998a, p. 79). They describe the program's vertical complexity as manifested by the interactions between the program's various stakeholders, the implementation of various interventions, and the varying outcomes of these interventions. Ultimately, they claim that this "program complexity leads to process [(or horizontal)] complexity" (1998a, p. 79) by triggering an array of mechanisms that occur in the presence of particular contextual factors. The authors state that, when using a realist approach to evaluation, the evaluator's job, is to untangle which contextual factors result in which program mechanisms; these results provide specific information regarding what did and did not "work" in the program and why, making them valuable for future studies and subsequent iterations of the program.

Farrington's response. Farrington responds that "CTC is not about community variables having causal effects" (1998, p. 206). Rather, the community is the location in which the program is implemented; such "risk-focused prevention could equally be implemented in schools or prisons" (1998, p 206). Therefore, he maintains that adopting the proposed OXO design would control for influencing factors from the community and combat threats to internal validity. In contrast, he perceives most Pawson and Tilley's examples of realist evaluations as one-group pre-test-posttest designs, which he considers to have "low internal validity [that] fails to control for extraneous variables or exclude plausible alternative explanations" (1998, p. 208).

Strength #4: Transferability and Cumulative Realist Learning

Overview. Pawson and Tilley refer to the empirical methodology of the natural sciences as a foundation for RE, as it "stresses theory and the scope for generalization that comes from attention to explanatory theory" (1997, p. 22). Realist evaluators begin by developing program theories (i.e., hypotheses) about how outcomes come about, based on existing evidence about the relationships between underlying mechanisms and contextual factors. These hypotheses are then tested, revised, and refined through targeted data collection and analysis methods, resulting in final program theories that indicate how and why causal mechanisms are activated in specific contexts. Consequently, elements of the final program theories can be transferred as middle range theories to programs operating in similar contexts or with similar interventions. Ultimately, they serve as valuable contributions to a growing body of evidence that can be referenced in future evaluations (Pawson & Tilley, 1997; King et al., 2016).

Pawson (2006) claims that, at their roots, many interventions consist of similar underlying mechanisms implemented in different contexts. Thus, mechanisms can act as "building blocks for middle-range program theories, which may be transferrable to different

contexts” (Astbury & Leeuw, 2010, p. 374). As an example, Pawson (2006) refers to the “naming and shaming-type” intervention that is manifested in various contexts, including the sex offender registry, school report cards, and hospital inspection results. The underlying mechanism is the same, but it is implemented in a variety of contexts and results in a variety of outcomes. As King et al. state, realist evaluators “should both learn from and build upon previous investigations and contribute findings to shared evidence bases” (2016, p. 34). In this way, they do not always have to start from scratch or “reinvent the wheel” when developing program theories.

Pawson and Tilley’s arguments. Pawson and Tilley state that a realist design would treat the CTC program itself as a theory with multiple sub-theories that would be tested and refined through an iterative process of data collection and analysis. Researchers would select communities for participation in “three parallel exploratory case studies” (1998a, p. 87), with full acknowledgement of their differences. By operating with this understanding, the researchers could observe the effects of a range of contextual factors on program mechanisms and outcomes, eschewing what Pawson and Tilley consider the “bogus [notion of] initial equivalence” (1998a, p. 87). Thus, the research design “would not attempt to evaluate the whole CTC initiative at one fell swoop...[but] would examine the detailed theory underlying each step...with pragmatic decisions being made upon which and how many links in the chain should receive research attention” (1998a, p. 85).

The findings of the evaluation would be transferable in that the final program theories can serve as middle-range theories for reference and testing in subsequent evaluations of similar programs. Pawson and Tilley provide several examples to illustrate the perceived effectiveness of this approach in which, “instead of comparison with some illusory control group,

measurement is directed at expected impacts which would follow if the working theories [were] correct" (1998a, p. 89).

Farrington's response. Farrington defends the proposal to evaluate the program as a whole, as his goal is to "change human behavior rather than test [particular theories] about context-mechanism-outcome linkages" (1998, p. 207) for multiple intervention strategies. Using an OXO approach, the researcher would first determine "*whether* CTC works... [by establishing] if CTC has its intended impact on behavior" (1998, p. 207, emphasis in original). If so, then the researcher should focus on *why* the behavior change occurred. Furthermore, he emphasizes that the numerous evaluations of CTC in the United States provide sufficient evidence for the success of CTC's specific prevention strategies. While he admits that he is assuming the "cross-national replicability of results" (p. 208), he sees little need to expend resources re-evaluating them in the UK, as advocated by Pawson and Tilley (1998a).

Farrington also challenges the realist tenet that the success of a program "is always contingent on context" (1998, p. 207), which implies that the same program will have different outcomes in different locations. To illustrate this point, he lists several examples of programs that have been implemented in various locations with similar results. From his perspective, because the OXO design would control for extraneous influences, Farrington would be able to evaluate the program more comprehensively than Pawson and Tilley. Furthermore, this would help maximize his allotted time and financial resources.

Strength #5: Multiple Methods and Validity

As a member of the theory-based evaluation family, RE is considered to be "method neutral" (Carter, 2012; Coryn et al, 2010; Pawson & Tilley, 1997). RE methodology facilitates the development, testing, and refinement of theories; evaluators use various methods and sources

of data to adjudicate between rival theories as they continue to narrow down explanations for given outcomes. Hence, theories “act as a template for method choice” (Carter, 2012, p. 9) and evaluators select data collection and analysis methods that are most appropriate for accomplishing the aforementioned tasks. REs typically incorporate a mixed method approach, as it provides a holistic perspective of phenomena and facilitates the triangulation of data, which reduces bias and increases validity (Creswell & Clark, 2011). As Manzano surmises, “realist hypotheses are not confirmed or abandoned through saturation...but through relevance and rigor obtained in a mixed-method strategy” (2016, p. 347).

Pawson and Tilley’s arguments. Pawson and Tilley’s realist re-imagination of Farrington’s study would incorporate multiple research methods and sources of data into a research design of three concurrent cases studies. As a theory-driven (as opposed to methods-based) approach, specific methods would be selected only after finalizing the initial program theories. Researchers would use a “teaching and learning process” in order to finalize initial theories, in which they would describe their understanding of the program to primary stakeholders, who would help to identify misconceptions and provide further details. This would increase the internal validity of the study by factoring the perspectives of program stakeholders into causal analysis.

Researchers would then use data gathered through multiple methods to iteratively test and refine these theories until either saturation or practical constraints required that final program theories be established. According to the authors’ research design, quantitative data would be used to measure changes in risk factors for youth crime, such as the rates of teenage pregnancies and reported incidents of domestic violence. Qualitative data “would be primarily ethnographic” (1998a, p. 87) and include observations and interviews with those implementing the program and

program participants; this data would be used to identify the underlying mechanisms and contextual factors influencing program outcomes. Thus, Pawson and Tilley assert that the use of multiple methods would both measure and provide explanations for the outcomes of targeted program interventions, including differences in outcomes within and between communities. Furthermore, using multiple methods and sources of data would increase the study's internal validity.

Farrington's response. Farrington considers Pawson and Tilley's proposed research design as "essentially a correlational design, with all the attendant problems...of inferring causality and excluding possible alternative explanations" (1998, p. 208). Thus, despite Pawson and Tilley's integration of multiple methods into their proposed research design, he still perceives it to have low internal validity. Moreover, he views the realist approach as "primarily concerned with documenting contexts and mechanisms, using qualitative, narrative and ethnographic methods, rather than with evaluating the effectiveness of a prevention program" (1998, p. 208). He concludes that this approach would be inappropriate for the CTC program, as it operates based on too many assumptions, focuses its energy on "testing specific hunches" (1998, p. 208), and is overly time-consuming.

Challenge #1: Propensity for Straw Man Arguments

A common criticism of RE proponents is the misinterpretation or inaccurate portrayal of "rival" approaches in order to discount them in favor of the realist position (Porter, 2015b). In other words, they are perceived to use straw man arguments to bolster the RE approach. Based on my review of the literature, in particular the exchanges involved in the "paradigm wars," I have identified four potential straw man arguments that have been used to support the realist position. I describe these arguments in Appendix I, along with their corresponding assumptions

for the RE position. I also identify which of the strengths of the RE approach (also listed in Appendix I) that each straw man argument has been used to support. This is not a comprehensive list and other arguments may be identified by either side. I referred to this list to critique my own work and ensure that I do not make unsubstantiated claims that favor the RE approach. Rather, I should examine each approach from a critical perspective and use sufficient evidence when constructing arguments.

Challenge #2: Lack of Standardized Methodology

While RE consists of a standard cycle of phases (develop theory, test theory, refine theory), the specifics of these phases vary according to the needs of each evaluation (Astbury, 2013; Pawson & Manzano-Santaella, 2012). Additionally, some evaluations require more cycles through the RE phases than others. Thus, as Salter and Kothari note, many evaluators struggle to conduct REs because “there are no simple steps...to follow and no standardized approach to take” (2014, p. 9). Thus, many evaluations that are labeled as realist vary considerably in their approaches and may not accurately reflect realist principles (Pawson & Manzano-Santaella, 2012). Pawson acknowledges the lack of specific methodological guidelines for RE (2013, 2016). However, he contends that he and Tilley “opted deliberately for the simplicity of the original CMO formula as an agile aide memoire for researchers on how they should begin to construct causal explanations” (2016, p. 137).

Challenge #3: Ambiguity in CMOCs

In multiple case studies, realist evaluators report difficulties with identifying the components of the CMO configurations, in particular the differentiation between context and mechanism (Dalkin et al., 2015; Jolly & Jolly, 2014; Salter & Kothari, 2014; Pawson & Manzano-Santaella, 2012; Timmins & Miller, 2007; Byng et al., 2005). Some of this ambiguity

is due to the fact that, depending on the purpose of the evaluation and the program theories being tested, an element can function either as a contextual factor or a mechanism. For example, pre-existing mechanisms are often considered part of the context into which a program will be introduced, reflecting the stratified nature of reality. Furthermore, programs are actively transforming and being transformed by the contexts in which they are implemented (de Souza, 2013; Pawson 2013, 2017).

Several researchers have attempted to lessen the ambiguity between context and mechanisms by more explicitly defining or revising the $C + M = O$ formula. Dalkin et al. (2015) separate mechanisms into two components, namely resources and reasoning. Similarly, de Souza separates contextual factors into four categories: structure, culture, agency, and relations (2013). Punton et al. (2016) add the component of intervention, or “I,” to CMO configurations. Alternatively, Porter (2015b) adds several elements to the configuration, resulting in $CM + PM + A = O$, in which CM represents contextual mechanisms, PM represents program mechanisms, and A represents individual agency. Most recently, Westhorp (2018) identified five potential types of mechanisms: Powers and liabilities, forces, interactions, feedback or feedforward processes, and reasoning and resources. While some realist evaluators have adopted and benefited from these new perspectives of CMOCs and their constituent elements, there remains confusion regarding this issue (Van Belle et al., 2016; King et al., 2016; Westhorp, 2014).

Challenge #4: Attribution

While Pawson and Tilley criticize traditional evaluation for oversimplifying attribution, they concede that RE has the potential for overcomplicating attribution (1997). They attempt to resolve this issue with the process of “theory adjudication,” which involves judging amongst rival program theories to determine which are closest to the real explanation (Pawson & Tilley,

1997, 2013; Astbury 2013; Stame, 2004). However, as Pawson & Tilley state, “in general, there are an infinite number of explanations for why, when, and how a program works” (1997, p. 17). Consequently, realist evaluators sometimes struggle to distill the most relevant and evidence-supported theories from the entire realm of possibilities (Tan & Harvey, 2016; Timmins & Miller, 2007; Byng et al., 2005). Pawson and Tilley recommend that evaluators employ a pragmatic approach to narrow their focus to a particular set of contextual factors for each evaluation; they should consider which factors are most relevant to the program’s operations, which factors align with stakeholders’ interests, what would contribute most to the literature, and what is feasible considering funding constraints and access to resources. If desired, they (or other evaluators) can revisit the program from a different perspective in subsequent evaluations (Pawson & Tilley, 1997; Tan & Harvey, 2016).

Challenge #5: Resource Intensive

In general, TBE approaches are more complex, expensive, labor-demanding, and time consuming than their traditional counterparts (Salter & Kothari, 2014; Greenhalgh et al., 2017; Porter, 2015b). As Astbury states, the “world in which evaluators operate is governed largely by market demands and contractual obligations” (2013, p. 397). Thus, realist evaluators may face difficulties in finding or convincing stakeholders to dedicate extra funding to an evaluation approach that will likely take more time and manpower (Punton et al., 2016).

Realist evaluators spend a considerable amount of time in the cycle of theory development, testing, and refinement. The amount of times they repeat this cycle often depends on impending time and resource constraints (Salter & Kothari, 2014; Manzano, 2016; Punton et al., 2016). For example, certain data collection methods, such as in-depth interviews and focus groups, require a considerable amount of time and resources. Hence, evaluators may have to

limit the number of individuals they can interview in order to work within time and budgetary constraints (Pawson & Manzano-Santaella, 2012; Manzano, 2016). This is challenging for realist evaluators, as the RE approach is somewhat emergent, and decisions about whom to interview and what questions to ask evolve throughout the evaluation (Manzano, 2016; Greenhalgh et al., 2017).

Challenge #6: Ethical Considerations

Realist evaluators must obtain the necessary ethical approvals prior to data collection. However, this can be difficult due to the “methodologically iterative and emergent” (Greenhalgh et al., 2017, p. 1) nature of RE. For example, the number of interviews, identify of interviewees, and interview questions may expand or change over the course of the evaluation. From the outset, evaluators need to communicate this possibility to their stakeholders and whomever grants ethical approvals; they should do their best to identify potential alterations to the evaluation process (Manzano, 2016; Greenhalgh et al., 2017).

Additionally, as realist evaluators’ primary role is to develop, test, refine, and finalize program theories, they must acknowledge and address potential biases and conflicts of interest when conducting evaluations (Astbury, 2013; Lin & Wu, 2016). They must also maintain an awareness of the implications that their role has for engaging with stakeholders. In this respect, realist evaluators hold neither an insider or external perspective, instead viewing stakeholders as “key sources,” but not all-knowing (Pawson & Tilley, 1997). Thus, as they question stakeholders, evaluators must consider the researcher-subject dynamic and how it may be affected by issues of power or social pressure to give the “right” answers (Greenhalgh et al., 2017).

Conclusion

A review of the literature helped me to evaluate various evaluative research approaches, justify my decision to use an RE approach for this study, and identify challenges that I might encounter. This involved researching the historical, philosophical, and theoretical foundations of RE and identifying how its primary strengths and challenges compare to those of other approaches. I found TBE, which grew out of discontentment with traditional methods-based evaluation approaches, to be a good fit for this study. TBE develops and tests theories to determine not only whether, but also why interventions succeed or fail. RE, which is a subset of TBE, is unique in its underlying realist philosophy and emphasis on the role of contextual factors in influencing the outcomes of an intervention. REs are not always feasible as they can be resource-intensive and researchers may struggle with their lack of standardized methodology. However, given the purpose and research questions of my study, I concluded that an RE approach is most appropriate.

CHAPTER 3: MATERIALS AND METHODS

This chapter details the methodology for this study. I begin with an overview of my research questions, followed by descriptions of the research design, participants, data collection procedures, and data analysis methods. I conclude with a discussion on my efforts to maintain credibility and trustworthiness throughout this study.

Research Questions

In this study, I addressed questions regarding the extent to which the desired course outcomes were being achieved, for which teachers (or subgroups of teachers) within the BCSD, in relation to what contextual factors, and how. The specific guiding questions for the study included:

What works?

- Which of the desired outcomes (e.g., teachers develop the ability to adapt VTS for diverse groups of students, teachers implement VTS in the classroom) are achieved and to what extent?
- What are the unintended outcomes of the course?

For whom?

- To what extent do participants/sub-groups of participants (e.g., teachers in different locations or who completed different stages of the course) benefit from this course, as indicated by achieving desired or unintended outcomes?

In what circumstances?

- What contextual factors influence the achievement of desired outcomes and to what extent (e.g., level of administrative support, relevance to instructional goals)?

How?

- What types of participant reactions/responses do course resources (e.g., course interface, structure) and activities (e.g., course annotations, reflection, applying content) elicit? To what extent do these responses relate to the achievement of desired outcomes?

RAMESES II Quality Standards for Realist Evaluation

The RAMESES Quality Standards for Realist Evaluation informed the methodology of this study. The standards were created by a group of prominent realist evaluators in 2017 as part of the RAMESES II (Realist and Meta-Narrative Evidence Syntheses) Project (Greenhalgh et al., 2015, 2017; Wong et al., 2017). There are eight standards (see Figure 12 and Appendix K) that each list criteria required for an inadequate, adequate, good, or excellent rating. Detailed criteria for each standard can be found in Appendix K.

Design of the Study

I designed this study using a RE approach, which incorporates multiple methods of data collection and analysis. As the RE approach is retroductive, I analyzed the data as it was being collected in order to identify patterns and see if they aligned with the program theories. Based on these findings, I refined and eventually finalized the initial program theories, from which middle-range theories were derived. Three iterative phases of data collection, analysis, and theory refinement were conducted (see Figure 13), including

- Phase I: Develop Initial Program Theories,
- Phase II: Test and Revise Program Theories, and

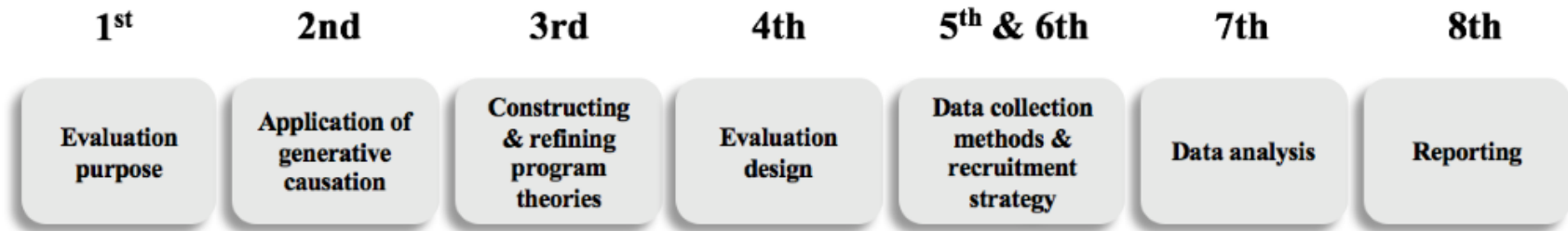


Figure 12. RAMESES II Quality Standards for Realist Evaluation

- Phase III: Refine and Finalize Program Theories.

These phases are described in detail in the subsequent sections.

Institutional Review Board Compliance

I am responsible for protecting participants and keeping their best interests central to the study. Thus, I obtained approval from The Institutional Review Board (IRB) at The University of Tennessee prior to recruiting participants and collecting data for this study. I also obtained informed consent from all participants through written consent forms that include a description of the study objectives and procedures, potential benefits and risks of participation, and a reminder that participation is voluntary and unbinding (Bordens & Abbot, 2014). Additionally, participants were informed that they were under no compulsion to answer every question or provide every piece of requested information and may terminate their participation at any time for any reason. I ensured that all participants' personal information was kept confidential throughout data collection, analyses, and in the final product by providing them pseudonyms and altering or completely removing any personal or identifying information.

Participants

Sampling in RE studies is purposive, as researchers need to recruit “respondents who can provide information about contexts, mechanisms and/or outcomes and the program theory” (RAMESES Project Team, 2017, p. 7). The participants for the study included the Executive, Operations, and Creative Directors of Watershed Collaborative, as well as the individuals I recruited from a pool of K-12 teachers from BCSD who enrolled in the *VTS Basics* OPD course during the fall of 2018 and spring of 2019. The BC Department of Education provided Watershed with a grant for any BCSD K-12 teacher to enroll in the *VTS Basics* OPD course for free. The course also counted towards teachers' required PD credits.

Thirteen course participants consented to be involved in the study; however, five of these individuals either did not actually start the course or did not complete enough to provide usable data. As a result, I collected data from eight course participants, including course progress rates, course annotations, and responses to question and reflection prompts. I also interviewed five of these participants. I interviewed two of the teachers (who are co-workers) together and two other teachers participated in two interviews, for a total of six interviews.

To obtain permission to access course progress rates, gather written data, and conduct interviews, I sent a study recruitment letter in an email (See Appendix O) with the informed consent form (see Appendix P), which described the purpose of my study, the need for the course and interview data, and information regarding participants' roles and rights should they choose to be part of this study. I sent the email to the Watershed Operations Director, who then sent it to the course participants from BC schools. She then forwarded their responses to me using the UT Vault, which is a secure file transfer service. In this way, I was not in direct contact of potential participants until after they volunteered to participate in this study, as per IRB requirements.

Data Collection & Analysis

In the *RAMESES Quality Standards for Realist Evaluation*, the authors define data analysis as “not a specific method but a way of interrogating program theory (or theories) with data and a way of using theory to understand patterns in data” (RAMESES Project Team, 2017, p. 8). Thus, I did not wait to engage in analysis processes until data collection was complete; rather, I worked from a retroductive approach throughout data collection (Greenhalgh et al., 2017; King et al., 2016), using both inductive and deductive methods to test and refine program theories that “[provided] the best possible explanation of acknowledged-to-be-incomplete data” (RAMESES Project Team, 2017, p. 8). According to the previously cited *Quality Standards*,

“excellent” RE studies use a retroductive approach to test and align program theories with data. Final program theories should clearly indicate the C-M-O components and present them in configurations that align with the realist tenet of generative causation (RAMESES Project Team, 2017).

I tested and aligned the program theories with data as it was collected throughout the three phases of my study: Develop Initial Program Theories, Test and Revise Program Theories, and Refine and Finalize Program Theories (see Figure 12). As a member of the theory-based evaluation family, RE is considered to be “method neutral” (Carter, 2012; Coryn et al, 2010; Pawson & Tilley, 1997) and theories “act as a template for method choice” (Carter, 2012, p. 9). REs typically incorporate multiple methods; this provides a holistic perspective of phenomena and facilitates the triangulation of data, which reduces bias and increases validity (Creswell & Clark, 2011). As Manzano surmises, “realist hypotheses are not confirmed or abandoned through saturation...but through relevance and rigor obtained in a mixed-method strategy” (2016, p. 347). Thus, I incorporated multiple forms of data in this study, in order to “scavenge for the best data to test out the [program] theories” (Pawson & Tilley, 1997, p. 11). I gathered data from online course progress rates, course annotations, responses to question and reflection prompts, a realist literature review, and interviews with Watershed Directors and course participants. Collectively, this data helped to address the extent to which desired course outcomes are being achieved, for which teachers, in what types of contexts, and how.

Phase I: Develop Initial Program Theories

Complexity Mapping

From a RE perspective, the *VTSB* course is a complex intervention embedded in a stratified social context, consisting of individual/intrapersonal, interpersonal, institutional, and

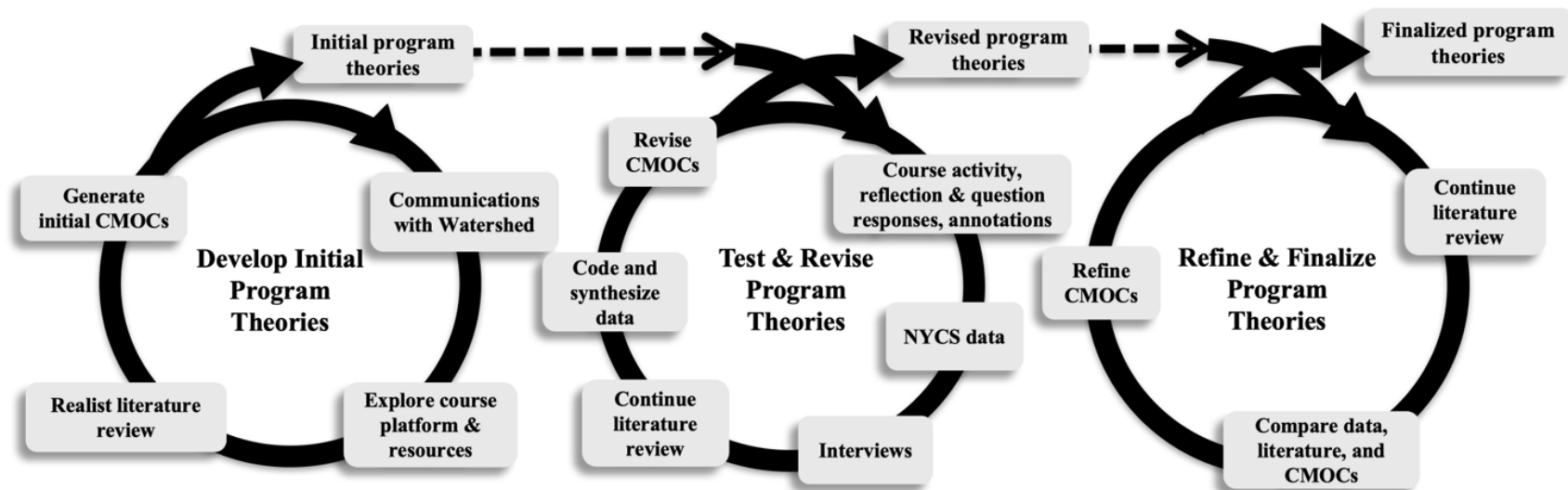


Figure 13. Three Phases of Study

infrastructural layers that reciprocally influence one another (see Figure 14); the intervention and social layers represent horizontal and vertical layers of complexity, respectively (Pawson & Tilley, 1997; Pawson & Tilley, 1998a). Causal mechanisms, which generate course outcomes, are enabled or constrained by various elements within this complex social context. Thus, as discussed in Chapter two, prior to creating program theories I needed to identify the elements of complexity that had the potential to affect causal mechanisms and, ultimately, course outcomes.

I used Pawson's (2013) acronym VICTORE (Volitions, Implementation, Contexts, Time, Outcomes, Rivalry, Emergence) to identify and map these elements. As stated by Pawson, there is "no correct way to map complexity" (2013, p. 45), so I used a combination of concept maps, flow charts, and tables. Mapping these elements ensured that I had a comprehensive overview of the pre-existing social context into which the *VTSB* course was introduced. Using this information, along with my list of available data sources, I determined which contextual elements were most relevant to and within the scope of this study. This served as a reference when generating my initial program theories.

Collecting and Analyzing Initial Data

In Phase I, I collected and analyzed data from several sources to develop my initial program theories, which hypothesize the causal relationships between contextual factors, mechanisms, and outcomes related to participants' engagement in the *VTSB* course. I organized these sources of data in a table, along with their corresponding methods of collection and analysis. I updated the table with new sources of data throughout Phases I, II, and III (see Appendix M). Initial sources of data (Phase I) included written and verbal conversations with Watershed staff, exploration of the *VTSB* course content and structure, and a review of the literature. Subsequent sources of data (Phases II & III) included participant interview transcripts

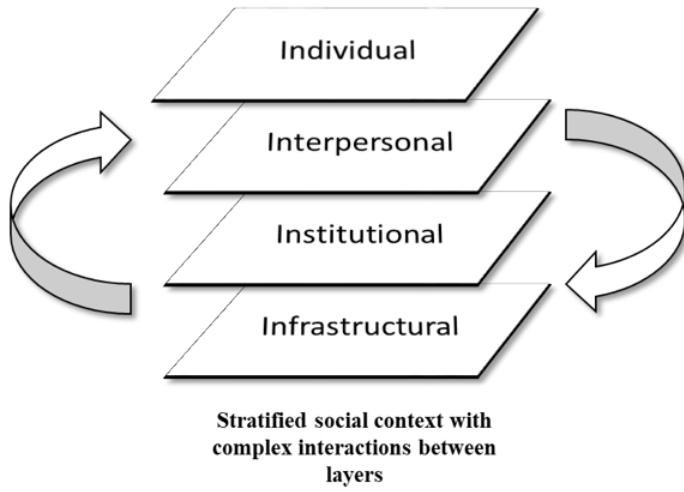


Figure 14. Stratified Social Context

and course annotations, course analytics, and materials from BCSD Open Data, such as the School Survey for Teachers, School Quality Reports, and various school websites.

I followed several of Pawson and Tilley's recommendations for developing my initial program theories (see Phase I in Figure 12), as described by Marchal, Kegels, & Van Belle (2019):

- Engage in discussions and document reviews to elicit the program designers' and implementers' "folk theories," or the assumptions and reasoning that affect their understanding of how the intervention works. For this step, I engaged in verbal and written communications with the Watershed staff and explored the VSTB course interface, instructional content, and supplemental resources.
- Review the literature to identify previous evaluations and studies that indicate what has worked for whom in what circumstances in similar interventions within relevant field(s). Identify any substantive theories, or "relevant theories used in the field in question or in other disciplines" (p. 85) within these publications that can help explain *why* or *how* mechanisms are triggered in certain contexts to generate certain outcomes. For this step, I conducted an initial search of the literature to identify factors most commonly associated with effective OPD, or OPD that successfully achieves its desired outcomes. I noted substantive theories that were evidenced in these publications. Next, I searched the literature to further explore these theories and determine whether they could help explain why or how the identified PD features were or were not effective in certain contexts.
- Synthesize data to generate program theories using CMO tables, which indicate the Context (C), Mechanism (M), and Outcome (O) components of the theories (see Figure 13 for an example).

All of these steps were iterative and became more targeted as I moved from developing initial program theories to testing, revising, refining, and finalizing them in Phases II and III.

Communication with Watershed. I maintained ongoing communications with the staff at Watershed Collaborative throughout this study. Data collected from these interactions includes notes from phone conversations, archived emails, and an agreed-upon project charter. My primary point of contact was the Operations Director, who answered my questions about the organization, the course platform, course content, and course analytics. She also facilitated the initial contact between myself and the course participants. At the outset, I spoke with the Executive, Creative, and Operations directors about my goals for the study, how Watershed would benefit from the study, and how we could work together to accomplish these goals. During these conversations, I learned about the history, vision, and trajectory of the organization, which helped me to construct a diagrammatic representation of Watershed's own program theory, or its "set of assumptions...that explain how and why they expect [the *VTSB* course] to reach its objectives and in which conditions" (Marchal, Kegels, & Van Belle, 2019, p. 83). This information helped me with complexity mapping and informed my generation of the initial program theories.

Exploration of *VTSB* course platform and resources. The Watershed staff enrolled me as a participant in the *VTSB* course, so I explored the course interface, organization, instructional content, and supplemental resources thoroughly at the outset of the study. I began by creating an outline of the course organization that listed and described the instructional components of each module, including the videos and their associated questions and reflection prompts. I listed the length of each video and the timestamps for each question and reflection prompt, so that I could later compare them to the timestamps of participants' written annotations. I also outlined the

assignments and supplemental resources for each module, which included PDFs and links to external websites and articles.

I then engaged in the course as a participant, navigating through each module, watching the instructional videos, and accessing the supplemental resources. I noted where question and reflection prompts were but did not make annotations myself. However, I took notes regarding content-related, structural, or didactic features that did or did not align with the effective features of OPD identified in my searches of the literature. For example, the literature indicates that effective OPD incorporates collaboration. Thus, I noted when I encountered assignments or statements in the videos encouraging collaborative participation. These findings informed my generation of the initial program theories.

Realist literature review. Realist reviews “enhance the interpretive power of a realist evaluation. Published accounts of similar programs, from similar or different contexts, may help [the researcher] to unearth program theories to compare with [her own program theories] ...and connect [them] to wider mid-range theories” (Booth, Wright, & Briscoe, 2019, p. 164). Realist reviews differ from systematic reviews in several ways, in that their searches of the literature are structured “according to emergent criteria that develop as theories are proposed, tested and refined”, they are “conducted iteratively throughout the review”, and they “seek to sample literature and attain modest forms of theoretical generalizability from evidence” (Booth, Wright, and Briscoe, 2019, p. 160). Thus, literature searches for realist reviews are not comprehensive and do not have specific in/exclusion criteria.

Specific guidelines for realist searches were lacking until Booth, Wright and Briscoe (2019) developed six stages of the realist search. These stages were derived from the six stages of a realist synthesis in Pawson’s *Evidence-Based Policy* (2006) and the *RAMESES Standards*

for Realist Reviews, which list criteria for an “excellent” realist search in Standards 7 and 8 (Wong et al., 2013). These criteria stipulate that the search be “driven by [the] objectives and focus of the review”, that the search does not restrict study/document types, that searches are iteratively conducted “for theory development, refinement, and testing as [the researcher’s] understanding of the topic increases”, and that searches are conducted in a variety of fields (Wong et al., 2013).

