ABSTRACT

Title of dissertation	An Exploration of Motivation, Transfer, and Implementation During Self-Regulated Instruction and Cognitive Apprenticeships in Secondary Science and Social Studies Classrooms
	Cameron Butler Doctor of Philosophy. 2021
Dissertation directed by	Professor Susan De La Paz Department of Counseling, Higher Education, and Special Education

The Next Generation Science Standards and the Common Core State Standards for Social Studies emphasize the importance of disciplinary literacy in helping students to think and interact with texts like experts (NCSS, 2013; NRC, 2013). Developing these skills, however, is difficult and students do not naturally display disciplinary thinking and literacy skills (Hogan & Maglienti, 2001; Wineburg, 1991). Researchers have determined several effective practices for addressing these cognitively demanding skills including the cognitive apprenticeship model with embedded self-regulated instruction (SRI; see De La Paz et al., 2017; Levin et al., 2021).

In this dissertation, I discuss a range of topics that focus broadly on supporting students and teachers while they learn and teach cognitively demanding disciplinary literacy skills. I begin by examining SRI and its effects on motivation for students with or at risk for learning disabilities (LD). Next, I analyze the impact of a cognitive apprenticeship model that incorporates SRI on students' literacy skills in science. Finally, I analyze the implementation of a similar cognitive apprenticeship model in history by observing fidelity, chronicling teacher curricular adaptations, and documenting the important factors that influence teachers' instructional decisions.

The first study identified that SRI has a positive impact on student motivation in addition to its positive impact on academics (see Antononiou & Souvignier, 2007; Graham & Harris 1989a; Schunk & Cox, 1986). The second study corroborates research on the cognitive apprenticeship model in science (Lee et al., 2021a; Levin et al., 2021) by demonstrating that the model can support the development of literacy skills in science inquiry settings. This study also shows how the model helps students to transfer learning to functionally different scenarios. The final study details the implementation of a cognitive apprenticeship intervention that supported the use of historical literacy skills during historical inquiries. This study establishes that with online professional development (PD) and continued coaching, teachers showed high degrees of fidelity and made adaptations that adhered to the tenants of the model and supported their specific students. Additionally, this study demonstrated how important factors such as teachers' beliefs and pedagogical content knowledge (PCK) affect teacher implementation.

AN EXPLORATION OF MOTIVATION, TRANSFER, AND IMPLEMENTATION DRUING SELF-REGULATED INSTRUCTION AND COGNITIVE APPRENTICESHIPS IN SECONDARY SCOANEC AND SOCIAL STUDIES

by

Cameron Butler

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:

Professor Susan De La Paz, Chair
Associate Professor Kelly Cummings
Associate Clinical Professor Daniel Levin
Assistant Professor Sarah McGrew
Professor Ana Taboada-Barber

Dedication

I dedicate this dissertation to all those who have supported me throughout this process. Specifically, I would like to acknowledge my dissertation advisor Dr. Susan De La Paz who has been a mentor, support system, and friend during my time working with her. Additionally, I would like to recognize my wife Bailey Rose who has been my emotional and intellectual partner throughout this entire process.

Table of Contents

Chapter 1: Introduction and Overview

Recent policy initiatives in both science, such as the Next Generation Science Standards (NGSS), and history, such as the College, Career, and Civic Life (C3) Framework for Social Studies Standards, emphasize the practical importance of disciplinary skills (NCSS, 2013; NGSS, 2013; NRC, 2013). Paramount among these skills is disciplinary literacy, which entails thinking, reading, and writing. There is an important distinction between general literacy and disciplinary literacy (LaDuke et al., 2016; Moje, 2008; Sandoval & Millwood, 2005, Shanahan & Shanahan, 2008). Disciplinary literacy is not merely thinking, reading, and writing within a given discipline like science or history; instead, it requires unique skills that emulate the way that experts in these fields such as scientists or historians think and interact with text (Seixas, 2006; Russ et al., 2006). Disciplinary literacy is the "ways of thinking, reading, and writing that are embedded in the production, consumption, and communication of knowledge in a discipline" (Monte-Sano, et al., 2017). Researchers and policy makers alike recognize the critical importance of disciplinary literacy. For example, argumentative writing, an important form of disciplinary literacy, can help students to build disciplinary thinking skills and even improve student content knowledge (De La Paz & Felton, 2010; Kuhn, 2010; Sampson, 2013a). By developing disciplinary literacy skills, students begin to think and reason in similar ways to experts in specific fields and can build skills that will be important in post-secondary settings.

For students to master disciplinary literacy, they are required to learn a series of complex and cognitively demanding skills (Monte-Sano et al., 2014; Osborne, J. 2010; Sampson et al., 2013b, Shanahan & Shanahan, 2012). Becoming proficient in many of these skills also means having strong fundamental skills such as reading comprehension, a proficient vocabulary, transcription skills, grammar skills, and even self-regulatory skills (Vacca et al., 2014). As a

result, the pursuit of disciplinary literacy is no small task. Researchers have found that students do not naturally exhibit many of the skills needed to engage in disciplinary literacy; however, they do exhibit the capability to learn these skills (De La Paz & Levin, 2018; Hogan & Maglienti, 2001; Wineburg, 1991).

The research community has identified several effective instructional supports to guide students as they develop disciplinary literacy skills (see Reisman, 2012; Sampson et al., 2013a,b). One such support is the cognitive apprenticeship model. This model is one of the primary foci of this dissertation and I explore both its effectiveness and implementation in the coming chapters. In what remains of this chapter, I first discuss the core tenants of the cognitive apprenticeship. Next, I differentiate disciplinary literacy and discuss the unique qualities of both scientific[1] and historical literacy. Within each discipline I outline research on effective instruction. Furthermore, when discussing science, I provide an overview of the field of transfer research, which I hoped to evaluate when studying a cognitive apprenticeship with students who were learning middle school science. When discussing history, I discuss teachers' beliefs, pedagogical content knowledge (PCK), and their implementation of historical reading and writing instruction, when exposed to online professional development (PD). Finally, I will outline my remaining chapters and discuss the relatedness of all three chapters within my dissertation.

Cognitive Apprenticeship

The skills required to master disciplinary literacy are cognitively demanding. Furthermore, many of the practices are not obvious especially to students who may have very

[1] In this dissertation we use scientific literacy to mean literacy in science or the ways that scientists think, read, and write in scientific settings.

little experience engaging with discipline specific texts (Leming et al. 2000; Levin et al., 2021; Russell & Waters, 2010; Stahl, 1996). For these reasons, it is unfair to expect and unwise to assume that students have developed these skills without proper instruction. One model for instruction particularly well suited for these types of skills is the cognitive apprenticeship model (Brown et al., 1989). At its core, this model is similar to a traditional apprenticeship and provides authentic environments for students to learn new skills from experts. It differs, however, from traditional apprenticeships because the focus is on the often hidden cognitive processes and strategies that experts like scientists and historians use in their respective fields (Collins et al., 1991). For disciplinary literacy, the cognitive apprenticeship model helps educators to make expert thinking and literacy practices visible to students through modeling and think-alouds (Monte-Sano et al., 2017; Engle & Conant, 2002; Levin et al., 2021). As students practice and engage in new ways of thinking, reading, and writing that simulate the expert, less modeling is required. Educators model less and begin to systematically remove scaffolds to push students toward mastery. Targeted feedback remains an important student support throughout the entire model (De La Paz et al., 2014; Greenleaf et al., 2011).

The cognitive apprenticeship model serves two vital functions; first, as an organizational structure to help design an effective learning environment, and second, as an approach to learning that helps students recognize the processes involved in complex skills (De La Paz et al., 2017). Through the model, students learn how to use disciplinary skills and eventually achieve similar thinking and literacy habits to those experts in the field. When originally conceptualizing the cognitive apprenticeship, Brown and his colleagues (1989) envisioned the model as an instructional framework that emphasizes making higher-order complex skills more visible such that novice learners can better understand the thinking processes that are used by experts.

Brown et al. (1989) and other research on the cognitive apprenticeship model heavily influenced Harris and Graham when designing the evidence based practice, self-regulatory strategy development (SRSD; Harris & Graham, 2009). SRSD emphasizes self-regulated learning, and students are explicitly taught self-regulatory skills in addition to academic content (Graham & Harris, 1999). This approach has been identified as effective for teaching writing to students with learning disabilities (LD) and academically talented students (De La Paz, 1999; Graham et al., 2005; Graham et al., 2012). When researchers have used the cognitive apprenticeship model in secondary general education settings they have employed specific components of SRI including goal-setting, modeling self-talk, self-questioning, and self-evaluation (see De La Paz et al., 2017; Levin et al., 2021). Instruction that targets self-regulated learning has been effective in multiple academic domains including reading (Antononiou & Souvignier, 2007), writing (Harris et al., 2006), and mathematics (Schunk & Cox, 1986). In addition, instruction targeting self-regulated learning can also promote student motivation (De La Paz & Butler, 2021).

Science

Each discipline requires unique ways of thinking and interacting with text. A scientist, for example, must make claims to address scientific inquiries and use both evidence and reasoning to support the claim and connect it to the evidence (Ford, 2008; Kuhn, 2010; McNeill & Krajcik, 2012). In this way, disciplinary literacy in science (scientific literacy) is functionally different from other forms of literacy.

Scientific Literacy. The NGSS describes science learning as the development of three important dimensions; disciplinary core ideas (DCIs), cross cutting concepts (CCCs), and engagement in scientific and engineering practices (SEPs; NGSS, 2013). Scientific literacy plays

a particularly important role as students engage in scientific practices which require them to think, read, and write like scientists (NGSS, 2013; NRC, 2013). An important component of scientific literacy and the central focus of my second study is the use of argumentation and argumentative writing in science. Argumentation is particularly important because it supports students' understanding of three essential practices when engaging with a scientific inquiry, (a) constructing a claim, (b) investigating evidence, and (c) considering reasoning or justification (Berland & McNeill 2010; Duschl and Osborne 2002; Driver et al. 2000; Ford & Forman, 2006; Osborne et al. 2004). By engaging in arguments, students are required to not only address inquiries by making a claim, but also justify the claim through evidence and reasoning. In this way, students simulate how a scientist would approach a scientific inquiry and leverage scientific literacy skills (Driver et al., 2000; Felton et al., 2019; Sampson et al, 2013a, b). Ford (2008) contends that through argumentation students develop a "grasp of practice" and grapple with their roles as scientists.

Scientific literacy skills are not naturally occurring in students and take time and proper guidance to develop (Hogan and Maglienti, 2001). In fact, students often struggle to engage in scientific inquiry because in school, students are rarely given the opportunity to engage in authentic scientific inquiries (Levin et al., 2021). Students have trouble understanding the purpose of reading and writing in science and set inadequate goals (Ferretti et al., 2001). Their biggest challenge, however, is providing justification that can connect a claim to the evidence. For this skill, students must be able to articulate the reasoning that integrates scientific principles into the claim and evidence (McNeill & Krajcik, 2012; Russ et al., 2008).

Effective Instruction for Scientific Literacy. In response, several science education researchers have developed interventions that specifically focus on inquiry and scientific literacy.

Sampson and colleagues (2013a, b) pioneered Argument Driven Inquiry (ADI) to help students engage in authentic inquiries and support their writing. ADI transforms traditional laboratory activities into short units that present information as authentic science inquiries and require argumentation (Chu et al., 2019; Hutner et al., 2020). Students address these inquiries by executing experiments, collecting evidence in the form of data, and writing arguments with justification (Sampson & Murphy, 2019). In a recent report, Levin et al. (2021) modified ADI and taught key elements using a cognitive apprenticeship model. The authors focus on improving student scientific literacy through the use of classroom discussion and argumentation. Students who received the model wrote more effective arguments to address the central inquires. Further support for the benefits of this approach with struggling learners comes from Lee and De La Paz (2021a) who explored the use of the cognitive apprenticeship model in science for students with LD and English Language Learners (ELL). They showed that with effective instruction, students with LD and students who were ELL could develop scientific literacy skills and write stronger science explanations. The authors embedded cognitive, and language scaffolds into laboratory activities to support student explanation writing skills in ways that aligned with earlier researchers. All six students in the study wrote stronger explanations that suggested a better understanding of causal and mechanistic reasoning. Additionally, all students showed better lexical and grammatical abilities.

Transfer. Absent from the research on the use of a cognitive apprenticeship model to promote scientific literacy are considerations for how students transfer knowledge. Transfer is the ability of a person to take something learned in one setting and apply it to another setting (Kober, 2015; Mestre, 2003). At its simplest, transfer can be understood as a continuum with more similar environments requiring less transfer while more divergent environments require

more transfer (Barnett & Ceci, 2002). The ends of this spectrum are often called near transfer to describe transfer to a similar setting and far transfer to describe transfer to a different setting (Kober, 2015). The impact of an instructional practice can in theory be measured by how it supports student transfer. This is important as those practices that limit transfer are less impactful to a students' overall learning while those instructional practices that promote transfer are more impactful to a students' overall education.

Promoting transfer is a difficult undertaking as many students seem to struggle to apply knowledge gained in one setting to even relatively similar settings. Perkins and Saloman (2012), for example, found that students struggled to transfer mathematics skills from one problem to the next when the only difference was the scenario description. In the first problem, students were taught to calculate the velocity of a ball falling from a certain distance in the sky. These same students were asked to calculate the velocity for the same distance drop; however, in this problem, the ball was dropped into a well. Despite the same math equation and number values, most students could not complete the second problem.

To address this issue, researchers have attempted to focus on three main approaches: promoting content transfer, facilitating context transfer, or a combination of the two (Engle et al., 2011, 2012; Mestre, 2003; Reeves & Weisberg, 1994; Schwartz & Nasir, 2003). For content transfer, many researchers have explored the use of specific supports to initiate learning that is deep, strong, and lasting (Gentner et al., 2017; Gick & Holyoak, 1983; Reeves & Weisberg, 1994; Rittle-Johnson & Star, 2007). Researchers contend that by creating deep knowledge, students are more likely to retain key understanding regardless of the setting (Chi & VanLehn, 2012; Engle et al., 2012). To create initial deep learning, teachers can provide multiple examples and non-examples to help students recognize similarities between contexts and provide content

cues that may be generalizable to multiple settings (Anolli et al., 2001; Bransford et al., 1999; Chang, 2006). For context transfer, teachers can create environments with common contextual cues that students learn to recognize across settings. This means using structures that may be common between many settings and even using common social cues that can be recognized in different contexts (Barnett & Ceci, 2002; Engle et al., 2011, 2012; Thorndike, 2009). The most effective instruction likely incorporates supporting both content and context transfer.

History

Historians think and interact with text differently than scientists, in large part because the problems that they seek to solve are different and based on different forms of evidence (De La Paz & McCutchen, 2017). Historians use a series of disciplinary skills to draw meaning from sources and construct their own interpretations (Holt, 1990; Booth, 1983; Wineburg, 1991). Some of the most important skills include contextualizing (understanding a given text within its proper historical context), sourcing (evaluating the source including the author's reliability and motives), and corroborating (evaluating other texts from the same time period to recognize similarities and differences; Seixas, 1998, 2006; Wineburg and Wilson, 1991).