Each of Booth, Wright, and Briscoe’s six stages (2019) has a specific purpose and recommended types of search techniques and complementary search strategies. I aligned these six stages and search techniques with the phases of my study, as shown in Table 5. I elaborate on the steps taken within each stage in the following sections.

Stage 1: Formulate search questions. I first sought to identify the contextual factors most commonly associated with effective OPD or PD, which I defined as OPD or PD that successfully achieves its desired outcomes. Using the CIMO framework, my review question was: *What contextual factors (C) trigger mechanisms (M) that generate desirable outcomes (O) for a PD or OPD course (I)?* A secondary question was: *What substantive theories help explain why or how these contextual factors generate desirable outcomes for a PD or OPD course?*

Stage 2: Background search/scoping the literature. Pawson states that that are no specific guidelines for conducting a background search (2006), as the primary objective is for the researcher to familiarize herself with the relevance, quantity, and types of available sources so that she can begin defining boundaries for the scope of the review. I followed Booth, Wright, and Briscoe’s suggestion to conduct topic-based searches, which “combine words, phrases, and index terms (subject headings) and are used to retrieve a set of records of potential relevance to the search question” (2019, p. 155). I began by searching the Academic OneSearch database, using

Table 5

Six Stages of a Realist Search

Phase(s)	Search Stage	Purpose	Search Techniques & Complementary Strategies
I	Stage 1: Formulate search questions	Formulate review questions that will guide initial program theory searches.	CIMO (Context, intervention, mechanisms, outcomes) framework
I	Stage 2: Background search/scoping the literature	Get a “feel for” the literature, including the quantity, relevance, and types of potential sources; use to define initial boundaries for the scope of the review.	Topic-based searches
I & II	Stage 3: Search for program theories	Identify candidate program theories and substantive theories to be prioritized and refined for testing. Iterative; explore related literature as theories emerge.	Backward and forward chaining, journal runs, author searching
II & III	Stage 4: Search for empirical evidence	Identify research that can be used to test initial program theories and identify the most relevant substantive theories.	Targeted search terms from initial program theories
II & III	Stage 5: Final search to refine program theories	Link program theories across disciplines and to relevant substantive theories.	Search for specific substantive theories
III	Stage 6: Document and report realist research	Ensure review can be assessed.	Ensure alignment with standards 7 and 8 of the <i>RAMESES Standards for Realist Reviews</i>

Note. Adapted from “Scoping and searching to support realist approaches,” by A. Booth, J. Wright, and S. Briscoe, 2019, In N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, & Dalkin, S. (Eds.), *Doing realist research* (pp. 154). London: SAGE Publications. Copyright 2018 by A. Booth, J. Wright, and S. Briscoe.

combinations of terms (and their variants) such as professional development, online professional development, professional learning, teacher learning, evaluation, realist, education, effective, outcomes, context, elements, and factors. I also did Google searches to get an overview of relevant non-scholarly sources.

Based on the search results, I determined that there was a sufficient amount and variety of relevant literature to answer my review questions. I used the Paperpile reference management system to track my progress and organize my files in a series of folders, according to topic, field of study, and type of reference. I referenced the search results, as well my research questions and other sources of data, to help define the initial scope of the review. For example, while many sources addressed the connection between professional development and student performance, I did not have access to student performance data. Consequently, I did not retain these sources or add similar ones to my Paperpile folders during subsequent stages.

Stage 3: Search for program theories. For this stage, I began by reading and annotating some of the sources identified in the previous stage. In particular, I noted the contextual factors that affect PD, in what ways they affected PD, and substantive theories that helped explain why they affected PD in these ways. I used berry picking search techniques (Bates, 1989) throughout Stages 3,4, and 5 to find new sources. Berry picking is useful when conducting searches for topics that are not fully refined, including the identification of theories (Booth, Wright, & Briscoe, 2019). The six search techniques for berry picking include backward chaining, forward chaining, journal runs, area scanning, using subject-indexed databases, and author searching. In a realist review, initial searches begin with search techniques that reflect the researcher's needs for certain information; as the researcher encounters new information, her needs evolve, requiring new sets of techniques.

In this stage, I relied on backward and forward chaining, journal runs, and author searching to identify new sources. In every source, I noted in-text references that would expand on identified concepts or that introduced new concepts relevant to my review questions. I also scanned the reference section and highlighted sources to explore. If the source proved irrelevant, I removed the highlight; if the source was relevant, I changed the color of the highlight to indicate that I had located and scanned it. This helped me to avoid duplicating searches. For certain sources, I used a database's citation search tools to obtain a list of references citing the source. I browsed the list and selected the articles that appeared relevant for further exploration.

I also noted articles published in special editions of journals in which the theme of the edition was relevant to my queries. I scanned these editions and searched the journal's database for additional relevant articles. The journals that yielded the greatest number of articles included *Professional Development in Education* and *Teaching and Teacher Education*. I also found valuable resources by searching the websites of several professional and non-profit organizations, including the *International Association for K-12 Online Learning* (iNACOL), *North American Council for Online Learning* (NACOL), *itad*, the *RAMESES Projects*, and *Learning Forward*.

Throughout these search processes, I noted authors who were commonly cited and searched for additional works that they had authored, including past articles, conference papers, evaluation reports, and dissertations. I scanned these to determine which were relevant to my review questions. Additionally, I searched some authors' personal or professional websites for background on their perspectives as well as comprehensive lists of their publications and projects. This search technique was particularly valuable when exploring substantive theories.

I agree with Booth, Wright, and Briscoe that “it is tempting to select a large number of sources...but, if your realist search is to remain pragmatic, you must select a manageable and parsimonious set of resources” (2019, p. 159). Thus, when I started seeing the same information repeated in sources, with little new information being added, I ended my searches for this stage.

Throughout my searches, I used Microsoft OneNote to organize my findings in two tables. I created the first table with four columns, including: contextual factors that affect PD, how these factors can positively affect PD, how these factors can negatively affect PD, and excerpts from the sources from which this information was derived. I created a new row for each contextual factor; for example, I created one row for the factor “time.” I noted ways that time can positively or negatively affect PD according to the literature; for instance, short workshops were generally less effective than more sustained courses that took place over several weeks or months. I then went through each row and used different colors to highlight words or phrases that represented contexts (C), mechanisms (M), and outcomes (O). This helped me to generate my initial program theories using CMOCs.

I created a second table to track substantive theories. The table had a column that listed substantive theories and a column that included excerpts that addressed these theories. I added to and refined this table throughout subsequent stages.

Creating CMOCs and Initial Program Theories

I used the data from my communications with Watershed, my exploration of the *VTSB* course, and my review of the literature to generate Context-Mechanism-Outcome configurations, which comprised my initial program theories. This was an iterative process that included several rounds of synthesis and refinement. First, I created a table to list all of the contextual factors, mechanisms, and outcomes that I identified in my data. The contextual factors and outcomes

were derived from my communications with Watershed, my exploration of the *VTSB* course, and my review of the literature, while most of the mechanisms were derived from the literature review. I aligned each set of connected contextual factors, mechanisms, and outcomes (CMOC) in a new row, for a total of thirteen CMOCs.

I reduced and refined the CMOCs through several iterations of combining or eliminating some that were similar, reducing the number of variables in certain elements, and making the wording more succinct. Ultimately, I had nine CMOCs that contained the components of my initial program theories (See Table 9). I created another column in the table to list sources of data that I could use in Phase II to test the theories.

Phase II: Test & Revise Program Theories

In Phase II, I tested my program theories through further collection and analyses of data, including continued written and verbal communications with Watershed, continued exploration of the *VTSB* course organization and content, participants' annotations and responses to question/reflection prompts, course activity measurements, materials from the BC Schools database, and interviews with five participants. I continued my review of the literature throughout this phase, focusing on substantive theories that could help explain why or how the contextual factors identified in Phase I generate desirable outcomes for a PD or OPD course.

Course Activity

Course analytics for this LMS were limited, so I gathered data about participants' course activity through alternative means. Each course module has checkpoints for each component that the participant completes, including watching the videos, answering the questions, and viewing the assignment prompts and lists of supplemental sources, including PDFs and links to articles on external websites. When a participant completes one of these components, the checkpoint is

highlighted. The module states the percentage of the module that has been completed based on the number of highlighted checkpoints. I referred to the percentages as well as the number and type of highlighted checkpoints to gauge participants' progress.

There were several challenges with this method. First, there are checkpoints for some reflection prompts that do not require written responses. When a participant does not respond, the checkpoint is not highlighted, lowering the completion percentage for the module. As a result, the percentage listed does not accurately reflect the participant's level of completion. Second, the checkpoints for assignments and supplemental resources are highlighted if the participant looks at those components. However, this does not indicate the extent to which the participant completes the assignments or accesses the supplemental resources. Third, while the participant may receive a highlighted checkpoint for watching a video, this does not mean they were actively engaged in watching the video. The quality and number of annotations brings some insight to this problem but does not solve it. Participants may be fully engaged and not make any annotations. Conversely, they may make a couple of annotations but not be fully mentally present when watching the entire video.

Because of these challenges, I relied on realist interviews to fill in any gaps. I asked participants about their level of course completion, including watching the videos, completing assignments, and accessing resources. This provided a more complete and accurate representation of participants' course activity.

Annotations and Question/Reflection Responses

The *VTS Basics* course platform affords participants the ability to make annotations during the course videos and on assignment pages. Additionally, it introduces questions and reflection prompts throughout each module to keep participants engaged with the material.

Participant in groups that co-enroll can look at and respond to their peers' answers to questions and reflection prompts. Watershed's Creative Director also provided feedback to participants by commenting on their annotations or responding to their questions. Annotations or responses made during course videos are time stamped so that they can be matched with their corresponding place in the video, providing context for the comment.

I analyzed eight participants' annotations and responses to questions/reflection prompts. I copied and pasted their responses into a spreadsheet and organized them according to the place they occurred in the module. In addition to the response itself, I also recorded the date of each response and, if during a video, the time stamp of its occurrence. Recording this data helped me to gauge the pace at which participants completed each course component and provided context for comments made during videos, as I was able to watch what was said leading up to the time of their comments. If comments were made in response to another participant(s), I recorded those conversation threads to provide context for each comment in the thread.

This analysis provided information to help test program theories, as I could observe the content and timing of participants' individual and collaborative reactions to specific course components and to each other's insights (Pawson & Tilley, 2013; Dalkin, et al., 2015). I also used participants' responses to inform my interview questions. I coded this data following the same steps used for coding interviews, as outlined in the following section.

BCSD Resources

In 2015, the Big City Department of Education (BCDOE) adopted the *Framework of Great Schools* as the district-wide model for developing, evaluating, and improving school quality. The framework is comprised of six "essential elements of improving schools" (Fariña, 2015, p. 1), including: rigorous instruction, a supportive environment, collaborative teachers,

effective school leadership, strong family-community ties, and trust. The DOE provides teachers and administrators many resources for applying and evaluating these elements. Tools used to evaluate schools' application of the framework include annual School Surveys, School Quality Reports, and School Quality Reviews. These resources are available to the public on BCDOE's site for open data.

I explored BCDOE site for information that would help inform my interview questions and revise my program theories. In accordance with complexity theory, the intervention (*VTSB* course) is implemented within individual, interpersonal, institutional, and infrastructural layers of complexity that reciprocally influence each other. The BCDOE website provides information about the institutional and infrastructural layers of complexity, which provides context for the individual and interpersonal layers of complexity, in which teachers are individually and collaboratively engaging in the *VTSB* course. Thus, I gathered data that could provide contextual information about each participant's school of employment and includes their school's 2018 Survey for Teachers results, Quality Report, demographics, and website.

Survey for teachers. BCSD conducts annual surveys for teachers, students, and parents. The survey questions are categorized according to the six elements from the *Framework for Great Schools*. Schools use survey results to identify areas for improvement. I examined the list of questions for the 2017-2018 School Survey for Teachers and labeled them based on their relevance to specific program theories (e.g., PT 1 as a label if the question related to program theory #1). I then downloaded the spreadsheet provided by BCDOE with the survey results for all BCSD and filtered for the schools that employed study participants. I eliminated questions that were not relevant to any program theories and tagged the remaining questions with the labels identified earlier. The survey results helped inform my creation of interview questions and

revision of program theories. For example, the teachers from one of the participant's schools scored the level of teacher collaboration very highly. In my interview with this participant, I asked questions about her collaboration with others to complete the *VTSB* course and how this affected her experiences. Her responses helped inform program theories related to collaboration and PD outcomes.

Quality reports. Schools are expected to set annual goals based on their performance in the six categories of the framework, as recorded in their School Quality Reports from the previous year. I looked at each school's 2017-2018 quality reports to identify areas of strength and areas for improvement, as these results would affect the school's areas of focus for the 2018-2019 school year. This provided context for my interview questions and program theories. For example, one participant's school scored as "fair" in effective school leadership, so I asked her questions about the leadership's involvement in PD and how it affected her experience with *VTSB*. Her responses helped inform program theories related to leadership's involvement and PD outcomes.

Demographics. The BCDOE provides detailed demographic information for all of its schools, including the student body's size, racial and ethnic composition, percentage of English Language Learner (ELL) students, and percentage of economically disadvantaged students. I created a table listing the data for each of these categories for each of the participant's schools. This information helped to provide an overview of the school's diversity, as well as potential challenges that teachers might face that could affect their ability to engage in PD, their motivation for engaging in PD, or their interests in specific types of PD. For example, one participant's school had over twice the amount of ELL students as another participant's school. She was interested in using VTS to better engage ELL students, as VTS is centered around

images. Knowing this information helped inform program theories related to teacher's motivations for engaging in and applying VTS in the classroom.

Websites. I also explored those websites of participant's schools that were operational. In general, the websites stated the school's mission or vision, values, schedule, and curricular information. Along with the other BCDOE information, exploring these sites supplemented my understanding of a school's culture, which was a component of several program theories.

Realist Interviews

Realist interviews are a vital part of the testing and refinement process for the researcher's initial program theories. These interviews are unique in that they are theory-driven and retroductive, using both deductive and inductive elements to test hypotheses. Thus, "the subject matter of the interview is the researcher's theory and interviewees are there to confirm, falsify and, basically, refine the theory" (Manzano, 2016, p. 343-344). During this phase, I conducted six interviews with five of the individuals recruited from the pool of BCSD K-12 teachers who enrolled in the 2018-2019 *VTS Basics* OPD course. I used these interviews to elicit information from study participants that would test the program theories selected in Phase I.

I designed the interview questions and conducted the interviews to support the objective of testing and revising my program theories. I referred to these theories, as well as recommended realist interview question stems (Manzano, 2016; Greenhalgh et al, 2017h), to create an initial list of questions and prompts (see Appendix L for examples). Prior to the first interview, I systematically reviewed my list of questions and prompts to ensure that they would elicit responses relevant to the elements of each program theory. Nevertheless, the question list served as a guide, not a rigid protocol (Manzano, 2016). I adapted the content, order, and amount of questions during the interviews, based on participant responses. I was familiar enough with my

program theories to ensure that I would still elicit responses about each element, even if I altered the questions.

In a realist interview, the researcher “controls the direction of the conversation... [and engages in] assisted sensemaking” (Manzano, 2016, p. 351). Consequently, neither the interviewer nor the questions are neutral. During my interviews, I steered participants towards specific topics and wove their responses into program theory statements. For example, one of the program theories hypothesized that teachers incorporate VTS into their practice when they experience dissonance between their current practice and internalized images of "perfect teachers/teaching" and identify strategies within *VTSB* to help bridge the gap between their current practice and these internalized images. To test this theory, I first asked participants if they had any instructional practices they were trying to improve and noted their responses. Throughout the interview, I noted if they discussed their development these instructional practices in the context of their experiences in the course and application of course content. I then made a statement, such as “So, by engaging in _____ (course activity), you improved your ability to _____ (instructional strategy)” and asked if they agreed with it. This helped to confirm or falsify my theory. I followed up on unexpected responses or responses that were not covered in my program theories but that could be used to transform or generate new theories.

Literature Review

I continued my review of the literature throughout this phase, changing my search terms or exploring new topics based on the data collected from other sources. During this phase, I used Stages 4 and 5 of Booth, Wright, and Briscoe’s six stages of realist searches (2019).

Stage 4: Search for empirical evidence. The objective of this stage is to “aim for the point at which additional evidence does not add to, or contradict, evidence already identified”

(Booth, Wright, and Briscoe, 2019, p. 162). However, it is also important for the researcher “to sample the literature rather than identify an exhaustive body of literature, and [to] privilege the specific over the comprehensive” (p. 162). Thus, I created targeted search terms following Booth, Wright, and Briscoe’s (2019) recommendation to combine two or more of the following elements from the initial and refined versions of program theories:

1. A context or population group (e.g., K-12 teachers);
2. A mechanism (e.g., increased self-efficacy);
3. A phenomenon of interest (e.g., engagement in professional development).

I used these searches to identify and explore potential substantive theories that could help explain why certain outcomes occurred as the result of underlying mechanisms based in social science theory. Through an iterative process of reading, annotating, and comparing with incoming data, I narrowed my selection of substantive theories, which I finalized in Stage 5.

Stage 5: Final search to refine program theories. This stage “[includes] searches for specific, named [substantive] theories” (Booth, Wright, & Briscoe, 2019, p. 162) identified in Stage 4. The researcher evaluates and finalizes the selection of substantive theories during this stage. Accordingly, I conducted searches using the substantive theories identified in Stage 4 and finalized my selection after reading, annotating, and evaluating sources for theories that would best explain and articulate the relationships between program theory components.

Westhorp describes how concepts from complexity theory can be used to help select and analyze substantive theories that are “complexity-consistent.” Complexity-consistent substantive theories “reflect key tenets of complexity theory” (2012, p. 405), among which are the embeddedness of systems within larger systems; generative, contingent causation; non-linearity; emergence; and feedback. Considering the complex nature of the intervention and the context in

which it was introduced, I selected substantive theories that were complexity-consistent. These theories are discussed in detail in Chapter 4.

As established, from a RE perspective, the *VTSB* course is a complex intervention embedded in a stratified social context, creating horizontal and vertical layers of complexity (Pawson & Tilley, 1998a). Thus, Westhorp also contends that the relationships between elements comprising these layers can be understood by layering complexity-consistent substantive theories “to reflect multiple levels of systems” (2012, p. 405). I created a table to establish and visualize the layered relationships between substantive theory and program theory elements (Westhorp, 2012), with complexity theory serving as the “base layer.” Again, the layering of these theories is discussed in detail in Chapter 4.

Coding

I used theory-driven qualitative coding techniques to analyze interview transcripts and participants’ written annotations and question responses. I began by creating a priori codes based on the table created in Stage 5 of the literature search. The hierarchical, layered structure of the table translated nicely into a hierarchy of parent-child codes that I used to create a codebook.

Using Nvivo, I coded the interview and written data deductively using the pre-established codebook. However, as this was a retroductive process, I also relied on inductive coding to revise, add, or eliminate the existing codes.

Codebook

I analyzed the interview transcripts, question/reflection prompt responses, and course annotations using Nvivo and a codebook with a priori cases and codes identified from the literature and initial data. My codebook initially consisted of 23 parent codes and 170 child

codes. I used this codebook for my initial coding of the interviews, after which I eliminated codes that had not been used and added codes that were not in the original codebook but evidenced in the interviews. The final codebook contained 18 parent codes and 77 child codes (see Table 6).

I organized my codes according to the layers of complexity—consistent substantive theories. The base layer corresponds to the RE levels of complexity: Individual/intrapersonal, individual/*VTSB* course, interpersonal, institutional, and infrastructural. Next, each of these is broken down into constructs from substantive theories, other concepts identified in the literature review, or my analysis of *VTSB* Course materials, BCSD materials, and conversations with Watershed Collaborative staff.

Testing and Revising Program Theories

I tested the initial program theories by mapping the C, M, O components of each initial PT onto corresponding codes in the codebook (e.g., "motivation to learn" was labeled with PT 1 (C, M, O); PT 5 (O)). Through this process, I was able to test the relationships between the initial program theories and the data from Phase II. I used this, as well as the analyzed data from Phases I & II, to revise the nine initial program theories, resulting in five new theories.

Phase III: Refine and Finalize Program Theories

In Phase III, I refined my theories from Phase II to create the final program theories, which I derived from the finalized CMO configurations (RAMESES Project Team, 2017). I maintained written and verbal communications with Watershed and continued my review of the literature from the previous two phases, focusing on complexity-consistent substantive theories and their relationship to the program theory elements. Lastly, I derived middle-range theories from elements of the finalized program theories.

Table 6

Codebook Codes

COMPLEXITY LAYER	PARENT CODES	CODES
Individual (Intrapersonal)	Ontological Beliefs	Reflection General/prior knowledge Causal attribution beliefs Pedagogy Content knowledge
	Epistemological Beliefs	Learning philosophy
	Purpose & Goals	Motivation to learn General goals Specific goals Willingness to learn
	Self-Perceptions & Definitions	Self-efficacy Pedagogical discontentment Self-perceived abilities
	Perceived Action Possibilities	Improved instructional practices Achieving instructional goals Enhanced curriculum Positive student responses Recruitment of other teachers Increased school evaluation scores Achieving department/school goals
	Emotion (a sub-aspect of each component and of their relations)	Enthusiasm Overwhelmed
	Situational Factors	Class time Diversity ELL Other duties/official roles
	Role identities	Teacher Learner Employee
	Role formation/Change	Harmony /Alignment Integration Tension/instability

Table 6. Continued

COMPLEXITY LAYER	PARENT CODES	CODES
Individual (<i>VTSB</i> course related)	Structural	Time Flexibility Collaboration Delivery format Interface Modules Videos Annotations Assignments Question/reflection prompts
	Didactic	Active Learning Reflection
	Content-Related	Relevance Coherence
Interpersonal	Leadership relationships	Support Communication Grants autonomy Expectations Accountability
	Peer relationships	Collaboration Mentoring Communication
Institutional School climate	Academic	Shared/Collective beliefs & goals Instructional strategies Goals/vision Instructional quality Expectations Accountability Leadership Promotion of PD Communication
	Social	Organizational structure Collaboration Distributed leadership Social climate Community/connectedness Efficacy Communication Diversity Ethnic/racial diversity ELL

Table 6 .Continued

COMPLEXITY LAYER	PARENT CODES	CODES
	Physical	Resources Time/scheduling
Infrastructural	Accountability	PD requirements Evaluations

Literature Review

I continued with Stage 5 (*Final Search to Refine Program Theories*) of the realist search; however, these searches were less frequent and more targeted than in the previous phase. I searched for additional sources about specific substantive theories and program theory elements if I needed more information about them to finalize a program theory. For the final stage, Stage 6 (*Document and Report Realist Research*), I ensured that my execution of the six search stages satisfied the criteria for realist searches from Standards 7 and 8 of the *RAMESES Standards for Realist Reviews* (Wong et al., 2013).

Refining and Finalizing Program Theories

I repeated the CMO mapping process from Phase II with the five revised theories. Lastly, I reviewed my data and analyses from Phases I and II and compared them with my program theories to identify elements for revision, elimination, or elaboration. Once I could no longer identify data that contradicted or added to the program theories, I considered them finalized (see Table 14).

Middle-Range Theories

I derived several middle-range theories from the elements of the finalized theories. These theories can be transferred to programs operating in similar contexts or with similar interventions, as they are neither universal nor specific to a single context (Pawson & Tilley, 1997; King et al., 2016; Astbury, 2013).

Saturation

RE research approaches use Pawson's "test of saturation," as well as a consideration of practical constraints, to determine theoretical saturation (Pawson et al., 2005; Weetman et al., 2017). As the research approach itself is iterative and occurs in multiple stages or cycles, so is

the question of saturation. Thus, during each phase, I considered whether the collection and analysis of additional data was further informing the refinement of my program theories. At the point in each phase where I determined that saturation had been reached, I moved to the next phase.

In RE studies, saturation is also affected by time constraints and the number and availability of consenting participants. Pawson and Tilley state that RE researchers are “pragmatic where data sources or resources are limited” (2004, p. 23). Thus, my access to data was limited by the number of teachers who consented to engage in the study, their time constraints, and the timeline for completing this study.

Trustworthiness

My experiences as a former K-12 educator, support of VTS, and interest in furthering arts-based, inquiry-driven pedagogies could have potentially affected the way that I conducted the study and analyzed and interpreted the data I collected. Because of my background, I identified with some of the goals and constraints of the teachers I interviewed, including:

- A desire to cultivate students’ creative and critical thinking skills,
- A desire to incorporate student-centered, inquiry-based activities in the classroom,
- A desire to increase students’ access to and interaction with the arts,
- Pressure to complete prescribed amounts of the curriculum within a rigid timeframe,
- Pressure to increase the school’s performance ratings, and
- Difficulty collaborating with colleagues due to multiple commitments and time constraints.

As I empathized with the participants, I hoped the *VTS Basics* course would help them to achieve their goals and mitigate constraints. My desire for the *VTS Basics* course to be successful and for participants to find favorable might have influenced the interview questions I asked or how I observed and perceived the interviewees' responses. I might have focused more on positive responses than on negative ones.

I also might have assumptions based on my prior experiences in the field of education that would cause me to misconstrue or overlook various contextual elements. For example, I might unintentionally ascribe incorrect characteristics to the principal at a participant's school if he or she shared traits with the principal at my previous school.

However, I took certain steps to minimize the effects of my biases. I practiced introspection and reflexivity to keep myself accountable for recognizing and acknowledging these biases. Taking reflective memos, sharing my research processes with peers, and including multiple sources of data helped me identify and maintain an awareness of my biases in relation to the study (Flick, 2007).

CHAPTER 4: RESULTS AND DISCUSSION

The purpose of this study was to determine the extent to which the desired outcomes for select K-12 teachers' participation in the *VTSB* OPD course were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. The findings of this study are presented in this chapter and address the following questions:

What works?

- Which of the desired outcomes (e.g., teachers develop the ability to adapt VTS for diverse groups of students, teachers implement VTS in the classroom) are achieved and to what extent?
- What are the unintended outcomes of the course?

For whom?

- To what extent do participants/sub-groups of participants (e.g., teachers in different locations or who completed different stages of the course) benefit from this course, as indicated by achieving desired or unintended outcomes?

In what circumstances?

- What contextual factors influence the achievement of desired outcomes and to what extent (e.g., level of administrative support, relevance to instructional goals)?

How?

- What types of participant reactions/responses do course resources (e.g., course interface, structure) and activities (e.g., course annotations, reflection, applying content) elicit? To what extent do these responses relate to the achievement of desired outcomes?

This chapter summarizes the data collected and analyzed during the three phases of the study, which involved developing program theories (Phase I), testing and revising program theories (Phase II), and refining and finalizing program theories (Phase III). Data included written and verbal conversations with Watershed staff, exploration of the *VTSB* course content and structure, an ongoing review of the literature, participant interview transcripts and course annotations, course analytics, and materials from BCSD Open Data. The review of the literature was conducted using a realist review approach and realist search techniques (Pawson, 2006; Booth, Wright, & Briscoe, 2019). All of the data was coded using theory-driven qualitative coding techniques, including the use of a priori codes derived from substantive theories. Ultimately, six program theories were finalized, from which general conclusions are drawn in Chapter 5.

Phase One Overview

In Phase I, I collected and analyzed data from written and verbal conversations with Watershed staff, exploration of the *VTSB* course content and structure, and a review of the literature to identify contextual factors, mechanisms, and outcomes in the data. The contextual factors and outcomes were derived from my communications with Watershed, my exploration of the *VTSB* course, and my review of the literature, while most of the mechanisms were derived from the literature review. I used this data to generate thirteen Context-Mechanism-Outcome configurations (CMOCs), which I reduced to nine through an iterative process of synthesis and refinement. These nine CMOCs served as my initial program theories (see Table 9).

Review of the Literature

I reviewed the literature using the realist review and realist search techniques described in Chapter 3 (Pawson, 2006; Booth, Wright, & Briscoe, 2019). The review was conducted to

identify factors most commonly associated with effective OPD or PD, which I defined as OPD or PD that successfully achieves its desired outcomes. The results helped answer the following two review questions, which informed the generation of program theories:

- *What contextual factors (C) trigger mechanisms (M) that generate desirable outcomes (O) for a PD or OPD course (I)?*
- *What substantive theories help explain why or how these contextual factors generate desirable outcomes for a PD or OPD course?*

I will now discuss my findings from Phase I that are relevant to these questions and informed the generation of my initial program theories.

Question #1: What contextual factors (C) trigger mechanisms (M) that generate desirable outcomes (O) for a PD or OPD course (I)?

To answer the first question, I identified contextual factors that affect PD engagement (see Table 7), both positively and negatively. I grouped these factors into three categories based on the premise that the *VTSB* course is embedded in a stratified social context with horizontal and vertical layers of complexity (Pawson & Tilley, 1998a). The categories represent individual, school, and PD characteristics layers (see Figure 14).

Individual. The individual category includes three contextual factors that affect PD engagement: teacher identity, teacher experiences, beliefs, and attitudes, and teacher motivation to learn.

Teacher identity. Kelly (2006) defines teacher identity as “the ways in which practitioners see themselves in response to the actions of others towards them...Facets of such constructions include how teachers interpret their role, the meanings and understandings which they bring to their role, their beliefs and intentions, and so on” (p. 513). These facets affect

teachers' engagement in PD in several ways. According to Day and Sachs (2004), when teachers have a clear professional identity, they are more intrinsically motivated and confident in their ability to improve their practice. When paired with pedagogical discontentment, this increased motivation and self-efficacy spurs teachers to adopt ideas and practices from PD that will help them develop in the areas they hope to improve (Appova & Arbaugh, 2018). However, in some instances, teachers who are confident in their practice and grounded in their professional identity may be resistant to ideas that they perceive as challenging their identity (Kelly, 2006; Webster-Wright, 2009).

Teacher experiences, beliefs, and attitudes. While teachers' experiences, beliefs, and attitudes are parts of their identity, they were treated in the literature as distinct aspects with distinct influences, warranting a separate category. Teachers' beliefs about teaching and learning are shaped by their cumulative experiences, including those as former students, as teachers in the classroom, as employees in the school, and professional learners in PD courses (Opfer & Pedder, 2011). As Opfer and Pedder (2011) state, this "intersection of experience and belief creates a powerful combination that determines not only the instructional decisions that teachers make but also, we would argue, what they themselves are willing to learn" (p. 2). Thus, teachers are often most likely to engage in PD courses that align with their experiences, beliefs, and attitudes about teaching and learning (Evans 2008, 2011). However, PD that incorporates active application of instructional strategies can change teachers' beliefs and attitudes if the new practices elicit positive student responses and improved performance (Lipowsky & Rzejak, 2014).

Teacher motivation to learn. Appova and Arbaugh identify six primary reasons that teachers are motivated to learn, including: to influence students and their learning, to learn with/from other teachers, to become a "better" teacher, to fulfill PD requirements, to constantly

seek and engage in learning as a “habit”, to gain knowledge about topics of their own interests, and to pursue further learning if funds, time, and resources are available (2019, p. 10). Thus, teachers are most likely to engage in PD that they perceive will satisfy these desires. As addressed in the previous two categories, teachers’ motivation to learn is also affected by their experiences, beliefs, and attitudes towards teaching and learning (Opfer & Pedder, 2011; Evans, 2008, 2011).

Interpersonal and institutional. The interpersonal and institutional category includes six contextual factors that affect PD engagement: collaboration, supportive leadership, resource, school culture, accountability, and agency/autonomy.

Collaboration. Collaboration is frequently identified as a factor that positively affects the amount, duration, and quality of teacher learning (Desimone, 2009; Appova & Arbaugh, 2019; Labone & Long, 2016), although Opfer and Pedder (2011) stipulate that “how much collaboration is necessary for teachers to learn will [vary]” (p. 386). In particular, professional learning communities facilitate collaborative participation in PD, as teachers are working towards common goals and are more publicly accountable for their progress (Fraser et al., 2007; Mitchell, 2013; Desimone & Garet, 2015). Leadership can also facilitate or inhibit collaborative PD through their expectations, dedication of resources, and level of accommodation affect teachers’ motivation and ability to engage in PD together (Drago-Severson, 2012; Day & Sachs, 2004).

Supportive leadership. School leadership “plays a key role in supporting and encouraging teachers to implement in the classroom the ideas and strategies they learned in the PD” (Desimone & Garet, 2015, p. 252). By showing interest in and monitoring teachers’ PD participation, allocating sufficient time and resources for teachers to engage in PD, granting

autonomy in PD selection, and maintaining a positive school culture, leadership creates an environment that facilitates teacher learning. However, establishing overly bureaucratic working environments, showing minimal interest in teacher PD progress, limiting allocation of time and resources, and making unilateral decisions about PD selection creates an environment that inhibits teacher learning (Mitchell, 2013; Lipowsky & Rzejak, 2014; Drago-Severson, 2012).

Resources. Engagement in PD requires time, funding, and a physical location. Often, teachers are not given additional time or space to participate in PD; they have to “find” the time before and after school, during their planning periods, and during their personal time, which constrains their completion of PD courses (Appova & Arbaugh, 2019). Granting more release time from class, creating spaces dedicated to teacher learning, and offering stipends “[provides] opportunities for teachers to pursue individual learning, including the PD opportunities available outside the district, and [helps] to initiate small-group collaborative teacher learning before and after school hours” (Appova & Arbaugh, 2019, p. 16). When teachers are less concerned about finding time, space, or funding, they can focus more on engaging in PD and implementing new practices in the classroom (Lipowsky & Rzejak, 2014).

School culture. School culture is one of the most frequently and extensively discussed factors that affects teachers’ engagement in PD. Opfer and Pedder discuss the relationship between the institutional and individual layers of complexity, stating that, “although an individual teacher’s [experiences, beliefs, and attitudes] may lead him or her to participate in professional learning activities, the access, support, and encouragement to participate are heavily determined by the school” (2011, p. 393). Schools that develop a culture of learning and inquiry, advocate student-centered practices, encourage reflective practice, communicate shared goals and expectations, and facilitate teacher collaboration increase teachers’ motivation to engage in

PD (Opfer & Pedder, 2011; Day & Sachs, 2004; Kelly, 2006). Conversely, schools that emphasize performance ratings, implement transmissive instructional practices, lack a shared vision, and promote teacher isolation decrease teachers' motivation to engage in PD.

Accountability. PD requirements keep teachers accountable for completing a certain number of PD hours within a specified time frame, which motivates teachers to engage in PD despite other demands on their time. While this aspect of accountability can be helpful, it can also be detrimental if more focus is placed on the quantity and not the quality of PD. Consequently, “accountability and PD requirements are helpful for motivating [teachers] to continue engaging in PD; however, it is the type of PD which is required that is disappointing and not motivating to the teachers to seek high-quality learning experiences” (Appova & Arbaugh, 2019, p. 13). Leadership's involvement in monitoring and evaluating PD also serves as a form of accountability; high expectations coupled with generous allocation of time and resources facilitates teacher professional learning (Mitchell, 2013; Lipowsky & Rzejak, 2014). Additionally, teachers keep each other accountable when sharing a common vision and working to achieve collective goals (Appova & Arbaugh, 2009; Opfer & Pedder, 2011; Drago-Severson, 2012).

Agency/autonomy. Teachers are more likely to adopt PD ideas and practices if they have input into their PD selection, engagement, and application (Kelly, 2006; Kennedy, 2014). When teachers feel that their input is valued and respected, their self-efficacy and motivation to improve their practice increases (Fraser et al., 2007); they are more invested in applying PD than they would be if they were “assigned” a PD course without being consulted. Teachers also need to exercise agency by having some control over when, where, and how they engage in PD (Kennedy, 2014).

PD characteristics. The PD characteristics category includes four contextual factors that affect PD engagement: active learning, time/flexibility, reflection/inquiry, and content focus/coherence.

Active learning. Active learning through the authentic application of content was often identified as one of the most influential factors of PD engagement (Desimone & Garet, 2015; Opfer & Pedder, 2011; Webster-Wright, 2009; Appova & Arbaugh, 2018). Opfer and Pedder found that “teachers learn most effectively when activities require them to engage with materials of practice, when activity is school based and integrated into the daily work of teachers, and when the pedagogy of professional development is active and requires teachers to learn in ways that reflect how they should teach pupils” (2011, p. 385). Additionally, engaging in active application of PD can change teachers’ attitudes and beliefs about new ideas and practices more effectively than receiving information through transmissive methods of instruction (Desimone & Garet, 2015; Opfer & Pedder, 2011).

Time/flexibility. A lack of time is one of the greatest constraints on teachers’ ability to engage in PD. Teachers need enough time to engage in PD that is “sustained and intensive rather than brief and sporadic” (Opfer & Pedder, 2011, p. 384), as ideas take time to process and skills take time to develop. Often, teachers’ engagement in PD is limited by many demands on their time and rigid scheduling. Alleviating some of these demands and providing more opportunities for teachers to participate in PD during their official working hours through flexible scheduling can increase their engagement. Furthermore, dedicating more time to PD affords more time for active practice and reflection on practice (Lipowsky & Rzejak, 2014; Appova & Arbaugh, 2018).

Reflection/inquiry. Ongoing reflection is an integral part of teachers’ professional learning and facilitate change in teachers’ practice (Keay et al., 2019). Through reflection,

teachers identify areas for improvement and question assumptions, leading to transformative change; Webster-Wright describe transformative change as a means by which “learners may conceive aspects of their world or themselves differently, allowing them to reinterpret experience from a new perspective and act to change situations” (2009, p. 722). Thus, PD that incorporates active learning and reflection is more likely to change teachers’ beliefs and result in lasting change (Lipowsky & Rzejak, 2014; Keay et al., 2019; Opfer & Pedder, 2011).

Content focus/coherence. Teachers are more likely to engage in PD in which the “content, goals, and activities...are consistent with the school curriculum and goals, teacher knowledge and beliefs, the needs of students, and school, district, and state reforms and policies” (Desimone & Garet, 2015, p. 253). When PD aligns with these factors, teachers perceive it as relevant and meaningful to their practice and are more motivated to participate (Lipowsky & Rzejak, 2014; Appova & Arbaugh, 2018; Lindvall & Ryve, 2019). Conversely, if teachers do not perceive PD as relevant to their practice, they are less motivated to engage in the PD.

Question #2: What substantive theories help explain why or how these contextual factors generate desirable outcomes for a PD or OPD course?

To begin answering this question, I identified substantive theories that help explain why the previously discussed contextual factors generate desirable outcomes for a PD or OPD course. In this phase, I identified the theories cited in the literature I reviewed for question 1. I tracked these theories in a table that listed the theories in one column and included excerpts that addressed these theories in an adjacent column. Theories and theoretical concepts included: Self-efficacy, possible selves, pedagogical (dis)contentment, identity theory (teacher role identity, professional, personal), teacher learning theories (social learning theories, constructivism), and

Table 7

Contextual Factors That Affect PD Engagement

Category	Contextual Factors
Individual	• Teacher Identity • Experiences, beliefs, and attitudes • Motivation to learn
School	• Collaboration • Supportive leadership • Resources • School culture • Accountability • Agency/autonomy
PD Participation	• Active learning • Time/flexibility • Reflection/inquiry • Content focus/coherence

professionalism/professional learning theories. In Phase II, I narrowed down the list of relevant theories to research in greater depth and used them to help revise my initial program theories.

Watershed Communications and Exploration of *VTSB* Course Platform and Resources

Through ongoing communications with the staff at Watershed Collaborative, particularly the Operations Director, I learned about the history, vision, and trajectory of the organization, which helped me to construct a diagrammatic representation of Watershed's own program theory (see Appendix N). I also explored the course interface, organization, instructional content, and supplemental resources. I compared the theory of change and course syllabus with the factors of effective PD that I identified in the literature review to determine which factors were present and how they were manifested in the course (see Table 8). Factors that were evidenced in the course include: active learning, time/flexibility, collaboration, reflection/inquiry, and content focus/coherence.

Active Learning and Reflection/Inquiry

The *VTSB* course modules consist of instructional or example videos and assignments requiring active application of content. Throughout the modules, teachers are prompted to respond to embedded question and reflection prompts. Thus, they are engaged in cycles of active learning and reflection.

Content Focus/Coherence

The *VTSB* course was designed to align with CCSS and to be easily integrated into curricula for various grade levels and subjects. The course provides documents that detail how VTS aligns with CCSS and prompts teachers to develop their own *Integrated Practice Plan*, in which they outline ways to integrate their learning about VTS, implementation of VTS, and reflection on VTS.

Time/Flexibility

Watershed designed the *VTSB* course to accommodate teachers' demanding and unpredictable schedules. Teachers can complete the course in anywhere from ten weeks to the entire school year. Watershed describes how to adapt the course activities to fit different completion schedules. Additionally, the course includes example videos of VTS sessions for various grade levels and subjects. Watershed encourages teachers to watch as many of these videos as they can, but to watch only the videos most relevant to their grade level/subject area if they are limited on time.

Collaboration

While the *VTSB* course can be completed individually, ideally it is completed collaboratively. To facilitate collaborative engagement, Watershed offers all participants a free course subscription to enroll two partners of their choosing. They also incorporate assignments that include peer observation, discussion, and reflection.

Generating Initial Program Theories

After gathering, analyzing, and synthesizing the data from Phase I, I created thirteen CMOCs. I based the C, M, O elements and the relationships between them on the PD factors identified in the literature review and the overlap between these factors and the *VTSB* course. I reduced and refined the CMOCs from thirteen to nine through several iterations of combining or eliminating some that were similar, reducing the number of variables in certain elements, and making the wording more succinct. Ultimately, I had nine CMOCs that contained the components of my initial program theories.

Table 8

Alignment between Contextual Factors That Affect PD Engagement and VTSB Course Content and Objectives

Factor	VTSB Course Content	Learning Objectives
Active learning	• Conduct VTS sessions in the classroom (Weeks 3-10) • Observe and be observed by a peer conducting VTS session (Week 6) • Implement VTS based on <i>Integrated Practice Plan</i> (Weeks 8 & 9) • Record video of VTS session in the classroom (Week 10) • Watch the video and complete <i>Observation Rubric</i> (Week 10)	• Gain greater understanding of, comfort with, and skill at facilitating VTS with students. • Gain greater comfort paraphrasing student comments.
Reflection/ Inquiry	• Respond to reflection prompts and questions in each module • Reflect on VTS implementation experiences individually and with a peer • Complete <i>Observation Rubric</i> during example videos, peer observations, and personal assessments • Watch videos and read articles on reflective practice (Week 6) • Complete <i>Reflective Practice Assessment</i> to develop a plan for cultivating a more reflective teaching practice with/through VTS (Week 6) • Read articles about reflective practice (Week 6) • Watch recording of VTS session and complete <i>Observation Rubric</i> (Week 10) • Complete Readiness Assessment before first VTS session and after final VTS session (Weeks 3 & 10)	• Prepare yourself for setting the appropriate expectations and environment for students and facilitating VTS in the classroom. • Develop understanding of the nature and potential benefits of developing a reflective practice. • Deepen understanding of deliberate practice and peer collaboration, and their potential impact on learning outcomes. • Identify strengths and areas for improvement in your VTS practice.

Table 8. Continued

Factor	VTSB Course Content	Learning Objectives
Content focus	• Develop and implement <i>Integrated Practice Plan</i> (Weeks 7-9) • Read <i>VTS and Anchor Standards</i> • Review VTS and state standards matrix. How can VTS practice help you meet subject-area and grade-level standards in your classroom? (Week 9)	• Deepen understanding of the relationship between VTS Elements, the quality in which they are employed, and the quality of discussion and learning produced. • Gain basic understanding of core concepts behind and capacities promoted through VTS: visual and critical thinking, evidential reasoning, and meaning-making. • Deepen understanding of inquiry, constructivism, and teaching for understanding, and their potential impact on learning outcomes.
Coherence	• Develop and implement <i>Integrated Practice Plan</i> (Weeks 7-9) • Read <i>VTS and Anchor Standards</i> • Review VTS and state standards matrix. How can VTS practice help you meet subject-area and grade-level standards in your classroom? (Week 9)	• Develop initial understanding of the relationship between visual literacy, VTS, literacy, and standards. • Deepen understanding of how VTS aligns with state standards and supports students' 'learning how to learn.' • When educators implement VTS, they support essential learning objectives, align instruction with standards, and complement school-wide priorities.
Collaboration	• Invite partners to share course enrollment for free (3-4-1 pricing) • Read collaboration articles (Week 2) • Introduce and explain VTS to a colleague, friend, or family member (Week 5) • Observe and be observed by a peer conducting VTS session (Week 6) • Reflect on VTS implementation experiences with a peer	• Watershed believes online professional learning works best when complemented by in-person collaboration and reflection with peers.

Table 8. Continued

Factor	<i>VTSB</i> Course Content	Learning Objectives
Time/ Flexibility	• Complete 60+ minutes of coursework and 30+ minutes of classroom instruction each week • Options for course completion ○ 10 weeks: 1 week for each step ○ Semester: 2 weeks for each step ○ School year: month for each step • Review Course Outline and create personal schedule for completing course content and implementing VTS in the classroom	• VTS is easy for time-constrained teachers to fit into existing instructional plans in part because implementation is flexible.

Table 9

Initial CMOCs and Program Theories

	#	CONTEXT	MECHANISM	OUTCOME
INDIVIDUAL	1	Teachers bring their prior experiences, knowledge, assumptions, and beliefs about teaching and learning to their selection of PD.	Teachers perceive <i>VTSB</i> content as meaningful and relevant and are motivated to engage in PD as a means of improving their practice and enhancing students' learning.	Teachers choose to enroll in <i>VTSB</i> as a PD course whose content and activities align with their prior knowledge, assumptions, and beliefs about teaching and learning.
		<i>When teachers bring their prior experiences, knowledge, assumptions, beliefs about teaching and learning to their selection of PD, they choose to enroll in VTSB, as they perceive its content to be meaningful and relevant.</i>		
	2	Teachers experience dissonance between their current practice and internalized images of "perfect teachers/teaching."	Teachers identify strategies within <i>VTSB</i> to help bridge the gap between their current practice and their internalized images of "perfect teachers/teaching."	Teachers incorporate identified strategies into their practice to bridge gap between their current state and desired levels of "perfection."
		<i>When teachers experience dissonance between their current practice and internalized images of "perfect teachers/teaching," they incorporate strategies into their practice that they identified within VTSB to help bridge the gap between their current practice and these internalized images.</i>		
SCHOOL	3	Leadership and teachers cultivate school-wide student-centered goals, norms, and practices.	Teachers perceive they are part of a team with shared goals, norms, and practices.	Teachers choose to engage in <i>VTSB</i> as a means of reinforcing student-centered goals, norms, and practices.
		<i>When leadership and teacher cultivate school-wide student-centered goals, norms, and practices, teachers perceive they are part of a team and choose to engage in VTSB as a means of reinforcing shared goals, norms, and practices.</i>		
	4	Leadership expresses personal interest in and monitor teachers' PD selections and progress	Teachers gain sense of accountability for applying VTS in the classroom.	Teachers treat <i>VTSB</i> as a priority.
		<i>When leadership expresses personal interest in and monitor teachers' PD selections and progress, teachers treat VTSB as a priority because they have a sense of accountability for applying course content in the classroom.</i>		
	5	Leadership encourages and accommodates teacher selection of and engagement in PD of their choice, including <i>VTSB</i> .	Teachers gain confidence in their decision to enroll in <i>VTSB</i> .	Teachers continually reflect on how VTS affect their practice and student learning
		<i>When leadership encourages and accommodates teacher selection of and engagement in PD of their choice, including VTSB, teachers gain confidence in their decision to enroll in VTSB and continually reflect on how VTS affect their practice and student learning.</i>		

Table 9. Continued

	#	CONTEXT	MECHANISM	OUTCOME
SCHOOL	6	Leadership provides sufficient physical resources to engage in <i>VTSB</i> in addition to meeting teachers' basic professional needs.	Teachers are "freed up" to concentrate on practice and not logistics.	Teachers invest time and effort into completing <i>VTSB</i> course and implementing VTS in the classroom.
		<i>When leadership provides sufficient physical resources to engage in VTSB, in addition to meeting teachers' basic professional needs, teachers are "freed" up to concentrate on practice and invest time and effort into completing the course and implementing course content in the classroom.</i>		
PD PARTICIPATION	7	Teachers who participate in <i>VTSB</i> at the same school share their progress and discuss their experiences with one another.	Teachers are motivated to "keep up" with colleagues' level of <i>VTSB</i> course participation and discuss/learn from each other's experiences (positive peer pressure).	Learning community emerges that encourages and keeps members accountable for <i>VTSB</i> participation and application.
		<i>When teachers who participate in VTSB at the same school share their progress and discuss their experiences with one another, teachers are motivated to "keep up" with their colleagues and a learning community emerges that encourages and keeps members accountable for course participation and application.</i>		
	8	The <i>VTSB</i> course format facilitates teachers' sustained, consistent engagement in course content and completion of assignments.	Teachers' self-efficacy for implementing VTS in the classroom rises.	Teachers' apply VTS into daily instruction.
		<i>If the VTSB course format facilitates teachers' sustained, consistent engagement in course content and completion of assignments, teachers' self-efficacy for implementing VTS in the classroom rises and they apply VTS into their daily instruction.</i>		
	9	Teachers reflect critically on their practice throughout engagement in <i>VTSB</i> course and application of VTS in the classroom.	Teachers evaluate how their application of VTS affects their teaching practice and students' learning.	Teachers revise how they use VTS in the classroom to improve their practice and enhance student learning.
		<i>When teachers reflect critically on their practice throughout their engagement in the VTSB course and application of course content, they revise how they use VTS in the classroom based on self-evaluations of how their application of content is affecting their teaching practice and students' learning.</i>		

Phase Two Overview

In Phase II, I continued collecting and analyzing data, which I used to test and revise my initial program theories. Data included continued written and verbal communications with Watershed, continued exploration of the *VTSB* course organization and content, participants' annotations and responses to question/reflection prompts, course activity measurements, materials from the BC Schools database, and six interviews with five participants. I also continued my review of the literature throughout this phase. I analyzed and synthesized this data to develop five revised program theories.

Review of the Literature

Question #2: What substantive theories help explain why or how these contextual factors generate desirable outcomes for a PD or OPD course?

To answer the second review question, I explored in depth the substantive theories identified in Stage 5 of my realist search, including: The Dynamic Systems Model of Role Identity (DSMRI), social learning theories, motivational theories, and identity theories. As described in Chapter three, these theories are complexity-consistent, in that they “reflect key tenets of complexity theory” (Westhorp, 2012, p. 405). Additionally, as the *VTSB* course is embedded in a stratified social context with horizontal and vertical layers of complexity, the relationships between elements comprising these layers can be understood by layering complexity-consistent substantive theories “to reflect the levels of reality with which the [RE] must deal.” (Westhorp, 2012, p. 405). Thus, complexity theory, realism, and the DSMRI are vertical layers of a nested complex system. Social learning theories, motivational theories, and identity theories are encompassed within the DSMRI's horizontal layer of complexity.

Dynamic Systems Model of Role Identity (DSMRI). During my initial searches for substantive theories regarding teacher learning and engagement in PD, I noted many theoretical concepts related to social learning theories, motivational theories, and identity theories, such as self-efficacy, self-regulation, self-determination, pedagogical discontentment, and possible selves. While pursuing these concepts separately, I came across the Dynamics Systems Model of Role Identity (DSMRI), which “incorporates constructs and processes that are commonly studied quantitatively and in isolation from others (e.g. teacher efficacy) into a formal model that uses these constructs as shared terminology while not obscuring or reducing the unique, rich, dynamic, and contextualized identity and learning experiences of individual teachers” (Garner & Kaplan, 2019, p. 14). The model can be used to inform changes in teachers’ role identities, motivations, and actions. Thus, the DSMRI is complexity-consistent (Westhorp, 2012) and relevant in identifying how teacher learning and identity help generate desirable outcomes for PD (Kaplan & Garner, 2015, 2017; Garner & Kaplan, 2019).

Teacher learning and role identity. Researchers have a limited ability to identify causal connections between factors of effective PD and teacher learning outcomes. Often, teacher learning has been assumed to occur in a linear fashion, in which the completion of PD sessions produces new knowledge or skills. However, as Garner and Kaplan point out, this transmissive model of teacher learning is too reductionist (2019; Kaplan & Garner, 2015, 2017). Instead, they advocate the DSMRI, in which teacher learning is perceived as changes in role identity through complex, dynamic, and contextualized processes.

Kaplan, Garner, & Semo state that, for a PD program to be successful, it “must address not only content and pedagogy but also the complex processes that promote the professional role identity development that drives teachers to learn, develop, and sustain strong commitments to

new practices and to the profession” (2015, p. 3). Role identity is the unit-of-analysis in DSMRI and is defined as “a phenomenon that comprises a network of interdependent elements that is continuously and dynamically emerging within its environment” (Garner & Kaplan, 2019, p. 10). Teachers have multiple role identities and sub-roles within those identities (e.g., sub-roles of teacher and PD participant within the role identity of a working professional) that exist within “a hierarchical complex dynamic system of identity” (2018, p. 12). Within this overall system, different role identities interact with and mutually influence one another. Such interactions often result in tension, or misalignment, within and between role identities. As in all complex dynamic systems, “the DSMRI aspires to self-organization”—that is, to a state of increased harmony, alignment, and integration” (Kaplan, Garner, & Sumo, 2015, p. 11) within and between role identities. Thus, they make changes, such as integrating or reforming identities, to alleviate tensions and restore harmony (e.g., engaging in PD to alleviate tension between a teacher’s desire for a more student-centered classroom and her current teacher-centered practices). Teacher learning occurs through these ongoing processes of change and emergence within and between identities (Garner & Kaplan, 2019).

DSMRI role identities, components, and elements. Role identities have “four multi-element components” (Garner & Kaplan, 2019, p. 10), including: self-perceptions and definitions, ontological and epistemological beliefs, purpose and goals, and perceived action possibilities (see Figure 15).

Each component is a category of elements, or theoretical constructs. The four constructs overlap to form the “central location of teacher action, which conveys that the relations of each component to teacher motivated action cannot be considered independently of other system components. (Kaplan et al., 2015, p. 7). For example, *Self-Perceptions and Definitions*

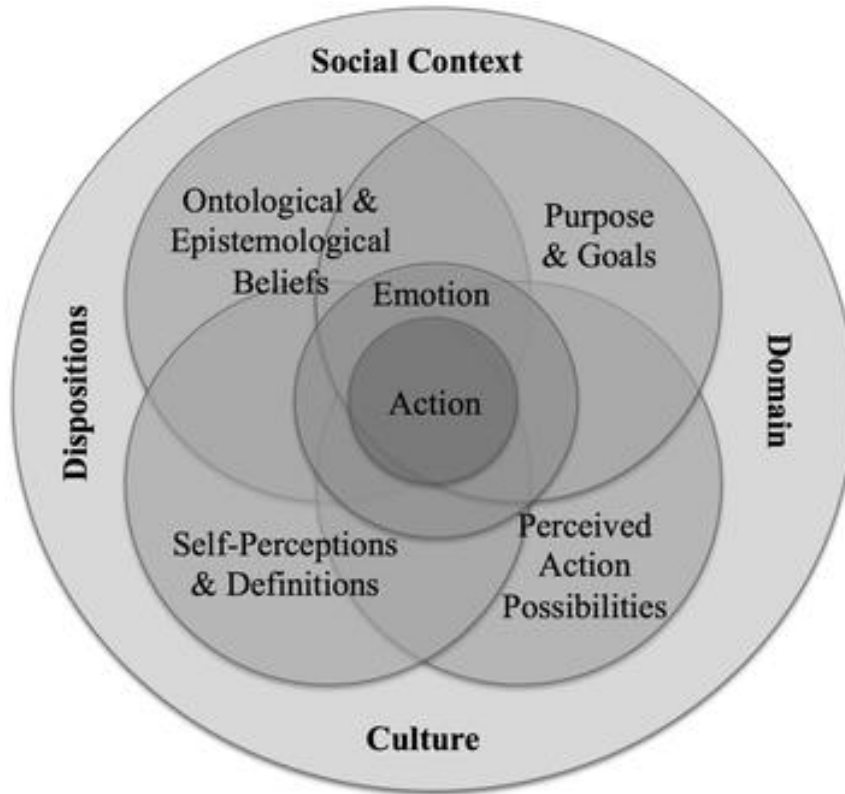


Figure 15. The DSMRI

Reprinted from "A complex dynamic systems perspective on identity and its development: The dynamic systems model of role identity," by A. Kaplan and J. Garner, 2017, *Developmental Psychology*, 53(11), p. 2041. Copyright 2017 by the American Psychological Association.

encompasses constructs such as self-efficacy, self-regulation, and self-perceived abilities. I used these components as parent codes when creating a codebook to analyze data in Phase II (see Table 6). Kaplan and Garner sent me their DSMRI codebook (2017), which I used for the descriptions of each component in my own codebook and are shown in Table 10.

The four components can be in harmony or tension with one another. When the components are aligned, teachers are content and motivated in their current practice; however, when the components are in tension, teachers seek ways to re-align them through altering, re-forming, or integrating them. (Kaplan & Garner, 2015; 2017. For example, a teacher's goal to increase student interaction in the classroom may align with her goal to conduct one VTS session in the classroom every week. Consequently, she is motivated to conduct a VTS session every week. Alternatively, a teacher's goal to engage in the *VTSB* course may be in conflict with her goal to spend more time co-planning with her PLC. This might result in the teacher integrating these two goals and recruiting her PLC members to engage in the *VTSB* course with her.

Similarly, there are also elements in each component that can be in harmony or tension with one another. For example, a teacher's goal to improve her student-centered instructional practice may align with her perceived ability to integrate VTS into the curriculum by the end of the semester. As a result, she is motivated to sign up for the *VTSB* course. Alternatively, a teacher's goal to implement VTS sessions by the end of the semester may not align with her perceived ability to finish the mandated curriculum by the end of that semester. This might result in the teacher revising her goal for implementing VTS sessions in the classroom by adjusting her timeline.

Additionally, "as any complex, dynamic system...the formation of each teacher's role identity is framed by the unique integration of four generalized control parameters" (Garner &

Table 10

DSMRI Components, Descriptions, and Elements

DSMRI Component	Description	Elements
Ontological Beliefs	Statements that indicate knowledge (e.g., theory), beliefs, assumptions, perceptions, and conceptions that the person expresses as true about the world.	Reflection General/prior knowledge Causal attribution beliefs Pedagogy Content knowledge
Epistemological Beliefs	Expressions that indicate the level of certainty, credibility, and complexity of the person's ontological beliefs. Often expressed in questions, or in hesitation regarding the definitiveness of statements.	Learning philosophy Growth, fixed mindset Beliefs about knowledge
Purpose & Goals	Statements that express the person's purpose for action in the role, as well as goals and objectives in the role. This category includes general goals of the role/domain/profession, personal goals, as well as specific objectives in particular contexts and situations.	Motivation to learn Willingness to learn Instructional goals General, specific goals Long, short term goals Overall purpose
Self-Perceptions & Definitions	Statements that include reference to the self in relation to the role. This includes how participants define themselves in relation to the domain/profession, what participants think about themselves in the role and as part of a role-related community, and how they think about their own functioning in the role (e.g., self-perceived abilities and efficacy, personal values, interests, personality attributes, self-characteristics and definitions).	Self-efficacy Self-regulation Pedagogical discontentment Possible selves Self-perceived abilities Initiative Time management Interests

Table 10. Continued

DSMRI Component	Description	Elements
Perceived Action Possibilities	Statements that indicate internal (e.g., thoughts, planning) and external behavior in relation to the role. This code includes practices and strategies that one is aware of as possibilities or that one has put into practice, as well as indications for those actions that the person perceived as not possible for him or her to enact.	Differentiating curriculum Enhancement of curriculum Instructional strategies Experimenting Student-centered Teacher-centered Visual learning Increased student motivation Increased student CT skills Increased student participation Students gain sense of freedom/reduced pressure to provide "right" answer Recruitment of other teachers

Kaplan, 2019, p. 13), which include: the social context, culture, subject domain, and personal dispositions. These parameters interact with and influence role identity components, elements, and formation processes. They provide context for the interactions within and between role identities (Kaplan & Garner, 2015, 2017). Ultimately, as Kaplan et al. recommend, I “[utilized] the DSMRI to [help] develop and test theories concerned with...individuals or groups of teachers of different characteristics and in different contexts” (2015, p. 13).

Interviews and Course Annotations

Background data: BCSD Docs

As described in Chapter three, I reviewed several sources of data from the BCDOE Open Database to help inform my interview questions and revise my program theories. The data provides contextual information about each participant’s school of employment and includes their school’s 2018 Survey for Teachers results, Quality Report, demographics, and website. I summarize my findings from exploring this data in Table 11 and refer to elements of it in the following discussion of participants’ contextual backgrounds.

Participants

There were eight study participants who provided different types of data (see Table 11). All of the participants are employed by Big City School District (BCSD), a large, urban school district in the Northeastern United States. Five participants engaged in interviews, ranging from twenty-six to thirty-eight minutes, and gave access to their course annotations and completion rates. The other three participants did not participate in interviews, but gave access to their course annotations and completion rates.

Olivia M taught second grade at PS 1 until December 2018. PS 1 is a large urban public elementary school in the southern part of BCSD. She is in her sixth year of teaching and this was

Table 11

Participants' Backgrounds and Level of Study Involvement

Name	School	Grade/Subject	Interviews	Annotations	Completion
Olivia M	PS 1	2 nd	2/22/19; 4/28/19	X	X
Madeline N	PS 2	11 th English	4/8/19; 5/9/19	X	X
Judith D	PS 3	Elementary Art	4/30/19	X	X
Marsha L	PS 4	3 rd	5/3/19	X	X
Shirley N	PS 4	3 rd	5/3/19	X	X
John T	PS 5	Middle School Teacher		X	X
Henry H	PS 5	Middle School Principal		X	X
Ron T	PS 6	Middle School Teacher		X	X

her fourth year at PS 1. The student population at the school is 94% economically disadvantaged, 59% ELL, and 24% special needs. As her master's degree is in Teachers of English to Speakers of Other Languages (TESOL), Olivia served as an ESL co-teacher in an Integrated Co-Teaching (ICT) classroom. At the end of 2018, she accepted a position teaching fourth grade at another elementary school in the same city.

Madeline N teaches eleventh grade English at PS 2, a large urban public high school in the southern part of BCSD. She is in her eighth year of teaching, but this is her first year at PS 2. She was a middle and high school SPED teacher for the previous seven years and always taught at least one English class. This school year she transitioned to teaching English full time. The student population at her school is 81 % economically disadvantaged, 20% ELL, and 16% special needs.

Judith D teaches art at PS 3, a large urban public elementary school in the southern part of BCSD. She teaches art for the whole school and visits each class once a week. She is in her seventeenth year of teaching and this is her seventeenth year at PS 3. The student population at her school is 82% economically disadvantaged, 43% ELL, and 20% special needs.

Marsha L and Shirley N co-teach a third-grade class at PS 4, a large urban public elementary school in the southern part of BCSD. Marsha is in her third year of teaching and this is her first year at PS 4. Shirley is in her second year of teaching and this is also her first year at PS 4. The student population at their school is 48% economically disadvantaged, 11% ELL, and 24% special needs.

Henry H and John T work at PS 5, which has an elementary school and just opened its middle school in 2018. Henry H is the founding Middle School Principal and John T teaches

there. The student population at their school is 74% economically disadvantaged, 18% ELL, and 22% special needs.

Ron T teaches at PS 6, a large middle school in the north of BCSD. The student population at his school is 93% economically disadvantaged, 19% ELL, and 24% special needs.

Interviews and Course Annotations

I coded the interviews, written responses to course questions and reflection prompts, and course annotations using the codebook described in Chapter 3 (see Table 6). The following is a summary of my findings from participant interviews and course annotations, including responses to embedded question and reflection prompts. The findings are organized according to the codebook in Table 6 (Chapter 3), which has three hierarchical layers of codes. The base layer corresponds to the RE levels of complexity: Individual/intrapersonal, individual/*VTSB* course, interpersonal, institutional, and infrastructural. Each of these is broken down into constructs from substantive theories, other concepts identified in the literature review, and my analysis of *VTSB* course materials, BCSD materials, and conversations with Watershed Collaborative staff.

Individual/Intrapersonal (Teacher Identity)

The first layer of complexity that is addressed in the hierarchy of codes is the individual layer, which is divided into intrapersonal and *VTSB*-related sections. The intrapersonal section examines interview and annotation transcripts in relation to the four components, formation processes, and changes in teachers' role identities.

Ontological & epistemological beliefs. The teachers indicated that they believe learning occurs most effectively through the facilitation of inquiry-based, collaborative, student-centered activities, which they incorporated into their practice to different extents. The beliefs and practices they described align with several of the primary tenets of a constructivist epistemology,

including: student-centered learning, co-construction of knowledge, connections to prior knowledge, teacher facilitation, and not eliciting “right/wrong” answers.