Historical Literacy. Historical literacy requires the mastery of a number of cognitively demanding skills (Monte Sano, 2008). Students do not naturally display many of these skills; however, each is important to properly engage with primary source documents and think like a historian (Nokes, 2010; Wineburg, 1991). One of the most important skills in historical literacy is the use of argumentation which allows students to address historical inquiries and create personal interpretations of historical events (VanSledright, 2010; VanSledright & Kelly, 1998). If students are taught important historical literacy skills such as contextualizing, sourcing, and corroboration, they can evaluate historical documents and construct effective arguments to

address historical inquiries. By using historical literacy skills, students can examine the author and their potential influences or biases, evaluate the source within the particular context, and decide how the source may support or rebut an argument (De La Paz et al., 2021).

Developing these skills can be challenging. Without instruction students rarely acknowledge discrepancies between sources and are likely to assume primary sources are reliable (Stahl, 1996). Students do not often evaluate authors and infrequently consider how the author's motives may bias a source (Wineburg, 1991). Additionally, students struggle with presentism, a phenomenon in which they apply modern standards to historical actors or sources (Huijgen et al., 2017). Adding to these concerns, teachers are often ill-prepared to teach inquiry and historical literacy in the classroom (Barton, 2005; Leming, et al., 2009; Levesque, 2009; Wineburg, 2006).

Effective Instruction for Historical Literacy. Researchers have identified argumentation and inquiry as the cornerstones for instruction that helps students think more like historians (Nokes, 2007; De La Paz, 2005). By allowing students to grapple with authentic historical inquiries and argue using primary sources as evidence, students can develop skills that emulate the tasks of historians (Afflerbach & VanSledright, 2001; De La Paz et al., 2021; VanSledright, 2010; VanSledright, et al., 2006). Studies have also found that when targeted specifically at supporting historical literacy skills, techniques such as providing direct instruction, facilitated guided practice, and providing feedback can help students develop more expert ways of thinking and interacting with texts (Leinhardt, 2000; Nokes et al., 2007; Reisman, 2012; Stoel et al., 2015). For example, De La Paz et al (2014, 2017) provided teachers and students with procedural and cognitive supports for reading and writing within a cognitive apprenticeship model to support middle school students' ability to write historical arguments.

Her team taught critical historical thinking skills such as contextualizing, sourcing, and corroborating to help students analyze sources and write arguments to address a central historical inquiry. Over the course of the cognitive apprenticeship, teachers modeled specific ways of thinking, provided guided practice for analyzing sources, had students engage in discussion, supported writing arguments, and provided targeted feedback. The authors found that this intervention resulted in stronger content knowledge as well as more persuasive and elaborate written arguments. Other studies have corroborated the effectiveness of this intervention for students with and without disabilities as young as fourth grade (Wissinger and De La Paz, 2020).

Teachers' Beliefs, PCK, and Implementation. The effectiveness of a cognitive apprenticeship model is well established in history. Multiple studies with students of different age groups and abilities have demonstrated that with proper instruction, students can develop historical literacy skills and emulate the skills used by historians (see De La Paz et al., 2017; De La Paz & Wissinger, 2017; Wissinger et al., 2021). The effects of even proven interventions, however, are largely determined by teacher fidelity and decisions related to implementation (Carroll et al., 2007). It has long been understanding that important factors such as beliefs and PCK influence teacher implementation. Fives and Buehl (2012) for example offer a workable framework to explain how teachers make instructional decisions, which I use as an organizational scheme in Chapter four. Unfortunately, there is little research on cognitive apprenticeship models in history instruction that examines the influences that determine teacher fidelity and implementation decisions. In this dissertation, I attempt to unpack how teachers' beliefs, content knowledge, PCK, and other factors influence how teachers implement a provided curriculum.

Teacher Professional Development. Researchers have identified several key components that can contribute to effective PD in history. PD should be content focused, incorporate active learning, include collaborative work, use models of practice, provide ongoing coaching, and include feedback (Darling-Hammond et al., 2017; Grossman et al., 2009). It is through these effective forms of instruction that researchers can hope to not only instruct a teacher on an intervention, but also change important factors that affect implementation such as beliefs and PCK. Monte-Sano et al. (2017), for example, describe the PD that introduced disciplinary cognitive apprenticeships by employing many of these best practices. The authors modeled key components of the intervention, allowed for cooperative planning and practice time, gave effective targeted feedback, and provided continued coaching throughout the study. Monte-Sano et al. (2017) documented the effects of their approach to PD on teacher learning. After teachers received PD on instruction to support students' historical literacy, teachers were more focused on key aspects of historical writing, they paid more attention to evaluating students' historical thinking, they considered the quality of student work (not just completion), and they were better able to identify student needs. In my final study, I consider the effects of an online PD series on the changing beliefs, PCK, and other factors of the teachers.

Summary

Recent changes in policy for both history and science place more emphasis on student ability to use disciplinary literacy (NCSS, 2013; NRC, 2013). Students do not naturally exhibit many skills needed for disciplinary literacy and face certain challenges when acquiring them (Nokes, 2010; Wineburg, 1991). Students must build several cognitively taxing skills to engage in disciplinary literacy such as constructing claims, using evidence, and providing reasoning in science, as well as contextualizing, sourcing, and corroborating in history (Seixas, 2006; Russ et

al., 2006). Having students engage in argumentation in science and history is an effective form of instruction for developing content knowledge and discipline specific literacy skills (Levin et al., 2021; Nokes & De La Paz, 2018). The cognitive apprenticeship model is a particularly promising model for helping students engage in argumentation and develop disciplinary literacy skills (see De La Paz et al., 2017, Levin et al., 2021). In this dissertation, I seek to expand the research on supporting teachers and students at they teach and learn disciplinary literacy. I address transfer in science as well as examine how teachers' beliefs and understandings influence curriculum implementation in history.

Synopsis of the Projects

This dissertation consists of three main aims. Each aim contributes to understanding how to support teachers and learners advanced literacy skills. Specifically, this dissertation considers the role of self-regulation that learners need to perform complex academic tasks and the use of cognitive apprenticeship models to support teachers and students with tasks that take a year (or more) to understand and accomplish. I conducted three studies in the pursuit of my goals. The aim for my first study was to determine the overall effectiveness of self-regulated learning in general on academic motivation for students with or at risk of LD. Though SRI has been shown to positively impact struggling students' reading, writing, and mathematics performance (see Antononiou & Souvignier, 2007; Graham & Harris 1989; Harris et al., 2006), prior literature did not conclusively establish if it also positively affected student motivation (De La Paz & Butler, 2018).

The aim for my second study was to explore how a cognitive apprenticeship model could support middle school students' scientific literacy skills. Researchers have identified the cognitive apprenticeship model as an effective model for instruction in history (see De La Paz,

2005; De La Paz et al., 2017; De La Paz & Wissinger, 2017; De La Paz et al., 2021), but there is considerably less research into this model's effectiveness in science (Levin et al., 2021). The aim for my final study was to better understand teacher implementation of a curriculum that used a cognitive apprenticeship model of instruction to teach historical literacy skills. Teacher fidelity, curricular adaptations, and different forms of implementation are important to the overall effectiveness of any intervention that relies on teachers' decision making to improve student learning, so I attempted to better understand these facets of instruction.

Significance

The goal of this dissertation is to better understand self-regulated instruction and whether it can be used to motivate student learning and how a cognitive apprenticeship model may promote transfer of student learning. I also seek to understand the role of teachers in the implementation of a cognitive apprenticeship. Through the dissertation, I probe these topics to determine how to support both teachers' and students' learning in secondary content area settings. The dissertation contributes to the field of research in three important ways. First, my synthesis expands the field's understanding of self-regulated instruction in academic content areas of reading, mathematics, and writing for students with learning disabilities and students who are at risk for LD. My findings confirm that this form of instruction can help struggling students' academic motivation. Second, this dissertation adds to the research on the cognitive apprenticeship model in science classrooms. My analysis of the independent and dependent variables when using this model show how it can support the transfer of students' scientific thinking and writing from one task to a functionally different and more complex task. Finally, I qualitatively study the implementation of the cognitive apprenticeship. I observe fidelity, detail how teachers make adaptations, and describe the factors that affect teacher decision making.

Description of Aims

Aim 1: Synthesize findings from the current literature to identify effective forms of self-regulated instruction that can improve academic motivation for students with or at risk for learning disabilities.

I have an in press systematic review (Butler & De La Paz, in press) of the existing literature on self-regulated learning (SRI) and student motivation for students with or at risk for LD. In this synthesis, we identify the components used in recent academic SRI interventions and examine if these components were effective at improving academic motivation for struggling students. We identify five common components of SRI instruction and report effect sizes to establish the impact of these intervention components. This synthesis further establishes the importance of SRI and helps to expand the research on what components of SRI are most impactful.

Aim 1 Research Questions:

RQ 1. What instructional components are present in the academic self-regulation intervention literature focused on motivation?

RQ 2. To what extent do academic interventions with self-regulated instructional elements improve motivation in struggling learners?

Aim 2: Analyze data gathered during an intervention study that employed the cognitive apprenticeship model in an 8th grade science classroom. Observe student outcomes related to scientific literacy including argumentative writing and detail if students can effectively perform near and far transfer tasks.

The second study examines the outcomes of an intervention that used the cognitive apprenticeship model to support student scientific thinking and argumentative writing in an 8th

grade science classroom. Researchers have identified many highly effective instruction models for supporting student disciplinary thinking in history education such as the cognitive apprenticeship model (see De La Paz et al., 2017 and Monte-Sano et al., 2017). Science, however, has a less extensive research base. This study expanded the research on the cognitive apprenticeship model and helps to illuminate how this model affects student learning. In this study, I observe both near transfer and far transfer outcomes and speculate as to how the model of instruction may support students' ability to transfer complex skills between settings. The results help to establish this model as effective in science settings and help to open the door for further examinations into student transfer of complex scientific thinking and writing skills.

Aim 2 Research Questions:

RQ 1: Do students demonstrate improved claims, evidence, and reasoning (CER) on a near transfer outcome of scientific argumentative writing?

RQ 2: Do students demonstrate improved claims, evidence, and reasoning (CER) on a far transfer outcome of scientific argumentative writing?

RQ 3: Do students show improved written argumentation outcomes when compared to a nonequivalent comparison group?

Aim 3: Describe teacher fidelity, how teachers make adaptations, and how important factors such as teachers' beliefs, content knowledge, and PCK affect the implementation of a yearlong 9th grade US history curriculum that uses the cognitive apprenticeship model.

Understanding teacher fidelity and how teachers make adaptations is important to successfully implementing any curriculum, especially one that focuses on complex skills such as historical literacy. Furthermore, to properly address fidelity, it is important to unpack the factors that affect teachers' decisions around implementation. To address these needs, I analyzed data

gathered during the roll out of a yearlong 9th grade US History curriculum that focuses specifically on inquiry learning and historical literacy. I collected data during coaching sessions, surveys, teacher artifacts, and other data sources to qualitatively address the goals of the study. I identify a number of key factors specific to our intervention including context, teacher and student characteristics, teacher beliefs, teacher content knowledge, teacher PCK, self-efficacy, and value that may affect implementation. I identify several themes and cross cutting explanations that help to explain teacher fidelity and implementation.

Aim 3 Research Questions:

RQ 1: To what extent do teachers' implement the curriculum with fidelity?

RQ 2: How do teachers adapt the yearlong curriculum? Are teachers' adaptations responsive to student needs? Do the adaptations facilitate or pose challenges for meeting the overall goals of the historical reading and writing curriculum?

RQ 3: In what ways do teachers' knowledge, goals, and beliefs of historical literacy influence how they implement the curriculum?

The following three chapters are presented as individual papers. This is to reflect that each paper has or will be submitted for publication as standalone manuscripts (Chapter 2 has been accepted for publication in *Learning Disabilities: Research and Practice* and Chapter 3 has been submitted for publication in *Research in Science Education*). Chapter 4 will be adapted further and likely streamlined for publication in a social studies journal such as *Theory and Research in Social Education* or the *Journal of Social Studies Research*, a teacher education journal (e.g., *Teacher Education Quarterly*) or a general literacy journal (e.g., *Journal of Literacy Research*). Individual titles are provided for each of the chapters as well as references (see Butler, 2021; Butler & De La Paz, in press; Butler et al., 2021) but abstracts are not

included. Chapter 5 will summarize the purpose, identify the major findings, make recommendations for future research, and explicate how the papers fit together.

Chapter 2: Synthesis on the Impact of Self-Regulated Instruction on Motivation Outcomes for Students with or at Risk for Learning Disabilities

The National Assessment of Academic Progress (NAEP) provides a sobering account of academic achievement in the United States. Only 37% of students reach proficiency in reading, 27% in writing, and 25% in mathematics by the twelfth grade (The Nation's Report Card, 2019). For students with or at risk for LD, who comprise many of the nation's struggling learners, this concern is amplified. Such groups lag well below age range peers across academic and adaptive skills (Wong, 2013). By definition, these students show deficits in one or more academic domains. In reading, fluent word recognition can be difficult, given students' foundational processing difficulties (Bryant et al., 2000). In writing, difficulties including simultaneously managing planning, text production, and revising are well known (De La Paz & Graham, 2002). Mathematics proves similarly difficult for many who struggle with basic computation and/or problem solving (Miller et al., 1998). In addition to difficulties specific to reading, writing, and mathematics, struggling learners experience decreased working memory, poor organizational, self-regulated or executive control skills, or they lack efficient learning strategies (Graham et al., 2018). As such, many of these students experience significant learning challenges that some suspect reflect underlying LD, that may not be formally diagnosed (c.f., Fuchs et al., 2004).

In addition, struggling learners[2] often show diminished academic motivation (Graham et al., 2017; O'Shea et al., 2017; Troia et al., 2012). In reading, a direct relationship between motivation and skill proficiency has been observed as early as first grade, as students who experience early failure become less motivated to read (Morgan et al., 2008). In writing, students

[2] We use this term throughout our manuscript, referring to students with or at risk for LD

with LD appear less engaged and attribute success to factors outside of their own control, which is less common for students with average literacy (García-Sánchez & de Caso-Fuertes, 2005). Similarly, in mathematics, differences exist between high and low achieving students with respect to self-efficacy, interest, and goal orientations (O'Shea et al., 2017).

Understanding Motivation

Though student deficits in motivation are commonly observed, the relationship between motivation and academic outcomes is not well understood. Conradi et al. (2014) cogently observed that despite the relevance of motivational factors to reading comprehension, the field of education remains plagued by difficulties in understanding essential constructs and developing accurate measures of those constructs. In addition, writing researchers have long theorized which motivation components are causally linked to writing outcomes and tried to determine how motivation variables influence short- and long-term engagement (Graham et al., 2018).

Schiefele et al. (2012) suggest there are both genuine motivation constructs and antecedents to motivation. Genuine motivation constructs include concepts of intrinsic and extrinsic motivation such as attitudes, intrinsic value, and task value. In contrast, antecedents to motivation include concepts related to students' expectancies for success in a given domain which can influence student motivation in themselves. Students' attitudes and self-efficacy are both components of motivation that bear further examination.