For example, when asked to describe her instructional style, Olivia responded, “I feel like I’m always trying to figure it out, but...in general, just like student-centered, discussion focused, inquiry based.” She also emphasized the importance of students and teachers co-constructing knowledge, stating that “I really prioritize the students learn from one another and that I’m learning alongside them.” Responding to the same question, Marsha said, “I think the goal is to have it more student-centered...And to have a lot of discussion and reflection and questioning.” Her co-teacher Shirley added, “Right, to have as much hands-on inquiry activities as possible.” In her course annotations, Shirley wrote about increasing students’ engagement and motivation by selecting course materials that are relevant to their prior experiences and knowledge,

Children are bringing their life experiences into their ideas of what they see. In our class, a child who rarely speaks about her family's ethnicity became very animated when images of people from that part of the world were shown in class. She showed pride in her expert knowledge where she often appears more embarrassed about that topic.

In both their interviews and course annotations, teachers emphasized the importance of maintaining the role of a facilitator and not trying to elicit specific “right” or “wrong” answers from students, as this may discourage them from participating in class discussions. Henry wrote that “reluctant learners gain confidence when they are able to contribute to a class and their answer is not "wrong." Similarly, Shirley noted in response to a VTS demonstration video, “when children have the opportunity to be heard without being right or wrong, children who often don't speak seem emotionally freed up to speak.” Reflecting on her classroom experiences, Madeline wrote that “it always amazes me how thrown off students are when they are asked to

think. They want so desperately to arrive at the right answer and for the teacher to tell them that they are right.” However, she thought that using VTS would help to change students’ mindsets because “ambiguity in teacher responses would force them to break the habit of needing to be right.”

Purpose and goals. When asked if there were any specific areas that they were working to improve in their personal practice, the teachers expressed an ongoing desire to expand their student-centered instructional practices. Madeline replied that she “would love for it to be more student centered,” Marsha thought “the goal is to have it more student-centered,” and Judith said she is “definitely working on making it more student-centered.”

The teachers also identified more specific areas in which they wanted to improve. Madeline wanted to use VTS to help her students decrease their focus on producing right or wrong answers and develop more creativity, ownership, and autonomy in their thinking, stating that,

In terms of performance, I really really want to get them away from thinking that they just have to be right. I’m so sick of students doing the work just to give it to me. I keep trying to explain to them. Like, this is not about me. You don’t need to be right to turn it in and for me to give you a check. I want you to think. I just want to see that you’re thinking.

Judith also wanted her students to worry less about giving right or wrong answers. She was conscious about the potential role of bias in this process and wrote about her hopes to

work on my ability to stay neutral and not try to drive the discussions in my class toward my opinions. All students deserve the chance to be heard and to think that their contribution to the conversation is valued and important. If they think of discussion as a

time to give correct and incorrect answers, there will be students who will be hesitant to participate.

Teachers' goals tied in with their motivation and willingness to learn; they enrolled and engaged in the *VTSB* course because they perceived it as a means of achieving their goals. Madeline said that she is "working on building rigor in instruction and creating opportunities where it can be challenging for all students," and described how she is "using the VTS to work on that." After implementing VTS a couple of times, she noticed that it was "definitely helping" make progress towards her goal of reducing students' hyper-focus on right and wrong answers. This motivated her to continue engaging in the course and experimenting with VTS in her classroom.

Like other teachers, Marsha and Shirley had observed how their students were reluctant to engage in discussions when there was a chance of being wrong in front of their peers, which made it "very easy for things to come down to a few vocal members of the class." Marsha and Shirley saw VTS as a way

to be able to foster more constructive classroom discussions so that it's not just the same three or four kids talking...it's giving kids who don't always have a chance to talk because they don't want to be wrong. This gives them a chance to be one of the thought makers of the class.

Judith identified specific goals related to classroom management, timing, and organization. These were areas she was especially concerned about as an art teacher who visited each of her students' classrooms once a week. Responding to the question "How VTS can support your goals?", she wrote that her "goals at this point are to establish discussion norms and to start using protocols for small group discussions during my art lessons." So, her goal was to

use VTS to establish behavioral expectations and incorporate more structured discussions into her practice.

Self-perceptions and definitions. The teachers' experience levels ranged from two to seventeen years of teaching within the BCSD system. Their previous roles included teaching various subjects and grade levels, teaching Integrated Co-Teaching (ICT) classes, teaching ELL classes, and teaching or serving as an aide for special education classes. These frequent transitions shaped their how they defined and evaluated their roles as teachers and employees and affected their self-efficacy in their current teaching roles.

Madeline, who came from a family of teachers, said, "I knew I wanted to be a teacher since I was able to realize what a future was." Prior to her current position as an eleventh grade English teacher, she worked in more ancillary roles in the classroom, which affirmed her decision that she

really wanted to be in front of the classroom. I felt like I was watching other people live out their dreams while I was passing out pencils...So, you know, I took a risk and I switched, basically it feels like careers, but it's not...I wasn't sure if it was gonna be the right move for a while and it was the right move. I finally remember like this is why I wanted to be a teacher.

She attributed her increased self-efficacy in her role as a teacher in large part to the encouragement of the leadership at her school, stating "it's so funny. I talked to my admin this week and I was like, 'There's a chance I might be a good teacher,' and he's like, 'Of course you are.'...it's crazy that I was never told. It's sad to think that nobody ever told me." Her increased confidence as a teacher made her more willing to engage in new initiatives like VTS to the extent that she began sharing VTS with other teachers and trying to recruit them to join the course.

Olivia also had experienced a lot of transition in teaching roles. Prior to her position teaching second grade, she had taught in various ELL and ICT classes, sometimes as a co-teacher, sometimes in a team of teachers. As a result, she stated that, “yeah, you know, I’m still trying to figure out what teaching is.” At the time of our first interview, she was also preparing to start a new job as a fourth-grade teacher at a different school. However, she was hopeful that her new school would provide more stability and clarity for her role in the classroom.

Shirley and Marsha also worked in a variety of other roles before teaching third grade at their current school. Shirley worked “in the performing arts and as a costumer and creative movement teacher,” and Marsha “did educational media before I became a teacher.” Their backgrounds in these areas influenced their interests as teachers; Marsha stated that one thing that attracted them to the VTS course was “maybe just the visual nature of it. We’re both into that kind of thing.” Additionally, they had a high self-efficacy in their ability to implement new initiatives based on mastery experiences such as taking part in maker spaces. Marsha said that “Shirley and I are both really comfortable diving into it and just doing it.”

However, teachers’ self-efficacy decreased when they encountered areas of uncertainty during the course. For example, some of the teachers were unsure of how to apply VTS in specific ways that would achieve their instructional goals and meet the needs of their diverse groups of students. In particular, Marsha, Shirley, and Madeline struggled to identify with the teachers, students, and instructional contexts in the example videos, as they did not resemble students and classroom context. They adapted VTS to fit into their daily instruction, but they did not have high self-efficacy in their ability to implement VTS properly while accommodating their students’ diverse needs.

As evidenced in their goal statements, all of the teachers expressed pedagogical discontentment in some area, which they hoped to address through adopting VTS practices, including expanding student-centered practices, increasing rigorous instruction, increasing student participation, and improving classroom and time management skills. Their self-efficacy was raised when there was alignment between their beliefs, practices, goals, and *VTSB* content. However, there was also increased tension when their goals were in conflict with other responsibilities or requirements.

Perceived action possibilities. In addition to general goals, teachers also described specific ways in which they hoped to apply VTS in their classrooms, especially in order to make classroom discussions more inclusive and accessible for all of their students. The teachers indicated that diversity was a primary consideration when planning and facilitating lessons and when deciding to enroll in and apply content from the *VTSB* course. As their school district is located in a part of the country with a high population of immigrants and non-native English speakers, the teachers saw potential in the VTS model for accommodating her ELL students. Olivia, an ESL teacher, stated

I'm always trying to think what will they have access to that isn't dependent on text comprehension? Something that's visual and really stimulating is something that even our struggling readers would have no barrier and our struggling speakers, it would really push them to develop their discussion skills and the language and build their vocabulary—I just saw a lot of potential for it to be really useful and effective... I thought it could be really—could be more engaging than literature.

Judith thought that VTS would be beneficial for many of her students, as it specifically targets skills that they needed to develop. She wrote, “my students are non-English speaking, so it is

difficult for them to express what they are thinking and especially giving evidence to back up their thinking.” Madeline also said, “I do have a couple former English language learners that are now in my class that I can see have really benefited from it...I know visuals always help.”

Marsha said that she and Shirley “teach such a diverse group of kids. It’s an integrated co-teaching classroom. We have kids with special needs, learning needs, and it just seems like a really cool way for all the kids to access something.” Similarly, Ron wrote that he thought VTS would be “good for special education students to access material more easily, but still think deeply.”

The teachers also thought of ways to develop their student-centered practices by incorporating VTS into specific subject areas and instructional activities. As an elementary teacher, Olivia taught multiple subjects to her students. She thought that VTS could be valuable in social studies, “because you know, with little ones, it could be like paintings or with older ones who are actually studying history it could be like photographs.” Thinking about her English class, Madeline said,

I could see facilitating this type of discussion as a prereading activity. Maybe I could show a scene from a chapter that we are about to read and have students make inferences, then after we read we could discuss these inferences vs. what actually happened in the text.

When exploring one of the supplemental resources in one of the online course modules, Judith made connections to her art curriculum, writing that, “the images in the What's Going On in This Picture?” series are very vibrant and I think I will be able to use some of them with my students in the future.” Thus she, as well as the other teachers, were continually reflecting on

potential ways that VTS could help enhance their instructional practices and achieve their instructional goals.

Role formation/change.

Harmony/alignment. Participants' roles as teachers aligned with their roles as learners when engaging in the *VTSB* course because it aligned with their beliefs, experiences, practices, interests, and goals related to teaching and learning. When asked if she enrolled in the course because of it aligned with her interest in the topic and her teaching philosophy, Olivia replied, "Totally. Yup, that was totally what it was." She added that, prior to reading about the course, she

had definitely heard of visual thinking strategies in some point in my career as some kind of personal study. I was definitely very aware of it, I don't know how, but then I decided to do it just on my own, because I saw the offer and I was like "Oh, cool."

Marsha and Shirley found alignment between the course content and their instructional and classroom management goals. Marsha said that

I think that the method and the content go very well with what we're doing in our class. It's part of what we want to be able to do. Like first of all, the classroom management aspect of it, how to organize the conversation and listening to one another is definitely a goal of ours and we think it would fit very nicely into the class.

Judith enrolled in the course "because it was about questioning and discussion and that's the focus of the school building and goes right along with Danielson, I thought it would help. And plus its art and I'm a visual arts teacher." The district adopted the Danielson Framework for Teaching in 2015 and uses it to evaluate teachers; Judith said that, "when we're observed, the administration will let us know there was feedback on how to be more student-centered, how to

have more questioning, just more student-centered classes, because that's what Danielson works towards...so VTS, it lends itself to it." So, the course aligned with her roles as an art teacher an employee with accountability measures to fulfill.

Integration. While teachers had many demands on their time, they found ways to integrate their roles as teachers and employees with their roles as learners in order to engage in and apply course content. They described how they made time to engage in the course and how they adapted the VTS content to fit into their curricula.

When asked if she was given extra time to complete the *VTSB* course, Madeline replied that "I make the time." Marsha said that she and Shirley "do it either in our prep period or after school" and Shirley added that they have to make time for it, "along with everything else we shoehorn it into our time." Similarly, Judith said, "I have to find my own time...but I'm going to get it done." The teachers indicated that, despite the challenges of time management, they were committed to engaging in the course.

Looking at the timestamps of teachers' annotations, they did not have consistent patterns of course engagement. They completed large sections in one- or two-day periods, followed by periods of no engagement (as indicated by course annotation timestamps; they may still have accessed the course during these time periods, but this information was not recorded). Comparing their times of activity to the school calendar, they often coincided with weekends or at the beginning or end of vacation periods. This indicates that they completed what they could when they could.

Teachers experienced tension in their understanding of how to incorporate VTS in the context of a lesson, or unit of instruction, to help meet their instructional requirements and goals.

However, they tried to alleviate this tension by adapting VTS to accommodate their needs and integrating their roles as employees, teachers, and learners.

Shirley said that she and Marsha were not sure whether VTS “is its own discrete curricular model with its own curriculum pieces or is it designed to be integrated with our own curriculum and with our own imagery that we want to pull out and have more discussions around?” They said that up to that point they “haven’t used the pictures on the program” but they had adapted VTS into lessons about China and during Black History Month. Judith also said, “I haven’t been using the actual pictures from the program along with the questions, but I’ve been asking those types of questions in some of my classes.”

Madeline also was not sure how to embed VTS into the context of an entire class period. She had begun incorporating VTS into her “do-now” activities at the beginning of class, as well as discussions and assignments for their unit on *Of Mice and Men*. However, she said, “I’m just doing this stuff on my own... I would love to see more of a full lesson with it, with like a writing-how you would do a writing activity after and stuff like that.”

Tension/instability. There were areas of tension between participants’ roles as teachers, employees, and learners. As teachers, they had goals to incorporate more student-centered practices into their instruction. They perceived ways to achieve these goals as learners in the *VTSB* course. However, as public school employees they also had time limitations, diverse student needs, and accountability standards to meet. Thus, teachers discussed time management, classroom management, students’ beliefs about learning, and accountability measures as some of the primary factors constraining their ability to implement student-centered practices, such as VTS, in the classroom.

Two common, and interrelated, areas in which teachers expressed concerns were classroom management and time management. As public school teachers, they had mandated schedules and curricula that made it challenging to engage in new initiatives. As Madeline expressed, “I would love for it to be more student centered, but in terms of classroom management, I feel like I have to keep it tight...it’s hard because we have forty-four-minute periods.” Judith felt similarly, and said she was “definitely working on making it more student-centered. It’s a process...it’s just getting the students to be able to do these things and still have management, like classroom management.” It was challenging for her to implement and consistently enforce rules and routines, including the VTS dialogue and etiquette, because she only saw her students once a week for forty-five minutes. Additionally, as her school did not have enough space for a dedicated art room, she was visiting each class with an “art cart,” which further complicated her schedule. She stated that, “I can never fit everything in that needs to get done...just having ten minutes out of a lesson is a huge chunk.” She was concerned about having enough time to incorporate VTS sessions while also teaching her required material and taking care of logistical issues.

Marsha and Shirley discussed how their larger class size affected their ability to implement more student-centered activities and maintain classroom management, saying that their

goal is to have it more student-centered...But with twenty-six kids in a room, it’s way different than when you’ve got a class of fifteen, where it’s a lot easier to have that be more effective in a smaller class size.

Madeline described how her goals of incorporating more student-centered practices through VTS did not align with her eleventh-grade students' long ingrained beliefs about learning and need for validation, stating that

It's going to take time to shift the mindset, cause you probably grow up—even me now, I want someone to tell me, 'Yes, you're right.' And the hardest part out of VTS...is I say to them before, 'I'm not going to give you any feedback. Like I'm not going to tell you that's a good answer' ...And sometimes we move on and they're like, 'Wait. Miss. My answer was the best, right?' and I'm like, 'Your answers were all good in different ways' ...it really upsets me how students don't know how to be creative.

Judith was concerned about meeting evaluation criteria during official observations. She said that, for these observations, the administration was

thinking more along the lines of asking an open-ended, quality question and doing an activity—it's structured different, what they want to see. And if I'm observed, and observations can be at any time, I can just tell her I'm doing a VTS thing for a video and they won't come in, but if I was going to be observed, VTS wouldn't be exactly what they want to see.

Additionally, events such as state testing and vacation periods affected teachers' ability to engage in the course and implement VTS in the classroom. John and Henry, who work at the same school, told the Watershed Operations Director that their participation in the course would be decreasing for a couple of weeks during state testing.

Individual/*VTSB*-Related

The *VTSB* related section of the individual layer examines interview and annotation transcripts in relation to the structural, didactic, and content-related aspects of the *VTSB* course.

Structural. Teachers discussed ways in which different structural aspects of the *VTSB* course affected their desire and ability to engage in the course and apply course content. When asked about the online delivery format, teachers responded that they liked the flexibility that the online, self-paced course format provided. As they were not given extra time during their work day to complete the course, they had to “find” or “make” time to participate when opportunities arose, so they appreciated being able to access the course from their choice of time and location. As Shirley stated, “I think the fact that it’s accessible wherever and however, that makes it an easy way to fit it in. That’s helpful.” The online format also helped them to balance other roles and responsibilities, such as parenting. Olivia stated that, “because I have a child, really the only way that PD generally would work for me is if it’s either during the school day or if it’s online.” Thus, the *VTSB* online, self-paced format helped accommodate her roles as a teacher and parent.

Teachers found navigating the modules, including accessing the videos, question/reflection prompts, and annotations to be mostly intuitive and straightforward. Madeline said, “I like the format of the course...I’m not technologically savvy and I feel like it was very user friendly.” Olivia also thought that the format “was pretty straightforward.” Judith said she “thought it was very simple. Very user-friendly. You know, takes a few minutes just to get acquainted with it, to explore...It’s easy to navigate. I didn’t have a problem with it.” However, Marsha and Shirley struggled to use the course in the context of a class lesson. Marsha recounted that one of the videos said that

...it’s such a great program because you can — in the interface you can click on the image and show it to your class and take notes on the interface, but I don’t know where that is.

Like what button do you have to click on to do that? Because we have a smart board and it’s super easy to hook up to our laptops, you know.

Teachers appreciated the inclusion of questions and reflection prompts, as well as the ability to annotate, during the videos, as they helped them to stay engaged and focused on key points. When asked if she found these features helpful, Judith responded, “Absolutely. If you don’t have anything, I probably wouldn’t have finished the videos if I didn’t have that. You need that.” Madeline had a comparable response, saying “I like that there are questions to respond to, so I know what I’m looking for.” Olivia particularly appreciated the annotation feature, as

In general, as a reader and a participant in PD, I’m trying to get better at annotating...I liked that it was just there and almost being like, “Come on, annotate.” So I did use it and you now, writing obviously helps you process thinking and it was a good reminder that I want to do that.

Teachers indicated that the primary purpose of written responses was to help them actively process information. However, they saw potential benefits in revisiting their responses to help them review their thoughts about specific concepts. When asked about this, Olivia replied that the annotations “helped me in the moment...You know, I think I would revisit the annotations, because I remember having very concrete thoughts about either teaching strategies or ideas or things I wanted to make work for me.” Madeline said, “I’ve come back to them, because whenever I log into the course, I always go through and see like if I’ve already answered any of my own questions.” She also reviewed questions to which the course facilitator had provided feedback.

The teachers made both positive comments and critiques about the videos embedded in the modules, including the instructional videos with the course facilitator, Philip, and the recorded examples of teachers facilitating VTS sessions for different grade levels and different subjects. When asked if she found the videos helpful, Judith responded, “Yeah. The teacher

videos, his videos. They were easy to listen to. I was interested and I followed along pretty well.” She also wrote in her course annotations that “I am definitely getting a good idea of how VTS works and how it should look and sound in a classroom, by watching the videos.” Olivia said, “Phillip. I like his videos. I thought he was pretty engaging.”

However, teachers also were unsure of which videos to watch and in what order. The modules are not arranged in completely linear fashion, as some are video and resource banks that are to be accessed throughout the course. Olivia said that,

I did start to get a little bit confused at like—there was sort of a bank of videos that span I think day twelve or whatever and I remember I looked at some of them, but yeah, I can’t remember how they connected to what I’d already done versus like the other, the future modules I hadn’t done yet. I don’t know.

Similarly, Marsha stated that, “There are some confusing aspects. For example, are we supposed to watch all the videos? That is not clear to me. So, we focused on the third-grade class. We watched two third-grade classes.”

They expressed further confusion about course pacing and requirements, including being unsure of how much of each module needed to be completed in order to satisfy course requirements. The icon displaying the percentage complete for each module only shows 100% if every video is watched and every question/reflection has a response. However, some of the modules include optional videos and some of the reflections do not have spaces to respond. Consequently, the “percentage complete” that is displayed is not representative of the participant’s actual accomplishments.

Shirley thought,

it seems like it could maybe be—it could use a little more user interface work in terms of maybe just even having here’s a map of what you need to do... Yeah, like okay, we watched the video for our third grade and we’ve watched a video from another grade and it says we’re only seventeen percent done. I thought that was all we had to do for this module... It seems like the program, in order for it to think you’re done, expects you to do more than you actually need to do kind of thing.

On a related note, while the teachers found the supplemental resources interesting and useful, they were unsure how many, and how to decide which, of the resources they were supposed to access and what they were supposed to “do” with them (i.e., read them, answer questions about them, use them for an assignment, etc.). Madeline came up with her own methods of determining which resources to access

I mean I’m sure there’s not a right or wrong approach, but to be honest, I looked at the resources with the titles or pictures that stuck out to me. And there were some titles, and without looking and I could be wrong, I just didn’t think would apply to my content or my students, whether it would be culturally relevant or whatever...like when I’m looking through movie titles and I don’t know what the movie is about, because I just like going to the movies, I’ll be like, “Oh, that might be interesting.”

Judith said she “definitely clicked on each one. Like, time is a little scarce right now, so being able to read each one didn’t happen. I saved a few that I thought were interesting for when I do have time this summer.”

Lastly, some teachers were unsure of the process of completing assignments. They wondered if the assignments were mandatory and if there were things they needed to “turn in.” When asked about completing the assignments, Madeline said, “That I was a little confused

about, because it was like, ‘Am I turning something in? Are these just things I’m doing on my own?’” Marsha added that they would appreciate “a checklist of what you need to do. It’s unclear to us what we need to do.”

The course syllabus provides guidelines for which videos to watch and which activities compete for each module, so perhaps teachers did not read the syllabus or read it but did not use it as a course guide. However, as it is not specific to BCSD, it does not discuss what needs to be completed to satisfy PD requirements. It implies that assignments do not need to be officially “turned in,” but does not explicitly state this.

Didactic. Teachers described how engaging in both the course content and active learning components, including application and reflection, played roles in developing their skills related to student-centered instructional practices. For example, Madeline described the integral roles of both exploring and applying the course content to develop these skills. Recently, she has been making efforts to recruit more teachers to the *VTSB* course, including giving a formal presentation to her colleagues and opening her classroom to them for observations. Summarizing their reactions, she stated,

it seems like they have so many questions. I want to be like, ‘Just take the course and all of your questions will be answered’... I have seen a couple of them looking for images to use to display visual thinking, but you know, to the best of my knowledge, none of them have taken the course yet. Which I would think you should do first. But no, to my knowledge, there’s no one taking the course before they just try it on their own.

Otherwise, it’s not going to be the correct way.

She added that she herself had been skeptical prior to applying VTS in her classroom, admitting “I even poked holes in it until I tried it myself.” Thus, based on her experiences, she indicated

that going through the course content prepared her to apply VTS in the classroom. Alternatively, she did not think that implementing VTS before going through the course content would be as effective.

While Marsha and Shirley had some critiques about *VTSB*, they also described how their experiences with implementing VTS had improved as they progressed through the course.

Shirley said,

So I guess we've really experimented with it. Before we've gotten through the whole course...As we've done more of the course, we've seen how there's a lot of silence and more stillness and how that's what prompts more discussion rather than like 'and here's something else and who wants to talk about it' because that's when we started again seeing—it was a different five or six kids that contribute to our regular academic discussions, but it was the same five or six kids that were contributing to this discussion.

So that was interesting.

Consequently, as they completed more of the course and practiced applying VTS more frequently, they developed their skills in facilitating inclusive classroom discussions.

In both their interviews and course annotations, the participants frequently mentioned how the teachers and students in the example videos were well-versed in VTS facilitation and discussion techniques. They found these videos to be aspirational models, but wondered how the teachers had reached that level of ability. One of the supplemental resources, *VTS Spotlight #2: Introducing VTS to your Classroom*, briefly addresses this topic, but none of the teachers had accessed it yet according to their course activity; perhaps as this resource was not embedded in one of the instructional modules but in a separate resource bank that was added to the course after teachers had started the course.

Shirley mused, “I wonder how long it took them to use the proper talking stems. Has there ever been fallout over agreements and disagreements?” Similarly, Madeline wrote, “I wonder if the students have been taught lessons on accountable talk/teaching with response stems “I disagree with _____ because _____.”” When asked about changes she would make to the course, Marsha said,

How do you teach it?...Like how do teachers use that vocabulary with the kids. Do they teach them the phrase? Do they teach them the acronym? How do they define it? What’s the conversation and how is it taught explicitly with the kids? ...Where are the videos on that? Are there videos of that? The videos in action are great but— Is there a lesson plan that introduces this methodology?

In response to a course question asking what a teacher could do to “get to the point” of the teacher in the video, Judith wrote that

I think to get to this point, a teacher would have to gain experience by facilitating a VTS discussion in their own classroom. Another thing that could help a teacher who is new to this type of teaching is reflecting on your lessons. One good way to do this is by video taping your own lessons and playing them back to see where you need improvements.

The more experience a teacher has, the easier the techniques will become.

Her statement combined reflection with active practice as a means of improving instructional practices in the classroom. Teachers also used reflective practices when responding to course question/reflection prompts and making annotations that connected the course material to their own classrooms. Module 4, entitled *Reflective Practice*, included instructional and supplemental materials, a self-assessment assignment, and guidelines for creating an *Integrated Practice Plan*, in which teachers planned “how the VTS Basics online learning experience should be integrated

within [their] existing classroom instruction and broader professional learning plan” (Watershed, 2018). The course also provided an *Observation Rubric* as a tool for reflecting on teachers’ practices in example videos. Judith wrote that, “The Observation Rubric definitely helps me focus on the main aspects of what a VTS discussion should consist of.” She also anticipated that using the rubric to reflect on her video-recorded lesson would be “a great way to self-reflect on my teaching.”

Content-related. Teachers found the *VTSB* course content relevant to their instructional beliefs, goals, practices, and interests. They also perceived the course as coherent in the sense that it aligned with school initiatives and state standards. As discussed in the ontological and epistemological beliefs section, the teachers’ beliefs about teaching and learning were based in constructivist learning theory. A primary reason they were attracted to the *VTSB* course was because their beliefs and goals aligned with the course’s focus on student-centered and inquiry-based instructional practices. Course materials specifically address constructivism and tie it to VTS in Module 4’s videos and supplemental resources. However, as constructivist principles are inherently embedded in VTS, they are evidenced throughout course content and application activities.

While teachers had been using VTS to different extents in their classrooms, they expressed uncertainty about how to fit VTS into the context of an entire lesson or unit and how to adapt it to different content areas. After watching an example VTS session, Olivia reflected,

Ah. Now I see. The "point" of VTS is more about the ambiguity that sparks discussion and the idea of social learning and discussion. I can see it has tons of value for its own sake...just still wondering how to connect it to content areas.

As an English teacher, Madeline wanted guidance on different ways to incorporate VTS in her daily lesson plans. Her classes were very structured, as she only had forty-four-minute class periods, so she was judicious with her time. She stated,

I use it as my do-now, but I would love to see if something could come before that and I could use it as an activity instead of a do now. What could come after it...with like a writing--how you would do a writing activity after and stuff like that.

Asked if it would help to watch a VTS session within the context of an entire lesson, Judith responded, “Yeah, that would help. Rather than getting that one piece of it.” Again, one of the supplemental resources, *VTS Spotlight #7: Connecting VTS to Content Areas*, briefly addresses this topic, but none of the teachers had accessed it yet according to their course activity.

Marsha, Shirley, and Madeline also stated that the VTS example videos were not very representative of their classrooms. Marsha stated that, while she understood that the videos were meant to be models, “it doesn’t resemble our classroom in any way because it’s so formal. The culture in our room just doesn’t look or sound like that.” Similarly, Madeline said,

when I first saw the video of the twelfth-grade class, I think I looked at the twelfth-grade class, I don’t know if there was an eleventh-grade class. But every student was like hands up and then the responses were so sophisticated, and it was like a lot of white kids, just to be blunt. And there’s not one white kid in my school. It’s like ninety percent black, a little bit Spanish.

They still found VTS to be relevant and adaptable to their instructional goals and students’ needs, but would appreciate having examples that better represent the diversity in their classrooms.

As previously mentioned, Olivia and Madeline had attempted to recruit other teachers to the course based on its coherence with school-wide initiatives and state standards. Olivia, who was part of Literacy Liaisons, a “cross grade literacy focus group that meets weekly and facilitates PD and does book studies,” told the group that she thought VTS “would really apply to our population of students and already connects to some of the teaching moves that we already use in reading writing and math.”

Madeline’s school had a new initiative called TIDE (i.e., topic sentence, introducing evidence, evidence, describe evidence) that the administration wanted teachers to incorporate into every lesson. She saw clear parallels between the TIDE initiative and VTS, which she shared with her administrators and colleagues. She said, “I could see more people getting on board...because they want people to ask more open-ended questions and have students not just be focused on getting a right answer and all this stuff and just--it ties in perfectly.”

Lastly, VTS aligned with district and state standards. As Judith described earlier, VTS “goes right along with” the Danielson Framework for Teaching, which was used by the school district for school and teacher evaluations. The course also provides a document that describes in detail how VTS supports specific Common Core State Standards.

Interpersonal

The second layer of complexity that is addressed in the hierarchy of codes is the interpersonal layer. The intrapersonal section examines interview and annotation transcripts in relation to teachers’ relationships with leadership and peers.

Leadership relationships. The teachers provided examples of their personal relationships with the leadership at their schools, including principals and assistant principals. They were all granted autonomy to engage in the PD opportunities of their choice, so they were

able to enroll in *VTSB* of their own volition. However, leadership's involvement and interest in their course progress varied.

Madeline talked extensively about how the administration at her current school had bolstered her self-efficacy and dedication, in the roles of both a teacher and an employee. She felt that she was finally trusted as a professional who knew what she was doing. In particular, she had formed a strong relationship with her principal, stating

I've only been teaching for eight years, but in my life, I've never felt that I'm good at what I do or that anyone believes in me and this principal, I don't know if it's his strategy or what, I've always felt like if someone makes you feel good, for me it makes me want to work harder.

She also appreciated being able to discuss ideas with her assistant principal (AP), who "was an English teacher for years, so her feedback is actionable and she gets, you know, the ins and outs of teaching English, because she did it." In her second interview, Madeline described how she debriefed about VTS sessions with her AP, who helped her reflect and brainstorm ideas for Madeline's VTS-related goals. She also stated that this AP was the one who sent the information about *VTSB* to all of the teachers and "wanted everyone to take the course."

Olivia was not as well-connected to her administration. She indicated that the large size of the school and fundamental differences in teachers' pedagogical beliefs made it challenging for those in leadership positions to cultivate a student-centered school-wide culture, despite their desire to do so. Nevertheless, Olivia said that she tried to communicate her hopes for VTS to the principal by emailing her and discussing VTS at a Literacy Liaisons meeting the she attended. However, the principal did not respond. Instead, "the only person who really responded was the

facilitator and she was like, ‘Maybe you can tell us more next time we meet.’ But then, that was pretty much the extent of it.”

Even though her school’s administration promoted an inquiry-based, student-centered academic environment, Judith’s relationship with her administrators seemed more focused on her role as an employee with accountability standards to fulfill. Considering whether she would be able to facilitate VTS sessions on a consistent basis, she said, “I don’t know if I’m going to be able to because of what the administration wants to see in each lesson.” Later, she repeated that, prior to implementing a VTS session, she would “have to talk to an assistant principle before I do that because it’s not exactly aligned with what they want to see.” She was concerned that, if she were being evaluated, taking time to engage in VTS might affect her ability to meet the required criteria, as “There’s a lot of stuff we have to fit in and I only have forty-five minutes. There’s a whole bunch of stuff that I have to fit in there.”

The leadership at Marsha and Shirley’s school was actively involved in progressive PD initiatives. Shirley recounted how,

when we first approached the maker space this year, one of our assistant principals was right in there, completely getting her hands dirty with everybody. Not just sitting back and watching. They’re in there with us. And actually it was another administer who pointed us to the whole VTS course in the first place.

Thus, they had no reservations about experimenting with VTS, as far as expectations from leadership were concerned.

Peer relationships. Besides Marsha and Shirley, none of the teachers collaborated with their peers to complete the *VTSB* course. There was little communication between colleagues regarding who had or had not signed up for the course, so some participants did not know

whether there was potential for collaboration or not. When asked if any other teachers were taking the course at their school, Marsha said, “No one on else our team, as far as we know.” This indicated that there might have been other participants at their school, but the co-teachers were unaware of them.

Madeline also said she did not think anyone else from her school had signed up, as she “mentioned it the other day as something you should already know and I was like, ‘You guys know, the VTS stuff.’ And they’re like, ‘Yeah, yeah, yeah.’ But I could tell they didn’t know what I was talking about.” Similarly, Judith asked some colleagues if they had signed up for the course; “They said yes in the beginning. They were like ‘Oh, the CTLE hours, sure.’ And then, they never really followed up and it didn’t happen.” Both she and Madeline indicated that they did not think the other teachers had enough time to commit to the course.

However, the teachers still desired collaborative participation; they wanted more teachers to be aware of VTS, its relevance to school initiatives, coherence with curricular standards, and benefits for students. As mentioned previously, Olivia and Madeline took actions to recruit other teachers to the course or implement VTS in their classrooms. Olivia saw great potential in collaboration and thought that, “if I even go through the course with one other literacy liaison this could be—we could facilitate our own kind of PD related to this a few months down the road.” Judith also saw benefits in collaboration and shared information about VTS with her peer collaborative group. She stated, “I think it would be a lot better with other people doing it with me. It would be a lot more interesting.”

Watershed designed the *VTSB* course for teams using research-based principles; they intended for the course to be collaborative (Watershed, 2018). Ultimately, teachers agreed with

Watershed that collaborating with others on the course would have been ideal. However, they adapted to their situations as many of their peers were uninterested or unable to participate.

Institutional

The third layer of complexity that is addressed in the hierarchy of codes is the institutional layer. The institutional section examines interview and annotation transcripts in relation to the academic, social, and physical aspects of the school culture.

Academic. As previously discussed, the teachers held student-centered perspectives on teaching and learning that were shared by leadership. This was evidenced in leadership's expectations in relation to their curricula, instructional practices, and student performance. Each of their principals had implemented school-wide initiatives and established community partnerships to help teachers develop and implement student-centered teaching practices. For example, Marsha and Shirley's school partnered with consultants who help design and provide targeted PD for various sections of the curriculum. Shirley said that their school was "definitely a more progressive model than many other public schools."

However, the teachers also communicated that these initiatives were works-in-progress. Describing her school's attempts to create a more student-centered culture, Judith said that they "definitely heading in that direction. We're not there yet, but we're getting there." Olivia's school partnered with a prestigious college to design curricula and provide high quality PD, as her leadership wanted to cultivate "more of a progressive curriculum and culture." However, this desire was constrained by the large size of the school, the number of ELL and SPED students, and divisions between more traditional and progressive teachers. She stated that,

our principal and some teacher leaders in the building really do want the school to be more project based and more inquiry based, but there's also a pretty strong cohort of

more, you know, really traditional public school teachers. And in a school with a thousand kids, inevitably you can't get away from that. Just the sheer number of teachers you need.

Madeline's school went through numerous changes this year, as they were a renewal school and almost all of their teachers were new to both the school and to teaching in general. Madeline indicated that, as this had been a year of rebuilding, she had a significant degree of autonomy in designing and implementing her curriculum, stating that the leadership was "really supportive of all kinds of ways to reach students." Nevertheless, there were also "a lot of new initiatives that start off that don't end up sticking."

While they were not given extra time or resources to engage in *VTSB*, teachers did not seem to expect this. Leadership communicated expectations for PD requirements and provided information about opportunities, including VTS, but did not dictate teachers' PD selections. Teachers indicated that they thought VTS would be more effective if it were consistently implemented throughout the school. Henry wrote that, "Because this is a part of the method, the expectation needs to be that we are adopting these processes and skills into all areas of the school." Judith also wrote that

I think it is very possible in all classrooms, however this takes practice on the part of the teacher, and the students. It also takes participation by the whole school. If students are involved in this type of discussion year after year, they become comfortable with the process and can therefor continue to build and grow their skills.

Social. In previous sections, teachers discussed their interpersonal relationships with leadership and their peers, which has implications for the collective perceptions of leadership and connectedness with other teachers. For instance, while Madeline's personal perception of her

school and her school's leadership were positive, this was not the school's consensus. As the school was officially failing the previous year, most of the faculty were replaced and major procedures, policies, and curricula were revised. Thus, not all of Madeline's colleagues viewed the principal or the school favorably. She stated, "To be honest, I think it's different for everyone. I know our principal had kind of a bad reputation... [PS 2] has a really bad reputation." Nevertheless, she said that the principal encourages and supports new initiatives and "tells us all the time, all of us, how great we are, how great we're doing and.... 'You guys were chosen. You guys are here for a reason.'"

Olivia discussed how, initially she

taught with a really wonderful, incredible partner, who was very aligned philosophically.

We had such a similar vision and similar demeanor and similar beliefs about children, and so we kind of created our own little island and I didn't realize it was an island until she was gone and then I was like, "Wow." Yup. I was really—I think I was very protected from more of a traditional public school culture.

Because of the geographical location of the school, teachers commuted from two separate boroughs, whose populations hold different values and socio-economic statuses. Consequently, the two groups of teachers had conflicting pedagogies, which created divisions in the school and brought resistance to some of the progressive initiatives introduced by leadership and teachers like Olivia. She said that this lack of unity was "a big reason that I left."

In their attempts to expand student-centered practices, the schools also supported collaborative teacher groups. Judith was part of a "study group with the other teachers on high quality questions and we're just looking at norms and different protocols...we talk about making sure that every student is included." As previously mentioned, Olivia took part in a literacy focus

group called Literacy Liaisons. Marsha and Shirley said that they meet with their fellow third grade teachers every week to determine “‘How are we dealing with this aspect?’ Or planning field trips that we’re doing as a grade that go with the curriculum...What are we going to do as homework for the next week so we are unified as a grade?” Lastly, Madeline had a role as a model teacher, in which she mentored five junior teachers and kept her classroom open for observations. She expressed frustration that the initiative had not been more supported by the administration, saying “I’ve done everything I can to structure it, but it really needs to come from admin... Hopefully next year it’s something that I can make it more structured.”

Physical. Judith was the only teacher who discussed the effects of physical resources on her ability to engage in VTS. She did not have her own classroom, so she used an “art cart” to travel to her students’ classrooms once a week. This physical limitation made it challenging to create and maintain classroom rules and norms, which she thinks are prerequisites to engaging in fruitful VTS discussions.

Infrastructural

The final layer of complexity that is addressed in the hierarchy of codes is the infrastructural layer. The infrastructural section examines interview and annotation transcripts in relation to school and district-level accountability measures, which includes PD requirements. Teachers’ awareness and concern regarding accountability measures are also addressed in the preceding sections, as infrastructural factors permeate the other layers of complexity.

PD requirements. Teachers in BCSD are required to obtain a certain number of PD credits every year. As *VTSB* counted towards these required credits, this was a factor in teachers’ decisions to enroll in and complete the course. Shirley said, “honestly, yeah. CTLE credits are very motivating...I think the CTLE credits are keeping us to stick with it...I think having the

carrot keeps us to stick with it.” Olivia and Judith thought that the PD credit hours would attract their peers to enroll in the course as indicated in Olivia's statement that, “In my head, I was like, ‘Oh, who doesn’t want that?’”

Revised Program Theories

At the end of this phase, I revised the nine initial program theories to develop five new theories (see Table 12).

Phase Three Overview

In Phase III, I refined my theories from Phase II to create the final program theories. I maintained written and verbal communications with Watershed and continued my review of the literature from the previous two phases, focusing on complexity-consistent substantive theories and their relationship to the program theory elements.

Review of the Literature

I looked more deeply at the concept of trust relationships between teachers and school leadership, how they affect teachers’ role identities, and ultimately how they affect teachers’ motivation to engage in and apply PD. My findings helped me to construct a sixth program theory.

According to Tschannen-Moran (2014), trust is a multifaceted, dynamic construct that mediates relationships. Trust is particularly important in interdependent relationships, such as those between teachers and principals, in which both parties rely on each other to make the

Table 12

Revised CMOCs and Program Theories

	#	CONTEXT	MECHANISM	OUTCOME
INDIVIDUAL	1	Leadership sends information about PD opportunities, including <i>VTSB</i> , to teachers. Teachers with student-centered pedagogies and instructional goals perceive <i>VTSB</i> course content as aligning with these beliefs and goals.	Because they perceive <i>VTSB</i> content as relevant to their instructional beliefs and goals, teachers are motivated to engage in the course as a means of developing student-centered instructional strategies and achieving student-centered instructional goals.	Teachers enroll in <i>VTSB</i> course, anticipating the development of student-centered instructional strategies and achievement of student-centered instructional goals.
		<i>When teachers with student-centered pedagogies and instructional goals perceive VTSB course content as aligning with their beliefs and goals, they enroll in the course because they perceive it as a relevant means of developing student-centered instructional strategies and achieving their instructional goals.</i>		
	2	Teachers experience dissonance between their current practice and their potential practice while engaging in <i>VTSB</i> course activities.	Teachers identify strategies within <i>VTSB</i> to help bridge the gap between their current practice and their potential practice.	Teachers plan how to incorporate identified VTS into their practice to bridge the gap between their current practice and possible practice.
		<i>When teachers who are engaging in VTSB course activities experience dissonance between their current practice and their potential practice, they identify strategies within the course to help bridge this gap and plan how to incorporate these strategies into their practice.</i>		
INTERPERSONAL	3	Leadership evaluates teacher practice based on district standards. They encourage school and department-level initiatives to cultivate instructional practices that align with district standards.	Teachers identify alignment between <i>VTSB</i> content, district standards, and goals of school/department initiatives. They perceive the incorporation of VTS into instruction as a means of achieving school-wide or department-level goals.	<i>VTSB</i> participants share VTS with colleagues and leadership and try to integrate them into existing initiatives.
		<i>When leadership encourages school and department-level initiatives to cultivate instructional practices that align with district standards used in teacher evaluations, teachers identify alignment between VTSB content, district standards, and goals of the school and department-level initiatives. They perceive the incorporation of VTS into instruction as a means of achieving school-wide and department-level goals, they share VTS with colleagues and leadership and try to integrate them into existing initiatives.</i>		

Table 12. Continued

	#	CONTEXT	MECHANISM	OUTCOME
INSTITUTIONAL	4	Situational factors affect teachers' ability to implement both mandated curricula and VTS within mandated timeframes.	Teachers experience tension between incorporating mandated curricula and VTS into their instruction within mandated timeframes.	Teachers incorporate VTS into instruction, but adapt or integrate it to support the implementation of their mandated curricula within mandated timeframes.
		<i>When situational factors affect teachers' ability to implement both mandated curricula and VTS within mandated timeframes, teachers experience tension and incorporate VTS into their instruction, but adapt or integrate it to support the implementation of their mandated curricula within mandated timeframes.</i>		
	5	Teachers have diverse classrooms and daily instructional routines.	Teachers struggle to identify with teachers, students, and instructional contexts in VTSB example videos.	Teachers integrate VTS into daily routines and adapt it in ways that accommodate their students' diverse needs.
		<i>When teachers have diverse classrooms and daily instructional routines, they struggle to identify with teachers, students, and instructional contexts in the VTSB examples videos, and adapt VTS in ways that accommodate their students' diverse needs.</i>		

school function properly. Principals trust teachers as educational professionals to educate students and teachers trust principals to support them in this process. Trusting relationships between teachers and principals increase teacher motivation, professionalism, and efficacy (Tschannen-Moran & Gareis, 2014; Tschannen-Moran, 2009; Angelle, Nixon, Norton, & Niles, 2011).

Five Facets of Trust

The five facets of trust are: benevolence, honesty, openness, reliability, and competency. Tschannen-Moran (2014) defines characteristics of these facets in relation to teachers and principals, as represented in Table 13.

Benevolence. Benevolence is a fundamental component of trust; teachers are “more willing to go the extra mile” (Tschannen-Moran, 2014, p. 23) for leaders who demonstrate personal care and appreciation for them not only as employees, but also as holistic individuals. Madeline discussed how her principal encouraged the faculty, “he tells us all the time, all of us, how great we are, how great we’re doing and he’s really like, ‘You guys were chosen. You guys are here for a reason.’” She also described how he had encouraged her individually and validated her identity role as a professional teacher, saying “in my life, I’ve never felt that I’m good at what I do or that anyone believes in me and this principal, I don’t know if it’s his strategy or what, I’ve always felt like if someone makes you feel good, for me it makes me want to work harder.”

Honesty. Honesty includes matching words and actions. When leadership expresses interest in or commits to doing something, they need to follow through in order to maintain trust. Olivia described how the school had a “purported focus on discussion, accountable talk and you know, inquiry approaches to learning” but failed to support her attempts to promote student-

Table 13

Five Facets of Trust

Facet	Characteristics
Benevolence	Caring, extending goodwill, demonstrating positive intentions, supporting teachers, expressing appreciation for faculty and staff efforts, being fair, guarding confidential information
Honesty	Showing integrity, telling the truth, keeping promises, honoring agreements, being authentic, accepting responsibility, avoiding manipulation, being real, being true to oneself
Openness	Maintaining open communication, sharing important information, delegating, sharing decision making, sharing power
Reliability	Being consistent, being dependable, showing commitment, expressing dedication, exercising diligence
Competency	Buffering teachers from outside disruptions, handling difficult situations, setting standards, pressing for results, working hard, setting an example, problem solving, resolving conflict, being flexible

Adapted from *Trust matters: Leadership for successful schools* (p. 39), by M. Tschannen-Moran, 2014, San Francisco, CA: Jossey Bass. Copyright 2014 by Megan Tschannen-Moran.

centered initiatives, such as VTS. Olivia recounted how the facilitator had responded when she told her Literacy Liaisons group about VTS: ““Oh this sounds cool. Do you want to tell us more about it or share something from it the next time the Liaisons meet?”” Olivia followed up with an email to the group and the principal, who had attended that meeting. However, Olivia said that there was no further response, adding to her resolve to leave the school for a new teaching position.

Openness. In trusting teacher-leader relationships, the teacher feels comfortable discussing both successes and failures with her leaders; she feels that they are approachable and will provide constructive feedback. Madeline tried different ways to get more students to participate in VTS. After an unsuccessful attempt, she spoke to her AP about her goals and potential strategies to achieve those goals. Her AP gave her a suggestion for incorporating VTS into a writing activity, which she later successfully implemented. Because she felt comfortable approaching her AP and trusted her to give her honest feedback, Madeline was able to improve her instruction.

Reliability. Teachers want to know that they can count on their leadership to consistently uphold and model certain standards. Shirley and Marsha’s school leadership had established a track-record of supporting student-centered initiatives like VTS. Shirley said, “when we first approached the maker space this year, one of our assistant principals was right in there, completely getting her hands dirty with everybody. Not just sitting back and watching. They’re in there with us. And actually, it was another administrator who pointed us to the whole VTS course in the first place.” Thus, they were confident experimenting with VTS in their classroom, knowing that their leadership would support their efforts.

Competency. Teachers not only want leaders to trust them as professionals, but they also want to trust leaders to fulfill their own responsibilities. Olivia described how the leadership at her school “aspired for it to be student-centered” but struggled to achieve this goal due to the large size of the school, the percentage of SPED and ELL students, and divisions in faculty members’ pedagogies. However, the administration’s job is to work with the student and faculty populations that exist at the school. The failure of Olivia’s administration to do so demonstrated a lack of competency that affected her level of trust in their ability to implement and sustain student-centered initiatives.

Trust and Power Dynamics

There is inherently a “hierarchical nature of the relationships within a school” (Tschannen-Moran, 2014, p. 41), with administrators occupying positions of greater power than the teachers. This dynamic can result in distrust if teachers think administrators are more concerned about school performance than supporting and equipping teachers to educate students. This kind of environment emphasizes the employee-employer relationship dynamic, which increases teachers’ anxiety and pressures them to focus more on meeting a checklist of criteria than on designing student-centered lessons. For example, Judith frequently expressed concerns about delivering “what the administration wants to see in each lesson” She was hesitant to implement VTS more frequently because, “if I was going to be observed, VTS wouldn’t be exactly what they want to see. So, I don’t know if I could do it here.”

Alternatively, schools in which there is distributed leadership and shared decision-making are more likely to cultivate trust between teachers and administrators, leading to increased teacher motivation, professionalism, and efficacy (Tschannen-Moran & Gareis, 2014; Tschannen-Moran, 2009; Angelle et al., 2011). In this kind of an environment, teachers do not

feel like they are operating in an employee-employer relationship, but a professional-professional and collegial relationship. Madeline described how her school trusted teachers to make instructional decisions, even if every lesson was not “rigorous” or aligned to specific

They really believe that whatever the teacher thinks is best. Like if you need an extra day to review something. I had an organization day earlier in the year and they’re not like, ‘Oh my gosh. That’s not a rigorous lesson.’ They’re really supportive of all kinds of ways to reach students.

This made Madeline feel like she was being treated as a professional and gave her confidence to engage in initiatives like *VTSB*.

Finalized Program Theories

At the end of this phase, I used data from Phases I, II, and III to refine the five theories from Phase II. I also added one program theory, based on my review of the literature in Phase III and comprehensive review of data from Phases I, II, and III.

This chapter summarized the data collected and analyzed during the three phases of the study, which involved developing program theories (Phase I), testing and revising program theories (Phase II), and refining and finalizing program theories (Phase III). Ultimately, I finalized six program theories, which I use in the next and final chapter to help summarize major findings, present middle-range theories and their implications, and recommend future research possibilities.

Table 14

Finalized CMOCs and Program Theories

	#	CONTEXT	MECHANISM	OUTCOME
INDIVIDUAL	1	Leadership sends information about PD opportunities, including <i>VTSB</i> , to teachers. Teachers with student-centered pedagogies and instructional goals perceive <i>VTSB</i> course content as aligning with these beliefs and goals.	Because they perceive <i>VTSB</i> content as relevant to their instructional beliefs and goals, teachers are motivated to engage in the course as a means of developing student-centered instructional strategies and achieving student-centered instructional goals.	Teachers enroll in <i>VTSB</i> course, anticipating the development of student-centered instructional strategies and achievement of student-centered instructional goals.
		<i>When teachers with student-centered pedagogies and instructional goals perceive VTSB course content as aligning with their beliefs and goals, they enroll in the course because they perceive it as a relevant means of developing student-centered instructional strategies and achieving their instructional goals.</i>		
	2	Teachers experience dissonance between their current practice and their potential practice while engaging in <i>VTSB</i> course activities.	Teachers identify strategies within <i>VTSB</i> to help bridge the gap between their current practice and their potential practice.	Teachers plan how to incorporate identified VTS into their practice to bridge the gap between their current practice and possible practice.
		<i>When teachers who are engaging in VTSB course activities experience dissonance between their current practice and their potential practice, they identify strategies within the course to help bridge this gap and plan how to incorporate these strategies into their practice.</i>		
INTERPERSONAL	3	Leadership evaluates teacher practice based on district standards. They encourage school and department-level initiatives to cultivate instructional practices that align with district standards.	Teachers identify alignment between <i>VTSB</i> content, district standards, and goals of school/department initiatives. They perceive the incorporation of VTS into instruction as a means of achieving school-wide or department-level goals.	<i>VTSB</i> participants share VTS with colleagues and leadership and try to integrate them into existing initiatives.
		When leadership encourages school and department-level initiatives to cultivate instructional practices that align with district standards used in teacher evaluations, teachers identify alignment between <i>VTSB</i> content, district standards, and goals of the school and department-level initiatives. They perceive the incorporation of VTS into instruction as a means of achieving school-wide and department-level goals, they share VTS with colleagues and leadership and try to integrate them into existing initiatives.		

Table 14. Continued

	#	CONTEXT	MECHANISM	OUTCOME
INSTITUTIONAL	4	Situational factors affect teachers' ability to implement both mandated curricula and VTS within mandated timeframes.	Teachers experience tension between incorporating mandated curricula and VTS into their instruction within mandated timeframes.	Teachers incorporate VTS into instruction, but adapt or integrate it to support the implementation of their mandated curricula within mandated timeframes.
		When situational factors affect teachers' ability to implement both mandated curricula and VTS within mandated timeframes, teachers experience tension and incorporate VTS into their instruction, but adapt or integrate it to support the implementation of their mandated curricula within mandated timeframes.		
	5	Teachers have diverse classrooms and daily instructional routines.	Teachers struggle to identify with teachers, students, and instructional contexts in <i>VTSB</i> example videos.	Teachers integrate VTS into daily routines and adapt it in ways that accommodate their students' diverse needs.
		When teachers have diverse classrooms and daily instructional routines, they struggle to identify with teachers, students, and instructional contexts in the <i>VTSB</i> examples videos, and adapt VTS in ways that accommodate their students' diverse needs.		

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

Summary

The primary purpose of this study was to determine the extent to which the desired outcomes for select K-12 teachers' participation in Watershed Collaborative's K-12 OPD course, *VTSB*, were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. A secondary purpose of the study was to set a precedent for future studies of K-12 OPD courses, particularly those using a RE approach. Using a RE approach for this study provided a systematic way to determine not only the extent to which participants were applying course content, but also *how* and *why*, including why there were different outcomes for different participants.

The study was conducted in three phases: developing program theories (Phase I), testing and revising program theories (Phase II), and refining and finalizing program theories (Phase III). The specific guiding questions for the study included:

What works?

- Which of the desired outcomes (e.g., teachers develop the ability to adapt VTS for diverse groups of students, teachers implement VTS in the classroom) are achieved and to what extent?
- What are the unintended outcomes of the course?

For whom?

- To what extent do participants/sub-groups of participants (e.g., teachers in different locations or who completed different stages of the course) benefit from this course, as indicated by achieving desired or unintended outcomes?

In what circumstances?

- What contextual factors influence the achievement of desired outcomes and to what extent (e.g., level of administrative support, relevance to instructional goals)?

How?

- What types of participant reactions/responses do course resources (e.g., course interface, structure) and activities (e.g., course annotations, reflection, applying content) elicit? To what extent do these responses relate to the achievement of desired outcomes?

As mentioned in Chapter 1, these guiding questions cannot be answered separately, as they are interdependent; the answer to “what works” depends on the answers the other questions. Therefore, the answers to these questions, which constitute my major findings, are presented in this fifth and final chapter in the form of six final program theories, which describe the relationships between the identified contextual factors, causal mechanisms, and outcomes (see Table 14). This chapter also provides five middle range theories with implications for K-12 OPD course developers, evaluators, facilitators, school leadership, and course participants, and concludes with recommendations for future studies in this area of research. In Chapter 4, I discussed the different contextual factors that triggered various causal mechanisms to generate multiple *VTSB* course outcomes. I grouped these contextual factors into four categories: individual/intrapersonal, interpersonal, institutional, and infrastructural, which represent the layers of the stratified social context in which *VTSB* course is embedded. My final six program theories include two in each category. In this section, I discuss six major findings of this study, as represented by the six finalized program theories (PTs). The findings are organized according to their corresponding contextual categories.

Major Findings and Discussion

Individual/Intrapersonal

Program theory 1. *When teachers perceive the VTSB course as aligning with their pedagogical beliefs and objectives, they also perceive it as an action possibility for achieving their instructional goals and are motivated to enroll in the course.*

CONTEXT	MECHANISM	OUTCOME
Leadership sends information about PD opportunities, including <i>VTSB</i> , to teachers. Teachers with student-centered pedagogies and instructional goals perceive <i>VTSB</i> course objectives and content as aligning with their pedagogical beliefs and instructional goals.	Teachers' perceive engaging in <i>VTSB</i> as an action possibility for achieving their instructional goals.	Teachers with student-centered pedagogies and instructional goals are motivated to enroll in the course, anticipating the achievement of student-centered instructional goals.

All of the teachers I interviewed indicated that they found out about the *VTSB* course via emails from their leadership. Thus, leadership's dissemination of information about *VTSB* was a pre-requisite to teachers enrolling in the course. However, although they all received this information, not every teacher in these schools enrolled in the course. This first PT discusses which teachers were motivated to enroll in *VTSB*, under what circumstances, and why.

The teachers I spoke with stated that when they read the information about *VTSB*, they perceived the course objectives and content as aligning with their pedagogical beliefs and instructional goals. As discussed in Chapter 4, the teachers' pedagogical beliefs align with key constructivist principles, including student-centered learning, co-construction of knowledge, connections to prior knowledge, teacher facilitation, and not eliciting "right/wrong" answers, which are fundamental to VTS and are emphasized throughout the *VTSB* course. Teachers' instructional goals were related to their pedagogical beliefs, in that they desired to develop more student-centered instructional strategies, make classroom discussions more inclusive, and

encourage more divergent thinking, as opposed to focusing on providing the “right” answers; these topics are also addressed in *VTSB*.

Consequently, when teachers perceived their beliefs and goals as aligning with course objectives and content, they saw engaging in the course as an action possibility for achieving their instructional goals and were motivated to enroll in the course. Their role identity (RI) and goals as teachers aligned with their RI as learners in *VTSB*. This aligns with the DSMRI perspective that “teachers are more likely to endorse a new practice as a viable action possibility and enact it in their practice when they have knowledge of it and perceive it to correspond with their unique system of beliefs, purpose and goals” (Kaplan et al., 2015, p. 9). This PT is also consistent with Evans’ (2008, 2011) and Opfer and Pedder’s (2011) findings that teachers are often most likely to engage in PD courses that align with their beliefs and attitudes about teaching and learning. In their systematic review of coherence in PD programs, Lindvall and Ryve (2019) also write extensively about the importance of coherence between teachers’ beliefs and goals and PD content as a critical feature of effective PD. Alternatively, teachers whose beliefs and goals are not consistent with a PD program, such as the teacher-centered instructors at Olivia’s school, may be more resistant to initiatives like VTS.

Program theory 2. *When teachers engage in VTSTB, their pedagogical discontentment increases in specific areas. Teachers identify strategies within the course that will realign their self-perceptions and goals in these areas and integrate them into their practice.*

CONTEXT	MECHANISM	OUTCOME
<i>VTSTB</i> content highlights misalignments between teachers’ self-perceptions of their current instructional practices, classroom environments, and management skills and their goals in these areas.	Teachers’ pedagogical discontentment increases regarding their current instructional practices, classroom environments, and management skills, but they have high self-efficacy in their ability to engage in new initiatives.	Teachers decide to search for strategies within the <i>VTSTB</i> course that will help bridge the gap between their current instructional practices, classroom environments, and management skills and their goals in these areas.

Through their interviews and course annotations, teachers expressed pedagogical discontentment in various areas, including their instructional practices, classroom environments, and management skills. They often compared these areas of discontentment in their personal practice to examples from the *VTSB* course that demonstrated instructional practices, classroom environments, and management skills to which they aspired. For example, when watching one of the example VTS sessions, Madeline was impressed by the number of hands being raised and the intellectual quality of the students' responses. She also discussed how she was not satisfied with her current ability to engage all of her students, including those with varying intellectual abilities and levels of confidence, in class discussions. Thus, the video highlighted the misalignment between her self-perceptions of her current practice and goals as a teacher, increasing her pedagogical discontentment and instilling a desire to identify strategies within the course that would help her achieve her goals. However, she maintained a high self-efficacy in her ability to implement new initiatives as she received personal support and encouragement from her principal and assistant principal.

Marsha, Shirley, and Judith considered the *VTSB* example discussions as models, or ideals, for how to conduct inclusive class discussions and compared it to their own practice. For instance, Shirley reflected that, "giving the children time to think before the conversation begins seems obvious when watching the videos, though sometimes harder in the moment in a classroom, when we worry that the void can create side-conversations that are often not constructive." The videos highlighted differences between her current, practices, classroom environment, and management skills and those to which she aspired, as exhibited in the videos. Nevertheless, she and Marsha had a high self-efficacy in their ability to implement new initiatives based on mastery experiences.

Consequently, while engaging in the course, teachers looked for strategies that addressed their specific areas of discontentment and would help bridge the gap between their current and desired instructional practices, classroom environments, and management skills. This would help realign their self-perceptions and goals in these areas. For example, after watching a third-grade example discussion, Madeline wrote, “I need to practice doing this. I should practice showing the same emotion/facial expression with illogical and outstanding answers.” Other teachers discussed how they were looking for ways they could use VTS techniques to differentiate instruction and accommodate ELL and SPED students while challenging gifted learners.

This PT aligns with the DSMRI concept that teachers with a conflicted RI structure would experience uncertainty, ambiguity, or tension with regard to certain goals and practices and would be motivated to resolve these tensions. Thus, variation in the structure of teachers’ RI would predispose them to exhibit more or less contextual and pedagogical discontentment... (Kaplan et al., 2015, p. 10).

Appova and Arbaugh (2018) relate pedagogical discontentment to the construct of possible selves, which, along with pedagogical discontentment, is an element in the self-perceptions and definitions component of the DSMRI. They quote Markus and Nurius (1986), who devised the idea of possible selves, which they

viewed as cognitive bridges between the present and future, specifying how individuals may change from how they are now to what they will become. When certain current self-conceptions are challenged or supported, it is often the nature of the activated possible selves that determines how the individual feels and what course the subsequent action will take. (Markus and Nurius, 1986, p. 961)

Appova and Arbaugh (2018) discuss how teachers with pedagogical discontentment and high self-efficacy in their ability to improve their practice saw PD as action possibilities for bridging the gap between their current practice and their future possible selves. This PT aligns with the idea that the combination of pedagogical discontentment and self-efficacy motivates teachers to find ways to improve their practice in specific areas. However, this PT only addresses teachers' decision to seek strategies for improvement; PT 6 addresses the challenges that arose as teachers searched for and experimented with these specific strategies, and how it affected their self-efficacy in their ability to implement VTS strategies while accommodating their students' diverse needs.

Interpersonal

Program theory 3. *When teachers and leadership have trusting relationships, teachers' self-efficacy and motivation as professionals increases and they apply and experiment with VTS in innovative ways.*

CONTEXT	MECHANISM	OUTCOME
Teachers and leadership have established trust relationships and teachers are granted autonomy in their instructional decision-making.	Teachers' self-efficacy and motivation in their role as trusted professionals increases.	Teachers apply and experiment with VTS in the classroom in innovative ways.

As discussed in Chapter 4, trust is a multifaceted, dynamic construct that mediates relationships. The five facets of trust are: benevolence, honesty, openness, reliability, and competency (Tschannen-Moran, 2014). The teachers who were most comfortably and frequently applying VTS indicated that they had established trusting relationships with the leadership in their school and were granted autonomy in their instructional decision-making. Madeline described how her principal encouraged the faculty at her school and how she felt comfortable approaching her assistant principal with questions about instructional successes and failures in

her classroom. Shirley and Marsha described how they could rely on their assistant principal to “get in the trenches” with them. The treatment from leadership in these relationships increased teachers’ self-efficacy and motivation in their RI as professional educators, as their self-perceptions of their abilities as teachers were affirmed by trusted authority figures. As Madeline said, “I’ve always felt like if someone makes you feel good, for me it makes me want to work harder.” Thus, they felt motivated and free to apply and experiment with VTS in the classroom in innovative ways.

Conversely, teachers who did not have trusting relationships with leadership were less frequently and less comfortably applying new strategies in the classroom. Olivia’s administration had a “purported focus” on student-centered instruction, but did not support her efforts to promote *VTSB*, which she considered a student-centered initiative. Additionally, they did not display competency in their ability to enforce school-wide student-centered initiatives. These factors contributed to her decision to leave the school. Judith seemed overly concerned about meeting administration’s evaluation criteria, to the detriment of implementing student-centered VTS sessions in her classroom. In these instances, teachers’ goals were in tension with the level of support they received from their administration, decreasing their perceptions of VTS as an action possibility in their classrooms. In Judith’s case, there was also misalignment between her RIs as a teacher and an employee; her student-centered self-perceptions and goals as a teacher did not align with the expectations of her as an employee. Consequently, she and Olivia alleviated this tension by deciding not to implement VTS as frequently in their classrooms, despite a desire to do so.

Teachers’ RIs as professional educators are validated when leaders provide verbal encouragement, maintain a “balance between accountability and agency” (Webster-Wright,

2990, p. 727). (i.e., distributed leadership), encourage experimentation, and indicate that teachers' PD selections align with their priorities as both educators and employers (Desimone & Garet, 2015). This PT aligns with the DSMRI in that teachers "[experienced] a sense of identification with their vocation and clarity with regard to goals and courses of action (Kaplan et al., 2015, p.10) when their RIs as professional educators were affirmed by leadership.

This PT also aligns with Tschannen-Moran's description of trust and its effects in relationships between teachers and their leadership. When teachers' RI as professionals are validated through trusting relationships with leadership, they are "more willing to go the extra mile" (Tschannen-Moran, 2014, p. 23) for those leaders. Alternatively, when teachers do not feel trusted as professionals, they are less likely to apply PD strategies in their classrooms (Appova & Arbaugh, 2018).

Program theory 4. *Teachers who are members of collaborative peer groups perceive alignment between VTSB content and department/school goals and initiatives and see incorporation of VTS into department/school initiatives as action possibilities for achieving shared goals. They encourage peer groups to incorporate VTS into existing department/school initiatives.*

CONTEXT	MECHANISM	OUTCOME
Teachers who are members of collaborative peer groups that develop initiatives to achieve departmental and school goals perceive alignment between <i>VTSB</i> content and department/school goals and initiatives.	Teachers perceive the integration of VTS into departmental and school initiatives as action possibilities for achieving departmental and school goals.	<i>VTSB</i> participants share VTS with peer groups and encourage its integration into existing departmental and school initiatives.