In fact, in a review that attempted to determine if the theoretical importance of motivation on writing was supported by existing research, Graham (2006) found that the two motivational beliefs most studied in writing were attitudes toward writing and self-efficacy. Graham et al. (2017), extended this finding as he and his colleagues subsequently found that attitudes toward writing and writing self-efficacy collectively predicted the quality and length of narrative stories

written by fourth graders. Of special relevance to this paper, Graham's work, and that of Schiefele et al. (2012), describe attitudes as feelings that cause learners to approach or avoid academic situations. This is of importance as students who hold more positive attitudes are more likely to persist or put more effort into their work, with improved learning outcomes (Graham et al., 2017a; Schiefele et al., 2012). The predictive importance of these variables has been similarly identified in reading (c.f., Vansteenkist et al., 2006) and mathematics (Pajares, 1996).

Self-efficacy is important because it is the degree to which students' have confidence in their own ability or capacity to achieve an intended outcome (Bandura, 1994). Self-efficacy therefore appears to function as an antecedent to motivation. Self-efficacy can also change (increase or decrease) depending on student perceptions of success and failure after engaging in activities that are intended to reach specific outcomes. Thus, self-efficacy's role can function as a consequence or dependent measure after interventions.

Self-efficacy is also important as it relates to self-regulation (Zimmerman & Risenberg, 1997). Self-regulation is a process by which students activate cognition and systematically direct behaviors to when perusing academic goals. Students use self-regulatory processes to direct behaviors to achieve specific goals (Schunk & Zimmerman, 1994). Research shows that poor academic performance among students with LD is often attributable to limited self-regulated skills (Harris & Graham, 1999). Common self-regulated behaviors that are often targeted for instruction are goal setting, self-assessment, self-instruction, self-reinforcement, imagery, and managing the academic environment (Harris et al., 2005). The development and application of self-regulation requires an adequate level of motivation. Pintrich (1999) demonstrated that motivation promotes self-regulated learning, indicating that motivation is necessary to ensure persistence of effort and use of self-regulated behaviors. Interventions that improve students'

academic performance are likely to improve their self-efficacy, goal setting, and attributions; therefore, given the reciprocal nature between learning and motivation, it seems likely that interventions with self-regulatory elements may also have positive effects on struggling learners' motivation.

Schunk and Zimmerman (2012) believe that motivation supports cognitive ability. In reading, motivated students show more consistently goal-directed actions and higher levels of academic engagement (Vansteenkist et al., 2006). Highly motivated students persist in the face of obstacles and spend more time reading in non-academic settings (Wigfield & Guthrie, 1997). In writing, positive motivation is associated with greater task persistence (Zimmerman & Ringle, 1981) and with improved academic achievement (Troia et al., 2012). Furthermore, skilled writers typically hold more positive motivational beliefs than unskilled writers and differences in motivational beliefs could effectively predict writing achievements (Graham et al., 2018). Students who are motivated, work through challenges by using learning and coping strategies resulting in academic success.

Previous Reviews and Syntheses on Interventions designed to improve Motivation

To address the motivation challenges that struggling learners often exhibit, effective interventions need to be identified. Recently three literature reviews provide evidence that academic interventions can be used to improve motivation. Unrau et al. (2018) explored the relationship between interventions and self-efficacy for reading. Their meta-analysis of 30 studies found interventions improved students' self-efficacy, with small to moderate effect sizes (ES). Additionally, in interventions that reported on comprehension, there were again strong positive correlations with reading self-efficacy.

In a recent meta-analysis of technology-based mathematics interventions, Higgins et al. (2019) observed how technology-based mathematics interventions affected student achievement, motivation, and attitude. After a systematic review process, Higgins et al, (2019) identified 24 studies. They calculated average weighted ES and found a small but significant effect of technology-based mathematics interventions on student motivation. Higgins et al. (2019) found that students showed considerably less improvement in motivation than achievement. Additionally, motivation ES varied across dimensions such a study design, intervention type, and dosage. De La Paz and Butler (2018) examined the impacts of writing interventions on student motivation for struggling students. After a systematic search we identified 19 writing studies that specifically targeted struggling learners with empirical measures of participants' motivation. Our findings revealed inconsistency in how motivation was defined and measured. One surprising result was that limited or nonsignificant benefits were evident in several strategy or SRI studies regarding improvement in motivation.

These reviews examined how specific interventions can help to improve motivation outcomes for students. Yet, while each set of authors examined motivation, there were discrepancies in how motivation was defined and observed. Moreover, each paper focused on only one of the major academic domains and reported on different motivation constructs. Unrau et al. (2018) measured student self-efficacy while Higgins et al. (2019) and De La Paz and Butler (2018) both reported on multiple outcome measures related to motivation including self-efficacy, task value, attributions, etc. Additionally, only De La Paz and Butler (2018) focused on struggling students. Unrau et al. (2018) provided descriptive information on struggling students but did not disaggregate self-efficacy outcome data and Higgins et al. (2019) made no mention of learning status.

More importantly, despite the importance of SRI and its theoretical connections to motivation, we were unable to find an article other than our own prior review that specifically discussed how instruction in self-regulation may affect student motivation. These authors were surprised to find that studies that included self-regulation had mixed effects on student motivation outcomes. Unrau et al. (2018) discussed the relationship between self-regulation and self-efficacy, though did not explore how studies that used SRI affected self-efficacy. Higgins made no mention of self-regulation. Thus, past reviews provide little information about the impact of SRI on motivation for struggling students across academic domains.

Current Study

To address gaps in preceding literature reviews and meta-analyses, we investigate self-regulated learning interventions in reading, writing, and mathematics and examine the effectiveness of instructional components on student motivation. In doing so, we focus on struggling learners, partly due to estimates about under identification of LD (U.S. Department of Education, 2010). Prior work indicates students who are academically comparable but not formally identified as LD make up close to a quarter of students in classrooms (Fuchs et al., 2004), yet we do not know how many struggling learners belong to each group. Clearly, issues with over and under identification of LD are ongoing, leading us to search for studies with students who were struggling academically, not just those labeled as having LD.

Moreover, we extend our prior work to also include mathematics and reading interventions in the current study. Of note, we focus solely on interventions that include some form of SRI because this form of instruction has consistently been shown to have positive academic outcomes for struggling learners (Cuenca-Carlino et al., 2016; De La Paz & Graham 2002; Miranda et al., 1997). In our earlier literature review we explored motivation outcomes for

writing studies that included students with LD and noted two important findings that helped to influence the current synthesis. First, there were several different constructs to measure motivation, presumably as researchers held different views on what these constructs meant. Unfortunately, this complicated our interpretation of results Additionally, we were surprised that, many writing intervention studies involving explicit strategy instruction or SRSD resulted in limited or nonsignificant benefits with respect to motivation. Interestingly, each of these studies resulted in improved written language skills but not motivation.

These two findings influenced the current work which seeks to, among other things, document what constructs studies use to measure motivation and analyze the effect of studies using SRI across reading, writing, and mathematics on motivation for struggling students. The current study shares nine articles with the work by De La Paz and Butler (2018). These articles focused on writing and used SRI, meaning that the articles fit within both studies parameters. Our (2018) work was a literature review and did not have the level of rigor for search protocols as the current synthesis. We seek to determine the nature of self-regulatory instructional elements designed to promote learning across academic domains, to examine the effectiveness of these on motivation. The following questions guided out review:

RQ1.What instructional components are present in the academic self-regulation intervention literature focused on motivation?

RQ2. To what extent do academic interventions with self-regulatory instructional elements improve motivation in struggling learners?

Methods

Search Procedure

As shown in the PRISMA diagram (see Figure 1), authors performed a systematic search of the literature including an online database search, hand search, and ancestral search (Moher et al., 2009). First, we conducted an online search using Education Research Information Center (ERIC), Education Source, and EBSCOHOST to locate studies in peer reviewed journals or dissertations published in English without date restrictions.

We identified terms related to motivation, self-regulation, and disability status. In identifying search terms related to motivation we recognize motivation as a multidimensional construct that draws from many research traditions. We chose to define terms using important constructs such as those described in Conradi et al. (2014) hierarchy as well as certain widely adopted theories such as self-efficacy theory, self-determination theory, attributions theory, and expectancy-value theory. We define the following search terms for motivation: *motivation, expectancy, value, self-efficacy, attribution, goal-orientation, perceived control, interest, intrinsic value, utility value, attainment value, cost, anxiety, attitude.*

We adopted a social cognitive view of self-regulation and attempt to identify important self-regulatory processes to include as search terms. We looked to research on self-regulation (e.g., Zimmerman, 2012) to define terminology. Search terms related to self-regulation are as follows: *self-regulation, self-observation, self-judgment, self-reaction, forethought, goal setting, social modeling, visualization, self-monitoring, help-seeking, self-talk, self-verbalization, rehearsal, self-instruction, environmental structuring, self-selection, cognitive modeling, self-reflection, self-reinforcement, self-evaluation, strategy review, self-rewarding, and SRSD.*

For our final search parameter, disability status, we chose search terms related to common disability identifiers as well as common types of disabilities. Search terms for disability status are as follows: *disability, learning disability, special needs, remedial, struggling,*

emotional and behavioral, dysgraphia, dyslexia, dyscalculia, written language disorder, specific reading comprehension deficit.

We used Boolean search procedures to include all search terms. A total of 2,777 studies were identified. We also conducted a 10-year hand search of important journals (e.g., 2008-2018) in the field of special education and intervention research identifying four additional articles. Of the 2781 total articles 169 were removed as duplicates. The footnotes of the PRISMA diagram include the online search terms and the journals used for the hand search. For our initial search we included only one limiter, that studies are either peer reviewed or dissertations, as the objective of the initial search was full inclusion.

Inclusion and Exclusion Criteria

We screened titles and abstracts of 2612 articles according to the following criteria. First, studies incorporated interventions for teaching reading, writing, or mathematics. Second, we determined whether studies used experimental, quasi experimental, or single case designs (SCD). Third, studies included participants from primary or secondary grades (grades 1-12). Fourth, we included studies where struggling learners were participants. We included studies whether researchers reported identification using school records (e.g., Individualized Education Plan), documentation of pervasive or significant academic deficits through psychometric tests, or through pretesting procedures. Fifth, we included studies with some form of SRI. Finally, we included studies with at least one motivation outcome measure. Studies that did not meet one or more criteria were removed from our sample. Of the initial 2612 articles, 2563 were excluded. The resulting 49 full text articles were retrieved and reviewed. Of the 49 full text articles reviewed 22 were excluded for the following reasons: (1) studies referenced self-regulation but did not include a specific component in which students were instructed to self-regulate learning

(n=7; e.g., Wong et al., 1997), (2) motivation was discussed but not directly measured (n=7; e.g., Adkins et al., 2012), (3) further investigation revealed that studies were not experimental (n=6; e.g., Horn, 2007), or (4) a substantial portion of the sample was students not identified as struggling (n=2; e.g., Lyon-Wagner, 2010). References lists from the 27 studies were checked and two more studies were identified making our total sample 29 articles.

Coding Procedures

For each investigation we capture the following information in the code sheet: general study information, participant information, study setting, independent variables, dependent variables, study design, procedures, intervention duration (dosage), and motivation. First, information about the publication was recorded including journal outlet, year of publication, and authors. Narrative information was recorded for the sample such as sample size, grade level, gender, ethnicity, and social economic status. We recorded information on treatment and control conditions including setting, intervention agents, academic domain, length of instruction (dosage), how fidelity was measured, and academic outcomes. We recorded the following outcomes related to motivation, noting type of construct, how the outcome was measured, procedures for determining validity and reliability, and if it was researcher created or an existing validated measure.

Data were also recorded related to the rigor of the study based on eight quality indicators for experimental, quasi-experimental, and single case design research in special education identified by the Council for Exceptional Children (CEC, 2014). Included was information about (1) context and participants (i.e., whether sufficient information was provided about participant demographics, setting, and intervention agents, and whether sufficient information was provide to describe student identification as struggling), (2) Description of practice (i.e., whether

intervention procedures were properly documented and whether fidelity was addressed in a valid manner), (3) internal validity (i.e., threats to internal validity were properly controlled for by providing baseline equivalencies, demonstrating control of independent variables, and accounting for attrition), and (4) outcomes and data analysis (i.e., outcomes were recorded using valid measures and data analysis helped to address any potential design concerns).

In considering quality we paid particular attention to treatment fidelity and threats to internal validity. We graded each quality indicator on a scale from zero to two for not present (low quality), partially present (medium quality), and completely present (high quality). In addition, studies were automatically labeled as low quality if the study showed improper control of internal threats (e.g., large differences between baseline conditions that were not accounted for in data analysis, non-equivalent control groups, etc.). We identified five studies that were low, 12 studies that were of medium, and 12 studies that were of high quality. Three studies did not report fidelity along with other significant issues and two more had non-equivalent control groups that were not addressed in the statistical analysis, disqualifying these studies from quantitative analysis.

We also coded all articles for instructional components related to self-regulation. We realized that terminology describing SRI was inconsistent and varied considerably between studies prompting us to operationally define self-regulation for our study. To address this, we developed a coding scheme based on (1) the behavior being taught and (2) the expressed purpose of that behavior. These dimensions helped to categorize self-regulated behaviors regardless of the terminology used in the study. We also defined exactly what qualified as SRI. SRI was when the teacher engaged in some form of explicit teaching of self-regulated practices or how to use structural supports to regulate learning. This most often was in the form of teacher modeling

skills but also could be explicit explanations of skills or structural supports that regulate learning. For example, if a teacher explicitly taught the use of a self-regulated skill such as self-monitoring through modeling of a procedural checklist this would qualify. However, just presenting students with a graphic organizer, without explicit instruction on how to use the tool to regulate their learning would not qualify.

The first author coded all 29 studies and two secondary coders double coded 100% of studies articles. The coders were trained by the first author on five randomly selected articles from the available pool, and reached 92% agreement during training. The coders independently coded the remaining 22 studies. Interrater reliability was assessed for study quality as well as for the presence of self-regulated instructional components.

The secondary coders first independently scored all articles using the quality scale, rating each as high quality, medium quality, or low quality depending on the final score tabulation and paying particular attention to fidelity of implementation and control of threats to internal validity. Agreement on article ratings of high quality, medium quality, and low quality was established to be 100%. Any individual components that were different were discussed and resolved. The secondary coders then identified self-regulated components for each article. The secondary coders compared components with the first author (agreement was over 90%). Any components that differed were discussed and resolved.

Calculating Effect Size

The first author and a doctoral student unfamiliar with the study purpose independently calculated Hedges' g for 12 studies (Antononiou & Souvignier, 2007; Berkeley et al., 2011; Cuenca-Sanchez et al., 2012; Garcia & De Caso, 2006; Garcia-Sanchez & Fidalgo-Redondo, 2006; Graham et al., 2005; Harris et al., 2006; Miranda et al., 1997; Nelson & Manset-

Williamson, 2006; Page-Voth & Graham, 1999; Rooney, 1997; Schunk & Cox, 1986) to investigate the effectiveness of SRI on motivation. Seventeen studies were excluded from ES calculation because they were determined to be of low quality, not experimental or quasi-experimental, lacked a BAU control condition, had inequivalent baselines, or did not assess motivation in the control group. Hedges' g was calculated using the following formula: $Hedges'g = \frac{M_1 - M_2}{SD_{pooled}}$ (M1=treatment group posttest mean; M2=control group posttest mean; SDpool=pool standard deviation of both treatment and control group). ES was calculated in this way for any outcome measure on all treatment conditions. Corrections were made for sample sizes smaller than n=50 using $Hedges'g = \frac{M_1 - M_2}{SD_{pooled}} \times (\frac{N-3}{N-2.25}) \times \sqrt{\frac{N-2}{N}}$ (N=number of participants). We used (https://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html) to eliminate hand calculation errors. There was 100% agreement in ES calculation between the first author and the second coder. ES for maintenance was calculated when possible.