The teachers I interviewed were all involved in collaborative peer groups that worked towards achieving departmental and school-level goals through various initiatives. The groups created protocols, planned instruction, mentored junior teachers, and facilitated PD opportunities.

When engaging in *VTSB*, the teachers perceived the course objectives and content as aligning with departmental and school-wide goals and initiatives and considered integrating VTS into these initiatives as action possibilities for achieving departmental and school goals. Olivia, Madeline, and Judith presented information about *VTSB* to the colleagues in their collaborative peer groups, encouraging them to take part in the course and integrate VTS into existing departmental and school initiatives, such as the TIDE and rigor initiatives at Madeline's school or "some of the teaching moves that we already use in reading writing and math" at Olivia's school. These efforts did not gain much traction with the teachers at either school, but when asked if she was going to continue trying to recruit teachers next year, Madeline said, "I don't see why not. I mean, I think the fact that my principal kind of like signed off on it—she wanted everyone to take the course..." Thus, despite peer responses, she still considered VTS as an action possibility for meeting departmental and school goals because she perceived alignment between VTS and departmental and school goals and initiatives.

The underlying causal mechanism for this PT is the same as PT 1, but at a different level of complexity (i.e., interpersonal vs. Intrapersonal). At the interpersonal level, this PT aligns with Lipowsky & Rzejak's finding that teachers may perceive PD as more relevant "when the focus of the training is congruent with current, ongoing school-related reform processes" (2014, p. 31). The teachers attempted to recruit their peer groups to *VTSB* so that they would start integrating VTS into departmental and school initiatives, which they perceived as action possibilities for achieving departmental and school goals.

While I did not gather information about this at the infrastructural level, it would be interesting to see if the same causal mechanism also applied at the level of state and national initiatives.

Institutional

Program theory 5. *As teachers must meet curricular and instructional expectations within mandated timeframes, they experience tension when also trying to incorporate VTS into their instruction. As a result, they incorporate VTS into their instruction, but prioritize meeting curricular and instructional expectations within mandated timeframes.*

CONTEXT	MECHANISM	OUTCOME
Teachers must meet curricular and instructional expectations within mandated timeframes.	Teachers experience tension when trying to meet curricular and instructional expectations and incorporate VTS into their instruction within mandated timeframes.	Teachers incorporate VTS into their instruction, but prioritize meeting curricular and instructional expectations within mandated timeframes.

The teachers described how they had to plan and organize their instruction to meet curricular and instructional expectations within mandated timeframes. Madeline and Judith discussed how they only had forty-five-minute class periods, so they felt they need to be very structured in order to “fit in” all of the things they needed to address. Their schools also have initiatives, such as TIDE as Madeline’s school, and open-ended, quality questioning at Judith’s school, that they had to address in their instruction. Marsha and Shirley worked with all of the other third grade teachers to plan instructional units, which Shirley said “are pretty well defined.” They met weekly to discuss their progress and ensure they were “unified as a grade.” Furthermore, all of the teachers had to work according to the school district’s official calendar, so events such as state testing and vacation periods affected their instructional planning as well. These time constraints and curricular requirements affected teachers’ ability to incorporate VTS into their instruction.

As previously discussed, teachers perceived alignment between curricular and instructional requirements and *VTSB* course objectives and content (PT 1 and 4). However, they

were still experimenting with how to incorporate VTS into their instruction in ways that would meet these requirements. Experimenting with new instructional strategies takes additional time and continual revision based on trial and error; the teachers needed time to engage in iterative cycles of experimenting, reflecting, and revising these new strategies. While teachers “made time” to engage in these cycles, they remained unsure which ways of implementing VTS would accomplish their lesson objectives as effectively and efficiently as the techniques they were used to implementing. Madeline, Shirley, and Marsha discussed using VTS for warm-up activities, but wanted to know how to better integrate it into the body of their lessons. Madeline also experimented with using VTS in activities regarding their class novels; she said that students responded well, but that it would take time to shift their mindsets to the kind of open-ended, creative thinking that VTS elicits.

As a result, there was misalignment between their identities as student-centered teachers, as learners engaged in cycles of experimentation, reflection, and revision, and as school employees with certain responsibilities and expectations. Teachers experienced tension between wanting to incorporate VTS into their instruction, needing time to experiment, and needing to ensure that their instruction equipped students to meet certain curricular requirements within specific timeframes. As Judith expressed, “There’s a lot of stuff we have to fit in and I only have forty-five minutes. There’s a whole bunch of stuff that I have to fit in there.” Ultimately, teachers prioritized meeting their instructional and curricular requirements over incorporating VTS into their instruction. They did not find the two mutually exclusive; as Kaplan et al. state, “For a teacher to apply innovative practices despite pressures to do otherwise, and to persist in the vocation despite unfavorable conditions, these actions must cohere with their core beliefs, values, self-perceptions, self-definitions, and goals for teaching” (2015, p. 3). Rather, they

thought that VTS would eventually help them to meet these requirements. However, time constraints affected their ability to experiment with VTS to determine the most effective ways to incorporate it into their instruction. Thus, they tried to integrate their RIs as teachers, learners, and employees, but gave priority to their RI as employees.

Desimone and Garet (2015) reviewed several studies in which PD was found to be more effective when the pacing and content of the PD aligned with the pacing and content of teachers' curricula. Similarly, Lipowsky & Rzejak (2014) state that "research has clearly shown that effective professional development is content-focused and concentrates on domain-specific topics or domain-specific student competencies" (p. 35). Thus, perhaps increased alignment between the content and pacing of teachers' curricula and PD would alleviate some of the tension between teachers' RI as teachers, learners, and employees. However, as Webster-Wright (2009) states,

PD which is content specific is not enough to ensure productive teacher learning. Our findings suggest that content-specific PD needs to be differentiated to specifically address and accommodate the differences in teachers' professional and learning needs (e.g. topics/courses taught, curriculum and textbooks, years of teaching experience) as well as the differences in the student populations that teachers serve and teach (e.g. grade-levels, advanced/ honors' students, students with special needs) (p. 17).

This issue is addressed in PT 6.

Program theory 6. *When teachers need to accommodate the needs of diverse groups of students in their daily instruction, they struggle to identify with the teachers, students, and instructional contexts in VTSTB example videos, so they integrate VTS into their instructional routines by adapting it in ways that accommodate their students' diverse needs.*

CONTEXT	MECHANISM	OUTCOME
Teachers need to accommodate the needs of diverse groups of students in their daily instruction.	Teachers struggle to identify with the teachers, students, and instructional contexts in <i>VTSB</i> example videos.	Teachers integrate VTS into daily instruction, but have low self-efficacy in their ability to implement VTS in ways that accommodate their students' diverse needs,

Teachers described their desire to make their classrooms more inclusive, accessible, and engaging for all of their students, including those from different ethnic backgrounds, ELL students, SPED students, gifted students, and students with varying levels of motivation and self-efficacy. They identified specific ways in which they thought VTS would be beneficial for these students and ultimately perceived the application of VTS as an action possibility that aligned with their goals for creating a more inclusive, accessible, and engaging classroom.

However, some of the teachers were unsure of how to apply VTS in specific ways that would achieve these results for their diverse groups of students, as they struggled to identify with the teachers, students, and instructional contexts in the example videos. Marsha, Shirley, and Madeline noted that, while they understood that the videos were meant to be models, they were also not representative of their classrooms. Both work in schools with a high-minority student population, in which a significant percentage of students are economically disadvantaged, and there are many ELL and SPED students. Reflecting on one of the course videos, Madeline said that the ethnic composition of the students in the video were very different from that of her own classroom. Marsha and Shirley said that the formal and orderly classroom culture in another one of the videos did not resemble the nature of their classroom, in which students required more redirection.

Nevertheless, the teachers still found VTS to be relevant to their instructional goals, so they adapted it in ways they thought would best accommodate their students' diverse needs. For

example, Marsha and Shirley used images from their students' cultures and integrated it into their curriculum for Black History Month. However, they still had many questions about how to better integrate it into their instruction. Madeline used heterogeneous grouping strategies in which students practiced paraphrasing each other's comments. She thought it went well but her comment about the activity that, "I don't know if this is bad," indicated her uncertainty of whether this was an appropriate way to apply VTS. Therefore, teachers adapted VTS to fit into their daily instruction, but they did not have high self-efficacy in their ability to implement VTS properly and still accommodate their students' diverse needs. Their self-perceptions did not align with their goals to meet students' needs, creating tension. However, they tried to alleviate this tension by integrating their roles as teachers of diverse students and learners in *VTSB* to the best of their ability.

This PT aligns with Webster-Wright's finding cited in the discussion on PT 5, as well as Desimone and Garet's finding that "in both the study and design of PD, we would be well-served to take into account certain conditions that commonly occur in urban school contexts in the U.S." (2015, p. 256).

Middle-Range Theories (MRTs) and Implications

I derived five middle-range theories from the elements of the finalized theories. MRTs can be transferred to programs operating in similar contexts or with similar interventions, as they are neither universal nor specific to a single context (Pawson & Tilley, 1997; Astbury, 2013). These MRTs have implications for K-12 OPD course developers, evaluators, school leadership, course facilitators, and course participants. K-12 OPD developers, evaluators, and facilitators can reference these theories to help anticipate internal and external contextual factors that may affect participants' engagement in a particular course. Course developers and facilitators can then

design or facilitate courses to mitigate the effects of these factors, sustain teacher engagement, and achieve desired course outcomes. School leadership can use these MRTs to help anticipate and accommodate teachers' needs when engaging in OPD. Finally, teachers can use these theories to anticipate potential challenges for engaging in an OPD course and devise strategies for addressing these challenges.

Individual

MRT 1: Teachers' motivation is sustained through authentic activities that incorporate theory and practice. K-12 teachers are motivated to enroll in courses whose content aligns with their pedagogical beliefs and instructional goals. They perceive their participation in these courses as action possibilities for achieving their goals. However, teachers should be motivated beyond their course enrollment; they must remain motivated throughout the duration of the course or they will not complete it.

Teachers' motivation is sustained when they engage in ongoing, authentic learning activities that integrate theory and practice. As Timperley states, "Ongoing, subsequent engagement is promoted by worthwhile learning activities and by opportunities to negotiate the meaning of existing and new theories and explore their differing impacts on students" (2008, p. 16). Thus, OPD courses should incorporate learning activities that are grounded in theory and can be embedded in teachers' daily practice. As teachers develop and strengthen a theoretical framework to support new instructional strategies, they are better motivated and equipped to apply these strategies throughout and beyond the OPD course.

Teachers' motivation is further sustained when these authentic activities align with or can be easily adapted to the pacing and content of their mandated curricula. At the course outset, teachers could be prompted to pre-plan how to integrate application activities within a specific

instructional unit or units. Alternatively, an expert course facilitator could help teachers modify and align topics and activities at the outset of the course.

MRT 2: The structural components of an effective OPD course seamlessly mediate learners' experiences with course content. K-12 teachers appreciate the flexibility and convenience of OPD courses. They like being able choose when, where, and how to access information. However, this autonomy also requires structure if teachers are to complete courses and achieve desired course outcomes. To achieve these outcomes, teachers' access to and application of content should be the focus of K-12 OPD courses.

The purpose of a course's structural components, such as its mode of delivery, course platform, and organization of content, is to mediate teachers' experiences with course content. Thus, effective course platforms are intuitive, user-friendly, and make content easy to access. Likewise, effective course organization facilitates self-regulated learning by ensuring teachers know what learning activities and assignments to complete, in what order, and how they will be assessed. Units of instruction should be presented in a way that aligns with the course scope and sequence; if course modules are presented linearly, their arrangement should match the desired order of content completion. Standalone modules or resource banks should clearly be designated as such; if teachers need to navigate between modules for an activity or assignment, readily accessible navigation controls can make this process more efficient.

When a course is poorly organized or lacks clear expectations and directions for activities and assignments, teachers may grow confused or frustrated. They may not access certain resources or not complete certain activities and assignments. Ultimately, this restricts their access to content and distracts from their learning experiences. Thus, course requirements and expectations should be presented clearly at the outset of the course to help teachers develop

metacognitive strategies for course completion. Accurately displaying teachers' course completion status can help them to adjust these strategies based on their satisfaction with their progress. Additionally, providing help guides or a live chat option can help prevent confusion and frustration.

Interpersonal

MRT 3: Interpersonal relationships between teachers and leadership are key. K-12 teachers' interpersonal relationships with their leadership can support or constrain their experiences in OPD courses. When school leaders treat teachers as professionals, their motivation and self-efficacy increase. Thus, leaders should trust teachers to make decisions, such as enrolling and engaging in specific OPD courses, that will benefit the school. This validates their RIs as professionals, increases their motivation to engage in OPD courses, and increases their self-efficacy in their ability to apply course content and achieve instructional goals. School leaders should also provide verbal encouragement, demonstrate a personal interest in teachers' OPD progress, and convey a willingness to openly discuss problems teachers encounter when experimenting with OPD application. This helps to establish a professional-professional dynamic as opposed to an employer-employee dynamic, which can be intimidating and demotivating.

Moreover, trusting relationships with school leaders can offset an overall non-supportive or non-collaborative school culture. When teachers perceive that they have support from leadership, they have less reservations about taking risks and experimenting with new instructional strategies from OPD courses. If teachers are unable to recruit colleagues to participate with them in an OPD course, or to apply strategies from an OPD course, support from leadership helps to sustain their motivation and self-efficacy and they are more likely to continue with the course and continue applying course content after the course is complete.

Alternatively, non-trusting relationships with school leaders can offset an overall positive school culture. A school might be considered “student centered” but a lack of trust, support, or an overemphasis on accountability decreases teachers’ motivation and self-efficacy, making them less likely to take risks and experiment with new instructional strategies from OPD courses. They are also less likely to continue implementing OPD course content once the course is completed.

To help sustain these teachers’ engagement, OPD courses could provide course facilitators who are experts in their fields and whom teachers would consider as authority figures. The facilitators would work to establish personal relationships with course participants through discussion boards, online chats, or synchronous video sessions. They would encourage participants throughout the course, answer questions, and provide constructive feedback in a timely manner. This would help motivate teachers to remain active in the course and increase their self-efficacy in their ability to apply course content. For example, Watershed Collaborative recently began offering an option for organizations to customize the VTS course for their specific needs. Watershed tells organizations that it will

help you calibrate VTS Basics to your community of learners, and even situate VTS within larger professional learning objectives. With VTS Basics at its core, your adapted version of the course will bring an ideal mix of video content and resources that best serves the needs of your network and clientele.

Institutional

MRT 4: Collective participation is ideal, but challenging to coordinate. Collaboration enhances teachers’ engagement in K-12 OPD courses and provides the opportunity for them to learn from a variety of perspectives. Furthermore, most teachers prefer collective participation to

individual participation. Nevertheless, collective participation is difficult to coordinate, due to tension between teachers' desire to engage in ODP and their demanding schedules, multiple roles and responsibilities, and accountability requirements. Even when teachers act as ambassadors for an OPD course and try to recruit their peers, there may be little or no response.

OPD courses can address this challenge in several ways. First, they could provide “ambassador” kits that teachers can use to host a mini PD session for their colleagues during a regular departmental or school meeting. The kits would include a short lesson plan and digital materials for the ambassador to facilitate a brief authentic activity using instructional strategies from the course. When teachers get to experience the effectiveness of these strategies themselves, they will be more likely to enroll in the course.

If teachers cannot find any collaborators at their school, another option would be to provide a course option that groups teachers from different schools. This would require a more regimented course schedule, with designated timelines for course topics and deadlines for activities. Consequently, teachers would have to sacrifice some of the flexibility that makes OPD courses attractive.

While also challenging, the course could be integrated into a school-wide initiative. Not only would this facilitate collective participation, but it would create consistency between the application of course-related instructional and behavioral strategies in teachers' classrooms. However, school and department-level leadership would have to fully commit to the initiative, including giving teachers adequate time for participation, checking on teachers' progress, and aligning OPD content with accountability measures.

MRT 5: Differentiation and representation makes content more relevant. Teachers who enroll in K-12 OPD courses have a variety of needs of goals that they expect the course to

meet. Differentiating K-12 OPD course content can help accommodate the needs of teachers who have different levels of experience, teach different grade levels or subject areas, and have classrooms with diverse learners. It also affords the ability for all of these teachers to engage in the same course, so that they can learn from a variety of perspectives.

Differentiation of course content can be accomplished in different ways. In one way, materials and activities differentiated for a variety of subject areas, grade levels, and diverse learners would be made available to every course participant as they progress through each unit of instruction. Teachers would collaborate with teachers from a variety of instructional contexts. A second way would be to use an algorithm to select materials and activities that align with teachers' selections of experience level, grade level, subject area, etc. at the outset of the course. Teachers would engage collaboratively with participants who have similar backgrounds and needs. The course could also incorporate more general questions that would include a greater variety of teachers (i.e., homogenous and heterogeneous grouping). A third option would be to tailor course materials and activities to teachers' needs on an individualized basis. Teachers could discuss their backgrounds and needs with an expert course facilitator, who would use this information to customize their selection of instructional materials and activities and make modifications throughout the course. The facilitator would manually group teachers for collaboration on different discussions and activities.

OPD courses should also strive to represent diverse groups of students and instructional contexts when creating instructional examples. This makes the examples more relevant to teachers' specific needs and classroom contexts, so that they are more effectively able to apply course content. Different aspects of student diversity to consider include ethnicity, English proficiency, special needs, and socio-economic status. Different aspects of instructional contexts

to consider include teacher's level of experience, quality of classroom resources, and level of classroom management. Again, course examples can be differentiated based on teachers' needs and backgrounds using the three options for differentiation in the previous paragraph.

Watershed's recent offering of course customization, as discussed in MRT 3, is an example of how to address issues with differentiation of OPD course content.

Future Research

1. The participants in this study incorporated VTS into English, art, and history-related subject material. However, while VTS is arts-based, it can be applied in any field. In addition to K-12 and Higher Education settings, VTS has been successfully applied in medical schools, law enforcement programs, corporate PD trainings. Thus, it would be interesting to see how teachers apply *VTSB* course material in K-12 courses that prepare students for these kinds of professions, such as math, science, and government courses. How might their needs differ? What kinds of differentiations would benefit them? How would they align *VTSB* course content with their curricula?
2. With the exceptions of the two co-teachers, the participants in this study did not engage in collective course participation. As the course is designed to be completed collaboratively, future studies could explore the effects of collective participation and the extent to which it helps achieve desired course objectives. These findings could be compared to findings from studies like this one, in which course participants engaged in the course individually.
3. All of the participants in this study voluntarily elected to enroll and engage in the *VTSB* course. This raised the likelihood of the course aligning with their pedagogical beliefs and instructional goals. Future studies could examine contexts in which teachers are

required to complete the course, either as part of a school or department-level initiative. Not only would this afford the ability to compare voluntary versus obligatory participation, but it would also afford an opportunity to examine the effects of the course on participants who hold pedagogical beliefs that do not align with the course, such as a more teacher-centered pedagogy.

4. The participants in this study were employed in public schools. Thus, they had to satisfy accountability measures, such as achieving a required number of PD credit hours and scoring well on their teacher evaluations. In addition, they had to follow mandated curricula and mandated timelines that constrained their ability to engage in the course. Future studies might examine how the motivations and experiences of teachers at private schools or homeschool differ from those of public school teachers.
5. Lastly, this study used the DSMRI as a substantive theory to inform a RE approach. However, a study that focuses more exclusively on the DSMRI and changes in teachers' RIs could provide valuable information for OPD designers, teacher educators, course facilitators, and participants. Gathering information about teachers' changing RIs throughout the course could provide "evidence-based insight into possible effective strategies for making the program relevant to teachers' identities and for triggering identity processes that may promote constructive teacher change" (Kaplan et al., 2015, p. 14). Future studies could also examine the effects of teachers using the DSMRI as a reflective, metacognitive tool to help them reflect on their experiences engaging in and applying new OPD content. The DSMRI would help them

in mapping their own RI and considering the interrelations among its components in relation to a new practice, [which] can provide [them] with a powerful scaffold

for a systematic identity exploration process (e.g., “What are your goals for your students? How do your current practices promote your goals? Why do you believe that such practices promote these goals? How might the new practice relate to your goals?”) (2015, p. 15).

Concluding Remarks

With the increasing national focus on K-12 accountability measures, school administrators are relying on PD as a means of improving teacher and student performance. In particular, as technology has become more widely available, OPD has become increasingly prevalent. However, very few evaluative studies of K-12 OPD have been conducted, especially in non-STEM fields. Thus, the primary purpose of this study was to determine the extent to which the desired outcomes for select K-12 teachers’ participation in a specific arts-based OPD were being achieved, for which teachers (or subgroups of teachers), in what types of contexts, and how. A secondary purpose was to set a precedent for future studies of K-12 OPD courses, particularly those using a realist evaluation approach.

Ultimately, six program theories were finalized, from which five middle-range theories were derived. These findings identified relationships between the contextual factors, causal mechanisms, and outcomes from the *VTSB* course. They indicated that teachers engaged in, applied, and shared course content when the content aligned with their pedagogical beliefs, when they experienced pedagogical discontentment, and when they had strong interpersonal relationships with school leadership. They also indicated that teachers’ engagement in the course was constrained by having to complete mandated curricula within mandated timeframes, when there was an over focus on accountability measures, and when course examples were not representative of their diverse students and instructional contexts.

These findings can be used by K-12 OPD course designers, facilitators, evaluators, and school leadership to anticipate contextual factors that may affect teachers' engagement in OPD courses. This study can also be used by researchers interested in exploring or adopting a RE approach for future studies of K-12 OPD courses. The final program and middle-range theories serve as valuable contributions to a growing body of evidence that can inform future K-12 OPD REs.

LIST OF REFERENCES

- Adams, M., Foutz, S., Luke, J., & Stein, J. (2006, September). *Thinking through art Isabella Stewart Gardner museum school partnership program: Year 3 preliminary research results*. Retrieved from <http://vtshome.org/wp-content/uploads/2016/08/3Gardner-Museum-Thinking-Through-Art.pdf>.
- Alibali, M. W., & Nathan, M. J. (2010). Conducting research in schools: A practical guide. *Journal of Cognition and Development, 11*(4), 397-407.
- American Evaluation Association. (2011). *Public statement on cultural competence in evaluation*. Retrieved January 2018 from www.eval.org.
- Angelle, P. S., Nixon, T. J., Norton, E. M., & Niles, C. A. (2011). An examination of teacher leadership, collective efficacy, and trust in schools. Paper presented at the annual meeting of the University Council for Educational Administration, Pittsburgh, PA.
- Appova, A., & Arbaugh, F. (2018). Teachers' motivation to learn: Implications for supporting professional growth. *Professional Development in Education, 44*(1), 5-21.
- Association of College and Research Libraries. (2011). *ACRL Visual Literacy Competency Standards for Higher Education*. Retrieved June 2018 from http://www.ala.org/acrl/publications/keeping_up_with/visual_literacy.
- Astbury, B. (2013). Some reflections on Pawson's science of evaluation. *Evaluation, 19*(4), 383-401.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation, 31*(3), 363-381.
- Bates, M. S., Phalen, L., & Moran, C. (2016). Online professional development: A primer. *Phi Delta Kappan, 97*(5), 70-73.

Bennett, T. (1996). What's new in evaluation research? A note on the Pawson and Tilley article. *British Journal of Criminology*, 36(4), 567-573.

Better evaluation. (2017). *Evaluation framework*. Retrieved January 2018 from http://www.betterevaluation.org/evaluation-options/evaluation_framework_templates

Baylen, D. M. & D'Alba, A. (2015). *Essentials of teaching and integrating visual and media literacy*. Switzerland: Springer International Publishing.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, 13(4), 439-455.

Booth, A. (2006). Brimful of STARLITE: Towards standards for reporting literature searches. *Journal of the Medical Library Association*, 94(4), 421-430.

Booth, A., Wright, J. & Briscoe, S. (2019). Scoping and searching to support realist approaches. In N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, & Dalkin, S. (Eds.), *Doing realist research* (pp. 147-166). London: SAGE Publications.

Bordens, K. S., & Abbott, B. (2014) *Research design and methods: A process approach* (9th ed). San Francisco: McGraw Hill.

Borko, H., Jacobs, J., & Koeliner, K. (2010). Contemporary approaches to teacher professional development. In Peterson, P. B. E. (Ed.), *International Encyclopedia of Education* (Vol. 7, pp. 548–556). Oxford: Elsevier.

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018) Rethinking models of professional learning as tools: A conceptual analysis to inform research and practice. *Professional Development in Education*, 44(1), 120-139,

- Bradley, W. E. (2011). A conceptual framework for the design and evaluation of online learning modules in professional training and academic education in business. *Proceedings of ASBBS*, 18(1), 196-207.
- Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Byng, R., Norman, I., & Redfern, S. (2005). Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness. *Evaluation*, 11(1), 69-93.
- Chingos, M. M. (2012). *Strength in numbers: State spending on K-12 assessment systems*. Washington, D. C.: Brown Center on Education Policy at Brookings.
- Danitz, T. (2001). *Special report: States pay \$400 million for tests in 2001*. Philadelphia, PA: Stateline, an initiative of The Pew Charitable Trusts.
- Cappello, M., & Walker, N. T. (2016). Visual thinking strategies: Teachers' reflections on closely reading complex visual texts within the disciplines. *Reading Teacher*, 70(3), 317–325.
- Carter, R. (2012). Theory-based evaluation approach (GSDRC Helpdesk Research Report). Birmingham, UK: Governance and Social Development Resource Centre, University of Birmingham.
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: SAGE Publications, Inc.
- Chen H. T. (2012) Theory-driven evaluation: Conceptual framework, application and advancement. In: Strobl R., Lobermeier O., Heitmeyer W. (eds) *Evaluation von programn und projekten für eine demokratische kultur*. Wiesbaden: Springer VS.

- Chen, H. T. & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10(1), 95-103.
- Cochran-Smith, M., Ell, F., Ludlow, L., Grudnoff, L., & Aitken, G. (2014). The challenge and promise of complexity theory for teacher education research. *Teachers College Record*, 116(5), 1–38.
- Coldwell, M., & Simkins, T. (2011). *Professional Development in Education*, 37(1), 143-157.
- Connell, J. P., & Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects and problems. In A. Fulbright-Anderson, A. C. Kubisch and J. P. Connell (eds) NewCoryn, C. L. S., Noakes L. A., Westine C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice: from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.
- Coryn, C. L. S., Noakes L. A., Westine C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice: from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.
- Crandall, S. J., Caellegh, A. S., & Steinecke, A. (2001). Reference to the literature and documentation. *Academic Medicine*, 76(9), 925-927.
- Creswell, J. W., & Plano Clark, V. L. (2010). *Designing and conducting mixed methods research*. Washington, D. C.: SAGE Publications, Inc.
- Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10(49), 1-7.

- Dash, S., de Kramer, R. M., O'Dwyer, L. M., Masters, J., & Russell, M. (2012). Impact of online professional development on teacher quality and student achievement in fifth grade mathematics. *Journal of Research on Technology in Education*, 45(1), 1–26.
- Day, C., & Sachs, J. (Eds.). (2004). International handbook on the continuing professional development of teachers. Maidenhead, UK: Open University Press.
- DeSantis, K., & Housen, A. (2007, October 5). Highlights of findings – San Antonio: Aesthetic development and creative and critical thinking skills study. Retrieved from <http://www.issuelab.org/permalink/resource/15725>.
- Desimone, L. M. (2009). Improving impact studies of teachers' Professional Development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, and Education*, 7(3), 252–263.
- Drago-Severson, E. (2012). New opportunities for principal leadership: Shaping school climates for enhanced teacher development. *Teachers College Record*, 114, 1–44.
- Ebenso, B., Uzochukwu, B., Manzano, A., Etiaba, E., Huss, R., Ensor, T., ... Mirzoev, T. (2016). Reflections from using logic modelling as part of realist evaluation of community health worker program in Nigeria. In Fourth Global Symposium on Health Systems Research: Resilient and responsive health systems for a changing world, 14–18 Nov 2016, Vancouver, Canada.
- Elliott, J. C. (2017). The evolution from traditional to online professional development: A review. *Journal of Digital Learning in Teacher Education*, 33(3), 114–125.

- Enhancing professional development for teachers.* (2007). Washington, D.C.: National Academies Press.
- Evans, L. (2002) What is teacher development? *Oxford Review of Education*, 28(1), 123-137.
- Evans, L. (2008). Professionalism, professionalism and development of education professionals. *British Journal of Educational Studies*, 56(1), 20-38.
- Evans, L. (2011). 'The 'shape' of teacher professionalism in England: Professional standards, performance management, professional development and the changes proposed in the 2010 white paper'. *British Educational Research Journal*, 37(5), 851– 870.
- Farrington, D. P. (1997). Evaluating a community crime prevention program. *Evaluation*, 3(2), 157-173.
- Farrington, D. P. (1998). Evaluating 'communities that care': Realistic scientific considerations. *Evaluation*, 4(2), 204-210.
- Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education*, 64(5), 426– 438.
- Flick, U. (2014). *An introduction to qualitative research* (5th ed.). London, UK: Sage Publications Ltd.
- Franco & Unrath, M., & Unrath, K. (2014). Carpe diem: seizing the common core with visual thinking strategies in the visual arts classroom. *Art Education*, 67(1), 28–32.
- Fraser, C., Kennedy, A., Reid, L. & Mckinney, S. (2007). Teachers' continuing professional development: Contested concepts, understandings and models, *Journal of In-service Education*, 33(2), 153-169,

- Garner, J.K. & Kaplan, A. (2019). A complex dynamic systems perspective on teacher learning and identity formation: an instrumental case. *Teachers and Teaching: Theory and Practice*, 25(1), 7-33.
- Gold, J., Walton, J., Cureton, P., & Anderson, L. (2011). Theorising and practitioners in HRD: The role of abductive reasoning. *Journal of European Industrial Training*, 35(3), 230-246.
- Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017a). *Developing realist program theories*. Retrieved from http://www.ramesesproject.org/media/RAMESES_II_Developing_realist_program_theories.pdf.
- Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017b). *Frequently asked questions about realist evaluation*. Retrieved from http://www.ramesesproject.org/media/RAMESES_II_FAQs_about_realist_evaluation.pdf.
- Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017c). *Philosophies and evaluation design*. Retrieved from http://www.ramesesproject.org/media/RAMESES_II_Philosophies_and_evaluation_design.pdf.
- Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017d). *Protocols and realist evaluation*. Retrieved from http://www.ramesesproject.org/media/RAMESES_II_Protocols_and_realist_evaluation.pdf.
- Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017e). *Realist evaluation and ethical considerations*. Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_Realist_evaluation_and_ethical_considerations.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017f). *Realist evaluation, realist synthesis, realist research-what's in a name?* Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_RE_RS_RR_whats_in_a_name.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017g). *Retrodution in realist evaluation*. Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_Retrodution.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017h). *The realist interview*. Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_Realist_interviewing.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017i). *"Theory" in realist evaluation*. Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_Theory_in_realist_evaluation.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017j). *What is a mechanism? What is a program mechanism?* Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_What_is_a_mechanism.pdf.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., & Jagosh, J. (2017k). *What realists mean by context*. Retrieved from

http://www.ramesesproject.org/media/RAMESES_II_Context.pdf.

- Greenhalgh, T., Wong, G., Jagosh, J., Greenhalgh, J., Manzano, A., Westthorp, G., & Pawson, R. (2015). Protocol—the RAMESES II study: developing guidance and reporting standards for realist evaluation. *BMJ Open*, 5(8), 1-9.
- Grohe, M., & Egan, S. (2015, December 15). *School partnership program 2010-14 elementary longitudinal case study*. Retrieved from <http://vtshome.org/wp-content/uploads/2016/08/9ISGM2010-14SchoolProgramReport.pdf>.
- Gunersel, A. B., Kaplan, A., Barnett, P., Etienne, M., & Ponnock, A. R. (2014, April). Profiles of change in conceptions and motivation regarding teaching in higher education within professional development. Paper presented at the annual convention of the American Educational Research Association, Philadelphia, PA, USA.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Guskey, T. R. (2014). Evaluating professional learning. In S. Billett et al. (eds.), *International Handbook of Research in Professional and Practice-based Learning*, (pp. 1215-1235). Dordrecht: Springer.
- Hahs-Vaughn, D., Zygouris-Coe, V., & Fiedler, R. (2007). A hybrid evaluation model for evaluating online professional development. *Technology, Pedagogy, and Education*, 16(1), 5-20.
- Hansen, H. F. (2005). Choosing evaluation models. *Evaluation*, 11(4), 447-462.
- Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education*, 3(2000), 41-61.
- Harasim, L. (2012). *Learning theory and online technologies*. New York: Routledge.
- Hawkins, A.J. (2014). The case for experimental design in realist evaluation.

- Learning Communities: International Journal of Learning in Social Contexts, Special Issue: Evaluation*, 14, 46-59.
- Holmes, A., Singer, B., & MacLeod, A. (2010). Professional development at a distance: A mixed-method study exploring in-service teachers' views on presence online. *Journal of Digital Learning in Teacher Education*, 27(2), 76–85
- Hedström, P., & Swedberg, R. (1996). Social mechanisms. *Acta Sociologica*, 39, 281-308.
- Housen, A. C. (2001). Aesthetic thought, critical thinking and transfer. *Arts and Learning Research Journal*, 18(1), 99-131.
- Housen, A. C. (1999, September). Eye of the beholder: Research, theory, and practice. Paper presented at the Calouste Gulbenkian Foundation, Service of Education Conference, Lisbon, Portugal.
- Hung, J. & Yang, D. (2015). The Validation of an instrument for evaluating the effectiveness of professional development program on teaching online. *Journal of Educational Technology Development and Exchange*, 8(1), 17-38. Palo Alto, CA: Empirical Education Inc.
- Jaciw, A. P., Schellinger, A. M., Lin, L., Zacamy, J., & Toby, M. (2016). *Effectiveness of internet-based reading apprenticeship improving science education (iRAISE): A report of a randomized experiment in Michigan and Pennsylvania*.
- Jolly, H., & Jolly, L. (2014). Telling context from mechanism in realist evaluation: The role for theory. *Learning Communities: International Journal of Learning in Social Contexts*, 14, 28-45.
- Joksimovic, S., Kovanovic, V., Skrpnyk, O., Gasevic, D., Dawson, S., & Siemens, G. (2015). The history and state of online learning. In G. Siemens, D. Gasevic, & S. Dawson (Eds.),

- Preparing for the digital university: A review of the history and current state of distance, blended, and online learning* (pp. 93-131). Athabasca, Alberta: Athabasca University.
- Kapitan, T. (1992). Peirce and the autonomy of abductive reasoning. *Erkenntnis*, 37(1), 1-26.
- Kaplan, A. & Garner, J.K. (n.d.). Dynamic Systems Model of Role Identity (DSMRI): Analysis guide and codebook (Unpublished).
- Kaplan, A., & Garner, J. K. (2017). A complex dynamic systems perspective on identity and its development: The dynamic systems model of role identity. *Developmental Psychology*, 53(11), 2036–2051.
- Kaplan, A., Garner, J., & Semo, S. (2015). Teacher role-identity and motivation as a dynamic system. Paper presented at the annual convention of the American Education Research Association, Chicago, IL.
- Kaplan, A., Neuber, A., & Garner, J. K. (2019). An identity systems perspective on high ability in self-regulated learning. *High Ability Studies*, 30(1), 1-26.
- Keay, J. K., Carse, N., & Jess, M. (2019) Understanding teachers as complex professional learners, *Professional Development in Education*, 45(1), 125-137
- Kelly, P. (2006). What is teacher learning? A socio-cultural perspective. *Oxford Review of Education*, 32(4), 505-519.
- Kellogg, S., Corn, J., & Booth, S. (2012). *Race to the top online professional development education*. NC: Consortium for Educational Research and Evaluation.
- Kennedy, A. (2014) Understanding continuing professional development: The need for theory to impact on policy and practice. *Professional Development in Education*, 40(5), 688-697.

- Kim, J. (2017). School accountability and standard-based education reform: The recall of social efficiency movement and scientific management. *International Journal of Educational Development*, 60(1), 80-87.
- King, M. R. N., Rothberg, S. J., Dawson, R. J., & Batmaz, F. (2016). Bridging the edtech evidence gap: A realist evaluation framework refined for complex technology initiatives. *Journal of Systems and Information Technology*, 18(1), 18-40.
- Labone, E., & Long, J. (2016). Features of effective professional learning: a case study of the implementation of a system-based professional learning model. *Professional Development in Education*, 42(1), 54-77.
- Lebec, M., & Luft, J. (2007). A mixed methods analysis of learning in online teacher professional development: A case report. *Issues in Technology and Teacher Education*, 7(2), 554-574.
- Lin, S., & Wu, M. (2016). Applying program theory-driven approach to design and evaluate a teacher professional development program. *Journal of Education and Practice*, 7(24), 138-148.
- Lindvall, J. & Ryve, A. (2019). Coherence and the positioning of teachers in professional development programs. A systematic review. *Educational Research Review* 27, 140–154.
- Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programs for teachers. *Ricercazione*, 7(2), 27–51.
- Lubke, J., Britt, V. G., Paulus, T. M., & Atkins, D. P. (2017). Hacking the literature review: Opportunities and innovations to improve the research process. *Reference & User Services Quarterly*, 56(4), p. 285-295.

- Mackenzie, M., & Blamey, A. (2005). The practice and the theory: Lessons from the application of a theories of change approach. *Evaluation, 11*(2), 151-168.
- Manzano, A. (2016). The craft of interviewing in realist evaluation. *Evaluation, 22*(3), 342-360.
- Marchal, B., Kegels, G., & Van Belle, S. (2019). Theory and realist methods. In N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, & Dalkin, S. (Eds.), *Doing realist research* (pp. 79-90). London: SAGE Publications.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(1), 954-969).
- Masters, J., De Kramer, R. M., O' Dwyer, L. M., Dash, S., & Russell, M. (2010). The effects of online professional development on fourth grade english language arts teachers' knowledge and instructional practices. *Journal of Educational Computing Research, 43*(3), 355-375.
- Maxfield, J., Huynh, D., & Mueller, D. (2007). *Evaluation of professional development through technology: An initiative of the Minneapolis Public Schools*. MN: Wilder Research.
- McGuinn, Patrick. (2012). Simulating reform: Race to the top, competitive grants and the Obama education agenda. *Education Policy, 26*(1), 136-159.
- Merchie, E., Tuytens, M., Devos, G., & Vanderlinde, R. (2016). Evaluating teachers' professional development initiatives: Towards an extended framework. *Research Papers in Education, 1*-26.
- Mirzoev, T., Etiaba, E., Ebenso, B., Uzochukwu, B., Manzano, A., Onwujekwe, O., ... Ensor, T. (2016). Study protocol: realist evaluation of effectiveness and sustainability of a community health workers program in improving maternal and child health in Nigeria. *Implementation Science: IS, 11*, 83. <http://doi.org/10.1186/s13012-016-0443-1>

- Mitchell, R. (2013). What is professional development, how does it occur in individuals, and how may it be used by educational leaders and managers for the purpose of school improvement? *Professional Development in Education*, 39(3), 387-400.
- Moeller, M., Cutler, K., Fiedler, D., & Weier, L. (2013). Visual thinking strategies = creative and critical thinking. *Phi Delta Kappan Magazine*, 95(3), 56–60.
- Muijs, D., Day, C., Harris, A., & Lindsay, G. (2004). Evaluating CPD: An overview. In C. Day & J. Sachs (Eds.) *International Handbook on the Continuing Professional Development of Teachers*, (pp. 291-310). Columbus, OH: Open University Press.
- Muijs, D., & Lindsay, G. (2008). Where are we at? An empirical study of levels and methods of evaluating continuing professional development. *British Educational Research Journal*, 34(2), 195-211.
- National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards for english language arts: Science and technical subjects: Introduction. Retrieved from: <http://www.corestandards.org/ELA-Literacy/RST/introduction/>
- O’Leary, S. (2010, May). What do you see that makes you say that? The role of asking young children to provide evidence for their observations in visual thinking strategies discussions (Unpublished master’s thesis). Tufts University, Massachusetts.
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, 81(3), 376–407.
- Owston, R. D., Sinclair, M., & Wideman, H. (2008). Blended learning for professional development: An evaluation of a program for middle school mathematics and science teachers. *Teachers College Record*, 110(5), 1033-1064.

- Owston, R. D., Wideman, H., Murphy, J., & Lupshenyuk, D. (2008). Blended teacher professional development: A synthesis of three program evaluations. *Internet and Higher Education, 11*(3-4), 201-210.
- Patton, M. Q. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning, 12*(4), 375-377.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, UK: SAGE Publications, Inc.
- Pawson R. (2013). *The science of evaluation: a realist manifesto*. London, UK: SAGE Publications, Inc.
- Pawson, R. (2015). The ersatz realism of critical realism: A reply to Porter. *Evaluation, 22*(1), 49-57.
- Pawson, R. (2016). Realist evaluation caricatured: A reply to Porter. *Nursing Philosophy, 17*(2), 132-139.
- Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(1), 21-34.
- Pawson, R., & Manzano-Santaella, A. (2012). A realist diagnostic workshop. *Evaluation, 18*(2), 176-191.
- Pawson, R., & Tilley, N. (1996). What's crucial in evaluation research: A reply to Bennett. *British Journal of Criminology, 36*(4), 574-578.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: SAGE Publications, Inc.
- Pawson, R., & Tilley, N. (1998a). Caring communities, paradigm polemics, design debates. *Evaluation, 4*(1), 73-90.

- Pawson, R., & Tilley, N. (1998b). Cook-book methods and disastrous recipes: A rejoinder to Farrington. *Evaluation*, 4(2), 211-213.
- Pawson, R., & Tilley, N. (2013). An introduction to scientific realist evaluation. In: E. Chelmsky & W. Shadish (Eds.) *Evaluation for the 21st century: A resource book*. (pp. 405-418). Thousand Oaks, CA: SAGE Publications, Inc.
- Patton, M. Q. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning*, 12(4), 375-377.
- Pietarinen, A., & Bellucci, F. (2014). New light on Peirce's conceptions of retrodution, deduction, and scientific reasoning. *International Studies in the Philosophy of Science*, 28(4), 353-373.
- Porter, S. (2015a). The uncritical realism of realist evaluation. *Evaluation*, 21(1), 65-82.
- Porter, S. (2015b). Realist evaluation: An immanent critique. *Nursing Philosophy*, 16(4), 239-251.
- Porter, S. (2017). Evaluating realist evaluation: A response to Pawson's reply. *Nursing Philosophy*, 18(2), 1-4.
- Porter, S., & O'Halloran, P. (2012). The use and limitation of realistic evaluation as a tool for evidence-based practice: A critical realist perspective. *Nursing Inquiry*, 19(1), 18-28.
- Posavac, E. J. (2011). *Program evaluation: Methods and case studies*. New York, B: Routledge.
- Punton, M., Vogel, I., & Lloyd, R. (2016). *Reflections from a realist evaluation in progress: Scaling ladders and stitching theory*. Centre for Development Impact (CDI): Brighton, UK

- Rakap, S., Jones, H. A., & Emery, A. K. (2014). Evaluation of a web-based professional development program (project ACE) for teachers of children with autism spectrum disorders. *Teacher Education and Special Education, 38*(3), 221-239.
- RAMESES II Project Team. (2017). Quality standards for realist evaluation: For funders or commissioners of realist evaluations. Retrieved from http://www.ramesesproject.org/media/RE_Quality_Standards_for_evaluators_and_peer_reviewers.pdf.
- Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research and Evaluation, 14*(13), 1-13.
- Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development, 29*(2), 35-53.
- Rice, K., & Hung, J. (2015). Data mining in online professional development program: An exploratory case study. *International Journal of Technology in Teaching and Learning, 11*(1), 1-20.
- Rogers, P. J. (2000). Program theory evaluation: Not whether programs work but how they work. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 209-232). Boston, MA: Kluwer.
- Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsí, T. A. (2000). Program theory evaluation: Practice, promise, and problems. In P. J. Rogers, T. A. Hacsí, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities* (pp. 5-14). *New Directions for Evaluation, No. 87*. San Francisco, CA: Jossey-Bass.

- Roskos, K., Jarosewich, T., Lenhart, L., & Collins, L. (2007). Design of online teacher professional development in a statewide Reading First professional development system. *The Internet and Higher Education*, 10(3), 173–183.
- Royse, D., Thyer, B. A., & Padgett, D. K. (2010). *Program evaluation: An introduction*. Belmont, CA: Wadsworth.
- Russell, M., Carey, R., Kleiman, G., & Venable, J. D. (2009). Face-to-face and online professional development for mathematics teachers: A comparative study. *Journal of Asynchronous Learning Networks*, 13(2), 71–87.
- Russell, M., Kleiman, G., Carey, R., & Douglas, J. (2009). Comparing self-paced and cohort-based online courses for teachers. *Journal of Research on Technology in Education*, 41(4), 443–466.
- Salter, K. L., & Kothari, A. (2014). Using realist evaluation to open the black box of knowledge translation: a state-of-the-art review. *Implementation Science*, 9(115), 1-14.
- Sankar, L., & Sankar, C. S. (2010). Comparing the effectiveness of face-to-face and online training on teacher knowledge and confidence. Proceedings from *Informing Science & IT Education Conference (InSITE) 2010*. Cassino, Italy.
- Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed). London: Routledge
- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19, 57-70.
- Scriven, M. (1999). The fine line between evaluation and explanation. *Evaluation Practice*, 15, 75-77.

- Schroeder, S., Currine, E., & McCardle, T. (2016). Mother tongues: The Opt Out movement's vocal response to patriarchal, neoliberal education reform. *Gender and Education*, 28(1), 1-18.
- Sherman, G., Byers, A., & Rapp, S. (2007). Evaluation of online, on-demand science professional development material involving two different implementation models. *Journal of Science and Educational Technology*, 17(1), 19-31.
- Snyder, L. M. (2009). Using the improvement-focused model to evaluate an online teacher education program. *Journal of Educational Technology Systems*, 38(2), 145-153
- Sorinola, O. O., Thistlewaite, J., Davies, D., & Peile, E. (2015). Faculty development for educators: a realist evaluation. *Advances in Health Science Education*, 20(2), 385-401.
- Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation*, 10(1), 58-76.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development (DFID Working Paper 38)*. London, UK.
- Stufflebeam, D. (2002). Evaluation models. *New Directions for Evaluation*, 89, 7-98.
- Surette, T. N., & Johnson, C. C. (2015). Assessing the ability of an online environment to facilitate the critical features of teacher professional development. *School Science and Mathematics*, 115(6), 260-270.
- Tan, H. T. R., & Harvey, G. (2016). Unpacking the Black Box: A realist evaluation of performance management for social services. *Public Management Review*, 18(10), 1456-1478.
- The Joint Committee on Standards for Educational Evaluation. (2011). The joint committee on standards for educational evaluation. In R. A. Green (Ed.) *Case study research: A*

- program evaluation guide for librarians*. (pp. 105-113). Santa Barbara, CA: ABC-CLIO, LLC.
- Tilley, N (2000) *Realistic evaluation: an overview*. Presentation to the Danish Evaluation Society, Nottingham, UK.
- Timmins, P., Miller, C. (2007). Making evaluations realistic: The challenge of complexity. *Support for Learning*, 22(1), 9-16.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). Teacher professional learning and development: Best evidence synthesis iteration. Wellington, New Zealand: Ministry of Education.
- Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4), 404-428.
- Tschannen-Moran, (2009). Fostering teacher professionalism in schools: The role of leadership orientation and trust. *Education Administration Quarterly*, 45(2), 217-247.
- Tschannen-Moran, M. (2014). *Trust matters: Leadership for successful schools*. San Francisco, CA: Jossey Bass.
- Tschannen-Moran, M., & Gareis, C. R. (2015). Principals, trust, and cultivating vibrant schools. *Societies*, 5, 256–276.
- Tschannen-Moran, M., & Gareis, C. R. (2014). Faculty trust in the principal: an essential ingredient in high-performing schools. *Journal of Educational Administration*, 53(1), 66-92.
- Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A., & Marchal, B. (2016). Can “realist” randomised controlled trials be genuinely realist? *Trials*, 17(313), 1-6.

- Vrijhoef, H. J. M. (2017). *Realist evaluation: Opening the black box of evaluation*. The Taos Institute: Chagrin Falls, OH.
- VTS. (2017). About us. (Weblog comment). Retrieved from <https://vtshome.org/about/>.
- VTS. (2017). Research and theory. (Weblog comment). Retrieved from <https://vtshome.org/research/>.
- Watershed Collaborative. (2017). Our mission and vision. Watershed ed. Retrieved from <https://www.watershed-ed.org/about>.
- Webster-Wright, A. (2009). Reframing professional development through understanding authentic learning. *Review of Educational Research*, 79(2), 702-739.
- Weetman, K, Wong, G., Scott, E., Schnurr, S., & Dale, J. (2017). Improving best practice for patients receiving hospital discharge letters: A realist review protocol. *BMJ Open*, 7, 1-7.
- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In Connell, J., Kubisch, A., Schorr, L. B., & Weiss, C. H. (Eds.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts* (pp. 65-92). New York, B: Aspen Institute.
- Weiss, C. H. (1997a). How can theory-based evaluations make greater headway? *Evaluation Review*, 21, 501-524.
- Weiss, C. H. (1997b). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, (76), 41-55.
- Westthorp, G., Prins, E., Kusters, C. S. L., & Brouwers, J. H. A. M. (2011). *Realist Evaluation: an overview*. Centre for Development Innovation: Wageningen, NL.

- Westhorp, G. (2008). Development of realist evaluation models and methods for use in small-scale community based settings. Unpublished PhD thesis.
- Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation*, 18(4), 405–20.
- Westhorp, G. (2013). Developing complexity-consistent theory in a realist investigation. *Evaluation*, 19(4), 64–382.
- Westhorp, G. (2014). *Realist impact evaluation: An introduction*. London, UK: Methods Lab, Overseas Development Institute (ODI).
- Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J., & Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine*, 14(96), 1-18.
- Yenawine, P. (2003). Jump starting visual literacy: Thoughts on image selection. *Art Education*, 56(1), 6–12.
- Yenawine, P. (2013, September 12). Teaching to the (new) tests: The benefits of discussion. (Weblog comment). Retrieved from <http://www.wholechilddeducation.org/blog/teaching-to-the-new-tests-the-benefits-of-discussion>.
- Yenawine, P. (1998). Visual art and student-centered discussions. *Theory into Practice*, 37(4), 314–321.
- Yenawine, P. (2013). *Visual thinking strategies: Using art to deepen learning across school disciplines*. Cambridge, MA: Harvard Education Press.
- Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (No. 033). Regional Educational Laboratory Southwest.

APPENDICES

Appendix A
Realist Evaluation Key Concepts

Term	Definition	Key Characteristics	(Non)-Examples
Program	Sophisticated social interactions set amidst a complex social reality to bring about desired results.	1. Theories incarnate 2. Active 3. Embedded in social systems 4. Parts of open systems 5. Self-transformational	<i>Examples:</i> • Set of interventions offered after school, such as tutoring and extra instruction, that are intended to increase struggling students' reading proficiency. <i>Non-examples:</i> • Non-related tutoring sessions in various school districts across the state.
Mechanism (M)	What it is about programs and interventions that bring about any effects. Comprised of “resources” and “responses.”	1. Usually hidden 2. Sensitive to variations in context 3. Generate outcomes 4. Identified by thinking through how a program changes behavior & targeted data collection and analysis 5. Can be challenging to distinguish from Context (C)	<i>Examples:</i> • Gravity: Not tangible, but observed via effects. • Gunpowder: Requires the right contextual factors to observe its potential power. • Generating teachers' interest in student-centered pedagogy through online PD course exercises. <i>Non-examples:</i> • Program interventions • Program variables

Term	Definition	Key Characteristics	(Non)-Examples
Context (C)	Cultural, political, historical, psychological, economic, and organizational actors in which programs are embedded.	1. Tangible: Geographic location, physical surroundings 2. Intangible: Cultural, political, historical psychological, economic, organizational factors 3. Can interact and influence each other 4. Evaluators must focus on factors that have most influence on mechanisms 5. Can be challenging to distinguish from Mechanisms (M)	Examples: • Tennis ball dropped in the air versus in the water: Same action, but different outcomes due to different contextual factors. • Online PD program: Teachers of non-traditional subjects may enroll at a higher rate than teachers of traditional subjects. Non-examples: • Sole focus on locality • Program mechanisms
Outcomes, Outcome Patterns (O)	Intended and unintended consequences of programs, resulting from the activation of different mechanisms in different contexts.	1. Multiple outcomes, depending on the interactions between contextual factors and causal mechanisms 2. Focus on how/why outcome patterns occur 3. Identified via data and measures 4. Identifying patterns may involve implementation variations, impact variations, socio-demographic sub-group variations, temporal outcome variations, personal attribute outcome variations, regional outcome variations, biological make-up outcome variations, etc.	Examples: • An online PD program results in increased instances of inquiry-based learning in participants' classrooms. Non-examples: • Focusing only a single measure of success

Term	Definition	Key Characteristics	(Non)-Examples
Context-mechanism-outcome configurations (CMOCs)	Propositions that bring together mechanism-variation and relevant context-variation to predict and explain outcome pattern variation.	1. Similar to recipes: Specific ingredients are measured, combined, and cooked at specific temperatures to create the desired product 2. Requires data and analysis to identify 3. Forms the basis for developing and refining program theories	Examples: • An online PD program that offers the option of online or face-to-face delivery modes to teachers in rural K-12 schools (context) may result in more teachers facilitating more inquiry-based activities (outcome) by generating teachers' interest in student-centered pedagogy(reasoning). Non-examples: • Program theories with no explicit identification of Context, Mechanism, or Outcomes

Appendix B
Online Versus Traditional PD

Criteria	Sub criteria	Online	Traditional
Content	Topic • Selection • Subject specific • Aligned with standards	+ Great variety of courses, including for niche subjects + Teachers can select courses that meet their specific needs + May be aligned with Common Core national standards – May not align to a specific state’s standards	± Focus on content with broad applicability ± Often tailored to district or state standards ± Limited variety of courses/workshops
Structure	Pedagogy • Active learning • Inquiry-based learning Differentiation • Feedback • Customization Interaction • Collaborative/community • Self-directed • Instructor presence	+ May be synchronous, asynchronous, or both + May include participants from across the nation + Capacity for long-term inquiry-based projects + Capacity to build community of learners over extended period of time – Often does not differentiate content for different backgrounds/cultures ± Often self-directed or facilitated by an instructor	+ Face-to-face interaction + Can include collaborative discussions and activities – Inquiry-based activities limited to length of PD session(s) ± Often instructor-centered - Participants likely from same school or district

Criteria	Sub criteria	Online	Traditional
Access Factors	Availability • Location • Time Flexibility • Location • Time Resources • Cost • Knowledge/experience • Support	+ Asynchronous portions accessible from any (Internet-equipped) location at any time + Synchronous portions accessible from any (Internet-equipped) location + Cost effective + Ecological – Tech support not always available – Need certain technological skills to participate – Need stable Internet connection, required hardware/software	+ Required materials typically provided + PD instructors can answer questions and resolve technological/resource issues immediately – Times and locations pre-determined – May require travel and hiring of substitute teachers – Costly
Effectiveness/ Impact	Student achievement Teacher performance	– Little information on effectiveness of OPD ± No significant differences in outcomes between online and traditional PD	± No significant differences in outcomes between online and traditional PD

Appendix C
Literature Review Sources: Genre & Type of PD

	Online/Blended PD Evaluation	General PD Evaluation	General Evaluation	
	8	1		9
Studies	<i>Sherman, Byers, & Rapp, 2007;</i> <i>Owston, Sinclair, & Wideman, 2008;</i> <i>Hung & Yang, 2015;</i> <i>Owston, Wideman, Murphy, & Lupshenyuk, 2008;</i> <i>Snyder, 2009;</i> <i>Rice & Hung, 2015</i> <i>Lebec & Luft, 2007</i> <i>Rakap et al., 2014</i>	<i>Muijs & Lindsay, 2008</i>		
	2	2		4
Frameworks/ Models	<i>Bradley, 2011;</i> <i>Hahs-Vaughn, Zygouris-Coe, & Fielder, 2007</i>	<i>King, 2014;</i> <i>Merchie, Tuytens, Devos, & Vanderlinde, 2016</i>		
		2		2
Theoretical/ Descriptive		<i>Muijs, Day, Harris, & Lindsay, 2004;</i> <i>Guskey, 2014;</i>	;	
		2		2
Critiques		<i>Reio et al., 201;7</i> <i>Coldwell & Simkins, 2011</i>		
	3			3
Reports	<i>Maxfield et al., 2007;</i> <i>Kellogg, Corn, & Booth, 2012;</i> <i>Jaciw, Schellinger, Lin, Zacamy, & Toby, 2016</i>			
	1		1	2
Literature Reviews	<i>Ke & Hoadley, 2009</i>		<i>Hansen, 2005</i>	

Appendix D
Literature Review Sources: Studies & Reports

Source	Purpose	Design & Data Collection	Major Findings/Results
Jaciw, A. P., Schellinger, A. M., Lin, L., Zacamy, J., & Toby, M. (2016). <i>Effectiveness of internet-based reading apprenticeship improving science education (iRAISE): A report of a randomized experiment in Michigan and Pennsylvania.</i>	Impact Measure the impact of the iRAISE OPD program on classroom instructional practices and student reading literacy in high school science classes in 27 schools in MI and PA.	Experimental • Cluster randomized control trial • Focused on outcomes • Intent-to-treat design • 50 % of participants in treatment group (received OPD) and 50 % in control group (no OPD) • Follows levels 1-3 & 5 of Guskey's model Quantitative & Qualitative Methods • Surveys, observations, records, assessment	• Positive reactions from teachers in the treatment group • Changed classroom practice of teachers in the treatment group • No positive impact on general reading literacy • Differential impact on general reading literacy for students with lower SES
Kellogg, S., Corn, J., & Booth, S. (2012). <i>Race to the top online professional development education.</i> NC: Consortium for Educational Research and Evaluation.	Process & Outcome Provide detailed information about the extend to which the OPD components of the RttT application have been carried out, and the immediate impact of RttT OPD efforts on educators.	Non-experimental • Longitudinal study with purposeful sampling • Participants self-enrolled in OPD courses • Follows levels 1-3 & 5 of Guskey's model Quantitative & Qualitative Methods • Observations, surveys, interviews, focus groups	• Positive teacher reactions about OPD course structure and contents • Use of OPD resources at state and local level not yet wholly consistent with national standards for OPD • School districts may need additional guidance, training, support, technology tools, and/or content resources

Source	Purpose	Design & Data Collection	Major Findings/Results
Lebec, M., & Luft, J. (2007). A mixed methods analysis of learning in online teacher professional development: A case report. <i>Issues in Technology and Teacher Education</i> , 7(2), 554-574.	Process Analyze the nature of the knowledge learned by participants enrolled in an OPD biology course and determine how the online environment influenced this process.	Non-experimental • Case report with a selection of recruited participants • Participants selected from a pool of recruitments • Follows levels 1-3 of Guskey's model Quantitative & Qualitative Methods • Interviews, exams, assignments	• Participants gained knowledge of concepts and terms but did not use them any more efficiently after their online experience • Participants' attitudes and motivation affected their participation in the OPD course • Suggests that OPD designers consider factors maximizing engagement, personal accountability, and motivation
Maxfield, J., Huynh, D., & Mueller, D. (2007). <i>Evaluation of professional development through technology: An initiative of the Minneapolis Public Schools</i> . MN: Wilder Research.	Outcome Address effectiveness of the PDT2 OPD model in meeting PD needs of teachers and assess the extent to which PDT2 increased the level of technology integration in participating teachers' classrooms.	Non-experimental • Participants self-enrolled in OPD courses • Follows levels 1-3 & 5 of Guskey's model Quantitative & Qualitative Methods • Surveys, observations, interviews	• Increased 50% of teachers' technology knowledge and skill • Did not substantially change teachers' use of technology in their classrooms • Increased amount of class time that students spent using technology • Increased teachers' interplay of technology, content, and pedagogy • Suggests that OPD designers consider incorporating stronger incentives to participate, clearer expectations, more applied content, and more accountability • Teachers were interested in addition of face-to-face component
Owston, R. D., Sinclair, M., & Wideman, H. (2008). Blended learning for professional development: An evaluation of a program for middle school	Process & Outcome Understand how an OPD program for middle school math and science/ technology teachers affected teacher attitudes toward and	Non-experimental • Explicitly uses Guskey's model • Follows levels 1-4 of Kirkpatrick's model Quantitative & Qualitative Methods	• Teachers highly satisfied with OPD program, especially face-to-face components • Teachers had strong organizational support except for obtaining adequate release time

Source	Purpose	Design & Data Collection	Major Findings/Results
mathematics and science teachers. <i>Teachers College Record</i> , 110(5), 1033-1064.	pedagogical practices in these subjects. Additionally, to understand students' perceptions and learning of the subjects.	Pre/post tests, surveys, interviews	- Significant changes in pedagogical beliefs and practices - Few pre-post difference for student attitudes towards subjects - Low SES students benefitted more from new instructional practices
Owston, R. D., Wideman, H., Murphy, J., & Lupshenyuk, D. (2008). Blended teacher professional development: A synthesis of three program evaluations. <i>Internet and Higher Education</i> , 11(3-4), 201-210.	Process & Outcome Answer the following four questions: - How can blended learning programs be designed to emphasize situated, on-the-job professional learning that focuses on current curriculum? - How can blended programs strengthen teachers' sense of community and collaborative skills? - How can blended programs transform teachers' practice? - How can blended programs be designed to increase teachers' impact on students?	Non-experimental - Cross case comparative analysis of 3 OPD program evaluation reports - Follows levels 1-3 & 5 of Guskey's model Quantitative & Qualitative Methods - Evaluation reports	- More structure in an OPD environment gives less flexibility to teachers to experiment with new strategies in their classrooms. However, more flexibility negatively impacts OPD participation. - A blended approach that has face-to-face sessions throughout the year is likely to be more effective than 100% face-to-face PD. - The more relevant an OPD program is to teachers' everyday practice, the more likely teachers are to change their practice - OPD programs designed to bring immediate changes to teacher practice are more likely to have an impact on students
Rakap, S., Jones, H. A., & Emery, A. K. (2014). Evaluation of a web-based professional development program (project ACE) for teachers of children with autism spectrum disorders. <i>Teacher</i>	Process & Outcome Assess: The extent to which the OPD program, Project ACE, increased teachers' knowledge and skills about ASDs and	Quasi-Experimental - Purposeful selection of qualified participants from a group of applicants - Follows levels 1-3 & 5 of Guskey's model	- Teachers satisfied with OPD courses in the program - The program successfully helped develop teachers' competencies, knowledge and skills - Teachers comfortable using teaching strategies they learned in the program

Source	Purpose	Design & Data Collection	Major Findings/Results
<i>Education and Special Education</i> , 38(3), 221-239.	serving children with ASDs and their families - The extent to which project ACE increased teachers' level of comfort using evidence-based instructional practices for children with ASDs - Teachers' satisfaction with the OPD courses.	<i>Quantitative & Qualitative Methods</i> Surveys, online assignments, pre/post tests	- Moderate application of new teaching strategies in the classroom - Suggests that OPD programs incorporate same face-to-face time for interactions with other participants and course facilitators - Need for more empirical studies
Rice, K., & Hung, J. (2015). Data mining in online professional development program: An exploratory case study. <i>International Journal of Technology in Teaching and Learning</i> , 11(1), 1-20.	<i>Process</i> Provide information about the usage patterns, clusters, and predictive variables that emerge when mining teacher PD data, and determine what relationships exist between LMS behaviors, usage patterns, end of course evaluation data, learner outcomes, and demographics.	<i>Non-experimental</i> - Case study - Participants self-enrolled in case study OPD course - Focus on exploratory data analysis <i>Quantitative & Qualitative Methods</i> - Surveys, course analytics	- Participants report higher support and confidence ratings when they interact with course content, their peers, and the instructor - The more time participants spent on online and the more frequently they logged in, the higher their engagement and performance levels - Interaction and engagement were important factors for learning in this OPD course - Suggests that data mining be used to evaluate the impact of OPD against its intended outcomes
Sherman, G., Byers, A., & Rapp, S. (2007). Evaluation of online, on-demand science professional development material involving two different implementation models. <i>Journal of Science and Educational Technology</i> , 17(1), 19-31.	<i>Impact</i> Measure middle school teachers' level of satisfaction with the OPD course and confidence in teaching force and motion concepts, as well as impact on teachers' professional practice. Additionally, determine whether there are any differences	<i>Quasi-experimental</i> 3 groups with 2 different treatments (1 group blended PD and 2 groups 100% OPD) - Participants volunteered - Follows levels 1-3 & 5 of Guskey's model <i>Quantitative & Qualitative Methods</i>	- Participants in the 2 100% OPD groups showed significant improvement in their applied achievement scores - Increased teachers' confidence in helping their students learn about force and motion - Positively impacted teachers' professional practice