Results

To begin, 1283 students participated in the studies that were analyzed (See Table 1). This included randomized designs (N=19; 66%), SCD (N=9; 31%) or quasi-experimental designs (N=1; 3%) without randomization. Of this sample 47% of students were identified as LD, 20% were academically comparable to students with LD, with only 4% who were typically achieving students in comparison conditions. Garcia and his colleagues conducted three studies in Spain, accounting for 29% of the struggling learners in our sample (Garcia & de Caso, 2006; Garcia-Sanchez & de Caso-Fuertes, 2005; Garcia-Sanchez & Fidelgo-Redondo, 2006). Twenty-five studies (86%) used student LD status, provided by the school district, as the primary indicator of struggling. Authors of seventeen studies (58%) corroborated LD status with student IQ or other psychometric indicators, assessing pretest ability by using norm referenced assessments such as

the Test of Written Language (TOWL), or interviewing teachers. Four studies (Nelson, & Mansett-Williamson, 2006; Bandura & Schunk, 1981; Schunk, 1983; Schunk & Rice, 1993) administered pretest assessments and interviewed teachers as researchers were not granted information on LD status.

The studies in our sample included participants from first to tenth grade. Nineteen studies included students between 5th and 8th grade. Three studies focused on early elementary grades between 1st and 3rd. Two more studies focused exclusively on high school students between 9th and 10th grade. Of the remaining five studies one included students from 3rd to 5th grade, one included students from 4th to 5th grade, one included students from 7th to 9th grade, and the last two studies include larger grade ranges between 4th and 8th grade and 5th and 10th grade. Eighteen studies provided demographic information, of these, 37% of students were White, 48% African American, 9% Latinx, 5% Asian or Pacific Islander, and 1% other. Five studies provided information on socio-economic status, and 11 studies provided information on English as a second language status.

Most interventions (N=17) focused on writing, with substantially fewer focusing on mathematics (N=7) or reading (N=5). In eight studies classroom teachers or other specified school staff implemented the intervention. In thirteen studies graduate or undergraduate students conducted the intervention. In the remaining three studies the authors administered the intervention. Three studies employed whole group instruction, 25 studies employed small group instruction, and one study employed one-on-one instruction. The intervention lengths varied from being brief, one to five-hour interventions (28%), six to ten hour interventions (38%), or of moderate lengths of 10 to 20 hours (17%), or in rare cases, more than 20 hours (10%). Fifteen studies assessed fidelity of implementation.

RQ1.What instructional components in the academic self-regulation intervention literature focus on motivation?

We identified three distinct self-regulatory components that were part of the academic interventions: goal setting, self-monitoring, and self-talk. For the latter two we identified purposes related to instructional and emotional regulation and identified progress-monitoring. This resulted in six different component types: goal setting, self-monitoring for task completion, self-monitoring for emotional regulation, self-talk for task completion, self-talk for emotional regulation, and self-talk for progress monitoring. Below is a brief description that is organized thematically, with an example of each component. See Table 2 for relative frequency in self-regulated instructional components.

Goal Setting. Half of the treatment conditions (51%) included goal setting elements. Students were given explicit instructions on how to set goals for instructional purposes. For example, Page-Voth and Graham (1999) prompted students to write goals on how many parts of a strong essay they would include, after learning what those components were. Schunk (1985) also used goal setting to support as students completed subtraction problems. Of note, Schunk found that goals set by children, rather than those provided by instructors resulted in more positive mathematics outcomes.

Self-monitoring for Task Completion. In more than half of the treatment conditions (54%) students were explicitly taught to monitor completion of instructional components, strategies, or procedures through modeling, the use of physical scaffolds, or teacher guidance. To illustrate, Antoniou and Souvignier (2007) asked students to monitor the use of specific summarizing strategies. Teachers did this by introducing the most important components of a narrative (i.e., the characters, the characters' goal, the problem and the solution) then showed

students how to find these components. Students then used a checklist to track if they had looked for and found critical components of a summary.

Self-talk for Task Completion. Just over half of the treatment conditions (56%) included prompts for students to engage in self-talk for task completion. Studies included modeling, physical scaffolds, or teacher guidance to use self-verbalizations to help students to vocalize or subvocalize when executing strategies, or procedures related to problem solving. For example, Schunk (1993) taught students to self-describe steps to completing subtraction problems to themselves and found that even in conditions where overt verbalization was faded students continued to utilize self-talk. More recently Ennis and Jolivette (2014) prompted students to create self-statements that helped them to complete the correct sequence of procedures to produce a strong essay.

Self-talk for Progress Monitoring. In one third of treatment conditions self-talk focused on evaluation of the steps taken to complete tasks or strategies, or to rate the quality of one's work. Modeling, procedural supports, and teacher guidance were used to teach students this skill. Harris et al. (2006) for example prompted students to asked themselves-questions to evaluate progress and react to procedural steps toward completion in two genres: story writing and persuasive essays.

Self-talk for Emotional Regulation. Researchers taught students to monitor their emotional state about one third of the time (35%), again through modeling, procedural supports, and teacher guidance. Berkeley at al. (2011) for example, used self-talk for emotions to help students build more positive attributions for reading. Students were prompted to use positive self-talk and self-reinforcement to encourage more positive internalized attributions of success.

Self-monitoring for Emotional Regulation. In a few treatment conditions (5%), teachers instructed students how to monitor emotions such as motivation and anxiety level. Cuenca-Carlino et al. (2016) taught students how to monitor "self-determination behaviors," which included level of interest as well as perceived needs, strengths, and challenges during multi-step mathematics problem. Students learned to self-monitor through modeling, physical scaffolds, or teacher guidance.

RQ2: To what extent do academic interventions with self-regulated instructional elements improve motivation in struggling learners?

About 40% of the studies (N=12) could be analyzed quantitatively (Antononiou & Souvignier, 2007; Berkeley et al., 2011; Cuenca-Sanchez et al., 2012; Garcia & De Caso, 2006; Garcia-Sanchez & Fidalgo-Redondo, 2006; Graham et al., 2005; Harris et al., 2006; Miranda et al., 1997; Nelson & Manset-Williamson, 2006; Page-Voth & Graham, 1999; Rooney, 1997; Schunk & Cox, 1986). A total of 784 participants were included across the 12 studies. Of the total sample 764 were identified as either with or at risk for LD.

Measures. Authors used measures of self-efficacy, expectancy for success, self-esteem, value, intrinsic motivation, attributions for success and failure, effort, and affect; however, measures were not defined consistently. Representative definitions are as follows. Self-efficacy is an individual's belief in their ability to execute necessary skills to solve some problem or accomplish a task. Similarly, expectancies are an individual's beliefs about the expected outcomes of an upcoming activity (Bandura, 1994). Self-esteem describes is a comprehensive assessment of competence or self-worth (Eccles & Wigfield, 2002). Value is a measure of the importance students place on a specific task and can be understood through its four component parts: attainment value, utility value, intrinsic value, and cost (Troia et al., 2012). Intrinsic

motivation is the desire to engage in an activity for internalized reasons such as personal enjoyment or interest (Ryan & Deci, 2000). Attributions are causal explanations for achievement outcomes that in turn influence future achievement strivings. Student effort can be understood as the exertion a student is willing to expend to engage in a specific task. Finally, affect is the emotional response that students feel before, during, and after engaging in an activity (Eccles & Wigfield, 2002).

Seven studies used a single construct to measure motivation (Antononiou & Souvignier, 2007; Cuenca-Sanchez et al., 2012; Garcia-Sanchez & Fidalgo-Redondo, 2006; Graham et al., 2005; Miranda et al., 1997; Page-Voth & Graham, 1999; Schunk & Cox, 1986). The other five studies used multiple constructs to measure motivation, with Garcia and de Caso using four different scales to measure task value, expectancies, affect, self-esteem, and attributions.

In all, twenty different measures were used to observe constructs related to motivation; in addition, all were created by researchers. All scales except for the forced choice scale used in Maranda et al. (1996) used Likert scales ranging from 3 to 11 options. Scales ranged from 1 question to 32 questions. All but three studies (Harris et al., 2006; Rooney, 1997; Schunk & Cox, 1986) provided internal consistency estimates ranging from 0.53 to 0.97. Three studies reported test-retest reliability (Miranda et al., 1997; Nelson & Manset-Williamson, 2006; Schunk & Cox, 1986) including Schunk (1986) which had not provided an internal consistency estimate.

Effect Sizes. Across the twelve studies, there were 20 treatment conditions that used at least one self-regulated instructional component. Thirty-four total ES were calculated ranging from -0.46 to 1.14. See Table 3 for ES for all treatment conditions across all outcome measures. Eighty percent of ES on motivation measures at posttest were positive (i.e., students in treatment conditions showed improved motivation relative to students in control conditions), indicating

that self-regulated instructional components made a positive impact on students' motivation, irrespective of the impact on student learning.

Single v. Multi-component Interventions. Just over half of the studies (N=8) used a combination of self-regulatory elements in the treatment conditions. In all, 83% of the treatment conditions were multicomponent, with two to five elements to the interventions. Overall, the combination of self-regulatory components had a positive effect (in all but one multicomponent intervention study) as students had higher outcomes related to self-efficacy, intrinsic motivation, effort, attributions, and affect. In addition, writing intervention studies generally included multiple self-regulatory elements, in part due to the fact that many studies employed the SRSD instructional approach, which embeds goal setting and options for self-talk and self-monitoring. In contrast, six studies that relied on a single component. Self-talk for progress monitoring was relied on most frequently in these studies, followed by self-monitoring for task completion and finally, goal setting.

Effect Sizes by Academic Domains. Differences in average ES were evident when comparing academic domains. The five reading studies reported outcomes on self-efficacy, attributions for success and failure, and affect (Antononiou & Souvignier, 2007; Berkeley et al., 2011; Miranda et al., 1997; Nelson & Manset-Williamson, 2006; Rooney, 1997). In these studies, treatment conditions outperformed control conditions roughly half of the time (53%). Most of these studies were intended to help students to improve their reading comprehension, with instructional elements that targeted a variety of underlying reading skills, in addition to elements that aimed to improve motivation.

By comparison, ES revealed improved motivation in 86% of the writing intervention studies (Cuenca-Sanchez et al., 2012; Garcia & De Caso, 2006; Garcia-Sanchez & Fidalgo-

Redondo, 2006; Graham et al., 2005; Harris et al., 2006; Page-Voth & Graham, 1999) with students in treatment groups appearing more motivated at posttest than those in the control group, with small to large positive ES. Negative outcomes came from a single study (Page-Voth & Graham, 1999). In addition, one study reported on outcomes of intrinsic motivation and effort (Harris et al. 2006), and another (Garcia & De Caso, 2006) focused on multiple measures (self-efficacy, value, self-esteem, affect, causal attributions, and expectancy). The remaining studies focused solely on self-efficacy.

Only one study involving mathematics met the criteria necessary for the quantitative analysis (Schunk & Cox, 1986). This study measured self-efficacy across two conditions of self-talk for task completion, with ES of 0.02 and 0.91. In one condition, sixth to eighth grade students with LD verbalized each step in executing an algorithm for solving subtraction problems that involved simple to complex place value regrouping. The second condition initially involved the same procedure, but students were prompted to internalize self-talk after they had completed half of the subtraction problems.

Maintenance. Three reading (Antononiou & Souvignier, 2007; Berkeley et al., 2011; Miranda et al., 1997) and one writing (Garcia & De Caso, 2006) assessed motivational outcomes between one and a half and three months after posttest. Scores were reported on self-efficacy and attributions, all of which were positive; ES ranged from small to large values. Both Antononiou and Souvignier (2007) and Berkeley et al. (2011) reported positive ES in maintenance that had been negative at posttest.

Discussion

The purpose of this systematic literature review was to describe instructional components in academic self-regulation intervention studies that were designed to improve student learning

outcomes and designed to improve motivation for students with or who are at risk for LD. We also sought to determine the relative effectiveness of these elements for both populations of struggling learners.

Twenty-nine studies were initially identified for our first aim, and a smaller corpus of 12 studies were identified for ES calculation to explore questions of impact. Most interventions (58%) focused on writing, with fewer focusing on mathematics (24%) or reading (17%). Overall, 47% of the studies included students identified as LD, with 49% more who were academically comparable to students with LD, meaning in turn that the majority of participants in our sample were struggling learners. Students in fifth or sixth grade participated in about two-thirds of the studies, although researchers who conducted these studies often included younger or older students as well (i.e., students in upper grades in middle school were also included). In contrast, very few focused exclusively on lower elementary or high school-aged participants. Fidelity implementation data was provided in half of the studies.

This study is an extension of work we began in 2018. We identified 20 additional studies across reading, writing, and mathematics. Students in these studies learned between one and five different self-regulatory skills (goal setting, self-monitoring for task completion, self-talk for completion, self-monitoring for emotional regulation, and self-talk for emotional regulation), however, a subset (goal setting, self-monitoring for task completion, and self-talk for completion) were more commonly embedded in the interventions. Self-talk for task completion was the most common component across the current sample of 29 studies, occurring in 87% of the interventions focused on writing, 69% of those designed to improve reading, and in 30% of the math interventions. Self-monitoring, without use of speech to guide task completion occurred in 87% of the writing interventions, 60% of the mathematics interventions, and in 54% of the

reading intervention studies. Finally, goal setting occurred in 83% of the writing interventions, in 70% of the math interventions, and in 23% of the reading intervention studies.

We also determined the relative effectiveness of instructional components for promoting motivation in the 12 quantitative academic interventions. The most commonly employed components in this subset of studies were goal setting, self-monitoring or self-talk for task completion (each occurred in more than half of the studies), with self-talk for emotional regulation, and/or for progress-monitoring occurring about one third of the time. Self-monitoring for emotional regulation rarely occurred. A major finding from this study is that the use of these elements were overwhelmingly positive, as 80% of the ES were positive. Because nearly all of the participants had or were academically comparable to those with LD, it appears that academic interventions with self-regulatory components resulted in improved motivation, especially when measures of self-efficacy and causal attributions were used.

Our results suggest that inclusion of self-regulated components led to improved academic functioning, which in turn led to improved self-efficacy and more positive attitudes. Support for this contention comes from Eccles and Wigfield (2002) who point out that students who perform better and are more self-regulated are more likely to attribute success to internal causes leading to a greater sense of pride and self-esteem. In addition, self-regulated learners are more likely to rely on internalized skill rather than external agents such as teachers (Alexander, 1997). If learners view themselves as active contributors to the learning process, they may have more positive attitudes if they succeed.

Benefits regarding the use of SRI for improving learning outcomes for struggling learners in reading, writing, and mathematics have been documented in prior reviews (e.g., Berkeley et al., 2010; Graham et al., 2012, Miller et al., 1998), and although not the primary focus of this

review, studies in our review reported similar positive learning outcomes. Moreover, self-efficacy has been linked to improved academic skills with typically achieving learners in reading (Toste et al., 2020), writing (De La Paz & Butler, 2018; Graham, 2017; Troia, 2012), and mathematics (Higgins, 2019; Pajares & Graham, 1999). This study's findings extend the link between self-regulatory instructional components and improved motivation outcomes to struggling learners.