Source	Purpose	Design & Data Collection	Major Findings/Results
	between the experiences of teachers in the different treatment groups.	Course analytics, surveys, interviews, assessment	Teachers desired a more structured experience with an involved course facilitator
Snyder, L. M. (2009). Using the improvement-focused model to evaluate an online teacher education program. <i>Journal of Educational Technology Systems</i>, 38(2), 145-153.	Process Identify extent to which teachers who completed the OPD and were working at the elementary or secondary level felt effectively prepared to teach reading. Additionally, to identify areas of the program that could be improved/revised to better serve teachers' needs.	Non-experimental - Improvement-focused - Participants self-enrolled in OPD program - Follows levels 1 & 2 of Guskey's model Quantitative & Qualitative Methods - Surveys, interviews	- Teachers highly satisfied with OPD course format and convenience - Teachers desired a greater teaching presence and incorporation of face-to-face components

Appendix E

Literature Review: Codes

PART I

- Evaluation design
 - Barriers
 - Contextual factors
 - Data analysis
 - Data collection
 - Evolution
 - Formative
 - Logic model
 - Measurement instruments
 - Purpose
 - Results
 - Summative
 - Theory-based
- Evaluation levels
 - Change in behavior, use of knowledge and skills
 - Learner reactions, satisfaction
 - Organizational support and change
 - Participants' learning
 - Relationships
 - Student achievement
- Models/frameworks
 - Alternative
 - Categories
 - Guskey
 - Kirkpatrick
 - Level-hierarchy
 - New model/synthesis
 - Sloan
 - Variables
- Professional development
 - Characteristics
 - Definition
 - Effective
 - Format
 - Online learning community
 - Online/blended
 - Traditional/face-to-face
- Research
 - Critiques
 - Empirical studies
 - Data mining
 - Examples
 - Gaps
 - Survey
- Definitions/Terminology

PART II

- Non-realist evaluation approaches
- Methodology
 - Data analysis
 - Data collection
 - Mixed methods
 - Qualitative
 - Quantitative
 - Evaluation questions
- Programs
 - Program logic
 - Program theory
 - Program theory models
- Realist evaluation
 - Advantages
 - Challenges
 - CMOCs
 - Context
 - Definition
 - Description
 - Examples
 - Generative causation
 - Mechanism
 - Middle range theory
 - Outcomes
 - Synthesis
 - Unique, differentiating factors
- Theory-based/theory-driven evaluation
 - Ambiguity
 - Assumptions
 - Black Box
 - Causation
 - Comparisons
 - Critiques
 - Definition
 - History
 - Prominent players
 - TB benefits
 - TB challenges
 - TB examples
 - TB Mechanism
 - Terminology
 - Theory/philosophy
 - Uses
- Research
 - Gaps

- Lack of consensus
- Standards
- Terminology

Appendix F
Literature Review Part 2 Sources: Genres

Descriptive/Theoretical	Critiques	Studies/Reports	Standards/Protocol
Astbury, 2013	Pawson & Manzano-Santaella, 2012	Lin & Wu, 2016	RAMESES II Project Team, 2017a
Astbury & Leeuw, 2010	Van Belle et al., 2016	Stern et al., 2012	RAMESES II Project Team, 2017b
Bickman, 1989	Bennett, 1996	Punton et al., 2016	Wong et al., 2015
Rogers & Weiss, 2007	Davis, 2005	Jolly & Jolly, 2014	Wong et al., 2016
Stame, 2004	de Souza, 2013	Sorinola et al., 2015	
Westhorp, 2014	Farrington, 1998	Byng et al., 2005	
Westhorp et al., 2011	Pawson, 2015	Tan & Harvey, 2016	
Greenhalgh et al., 2017	Pawson, 2016	Bennett, 1991	
Pawson & Tilley, 1997	Pawson & Tilley, 1994		
Pawson & Tilley, 2004	Pawson & Tilley, 1996		
Tilley, 2000	Pawson & Tilley, 1998a		
Timmins & Miller, 2007	Pawson & Tilley, 1998b		
Pawson & Tilley, 2013	Porter, 2015a		
Dalkin et al., 2015	Porter, 2015b		
King et al., 2016	Porter & O'Halloran, 2012		
Manzano, 2016	Neilsen & Miraglia, 2017		

Descriptive/Theoretical	Critiques	Studies/Reports	Standards/Protocol
Crow, 2010	Bonnell et al., 2012		
Crow, 2011	Bonnell et al., 2016		
Crow & Semmens, 2007	Marchal et al., 2013		
Farrington, 1997			
Porter et al., 2017			
Tilley, 1996			
Tilley, 2011			
Coryn et al., 2010			
Blamey & Mackenzie, 2007			
Salter & Kothari, 2014			

Appendix G

Literature Review Search Terms

PD Mode of Delivery		PD Participants		PD Synonyms		Evaluation Concepts
Online OR Blended	AND	K-12 OR Teacher OR Educator OR Education	AND	“Professional Development” OR “Professional learning” OR In-service OR Training	AND	Evaluation OR “Evaluation report” OR “Evaluation standards” OR “Evaluation model” OR “Evaluation framework”

Primary Term		Conceptual/ Organizational Approach of Source	RE and TBE-Related Concepts
“Realist evaluation” OR “Theory-driven evaluation” OR “Theory-based evaluation”	AND	Approach OR Standards OR Model OR Framework OR Introduction OR Review OR Critique OR Design	“Middle-range theory” OR “Program theory” OR “Theory of change” OR “Logic Model” OR Causality OR Complexity OR Mechanisms OR Context

Appendix H
Realist Evaluation Strengths

Strength	Support	Critiques	Sources
Causal Attribution & Focus on Context & Explanatory Focus	- REs identify specific contextual factors affecting when, where, how, and why program mechanisms are activated. - REs address <i>how</i> or why desired outcomes were or were not achieved. - Subsequent RE studies build on one another to generate more generalizable conclusions over time.	- REs can be impractical for evaluators with specific time and budget constraints. - RE is not a mainstream evaluation approach, so stakeholders may not be familiar with it. - RE terminology, methodology and underlying epistemology might be challenging to explain to others involved in the evaluation. - RE questions (determining what works, for whom, and in what circumstances) are not as exclusive as often claimed. - Adopting a realist epistemology is not required to answer these questions.	- Astbury & Leeuw, 2010 - Astbury, 2013 - Blamey & Mackenzie, 2007 - Dieleman, Wong, & Marchal, 2012 - Gill & Turbin, 1998 - Greenhalgh et al., 2017 - King et al., 2016 - Pawson & Manzano-Santaella, 2012 - Pawson & Tilley, 1997, 2004, 2013 - Salter & Kothari, 2014 - Sorinola et al., 2015 - Tilley, 2000 - Tilley, 2000 - Timmins & Miller, 2007 - Van Belle et al., 2016 - Westthorp, 2014, 2018
Central Role of Theory/Theory-Driven	- Testing and refining program theories afford the ability to determine relationship between contextual factors, mechanisms, and outcomes. - REs are method-neutral. - Incorporation of multiple research methods and data sources enhances validity.	- Process of testing and refining program theories can be expensive and time consuming. - Some evaluations only need to measure outcomes, in which case identifying related contextual factors and mechanisms is not necessary and expends valuable resources.	- Blamey & Mackenzie, 2007 - Byng et al., 2005 - Coryn et al., 2010 - Dieleman, Wong, & Marchal, 2012 - Hewitt et al., 2012 - King et al., 2016 - Pawson & Manzano-Santaella, 2012

Strength	Support	Critiques	Sources
		Generating, testing, and refining researchers' own theories risks confirmation bias.	- Pawson & Tilley, 1997, 2004, 2013 - Salter & Kothari, 2014 - Tilley, 2000 - Timmins & Miller, 2007 - Westhorp et al., 2011 - Westhorp, 2014 - Wong et al., 2012 - Astbury & Leeuw, 2010 - Jolly & Jolly, 2014 - Manzano, 2016 - Punton et al., 2016 - Salter & Kothari, 2014 - Stame, 2004 - Van Belle et al., 2016
Handling Complexity & Targeting Focus	- REs target specific program areas for evaluation, generating specific findings. - REs consider the primary elements of program complexity to help determine data collection and analysis methods. - REs consider both horizontal and vertical complexity when "untangling" which contextual factors result in which causal mechanisms. - REs consider outcomes at different levels of a system, such as individual, program, organization, government, etc.	- Many REs can be considered one-group pre-test-posttest designs, which have low internal validity. - Low internal validity of REs can fail to control for extraneous variables and exclude alternative explanations. - REs do not evaluate programs comprehensively, so they miss the "big picture."	- Astbury, 2013 - Crow, - King et al., 2016 - Pawson & Tilley, 1997, 2004 - Pawson 2004, 2013 - Punton et al., 2016 - Salter & Kothari, 2014 - Vrithoef, 2017 - Westhorp, 2014

Strength	Support	Critiques	Sources
Transferability & Cumulative Realist Learning	• Elements of final program theories can be transferred as middle-range theories to programs operating in similar contexts or with similar interventions. • Evaluators use middle-range theories to help develop initial program theories; they do not have to start from scratch. • Middle-range theories are transferrable across diverse domains and programs.	• Evaluating a program as a whole is practical and efficient; researchers first determine whether a program works before committing resources to determining why it works. • By assuming replicability of results for the same program in similar areas, researchers do not need to expend resources re-evaluating the same programs multiple times.	• Astbury & Leeuw, 2010 • Astbury, 2013 • Blamey & Mackenzie, 2007 • Byng et al., 2005 • King et al., 2016 • Pawson & Tilley, 1997, 2004 • Pawson, 2006 • Punton et al., 2016 • Stern et al., 2012 • Tilley, 2000 • Timmins & Miller, 2007 • Van Belle et al., 2016 • Westthorp, 2014
Multiple Methods & Validity	• Using both quantitative and qualitative methods and sources of data both measures outcomes and helps to provide explanations for outcomes. • The use of multiple methods and sources of data increases validity. • Teaching and learning process enhances internal validity. • Using multiple methods facilitates theory adjudication.	• REs can be considered correlational designs that infer causality and exclude plausible alternatives, resulting in low internal validity. • REs operate on too many assumptions. • Gathering and analyzing various types of data can be expensive and time consuming. • RE methodology is not standardized enough to ensure reliability.	• Astbury, 2013 • Blamey & Mackenzie, 2007 • Byng et al., 2005 • Carter, 2012 • Coryn et al., 2010 • Manzano, 2016 • Pawson & Manzano-Santaella, 2012 • Pawson & Tilley, 1997, 2004 • Salter & Kothari, 2014 • Sorinola et al., 2015 • Timmins & Miller, 2007 • Van Belle et al., 2016 • Westthorp, 2014

Appendix I
Realist Evaluation Straw Man Arguments

Straw Man Argument	Corresponding Implication for RE	Used in Support of Strengths
A successionist view of causality does not maintain internal validity. The quasi-experimental design has a successionist view of causality. Quasi-experimental designs fail to maintain internal validity.	A generative view of causality maintains internal validity. The realist approach has a generative view of causality. Realist designs maintain internal validity.	5
Identifying the relationship between causal mechanisms and contextual factors is necessary and sufficient to explain why outcomes occur. The quasi-experimental approach does not identify causal mechanisms and contextual factors. The quasi-experimental approach is not sufficient to explain why outcomes occur.	Identifying the relationship between causal mechanisms and contextual factors is necessary and sufficient to explain why outcomes occur. The realist approach identifies causal mechanisms and contextual factors. The realist approach is sufficient to explain why outcomes occur.	1, 2, 3, & 4
Realist epistemology is necessary and sufficient to explain the fundamental role of theory in evaluation. Realist epistemology and the quasi-experimental approach are mutually exclusive. Therefore, the quasi-experimental approach is not sufficient to explain the fundamental role of theory in evaluation. (1998, p. 74)	Realist epistemology is necessary and sufficient to explain the fundamental role of theory in evaluation. The realist approach requires a realist epistemology. Therefore, the realist approach is sufficient to explain the fundamental role of theory in evaluation. (1998, p. 74)	1 & 4

Straw Man Argument	Corresponding Implication for RE	Used in Support of Strengths
The ability to explain why outcomes occur is necessary for progress in the field of evaluation. The quasi-experimental approach does not explain why outcomes occur. Ending the domination of the quasi-experimental approach will create the potential for/increase the possibility of progress in the field of evaluation.	The ability to explain why outcomes occur is necessary for progress in the field of evaluation. The realist approach explains why outcomes occur. Adopting a realist approach will create the potential for/increase the possibility of progress in the field of evaluation.	1

Appendix J

Realist Evaluation Challenges

Challenge	Sources	Response
Lack of Standardized Methodology	• Astbury, 2013 • Byng et al., 2005 • Jolly & Jolly, 2012 • Pawson & Manzano-Santaella, 2012 • Pawson, 2013, 2016 • Porter, 2015b • Punton et al., 2016 • Salter & Kothari, 2014 • Timmins & Miller, 2007 • Wong et al., 2015, 2016	• Adhere to RAMESES Quality Standards for Realist Evaluation (Greenhalgh et al., 2017) and RAMESES Reporting Standards for Realist Evaluation (Wong et al., 2016). • Consistent use of terminology to maintain clarity. • Full transparency in descriptions of researcher's positionality. • Detailed documentation of steps involved in the research process. • Transparent and upfront discussions with stakeholders regarding tentative nature of research process, including potential changes in methods and timing of data collection and analysis methods.
Ambiguity of CMOCs	• Astbury, 2013 • Byng et al., 2005 • Dalkin et al., 2015 • de Souza, 2013 • Jolly & Jolly, 2014 • King et al., 2016 • Pawson & Manzano-Santaella, 2012 • Pawson, 2013, 2017 • Porter, 2015b, 2017 • Punton et al., 2016 • Salter & Kothari, 2014 • Timmins & Miller, 2007 • Van Belle et al., 2016 • Westthorp, 2014, 2018	• Identify characteristics of program complexity using VICTORE complexity checklist to inform research design (Pawson, 2013). • Create logic model (admittedly non-comprehensive and intentionally non-binding; see Gasper, 1997, 2000) to present and revise during initial interviews as a visual aid. This will help establish initial program theories by indicating points on the program's implementation chain and identifying potentially relevant contextual factors. • Explicitly define types of mechanisms to distinguish from contextual factors (see Westthorp, 2018; de Souza, 2013, Dalkin et al., 2015). • Identify contextual factors related to structure, culture, agency, and relations to distinguish from mechanisms (de Souza, 2013). • Create CMO matrix to indicate hypothesized relationships between specific contextual factors, mechanisms, and outcomes (Pawson & Manzano-Santaella, 2012). • Use metaphors to help explain CMOCs (Punton et al., 2016).
Attribution	• Astbury, 2013 • Astbury, 2013 • Byng et al., 2005 • Dalkin et al., 2015 • Gill & Turbin, 1998 • Pawson & Tilley, 1997, 2013 • Sorinola et al., 2015 • Stame, 2004 • Tan & Harvey, 2016 • Timmins & Miller, 2007	• Narrow focus to contextual factors are most relevant to the research questions by identifying which factors align with stakeholders' interests, which would contribute most to the literature, and what is feasible given time constraints and access to resources (Pawson & Tilley, 1997). • Use iterative retroductive process to test and refine program theories (deductive reasoning to test theories, inductive reasoning to devise new theories).

		• Use the Relevant Explanation Finder as a framework for adjudicating between rival program theories (Lemire et al., 2012; Astbury, 2013).
Resource Intensive	• Astbury, 2013 • Gill & Turbin, 1998 • Greenhalgh et al., 2017 • Manzano, 2016 • Pawson & Manzano-Santaella, 2012 • Porter, 2015b • Punton et al., 2016 • Salter & Kothari, 2014	• Adhere to research schedule deadlines. • Ensure stakeholders understand the nature of the RE process and agree to proposed timeline and methods of data collection. • Limit iterations of theory testing and refinement to fit within scheduled deadlines.
Ethical Considerations	• Astbury, 2013 • Greenhalgh et al., 2017 • Lin & Wu, 2016 • Manzano, 2016 • Pawson & Tilley, 1997	• Adhere to IRB guidelines and IRB approved research plan. • Review IRB guidelines before engaging with consenting research participants. • Engage in continual self-reflection to identify any personal biases and how they might affect the research process.

Appendix K
RAMESES Realist Evaluation Standards

#	Standard	Criteria	Addressed?
1	The Evaluation Purpose	• A realist approach is suitable for the purposes of the evaluation. That is, it seeks to improve understanding of the core questions for realist evaluation. • The evaluation question(s) are framed to be suitable for a realist evaluation.	Chapters 1 and 3
2	Understanding and applying a realist principle of generative causation in realist evaluations	• A realist principle of generative causation is applied.	Chapter 3
3	Constructing and refining a realist program theory or theories	• An initial tentative program theory (or theories) is identified and developed. Programme theory is 're-cast' and refined as realist program theory.	Chapters 3 and 4
4	Evaluation Design	• The evaluation design is described and justified. • Ethical clearance is obtained if required.	Chapter 3
5	Data Collection Methods	• Data collection methods are suitable for capturing the data needed in a realist evaluation.	Chapter 3
6	Sample Recruitment Strategy	• The respondents or key informants recruited are able to provide sufficient data needed for a realist evaluation	Chapter 3
7	Data Analysis	• The overall approach to analysis is retroductive. • Data analyses processes applied to gathered data are consistent with a realist principle of generative causation. • A realist logic of analysis is applied to develop and refine theory.	Chapters 3, 4, and 5
8	Reporting	• The evaluation is reported using the items listed in the RAMESES II reporting standard for realist evaluations.	Chapters 4 and 5

#	Standard	Criteria	Addressed?
		Findings and implications are clear and reported in formats that are consistent with realist assumptions.	

Appendix L
Sample Realist Interview Questions

Term	Question Stems	Adapted for Practitioners/Facilitator	Adapted for Participants
Outcomes (O)	1. Can you tell me what your involvement in (or contact with) [program / element of program] has been? 2. What do you consider the outcomes of [program / element of program] to have been for [specific stakeholder group]? For program participants, this question can be asked “for yourself”, “for your family”, and so on. 3. Can you provide an example of [outcome named in previous question]? 4. Do you think that the outcomes have been the same for all [people within the specific stakeholder group – e.g. participants, workers]? In what ways have they been different? 5. Have the outcomes been the same for [specific sub-groups identified in the program theory – e.g. specific disadvantaged groups;	1. What has been your role in Watershed Collaborative and its <i>VTs Basics</i> OPD course? 2. What do you consider the outcomes of participating in <i>VTs Basics</i> to have been for course participants? 3. Can you give an example of how participants exercise reflective practice? 4. Do you think that the outcomes have been the same for all of the course participants? In what ways have they been different? 5. Have the outcomes been the same for participants who only complete [x%] of the assignments? In what ways have they been different?	1. What has been your involvement in the <i>VTs Basics</i> OPD course? 2. What do you consider the outcomes of participating in <i>VTs Basics</i> to have been for yourself? 3. Can you provide an example of your increased ability to facilitate inquiry-based discussions in your classroom? 4. Do you think that the outcomes have been the same for you and your colleagues? In what ways have they been different? 5. Have the outcomes been the same for your colleagues who teach [different subject]? In what ways have they been different?

Term	Question Stems	Adapted for Practitioners/Facilitator	Adapted for Participants
	different religions]. In what ways have they been different?		
Mechanisms (M)	1. We are curious about how [program/element of program] causes its outcomes. How do you think the program has caused, or helped to cause [outcome identified by respondent]? 2. Do you think [the program/element of program] changed the way [stakeholder group] thinks or feels about [program objective] in any way? In what ways? Can you provide examples? For participants: Has the program changed the way you think or feel about [x] in any way?	1. I am curious about how the reflection activities cause course outcomes. How do you think the program has caused, or helped to cause participants to become reflective practitioners? 2. Do you think the course participants' feedback changed the way Watershed team members think or feel about the action-based course assignments in any way? In what ways? Can you provide examples?	1. I am curious about how observing your colleagues implement VTS lessons causes course outcomes. How do you think the program has caused, or helped to cause your increased ability to facilitate inquiry-based discussions? 2. Has the program changed the way you think or feel about facilitating inquiry-based discussions with your students in any way?

Term	Question Stems	Adapted for Practitioners/Facilitator	Adapted for Participants
Context (C)	1. There are lots of ideas about how [program/element of program] actually works, and we think it probably works differently in different places or for different people. One of those ideas is [brief description of main mechanism]. Has it worked at all like that here/for you? Can you give an example? 2. What is it about the way [program] was implemented that made a difference to how it worked? Or ‘What is it about the way [Organisation] works that makes a difference to how it works? 3. We’ve seen that this [program] work differently in different places. What is it about this place that makes it work [so well, less well]?	1. There are lots of ideas about how the reflective course activities actually work, and we think they probably work differently in different places or for different people. One of those ideas is that these activities are successful with participants who teach [x subject]. Have they worked at all like that with the program participants in [BC schools]? Can you give an example? 2. What is it about the way Watershed Collaborative works that makes a difference to how the <i>VTS Basics</i> course works? 3. We’ve seen that the <i>VTS Basics</i> course works differently in different places. What is it about the participants from [BC schools] that makes it work [so well, less well]?	1. There are lots of ideas about how the action-based course activities actually work, and we think they probably work differently in different places or for different people. One of those ideas is that these activities are successful with participants who teach [x subject]. Have they worked at all like that with you and your colleagues in [BC schools]? Can you give an example? 2. What is it about the way the <i>VTS Basics</i> course was implemented that made a difference to how it worked? 3. We’ve seen that the <i>VTS Basics</i> course works differently in different places. What is it about [BC schools] that makes it work [so well, less well] here?
Recommended Changes	1. If you could change something about this program to make it work more effectively here, what would you change and why?	1. If you could change something about the <i>VTS Basics</i> course to make it work more effectively in [BC schools], what would you change and why?	1. If you could change something about the <i>VTS Basics</i> course to make it work more effectively here, what would you change and why?

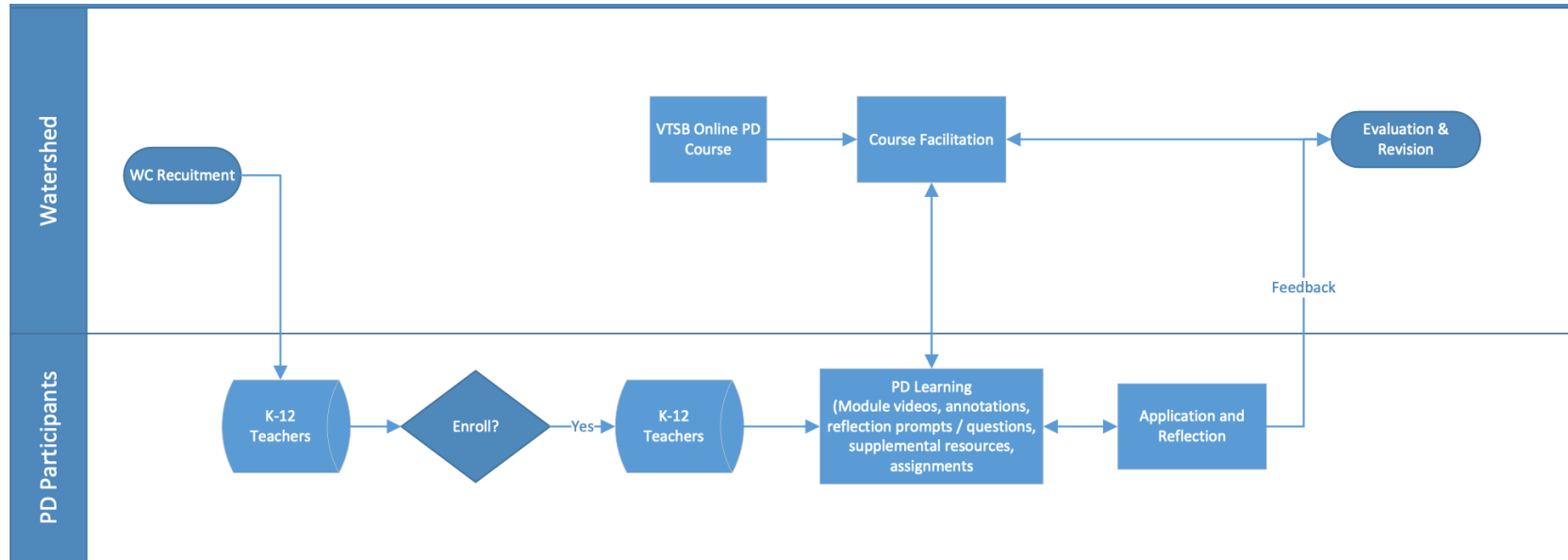
Term	Question Stems	Adapted for Practitioners/Facilitator	Adapted for Participants
	2. What else do you think we need to know, to really understand how this program has worked here?	2. What else do you think we need to know, to really understand how this course has worked in [BC schools]?	2. What else do you think we need to know, to really understand how this course has worked here?

Appendix M
Data Sources and Phases

Data Source	Phases	Description	Collection	Analysis
Relevant Literature	I, II, III	Theoretical literature review to help generate initial program theories and revise and refine subsequent program theories. Review Question: • What contextual factors (C) trigger mechanisms (M) that generate desirable outcomes (O) for a PD or OPD course (I)? • What substantive theories help explain why or how these contextual factors generate desirable outcomes for a PD or OPD course?	Realist Review • Six stages of the realist search (Booth, Wright, & Briscoe, 2019) ○ Stage 1: Formulate search questions ○ Stage 2: Background search/scoping the literature ○ Stage 3: Search for program theories ○ Stage 4: Search for empirical evidence ○ Stage 5: Final search to refine program theories ○ Stage 6: Document and report realist research	• Theory-driven qualitative coding techniques
Communications with Watershed Collaborative	I, II, III	Continual communication with the Watershed Creative and Operations directors throughout the project, from Fall 2018-Spring 2019.	• Archived emails • Project Charter • Notes from phone conversations	• Theory-driven qualitative coding techniques
Interviews	II, III	Six realist interviews with five participants. Two participants gave two interviews and two were interviewed together.	• Realist Interview transcripts	• Theory-driven qualitative coding techniques

Data Source	Phases	Description	Collection	Analysis
Course activity	II, III	Participants' completion percentages of modules, responses to question/reflection prompts, and course annotations.	• Excel spreadsheet ○ Videos watched ○ Question/reflection prompt responses ○ Annotations ○ Dates/times	• Theory-driven qualitative coding techniques •
<i>VTSB</i> course platform & resources	I, II	Exploration of course interface, organization, instructional content (syllabus, modules, videos, questions/reflection prompts, assignments), and supplemental resources.	• Outline of the course organization • Course materials • Usability test of course navigation, organization, content • Notes regarding content-related, structural, or didactic features that did or did not align with the effective features of OPD identified in literature.	• Theory-driven qualitative coding techniques (syllabus) • Summarize and synthesize notes
BCSD Data	II	Background information about participants' schools and district from BCSD Open Database.	• Relevant questions from 2018 School Survey for teachers • Demographic information • School Quality Snapshots, Guides, and Reviews • BC State PD Standards • School websites	• Summarize and synthesize notes • Compare with initial PTs

Appendix N
Watershed Collaborative Theory of Change Model



Appendix O
Email Invitation for Study Participation

Greetings,

My name is Stephanie Teague and I am a doctoral candidate in the Educational Psychology and Counseling Department at The University of Tennessee. I am writing to invite you to participate in my research study about the Watershed Collaborative's *VTs Basics* online professional development course. This study will gather and analyze information about participants' experiences, their perceptions of the course, and their implementation of course content. Research findings will be used to suggest course improvements.

You are eligible to be in this study because you are enrolled in the *VTs Basics* Fall 2018 course. If you decide to participate in this study, you will grant the researcher access to specific course data, including your end-of-course survey responses, course activity records, and course annotation transcripts. You may also agree to participate in an audio recorded 20-30-minute interview, with the possibility of a follow-up interview.

Your participation is completely voluntary. If you choose to participate, please read the attached consent form, which includes more detailed information about the study, potential risks and benefits of participation, confidentiality of personal information, and your rights as a participant. **After reading the consent form, please select one of the three options at the end of the form to indicate your desired level of participation in the study. Please email the completed form to the researcher at: steague4@vols.utk.edu.**

If you have other questions or concerns at any time about the study or procedures, you may contact the following:

Researcher: Stephanie Teague, steague4@vols.utk.edu.

Faculty Advisor: Lisa Yamagata-Lynch, lisayl@utk.edu

Institutional Review Board of The University of Tennessee, Knoxville: 865-974-7697 or utkirb@utk.edu.

Thank you very much.
Sincerely,
Stephanie Teague

Appendix P
Informed Consent Statement

INFORMED CONSENT

Realist Evaluative Study for Watershed Collaborative's *VTs Basics* Online Professional Development Course

Please email this completed form to the evaluator at: steague4@vols.utk.edu.

INTRODUCTION

You are invited to participate in a research study that is being conducted for the Watershed Collaborative's *VTs Basics* Online Professional Development Course. This study will gather information concerning course participants' experiences while completing the course, their perceptions of the course, and their implementation of course content. This information will be used to suggest course improvements.

PARTICIPATION IN THE STUDY

You were selected as a potential participant in this study because of your experience with the *VTs Basics* Online Professional Development Course. Your participation in this study is completely voluntary. Should you agree to participate, you will grant the researcher access to specific course data, including your end-of-course survey responses, course activity record, and course annotation transcript. You may also agree to participate in a 20-30-minute interview, with the possibility of a follow-up interview.

The interview(s) will be audio recorded so that the researcher can later transcribe and analyze your responses. Only those involved in the study will have access or listen to the audio recordings. They will be transcribed and destroyed prior to analysis. The course data, interview transcriptions, and observation notes will be encrypted and stored on a secure file hosting site using a secure password protected server. They will be erased when the study has been completed.

You may decline to answer any particular questions during an interview that you do not wish to answer for any reason. You may withdraw from participation from an interview at any time with no questions asked. If I withdraw from the study, the interview transcription, the accompanying audio recording, and observation notes will be destroyed.

RISKS

The risks for participating in this study are minimal. You may be uncomfortable answering certain questions about your experiences in the *VTs Basics* course. Please be aware that your responses will be kept confidential and that you may decline to answer any questions at any time.

BENEFITS

The information you provide may benefit you indirectly, if a colleague(s) decide to participate in the *VTS Basics* course in the future. At a minimum, the information learned in this study should provide general benefits for Watershed Collaborative and future course participants.

CONFIDENTIALITY

All of your responses will be kept confidential. Data will be securely stored and made available only to those involved in the study. The inadvertent release of the course data, interview recordings, and/or transcriptions may be a risk; however, confidentiality is ensured by giving participants pseudonyms. All audio recordings, transcriptions, observation notes, and course data will be stored on a secure file hosting site using a secure password protected server. No identifying references from the course data or information discussed during interviews will be included in written or oral reports. You will only be identified by pseudonyms.

CONTACT INFORMATION

If you have questions or concerns at any time about the study or procedures, you may contact the evaluator:

Stephanie Teague

steague4@vols.utk.edu.

CONSENT

I have read and understand the above information. I know who to contact regarding questions about my participation in this study. Highlighting the “Survey,” “Course Data,” and/or “Interview(s)” buttons indicates that I have read the above information and voluntarily agree to participate in these activities.

- ☐ Survey
- ☐ Course Data
- ☐ Interview(s)

Please email this completed form to the evaluator at: steague4@vols.utk.edu.

Please keep a copy of this consent form for your records.

VITA

Stephanie Jean Teague was born in New Orleans, Louisiana. She is the second of three children born to Richard and Cynthia Teague. She has an older brother, David, and a younger sister, Christianne. The family lived in Louisiana, Alaska, and England before settling in Texas.

After graduating from Oak Ridge High School, Stephanie attended Southern Methodist University in Dallas, Texas. She earned a Bachelor of Science degree in anthropology, with a minor in Classical Latin. Following her undergraduate degree, Stephanie earned a Master of Arts in Teaching degree at Rice University in Houston, Texas. As part of this program, she worked for a year as a full-time sixth grade Latin teacher in a Title I school in Houston.

Stephanie then moved to Knoxville, Tennessee to pursue a doctoral degree in Learning Environments and Educational Studies (LEEDS). Along the way, she earned a Master of Science degree in Instructional Design and Technology and served for three years as a graduate teaching associate for the course *Educational Psychology 401: Applied Educational Psychology*. She will graduate in August 2019 with her Doctor of Philosophy degree in LEEDS.