ES differences related to academic domain, for example, the stronger improvement in motivation for writing when compared to reading, are likely explained both by the differences in complexity of the underlying tasks, and by the associated use of multiple self-regulatory elements to teach more cognitively challenging tasks. In other words, writing is inherently more complex than reading (Graham et al., 2017). To illustrate, reading previously generated text is required for the processes of revision – students must be able to identify discrepancies between what they intended with content that they actually wrote (De La Paz & Sherman, 2013). Unfortunately, claims about the use of self-regulatory components to improve motivation for mathematics are not possible due to the fact that we were only able to locate one intervention study in mathematics for generating ES.

In response to the need for students to execute and coordinate multiple complex processes in writing (e.g., planning, text generation, and revision), approaches to writing instruction routinely embed multiple forms of self-regulation. Self-regulated strategy development (SRSD), a prominent and effective writing intervention for struggling learners, was developed to operationalize and demonstrate how strong writers execute the writing process (Harris & Graham, 2009). SRSD incorporates multiple self-regulatory elements into a single intervention which may be another underlying reason for its documented success Therefore,

there may be additive effects of different self-regulated instructional components on student motivation outcomes. Improving motivation is a multifaceted process that involves, among other avenues, improving self-efficacy and changing student attitudes (Graham et al., 2018). Multicomponent self-regulated interventions may be able to target multiple avenues for improving motivation.

Finally, it is interesting to note that struggling learners showed improved motivation on assessments at maintenance as compared to at posttest in several studies, and makes sense given that similar improvements in the quality of student writing at maintenance have been observed in writing intervention studies (e.g., De La Paz & Graham, 2002; Graham et al., 2005). With respect to self-efficacy, improved motivation may occur because as students become stronger self-regulated learners, they are more likely to persist on difficult tasks and succeed. This may lead through mastery experiences of success to greater beliefs of self-efficacy over time. Conversely, this explanation should be explored further as improved motivation may be gradual as struggling learners often have to contend with well-established negative self-efficacy beliefs that have been developed across years of academic turmoil (Troia et al., 2012).

Limitations and Future Research

This review has several limitations and suggestions for future research. First, outcome measures used to report motivation were inconsistent. Motivation was measured using multiple different constructs making comparison across studies difficult. Second, publication bias is a concern. Despite including dissertations, nearly all studies were peer reviewed articles. There is a bias toward positive results in peer reviewed publications. Finally, we could only determine effectiveness of the instructional components in 12 of the 29 studies (about 40%). To establish a more cohesive understanding on how self-regulatory instructional components may be used to

improve student motivation, we echo calls that researchers provide theoretical explanations about how the active ingredients in their interventions are expected to improve motivation. In addition, while perhaps difficult to achieve, it would be helpful for authors to provide explicit rationale regarding the use of specific motivation outcome measures, and to use constructs in ways that are consistent with prior use. Finally, we suggest that students are asked to complete motivation measures at different times during a study, such as after learning a writing strategy, and again after receiving an evaluation, to learn more about how various instructional elements work to promote motivation

Implications for Practice

Motivation is an extremely complicated set of constructs, and much of how it relates to student learning outcomes has yet to be established (Troia et al., 2012). Expectancy–value theory (Wigfield & Eccles, 2000), in which motivation is believed to be determined by what a learner believes will happen when engaged in a given task as well as how much worth they assign to the task can be used to understand motivation in students with LD and other struggling learners who commonly express low expectancy and low value in academic tasks. Fortunately, the use of self-regulatory instructional components can be used to improve students' motivation as well as their academic learning outcomes. Perhaps not surprising, it seems beneficial to embed more self-regulatory components for more complex tasks so that struggling learners can manage multiple or perhaps reciprocal cognitive processing demands (i.e., the use of reading during writing or in multi-step mathematical problem solving tasks). Goal setting and self-talk for task completion or for progress monitoring seem especially likely to improve the likelihood that struggling learners will demonstrate positive gains in motivation.

Tables

Table 1: Study Design and Domain

Study	Design	Domain	Study Quality	Sample Size	Percentage Struggling Learners	Grade	Dosage (hrs.)	Dependent Measure
Antononiou and Souvignier (2007)	RCT	Reading	Medium	73	100%	5th-8th	29	SE
Bandura and Schunk (1981)	RCT	Mathematics	Low	41	100%	1st-3rd	~3.5	SE
Berkeley, Mastropieri, and Scruggs (2011)	RCT	Reading	High	59	100%	7th-9th	6	CA
Buzza and Dol (2015)	QED	Mathematics	Low	16	100%	10th	NR	SE, I
Cerar (2012)	SCD	Writing	High	7	100%	6th-8th	22	SE
Cuenca-Carlino, Freeman-Green, Stephenson, & Hauth (2016)	SCD	Mathematics	High	6	100%	8th	10	SE
Cuenca-Carlino and Mustian (2013)	SCD	Writing	High	9	100%	6th-8th	~15	SE
Cuenca-Sanchez, Mastropieri, Scruggs, and Kidd (2012)	RCT	Writing	High	21	100%	6th	16.5	SE
Daniel (2003)	SCD	Mathematics	Low	18	100%	6th-8th	8	SE, CA
Ennis and Jolivette (2014)	SCD	Writing	High	6	100%	9th	~7	SE, IM
Garcia and De Caso (2006)	RCT	Writing	Medium	60	100%	5th-6th	~8	SE
García-Sánchez and Caso-Fuertes (2005)	RCT	Writing	Medium	191	100%	5th-6th	~23	SE
García-Sánchez and Fidalgo-Redondo (2006)	RCT	Writing	Medium	121	100%	5th-6th	~20	SE
Graham and Harris (1989)a	RCT	Writing	Medium	33	66%	5th-6th	6	SE
Graham and Harris (1989)b	SCD	Writing	Medium	3	100%	6th	4.5	SE
Graham, Harris, and Mason (2005)	RCT	Writing	High	73	100%	3rd	~5.8	SE

Harris, Graham, and Mason (2006)	RCT	Writing	High	66	100%	2nd	6.3	IM
Macarthur and Philippakos (2010)	SCD	Writing	Medium	6	100%	6th-8th	~7.5	SE
Mills (2012)	SCD	Writing	High	10	100%	8th	~7	SE
Miranda, Villaescusa, and Vidal-Abarca (1997)	RCT	Reading	Medium	80	75%	5th-6th	~17	CA
Nelson and Manset-Williamson (2006)	RCT	Reading	High	20	100%	4th-8th	20	SE, CA, A
Page-Voth and Graham (1999)	RCT	Writing	High	30	100%	7th-8th	NR	SE
Rooney (1997)	RCT	Reading	Medium	91	100%	3rd-5th	~2	SE
Sawyer, Graham, and Harris (1992)	RCT	Writing	High	33	66%	5th-6th	~5	SE
Schunk (1983)	RCT	Mathematics	Low	30	100%	4th-5th	5	SE
Schunk (1985)	RCT	Mathematics	Low	40	100%	6th	3.5	SE
Schunk and Cox (1986)	RCT	Mathematics	Medium	90	100%	5th-10th	3.5	CA
Schunk and Rice (1993)	RCT	Reading	Medium	44	100%	5th	5	SE
Sexton, Harris, and Graham (1998)	SCD	Writing	Medium	6	100%	5th-6th	~7	CA, E

*RCT: Randomized Control Trial, SCD: Single Case Design, QED: Quasi Experimental Design, SE: Self-Efficacy, CA: Causal Attributions, E: Effort, A: Affect, IM: Intrinsic Motivation, I: Interest NR: not reported

Table 2: Frequency of Occurrence of Self-Regulated Instructional Components

	Sample Size (Total Treatment Conditions)	Goal Setting	Self-Monitoring Task Completion	Self- Talk for Task Completion	Self-Talk for Progress Monitoring	Self-Monitoring for Emotional Regulation	Self-Talk for Emotional Regulation
All Studies	46	51%	54%	56%	33%	5%	35%
Reading	13	23%	54%	69%	15%	8%	8%
Writing	23	78%	78%	87%	70%	9%	83%
Math	10	71%	71%	14%	14%	0%	0%
Single Component	11	27%	18%	45%	0%	0%	0%
Multi Component	35	71%	77%	77%	54%	9%	83%

Note: Percentages were calculated by dividing number of treatment conditions in which a specific self-regulated instructional component is present by total number of treatment conditions.

Table 3: Effect Size by Treatment Condition

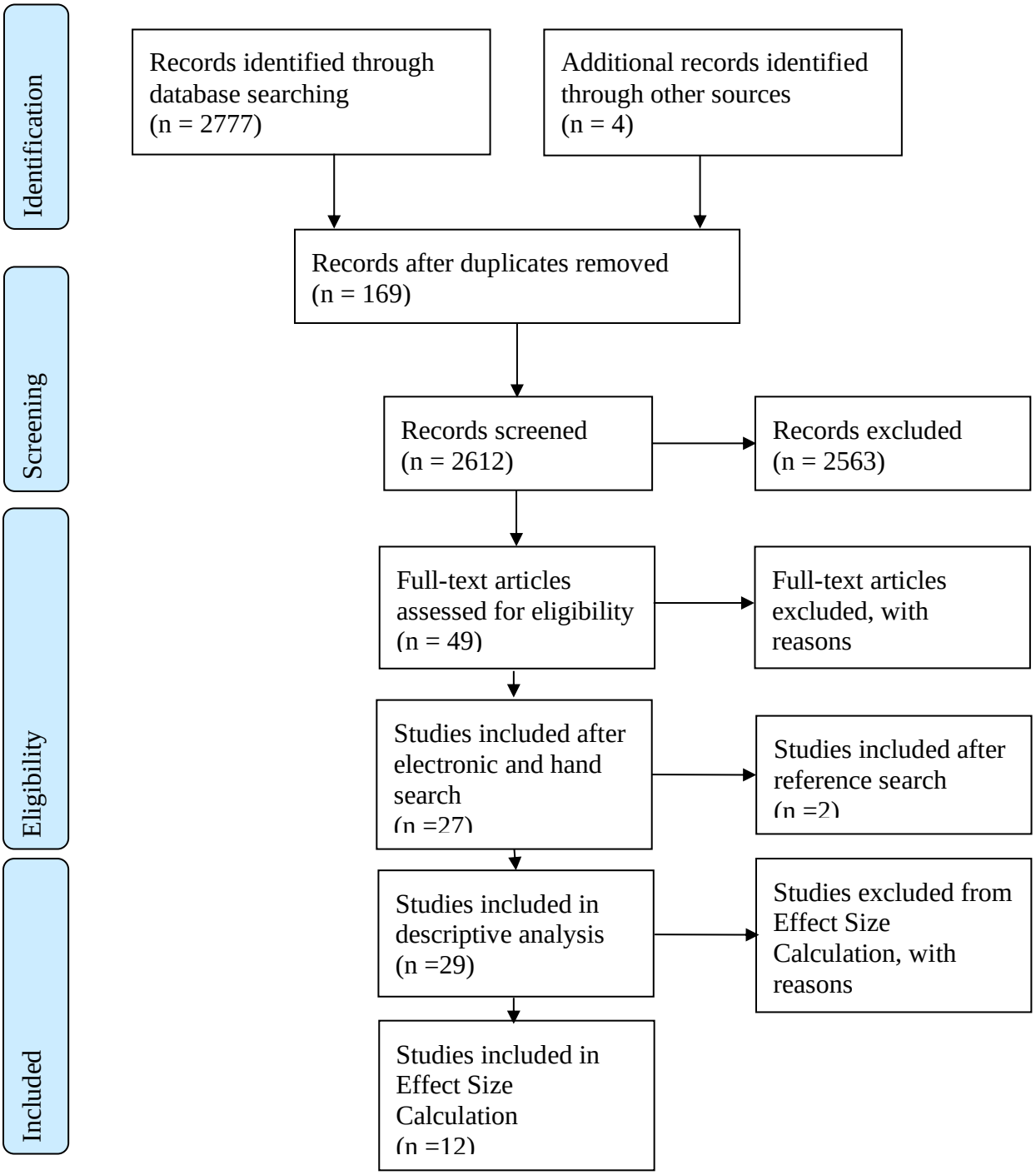
Study	Domain	Dependent Measure	IV	ES	Maintenance ES
Antononiou and Souvignier (2007)	Reading	SE	Treatment	-0.37	0.38
Berkeley, et al. (2011)	Reading	CA (success)	RCS	-0.03	0.09
			AT	0.83	0.39
		CA (failure)	RCS	-0.04	0.04
			AT	0.27	0.39
Cuenca-Sanchez, et al. (2012)	Writing	SE	Treatment	1.03	
*García-Sánchez and De Caso (2005)	Writing	SE	Treatment	0.51	0.31
García-Sánchez and Fidalgo-Redondo (2006)	Writing	SE	SRSD	0.76	
			SCM	1	
Graham, Harris, and Mason (2005)	Writing	SE (planning)	SRSD	0.2	
			SRSD+	0.38	
		SE (generating)	SRSD	0.25	
			SRSD+	0.1	
Harris, Graham, and Mason (2006)	Writing	IM	SRSD	1.14	
			SRSD+	0.35	
		E	SRSD	0.72	
			SRSD+	0.25	
Miranda, Villaescusa, and Vidal-Abarca (1997)	Reading	CA (for effort)	SI	0.63	1.17
			AT	0.86	1.2
Nelson and Manset-Williamson (2006)	Reading	SE	Treatment (EC)	-0.46	
		CA (success)	Treatment	0.58	
		CA (failure)	Treatment	1.3	
		A (positive)	Treatment	-0.12	
		A (negative)	Treatment	-0.45	
Page-Voth and Graham (1999)	Writing	SE	Goals	-0.3	
			Goals+Strategy	-0.07	
Rooney (1997)	Reading	SE	Self-Regulation	0.7	
			MF	0.26	
			SR+MF	0.36	

		CA	Self-Regulation	0.44
			MF	0.25
			SR+MF	0.27
Schunk and Cox (1986)	Math	SE	Continuous	0.02
			Discontinuous	0.91

Note: *Garcia-Sanchez and De Caso (2005) reported multiple other outcomes but did not provide information to calculate ES. Outcomes: SE=self-efficacy, CA=casual attributions, A=affect, E=Effort, IV: RCS= reading comprehension strategy, AT=attribution retraining, SRSD=self-regulated strategy development, SCM=social cognitive model, SRSD+=self-regulated strategy development plus peer support, SI=strategy instruction, SR= self-regulation, MF=model feedback, treatment=authors offered no specific name for the intervention

Figures

Figure 1: Search Protocol



Hand search was conducted using the following journals: Journal of Special Education, Learning Disabilities: A Contemporary Journal of Special Education, Exceptional Children, Behavior Disorders, Learning Disabilities Quarterly, Journal of Learning Disabilities, Contemporary Education Psychology

Chapter 3: The Effects of a Cognitive Apprenticeship on Transfer of Argumentative Writing in Middle School Science

The Next Generation Science Standards (NGSS) emphasize student engagement in scientific practices (National Research Council, 2013), and calls for an emphasis on argumentation in science classrooms have long existed (Berland and McNeill 2010; Driver et al. 2000; Duschl and Osborne 2002; Osborne et al. 2004). Moreover, practices of construction and critique are seen as central to scientific practices of explanation and argumentation (Ford, 2008). Many students struggle however to differentiate raw data vs. data transformed into evidence (Kuhn, 2010). A major challenge for students is justifying how evidence supports a claim, that is, the reasoning that integrates scientific principles in the match between the evidence and the claim (McNeill & Krajcik, 2012). Students also often have difficulty constructing and critiquing arguments with evidence, often confusing requests for evidence with requests for elaboration of theory, and sometimes only seeing evidence that fits with their a priori theories (Kuhn, 2010). Additionally, engaging students in writing integrated with scientific practices presents unique challenges (e.g., discerning the goals and purposes of writing tasks, learning new organizational structures and learning specialized linguistic features such as nominalizations and technicality; Levin et al., 2021).

There are advantages in using writing in science classrooms, nonetheless. Creating opportunities for students to construct reasoned arguments and causal explanations and support them with evidence in writing provides a context in which *all* students can participate, in contrast to oral discourse, where a few students may dominate the conversation (e.g., De La Paz et al. 2012). Attention to writing is important as many middle school students struggle with basic literacy skills (Graham et al., 2015), and because writing in academic subject areas can support

students' content learning (Bangert-Drowns et al. 2004; Carter et al. 2007). Fortunately, effective instructional approaches have been used successfully to help learners use evidence to substantiate scientific claims in writing. One such approach is the cognitive apprenticeship (described in detail below), which provides supports for learning conceptual knowledge and disciplinary ways of thinking when writing about science as well as procedures for regulating the writing process.

Effective Models of Argumentation and Writing Instruction

When supporting argumentative writing, students face unique challenges. Sampson et al. (2013a, 2013b) explored how to facilitate middle school students' argumentative writing by helping teachers provide instructional scaffolds as students participated in a series of science experiments. These supports included the use of written examples that served as models, teacher feedback about work quality, and opportunities to revise their writing, all of which have been found to be effective practices for teaching writing (e.g., Graham et al., 2015). A recent synthesis by Lee and De La Paz (2021b) confirmed the effectiveness of these practices finding that both cognitive and linguistic supports are beneficial when students are asked to write in science. This line of research informs our current approach, which adapts activities and materials from Argument Driven Inquiry (ADI; Grooms et al., 2016; Sampson et al., 2013a, 2013b) by introducing a cognitive apprenticeship that provides carefully sequenced instructional routines for teachers and students and cognitive tools that support student learning.

Cognitive Apprenticeship Approach to Instruction

An effective model for the instruction of argumentative writing is the cognitive apprenticeship, which makes expert thinking and literacy practices visible to novices (De La Paz et al., 2017; Levin et al., 2021). A cognitive apprenticeship is both a model for instruction that

helps teachers to organize lessons and an approach that supports students in understanding the processes involved in complex learning activities (Brown et al., 1989; De La Paz et al., 2014). As Brown and colleagues (1989) explain, by using the cognitive apprenticeship teachers can support higher order thinking and make visible heuristic strategies used by experts.

Work by De La Paz and colleagues has shown this approach to be successful in teaching disciplinary forms of writing in history (see Wissinger & De La Paz, 2020 for a review). In a series of studies focused on historical reading and writing, De La Paz and others (2014, 2017) teachers worked with students to articulate historical reading, thinking, and writing practices in the context of each investigation. In this work, the authors created inquiry lessons that gave students multiple, developmentally appropriate opportunities to practice historical reading, reasoning and writing. Teachers moved through several stages of instruction, helping students set learning goals and develop metacognitive strategies. Teachers taught historical reading and writing heuristics and modeled them by thinking aloud (Collins, et al, 1991), then gradually released responsibility to students who applied the heuristics with reduced scaffolding. Scaffolding included tools for students to use while reading, planning, and writing, and tools for teachers to monitor students' progress (Monte-Sano et al., 2014).

The authors found that with these supports students could better analyze evidence, acknowledge other perspectives, contextualize, and write substantive historical arguments (effect sizes = .45 on students' argumentative writing, .32 on their writing quality, and .60 on the degree to which papers were elaborated) which translates roughly to a 17% gain in argumentative writing, a 13% gain in writing quality, and a 23% increase in elaboration (De La Paz et al., 2017). More recently similar results have been shown in science. Lee and De La Paz (2021a) effectively used the cognitive apprenticeship with embedded cognitive and linguistic scaffolds to

support six struggling students. After instruction, students produced longer and higher quality explanations for scientific phenomenon.

Student Transfer of Knowledge

The current study addresses a gap in the literature on cognitive apprenticeships. Prior studies in history and science (De La Paz et al., 2017 , Lee & De La Paz 2021a, Levin et al., 2021) use near transfer measures (i.e., experimenter designed posttests). In contrast, the purpose of the current study is to demonstrate the effects of this instructional approach on far transfer tasks. We use Mestre's (2003) definition of transfer as the ability to apply knowledge or procedures learned in one context to a new, different context. Kober, (2015) further distinguished between near and far transfer. Near transfer is the demonstration of learning from the initial learning setting to one that is highly congruent, while far transfer refers to the application of learning to a distinctly different setting. While space limitations preclude further description of these constructs, interested readers may refer to Barnett and Ceci (2002) for additional information.

Content Focused Explanations of Transfer. The majority of research on transfer focuses on content that learners are expected to transfer (Mestre, 2003; Reeves & Weisberg, 1994; Schwartz & Nasir, 2003). Researchers contend that students are more likely to apply what they have learned from one setting to another if they have formed a content-based representation or a sufficiently complex schema that can be generalized (Chi & VahLehn, 2012; Engle et al., 2012). Early research on content transfer observed how inducing generalizations using surface similarities and cues could promote transfer (Gick & Holyoak, 1983). Subsequent research has focused on providing substantive content expecting that elaborated information would help learners to create expansive schemas that learners could more easily transfer (Gentner et al.,

2017; Reeves & Weisberg, 1994; Rittle-Johnson & Star, 2007). The most fundamental prerequisites for transfer are (a) initial learning that is sufficiently deep, strong, and lasting, (b) that teachers use multiple examples and nonexamples of transferable ideas, and (c) that teachers provide specific content-based cues that evoke transfer between analogous and non-analogous settings (Anolli et al., 2001; Bransford et al., 1999; Chang, 2006). From a conceptual perspective, the extent to which initial learning is deep, sufficient, and lasting and the extent that content cues are embedded should dictate the degree that learning, and transfer occur.

Context-Focused Explanations of Transfer. In contrast, research that focuses on contextual factors that promote learning transfer is less well represented in the field. Barnett and Ceci (2002) and Thorndike (2009) suggest that the context by which content learning occurs is important to consider. They theorize that students store contextual features in memory along with content suggesting that contextual similarities can prime or cue recall of the associated content. In addition to structural similarities, Engle (2006, 2011) asserted when students engage together meaningfully, the social interactions can also act as important cueing mechanisms. As such, discussions may facilitate students' understanding that a scientific principle that can then be generalized to other settings. Further, because transfer can be difficult to achieve, it is important to build in redundancies during instruction to ensure students recognize similarities between divergent settings (Englert, 2006). Thus, the degree to which contextual and social cues are embedded in learning affect far transfer (Barnett & Ceci. 2002).

The Cognitive Apprenticeship and Student Transfer

In the current study we explore how near and far transfer occurs when students write arguments in middle school science classes. We anticipate that specific components of our cognitive apprenticeship will promote the application of skills to far transfer contexts as several

instructional elements mediate learning. First, we developed four investigations providing students with multiple opportunities to explore scientific phenomena. Although topics varied, each retained important contextual features such as procedural steps for collecting and analyzing data. Second, explicit instruction on scientific thinking and the use of cognitive tools (Brown et al., 1989) are two important instructional elements. The teacher introduces scientific thinking, writing skills, and discussion skills through explicit instruction and gradually releases scaffolds to promote complex mental representations of critical information and procedures. The same lesson structure and materials helped students attend to critical content cues such as the presence of independent and dependent variables in experimentation.

Finally, our approach builds in purposeful social interactions that can act similarly as contextual cues to be transferred to novel contexts (Levin et al., 2021) by emphasizing the role of small and whole group discussion. These discussions provide opportunities to engage in multiple social interactions and gives students the chance to grapple with contrasting views to construct more complex understandings of scientific concepts.

Purpose

The purpose of this study was to gain data on the effectiveness of a cognitive apprenticeship to support academically and culturally diverse middle school students to compose written scientific arguments. This study also explores whether cognitive apprenticeships are effective in supporting student transfer. We evaluate transfer of knowledge by comparing how students are able to construct arguments from an initial learning context to a novel context. Moreover, we compared outcomes to a nonequivalent comparison group that did not take part in the intervention, to strengthen our research design. We ask several research questions to explore the relationship between the instructional intervention and student outcomes. Our questions are:

1. Do students demonstrate improved claims, evidence, and reasoning (CER) on a near transfer outcome of scientific argumentative writing?
2. Do students demonstrate improved claims, evidence, and reasoning (CER) on a far transfer outcome of scientific argumentative writing?
3. Do students show improved written argumentation outcomes when compared to a nonequivalent comparison group?

Methods

Research Design

We used a quasi-experimental pretest-posttest design, including a non-randomized, non-equivalent, business-as-usual (BAU) comparison group (Figure 1). The comparison condition occurred in the first year and the intervention occurred in the second year as the teacher's students served as a comparison for a different project with the same outcome measures initially. In the second year, the same teacher provided 12 days of instruction over the six months for the treatment condition.

Participants

Students in eighth grade science classes participated in this study, which was conducted at the same school in the Mid-Atlantic. We were able to gather pre and posttest data 110 students (38 students in the comparison and 72 students in the treatment condition). See Table 1 for demographic information. The school reported that 89% of students received free or reduced lunch. One teacher with seven years of experience teaching science taught eight 8th grade classes across two different years.

Investigations Overview

We chose four topics that aligned with the teacher's curriculum. We designed each investigation to take three 90 minute periods. Lessons drew materials from work by Sampson and colleagues (2013a, b) and were modeled after previous work conducted by Levin et al. (2021) which used the following structure.

1. Day 1: Students' design and execute an experiment that is meant to address the central question of the inquiry. Students collect and grapple with evidence to construct an initial claim, justified it with scientific principles, a Claims, Evidence, and Reasoning (CER) essay.
2. Day 2: Students begin in small group discussion to construct their first CER to address the central inquiry. The teacher structures an "argumentative discussion" in which students can share and critique CER using discussion supports such as adapted graphic organizers and prompts.
3. Day 3: Students either revise their initial CER or rewrite the CER based on information gathered during discussions.

During each of the four investigations the teacher first modeled, then supported new skills. In Investigation 1 students explored climate change and answered the question, "How do cloud cover and carbon dioxide concentration in the atmosphere affect the surface temperature of the Earth?" The teacher prepared students for learning by introducing the guiding question and providing students with background information to inform the construction of the argument. Readings and background information focused on helping students understand the underlying mechanism to help students build schemas for understanding the scientific phenomenon. The goal of this investigation was to model key skills for experimentation, discussion, and argumentative writing as well as the use of scaffolds to support scientific thinking.

During the second investigation, students examined electromagnetism and investigated how the strength of an electromagnet was affected by the number of turns of wire in a coil. Our goal was to guide students in using newly introduced skills and scaffolds needed to think scientifically and produce a scientific argument. The teacher modeled how to (a) determine what the central question was asking, (b) understand relevant information in the background reading, (c) conduct the experiment, and (d) collect data. The teacher modeled how to engage in productive discussion in small groups and led "argumentative discussions" in whole group, in which students critiqued and revised their initial CERs. These small and whole group discussions were further facilitated by prompts that encouraged students to consider their understanding of the central questions, analyze how they had collected data, and compare the conclusions they had drawn from that data. The teacher also introduced a simplified rubric and modeled judging evidence as a support to help students better evaluate their own initial CERs.

During the third and final investigation students took more responsibility for the major aspects of the investigations. In the third, students explored food chains and considered factors that affected the size of the wolf population in NetLogo's "sheep-wolves-grass" simulation, and in the fourth they examined kinetic energy, considering how mass and velocity influenced kinetic energy. Both investigations shared the similar goal of reducing supports and moving students toward independent use of acquired scientific thinking and argumentative writing skills. The teacher gave students more responsibility for small and whole class argumentative discussion (although he still led discussion for struggling learners in two classes).

Between the third and fourth investigations, the teacher taught a writing lesson. Students learned how to use a simplified version of a rubric designed by McNeil and Krajcik (2012) to grade CER for sample essays as well as their own work. The goal was to improve students' self-

evaluation and revision skills. In the last investigation, students conducted their science experiment with greater independence. They still worked in small groups to conduct the investigations collaboratively, but students completed CER evaluations and revisions independently. The teacher no longer provided supports and students worked without scaffolds.

Comparison Condition

We did not observe students in the comparison condition, however we did question the teacher as to the forms of instruction he used in the previous year. Though the teacher did not use a cognitive apprenticeship model in the comparison condition, he did describe effective forms of instruction. These included the use of scientific inquiries, experiments, student discussions, and writing. He covered the same topics in the comparison condition as in the treatment condition.

Professional Development

We designed professional development (PD) for the single participating teacher, implementing it the same way that we would have if working with several teachers. We introduced lessons and materials from each investigation, emphasizing specific features such as modeling, practice, feedback, and reflection. We discussed how he planned to implement these skills and gave him opportunities to practice key teaching elements with feedback. This approach enabled him to reflect on and share his understandings on what scientific thinking and writing meant for his culturally, linguistically, and academically diverse learners.

PD entailed both formal PD opportunities as well as continued coaching. All three authors met with the teacher before each investigation and introduced new content as well as new skills. The first author then visited the classroom multiple times during the implementation of each investigation to support the teacher and observe practice. At least once a week during the implementation of each investigation, the first author would meet with the teacher to discuss

instructional decisions and aid the teacher in addressing any problems. Both the formal PD and coaching were highly specialized to meet the needs of the specific teacher and his classroom.

Implementation Fidelity

Fidelity was measured as both a checklist of critical lesson components and a general measure of the quality of instruction. Of the 63 lessons across the four class treatment periods, 30% of lessons (n=19) were randomly chosen for evaluation. The first author developed a list of critical components inherent to each of the investigations and the cognitive apprenticeship model. The first author reviewed audio and identified if the teacher adhered to these critical components within each lesson. The teacher implemented lessons with 90% fidelity meaning that the teacher adhered to 90% of these critical components across the 19 lessons. The first author also created a quality scale rating the teacher's implementation of each of the critical components on a scale from one to four. A rating of one meant that the component was present but ineffectively implemented and a rating of four meant that the component was present and implemented with high effectiveness. The teacher demonstrated consistently high-quality instruction with an average rating of 3.8 out of 4 for the reviewed lessons. We did not observe in the comparison condition, but the teacher reported that he used traditional lecture-based instruction.

Data Collection and Analysis

Overview of Writing Measures. We administered a standardized writing test to assess students' initial writing ability for both comparison (Y1) and treatment (Y2) conditions. We collected written responses to the fourth investigation's central question to assess near transfer for the treatment group (Y2) only. Students completed a far transfer scientific reasoning task at the beginning and end of the academic year, for both comparison (Y1) and treatment (Y2), Our

design dictated that we compare near and far transfer outcomes within the treatment group (Y2). In order to gain more confidence in our results, we then compared far transfer outcomes of the comparison group to those of the treatment group.

Test of Written Language (TOWL 4th Edition). The TOWL is a norm-referenced, comprehensive test of written expression. Students compose a story based on a picture prompt. We analyzed two subtests, one for grammar (subtest 6) and one for quality (subtest 7). The first author scored all assessments. A second rater, unfamiliar with the project was trained and scored 30% of the sample (randomly selected), and Cronbach's alpha was calculated to measure agreement. A Cronbach's alpha of 0.98 was achieved across the two subtests.

Analysis of Near Transfer (Kinetic Energy Task). Students in the intervention condition constructed an argument by providing a claim to answer the central question, evidence to support the claim, and reasoning to connect the evidence to the claim and other pertinent scientific information. They then participated in two rounds of discussion, first in small groups and then in a whole group. Students revised their first CER after both discussions. We measured growth by comparing students' first to their second CER using a rubric adapted from McNeil and Krajcik (2012), calculating total scores by adding scores for C, E, and R together making six the maximum value. This outcome was considered a near transfer outcome because students were applying skills developed during the investigations to an outcome directly designed to reflect the skills being taught. The setting was familiar, and the outcome repeated throughout the investigations. The first author assessed all students' initial CER and revised CER. A second rater, unfamiliar with the project was trained and assessed 30% of assessments to ensure reliability. A Cronbach's alpha of 0.91 was calculated for initial CER and 0.96 for revised CER.

Analysis of Far Transfer (Purple Loosestrife Task). Students constructed arguments using a hypothetical teacher's observations about a change in marsh conditions as data, based on an interview protocol developed by Hogan and Maglienti (2001). We considered students' arguments about purple loosestrife to be a far transfer task for two reasons. First, it was given to students after completing the four investigations, thus representing a change in time, since we compared students' responses to their arguments that were written at the beginning of the school year. Second, the purple loosestrife task represented a change in the functional academic setting. The far transfer task was different from the learning setting in two important ways. First, the scenario was presented as a narrative explanation rather than a central question. Students needed to recognize that argumentation was the correct approach to the far transfer task. Second, data were presented as a series of observations rather than experimental data. Students needed to recognize that the observations were a form of data and translate this data into workable evidence. The same rubric was used for evaluating students' responses, and the same procedure was used to determine reliability. A Cronbach's alpha of 0.82 was calculated for pretest CER and 0.88 for posttest CER.

We first analyzed near and far transfer within the treatment group. We assessed growth on the Kinetic Energy task using dependent t-tests to compare initial CER to revised CER. Similarly, we used dependent t-tests to compare pretest scores on the Purple Loosestrife task to posttest scores within the treatment condition. Second, we analyzed far transfer outcomes comparing growth between the treatment and comparison groups by using a multilevel approach. In so doing, we first addressed baseline equivalence between the treatment and comparison groups by testing demographic factors as well as pretest indicators of student ability such as the TOWL. We assessed the experimental effects controlling for confounding student-level

variables. We determined variables for consideration through the baseline equivalency testing, which revealed Hispanic status, student pretest writing ability (as measure by the TOWL), and student pretest argumentation scores to be inequivalent at baseline and important covariates. We also included two other demographic variables of interest, gender, and special education status. The coefficient of interest is the indicator of the treatment cohort at Level 2.

Level 1: $Posttest_{ij} = \beta_{1j} pretest_{ij} + \beta_{2j} gender_{ij} + \beta_{3j} special\ ed_{ij} + \beta_{4j} hispanic_{ij} + \beta_{5j} TOWL6_{ij} + \beta_{6j} TOWL7_{ij} + e_{ij}$

Level 2: $Posttest_{0j} = \gamma_{00} + \gamma_{01} Treatment_j + \mu_{0j}$

The final combined model was used to examine the total argumentation scores as well as each of the components of CER. Finally, we examined the heterogeneity of treatment effects by adding interaction terms into the model.

Results

Research Question 1: Do students demonstrate improved claims, evidence, and reasoning (CER) on near transfer outcomes of scientific argumentative writing?

Dependent t-tests indicate that students produced better CER after class discussions and written revisions. The results reported in Table 2 indicate that the cognitive apprenticeship had significant effects on student ability to construct arguments (ES=1.08). The intervention had a small effect on the ability to construct a claim (ES=0.21), a small to moderate effect on using evidence (ES=0.34), and a moderate to large effect on providing reasoning (ES=0.67). These findings demonstrate that the entire class of students benefited from the small and whole class discussions, not simply the relatively small subset of students who participated orally.

Research Question 2: Do students demonstrate improved claims, evidence, and reasoning (CER) on a far transfer outcome of scientific argumentative writing?

To answer this question, we first analyzed growth within the treatment condition, to see if improved learning was evident from the beginning to end of an academic year. Dependent t-tests reveal that students in the treatment group made significant pretest to posttest growth on the Purple Loosestrife task (see Table 3) with moderate to large significant effects on students' overall ability to produce argumentative writing (ES=0.67), small effects on ability to construct a claim (ES=0.37), and moderate to large effects on both ability to use evidence (ES=0.71) and provide reasoning (ES=0.63). Therefore, students were able to apply what they had learned from the four scientific investigations to a different task given at the end of the year.

Research Question 3: Do students show improved written argumentation outcomes when compared to a nonequivalent comparison group?

We assessed pretest equivalence between students in the comparison group to students in the intervention group. Table 4 gives summaries for these measures. We found insignificant differences associated with student learning characteristics related to control variables of gender, special education status, and Hispanic status. Conversely, we found that the comparison group showed statistically stronger pretest writing ability on the TOWL. Additionally, we found that the comparison group also scored statistically higher for student pretest argumentation ability; therefore, we took these differences into account statistically when analyzing learning gains.

Next, we examined the variability between sections on posttest. To do this, I first fit an unconditional model (Table 5) that had no predictive variables to condition the estimate. Using this model, I found the estimates of covariate parameters for my outcome measure. Using these estimates I calculated the ICC using the following formula $ICC=\frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2}$. The intraclass correlation (ICC) for total argumentation scores was more than 23.6%. Because considerable variability was explained at the class level, we developed a multilevel model to explore

variability within and between sections. We fit the model, controlling for variables found to be inequivalent at baseline and variables of interest at Level 1. Condition (treatment v comparison) was the Level 2 indicator of learning outcomes.

The reference group in the model is a non-Hispanic male who did not receive the treatment and who did not have an IEP. The results in Table 6 demonstrate that the treatment had a large positive (ES=1.95) and statistically significant effect on total argumentation. There was a small to moderate, negative though statistically insignificant effect on student ability to construct a claim (ES= -0.31). Conversely, there was also a large, statistically significant, positive effect on students' ability to use evidence (ES=0.93), and a moderate to large, statistically significant, positive effect on their ability to provide reasoning (ES=0.57). These results represent positive effects for students in the treatment group, as compared to students who did not receive treatment after controlling for gender, Hispanic status, special education status, pretest writing ability, and measures of pretest argumentation ability.

Finally, we examined whether the intervention had differential effects for students in the treatment group depending on demographic factors, writing proficiency, or pretest argumentation skills. We ran models exploring cross-level interactions between Hispanic status, special education status, gender, TOWL 6 scores, TOWL 7 scores, pretest scores, and treatment condition. Results show that there is no evidence of differential effects of the treatment on any of the covariates of interests ($p>0.5$). This indicates that all students benefitted equally from the intervention, regardless of background.

Discussion

Previous research has shown the benefits of supporting argumentation and writing through a cognitive apprenticeship model for students in middle and high school in science and

history (De La Paz et al., 2017; Lee & De La Paz, 2021a; Levin et al., 2021). To our knowledge, this study is unique by theorizing how specific elements in cognitive apprenticeships promote transfer. We examined students' initial argumentative writing skills, and further determined whether our approach supported learning transfer to a novel science task. We were not surprised at our findings, however, as our approach allowed us to deepen the content of learning tasks, and we built in meaningful context cues to facilitate transfer. To elaborate, the teacher introduced important scientific thinking, writing, and discussion skills through explicit instruction and purposeful release of scaffolds helping students to build sufficient mental representation of highly complex skills. We employed similar lesson structures, tools, and materials with multiple examples to build mastery over time and evoke transferable ideas (e.g., the importance in using scientific principles when linking evidence to claims). Over the course of the four investigations, we retained key procedures as contextual cues and used discussions to help students to deepen their understanding of scientific ideas.

Our first set of results establish that the cognitive apprenticeship helped students to develop skills needed to make meaningful revisions to their CER. These findings are consistent with those reported in studies in which researchers provide explicit strategy instruction and scaffold learning to improve writing outcomes (De La Paz et al., 2017; Lee & De La Paz, 2021a; Levin et al., 2021) and may be explained in part by instructional differences between treatment and comparison conditions. The teacher reported the use of traditional lecture based forms of instruction in the comparison condition making it likely that students benefited in the treatment condition from the authentic scientific experiments, the extensive use of discussion, and dedicated structural supports.

Unique to our cognitive apprenticeship was the particular emphasis on discussion. This greater emphasis may have contributed to the particularly large effect size for the near transfer outcome (ES>1). By modeling small group interactions and supporting argumentative discussions, the teacher facilitated student engagement in meaningful disciplinary interactions (Felton et al., 2019). Additional scaffolding prompted students to consider critical disciplinary aspects such as comprehension of the central question, collection and analysis of data, and evaluation of conclusions. The discussion with scaffolds helped students to engage in disciplinary discussion and build more complex understandings of the content, methods, and reasoning needed for making meaningful revisions to their CER.

It is also worth noting that the effectiveness of discussion may be contingent on the strength of a given teacher's implementation. Felton et al., (in review) provide deep analysis of the discussion component of our current cognitive apprenticeship. In our analyses we found that the participating teacher shifted between "dialogic stances" as a way to support student thinking. At times he took a guiding stance for directing student thinking, a facilitating stance for prompting peer dialogue, and finally a coaching stance for ceding dialogue to students. The teacher juggled these different stances to help different groups of students develop a deep and complex comprehension of critical scientific thinking skills. This veteran teacher was responsive to the needs of his students during whole class discussions.

The results for our remaining research questions indicate that students who participated in the study were able to transfer skills to a markedly different task, and that gains observed in the treatment condition are evident when compared to students in a comparison condition. These results represent a more significant contribution. Previous research resulted in learning outcomes that closely reflected the context of the intervention (De La Paz et al., 2017, Lee & De La Paz,

2021a). In contrast, our current research conceptualizes the mechanisms in our cognitive apprenticeship that contribute to the successful transfer of skills, and outcome data that shows improved learning on a measure that is significantly different from the context of the intervention.

Our research confirms that a cognitive apprenticeship can produce positive outcomes on measures of far transfer. It facilitates learning by supporting the mechanisms that make transfer of skills possible. Transfer can happen through both physical content and social context (Barnett & Ceci, 2002; Chi & VahLehn, 2012; Engle, 2011). Students build complex mental representations of content that can be generalized to separate contexts (Chi & VahLehn, 2012). Additionally, students can learn to recognize physical and social cues that are consistent between two contexts (Barnett & Ceci, 2002; Engle, 2011). Our cognitive apprenticeship supports these mechanisms of transfer through explicit instruction and gradual release of scaffolds, consistent methods and supports, and the use of class discussion. While the field has long known that cognitive apprenticeships are multi-component interventions, in this study we attempted to collect data that could explain why our model promoted both initial learning and transfer.

Limitations

The current study had some limitations. First, because of difficulty recruiting science teachers interested in writing instruction, we were not able to use a comparison group at the same time the intervention occurred, and we had to use data from the same teacher across two years. Second, due to time constraints we were not able to conduct a maintenance probe with students in the intervention group, which would have provided us valuable information on transfer over time, in addition to the data that we collected on transfer of learning.

Future Directions

The current study extends research on the use of the cognitive apprenticeship in science however, further research is needed to fully establish how individual instructional elements cause students to develop more advanced argumentative writing skills. This line of research would benefit from an expansion of scope addressing different grades and scientific disciplines. However, our results provide initial data that this model is effective for supporting the meaningful transfer of needed skills to novel scientific settings.

Tables

Table 1: Student Demographic Information

	Full Sample (*N=110*)		Control Condition (*N=38*)		Treatment Condition (*N=72*)	
Variable	Mean	SD	Mean	SD	Mean	SD
Hispanic Status	79%		61%		89%	
Special Ed	29%		21%		33%	
Gender	43%		47%		42%	
TOWL6	15.35	4.8	17.15	5.26	14.4	4.31
TOWL7	8.45	3.23	9.97	2.59	7.642	3.28
Pretest Total	0.82	1	1.34	1.15	0.54	0.8
Pretest Claim	0.57	0.63	0.84	0.64	0.43	0.62
Pretest Evidence	0.17	0.4	0.29	0.52	0.11	0.32
Pretest Reasoning	0.07	0.26	0.21	0.41	0	0

Table 2: Near Transfer (Kinetic Energy) Outcomes

	Initial CER		Revised CER					
Outcome	Mean	SD	Mean	SD	DF	t-score	p-value	ES
CER Total	0.15	0.36	1.54	1.79	80	3.17	<0.003**	1.08
Claim	0.51	0.69	0.66	0.73	92	1.48	0.142	0.21
Evidence	0.22	0.51	0.43	0.69	90	2.88	0.006*	0.34
Reasoning	0.18	0.44	0.61	0.79	96	4.46	<0.001**	0.67

Table 3: Far Transfer (Purple Loosestrife) Outcomes

	Pretest		Posttest					
Outcome	Mean	SD	Mean	SD	DF	t-score	p-value	ES
CER								0.67
Total	0.54	0.8	1.39	1.61	142	4.74	<0.001***	
Claim	0.43	0.62	0.67	0.67	142	3.03	0.034*	0.37
Evidence	0.11	0.32	0.49	0.69	142	4.19	<0.001***	0.71
Reasoning	0	0	0.24	0.54	142	3.69	<0.001***	0.63

Table 4: Baseline Equivalence

	Control		Treatment				
Variable	Mean	SD	Mean	SD	DF	t-score	p-value
Hispanic Status	0.61	0.5	0.89	0.32	108	3.7	<0.001***
Special Ed	0.21	0.41	0.33	0.47	108	1.35	0.18
Gender	0.47	0.51	0.42	0.5	108	0.56	0.57
TOWL6	17.15	5.26	14.4	4.31	108	2.95	0.004***
TOWL7	9.97	2.59	7.642	3.28	108	3.8	<0.001***
Pretest Total	1.34	1.15	0.54	0.8	108	4.27	<0.001***
Pretest Claim	0.84	0.64	0.43	0.62	108	3.27	0.002**
Pretest Evidence	0.29	0.52	0.11	0.32	108	2.25	0.03*
Pretest Reasoning	0.21	0.41	0	0	108	4.34	<0.001***

Table 5: Unconditional Model

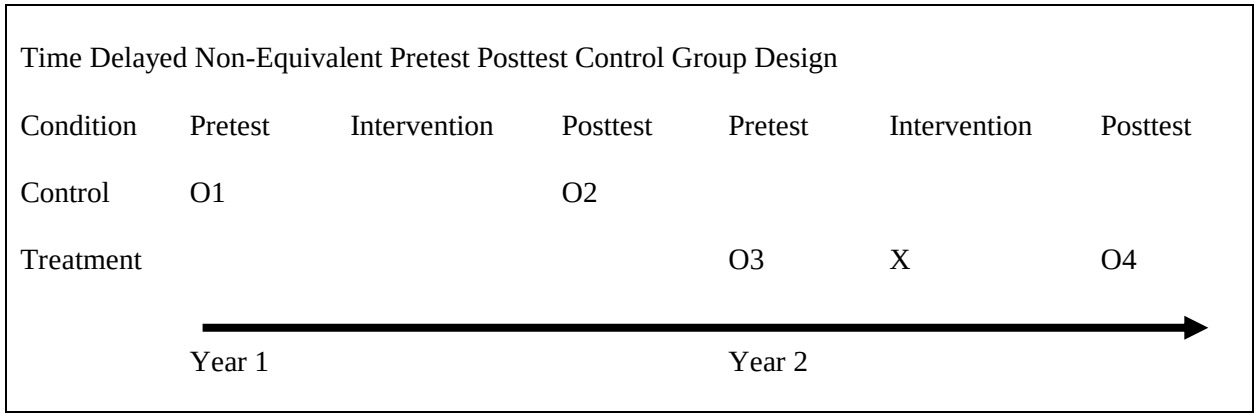
Parameter	Estimate	Std. Error	df	t	Significance	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.42	0.165	8.28	8.59	<0.001	1.04	1.79

Table 6: Treatment Control Comparison

	Total		Claim		Evidence		Reasoning	
Variable	RC	SE	RC	SE	RC	SE	RC	SE
Intercept	-0.602	0.574	0.245	0.26	-0.511	0.257	-0.242	0.205
Cohort	0.676*	0.303	-0.145	0.133	0.402**	0.129	0.261*	0.11
Hispanic	0.195	0.322	0.111	0.145	0.073	0.144	0.054	0.116
Special Ed	-0.031	0.286	-0.067	0.129	0.051	0.13	-0.083	0.102
Gender	-0.426	0.25	-0.1	0.113	-0.143	0.112	-0.128	0.089
TOWL6	-0.021	0.044	0.017	0.02	-0.003	-0.02	-0.015	0.016
TOWL7	0.171**	0.063	0.011	0.028	0.078**	0.028	0.061**	0.11
Pretest Total CER	0.621***	0.142						
Pretest Claim			0.477***	0.092				
Pretest Evidence					0.229	0.144		
Pretest Reasoning							0.443*	0.19
Goodness of Fit								
-2 Log Likelihood		360.76		186.48		186.07		135.76
Treatment ES		1.95		-0.31		0.93		0.57

Figures

Figure 1: Experimental Design



Note: O1: Control pretest, O2, Control posttest, O3: treatment pretest, X: intervention, O4: treatment posttest

Chapter 4: Implementation of a Historical Literacy Curriculum: Teacher Fidelity, Adaptations, and Practice

History education often receives less emphasis in schools than does reading and mathematics. A 2006 survey from the Center on Education Policy found that the average school only devoted 178 minutes per week to history compared to 612 minutes for reading and 457 minutes for mathematics (Center on Education Policy, 2007). Recent policy initiatives articulate the critical authentic skills that students can acquire through history education. The C3 framework is the most recent attempt to outline these skills and provide a meaningful guide to educators (CCSSO, 2010; NCSS, 2013).

C3 Framework and Best Practices

The C3 framework conceptualizes history and social studies as composed of four dimensions. The second and third dimensions focus specifically on historical literacy tools and strategies, and concepts related to evaluating evidence needed to argue during inquiries (NCSS, 2013). As outlined by the C3 framework, teaching historical literacy is vitally important. (CCSSO, 2010; NCSS, 2013).

Historical Literacy. Historical literacy is not simply using domain general literacy skills in a history classroom. Instead, historical literacy entails specific ways of thinking, reading, and writing that are integral to the production and communication of information within a historical setting (Monte-Sano, 2008). Each discipline has specific skills, tools, and ways of thinking that are unique and independent for that environment (Shanahan & Shanahan, 2008, 2012). Thus, within history, a historian reads and interprets text much differently than a mathematician or physicist. Students do not naturally demonstrate these discipline specific skills; however, they do

have the capability to develop these skills given time and proper instruction (Nokes, 2010; Wineburg, 1991).

Historical literacy is a concept that has been explored meaningfully for the past four decades. Various definitions all agree that critical to historical literacy is the use of a variety of primary sources to research and analyze accounts to construct interpretations of historical events (Seixas, 2006; Wineburg and Wilson, 1991; Holt, 1990; Booth, 1983). In the 1990s, researchers began to observe how students worked with primary and secondary sources. Students were able to use multiple sources and showed the ability to construct basic historical understandings (Seixas 1998, 1993; VanSledright & Kelly, 1998). Unfortunately, researchers also found that many teachers only used primary and secondary documents as a means of supporting a narrative version of history, often presented in textbooks, rather than allowing students to form their own interpretations (Stahl, 1996).

Wineburg (1991) examined the difference between students and historians by comparing eight high school AP students to eight historians. He had participants read a series of revolutionary war era texts and complete think-aloud activities. Wineburg (1991) found that while historians focused on the skills of contextualizing, corroborating, and sourcing to analyze the documents, students were likely to treat the sources as credible and factual retellings. Students did not attend to discrepancies between the sources and made little effort to critique the author or the author's purpose. More recent research has focused more specifically on students' abilities to read, think, and write like a historian. Students as young as nine, seem to be capable of grappling with multiple texts and develop historical interpretations (Afflerbach & VanSledright, 2001; De La Paz et al., 2021; VanSledright, 2010; VanSledright, et al., 2006; Wissinger et al., 2017).

With proper instruction students can grapple with sources and use skills such as sourcing, contextualizing, and corroborating to draw meaning and make judgments that can be translated into effective arguments (De La Paz et al., 2014; Monte-Sano, 2008; Reisman, 2009). For example, when reading a primary source on the U.S backed coup in Guatemala, students are able to consider who created the source and what were the creator's motivations. Additionally, students can place the source within its greater historical context and compare the source against other sources written about the topic. By using these strategies, the students can comprehend the source through a historical thinking lens and recreate the way that experts in the field interact with texts (Nokes & De La Paz, 2018). Students use this type of analysis to create their own interpretation of a historical event.

Learning historical literacy skills in the classroom is, however, particularly challenging. Students must acquire both domain general skills in reading and writing while also developing history specific strategies and tools that experts use in the field (LaDuke et al., 2016; Moje, 2008). Though historical literacy is functionally different than domain general literacy, it still requires that students develop general skills such as reading comprehension and vocabulary development (McConachie, 2010: Vacca et al., 2014).

Instruction in Historical Literacy

To combat these difficulties and support students in developing historical literacy skills, researchers have identified several effective forms of instruction. One such form that is particularly effective is inquiry based instruction.

Inquiry Based Instruction in History. Inquiry based instruction is defined as an active approach that is student centered and revolves around questioning, critical thinking, and problem solving (Savery, 2006). Students address a central question and teachers work to guide students

and provide effective scaffolding to build knowledge. Inquiry instruction does not ignore content; instead, it helps to contextualize content and adds an authentic purpose to the learning (Reisman, 2012; Saye, 2013).

Multiple studies have confirmed the effectiveness of inquiry instruction in history education (e.g., Parker et al., 2011; Reisman, 2012). Parker et al. (2011) observed the implementation of an inquiry learning intervention in AP US Government and Political classes. The study found that students in the AP class using the intervention scored higher on the end of the year AP exams than students who took the same class using a traditional lecture-based format. The treatment classroom scored an average of 3.46 out of 5 while the other two classrooms scored an average of 2.58 and 2.40 respectively.

In another example, Reisman (2012) conducted a study involving inquiry and documents-based curriculum with eleventh grade students in urban public schools. She trained teachers using the *Reading Like a Historian* (RLH) US History curriculum from the Stanford History Education Group (SHEG). The RLH curriculum frames lessons around a central historical question and consists of three main parts: a background information portion; a document analysis portion in which students practice and apply historical skills like sourcing, contextualizing, close reading, and corroboration; and finally, a discussion portion in which students cooperatively consider using documents as evidence. Students in the RLH condition recorded significantly higher factual knowledge as well as stronger historical reading, thinking, and writing skills.

Inquiry instruction can even be used to support struggling learners. De La Paz and Wissinger (2017) demonstrated that students with or at risk of LD could be taught to address a historical controversy by writing argumentative essays. The authors helped students to develop background knowledge, discuss the historical controversy, and write argumentative essays. The

discussions focused on two schemas, *argument from an expert opinion* and *argument from consequences,* each of which was supported by a series of critical questions to guide the discussion. Those students in the experimental condition demonstrated better historical knowledge, wrote essays that were more persuasive, and showed more sophisticated historical thinking skills than students in the control condition.

De La Paz and colleagues (2014, 2017) identified the cognitive apprenticeship model as a particularly promising model for instruction in historical literacy. The cognitive apprenticeship model incorporates explicit instruction as well as introduces and then gradually releases scaffolds allowing students to master new and complex skills (Brown et al., 1989; Collins et al., 1991). Using this framework, experts begin by modeling complex thinking processes and gradually shifting responsibility of task completion to students. In this expansive investigation, the authors used this instructional model to teach 1,029 ethnically and culturally diverse 8th grade students to think historically and write argumentative essays. The authors created six inquiry lessons that used explicit instruction and gradual reduction of scaffolds to reinforce historical thinking, general and historical literacy skills, and even metacognitive and self-regulatory skills. These supports helped students to develop important historical thinking skills such as perspective taking, contextualizing, sourcing, and corroborating. Students who received the intervention wrote essays that were more persuasive, higher quality, and more elaborate (effect sizes = .45 on students' argumentative writing, .32 on their writing quality, and .60 on the degree to which papers were elaborated). The current qualitative study is part of a larger implementation study that used a similar cognitive apprenticeship developed by De La Paz to meet the demands of students in high school.

Realities of Classroom Practice

Although there is meaningful evidence that supports the teaching of historical literacy in the classroom, researchers have found that when teachers use primary sources, they often do not expressly teach historic literacy skills. Barton (2005), Seixas (1998,) and Wineburg (2006) all found that when teachers used primary sources they often did not do so effectively. Teachers frequently present primary sources as inherently reliable rather than having students analyzing the documents for reliability. Stahl (1996) found that students often ignored contradictions in primary sources and did not consider sources within the proper context.

Though many teachers claim that they use inquiry instruction and state a desire to employ student centered teaching, researchers have found that teachers do not use these strategies as often as they report. In a recent largescale teacher survey of just under 3,000 teachers, most (80%) reported using student centered approaches as their most common instructional method. Additionally, 88% of teachers reported using primary documents at least once a week (Fitchett & Vanfossen, 2013). This is in direct contrast to several student surveys in which students reported that the majority of instruction in their US History course remained teacher centered (Leming et al. 200; Russell & Waters, 2010). Students reported that the most common teacher method was lecturing and worksheet completion, and that readings often consisted of memorizing facts and responding to multiple choice or short answer questions. This contrast may be explained by teachers' overarching desires being tempered by difficult barriers to implementing inquiry-based literacy instruction.

Researchers have identified a number of reasons why teachers may not use inquiry and historical literacy instruction. First, there is some reluctance to teaching historical literacy as history teachers may fear that they are being turned into English teachers and that teaching reading and writing is watering down the history curriculum (Lee & Swan, 2013). Second,

teachers often feel external pressure to get through the entirety of the school curriculum. Teachers often feel that provided curriculums are far too extensive and that it is their primary job to provide details and facts that students may need for required exams (Thacker et al., 2016). Third, teachers report that creating inquiry and historical literacy lessons is too time consuming and that they do not have access to the resources necessary to effectively plan and implement instruction (Leming, et al., 2009; Levesque, 2009; Wineburg, 2001). Fourth, many teachers avoid teaching historical literacy because they do not have the pedagogical skills needed and do not feel prepared to support students (Rangland, 2007). When teachers do not feel prepared to teach inquiry and historical literacy, they often fall back on what Shulman (2005) refers to as "signature pedagogies" which include lecturing, using textbooks, and using other teacher centered approaches (Bain, 2006; Kiuhara et al., 2009; Nokes, 2010). Finally, teachers may perceive the complex skills required to engage in historical literacy too difficult for students. Many teachers hold fixed mindsets about their students (Dweck, 2016) and believe that their students' innate abilities are not sufficient to engage with historical thinking and literacy.

Teacher Development

Teachers face considerable barriers in implementing inquiry and historical literacy instruction, so it is important that teacher professional development (PD) and support be effective. The Learning Policy Institute (Darling-Hammond et al., 2017) identified a series of effective practices for PD. These include that development (a) is content focused, (b) incorporates active learning, (c) includes collaborative work, (d) uses models of effective practice, (e) provides ongoing coaching and support, (f) includes feedback and reflection, and (g) is sustained in its duration. Grossman et al. (2009) provide a framework for PD implementation that includes many of these components. The authors place emphasis on sharing representations,

decompositions, and approximations of practice that show how to use specific scaffolds and approaches to instructional work.

Monte Sano et al. (2017) describes the PD employed in a recent intervention that used a cognitive apprenticeship model to support teachers in delivering instruction that could help students to develop historical literacy skills. The authors modeled the use of investigation materials, explained the key elements and goals, talked through the core tenants of each investigation, and gave teachers the opportunity to practice teaching important parts of each lesson with peers and researcher support. During PD sessions, teachers worked in small groups and took turns modeling strategies and materials while receiving feedback and cooperatively brainstorming how implementation may look in the classroom. In this way, teachers were able to practice in a low stress, authentic environment and receive meaningful and targeted feedback. Once instruction began, the authors also had participants review students' work and reflect on how instruction could better respond to students' specific needs. De La Paz et al. (2017) found that when using this PD framework teachers were able to improve student historical literacy outcomes. The larger study within which my study was conducted used a similar model of PD; however, unlike previous studies, the current PD was conducted completely online.

Conceptual Framework: Teacher Implementation and Professional Development

In the current study, I attempt to understand how teachers implement a historical literacy curriculum provided by the University of Maryland (UMD) and how PD provided by UMD may affect this implementation. As part of this investigation, I observe fidelity, chronicle teacher adaptations to the curriculum, and try to analyze what factors influence teachers' implementation decisions. Additionally, I document the changes in fidelity, adaptations, and important factors as teachers attend PD and speculate as to the effects of the PD supports. I report outcomes from a

yearlong PD series meant to support the initial implementation of a curriculum that focuses on inquiry-oriented instruction and historical literacy. Researchers have long recognized that factors such as teacher beliefs, PCK, self-efficacy, and other factors influence teacher's decisions about practice. Fives and Buehl (2012; Figure 1) provide a model for how important factors influence teacher practice which I use during my coding as a helpful organizational scheme.

Moreover, researchers have documented how PD can affect important factors which influence teacher's decisions. Studies have even looked specifically at how PD affects teacher's beliefs and PCK when using cognitive apprenticeships, the model used during this study. Lee et al., (2021) demonstrated that with effective PD two teachers significantly changed their beliefs about the purpose of writing within science education during a cognitive apprenticeship. Lee et al., (2021) drew from Guskey's (1986) model of teacher change which speculated that through PD and meaningful application of effective strategies teachers could change their internalized beliefs. Monte-Sano et al., (2017) documented how teachers, through PD, developed a series of new pedagogical skills which helped them to properly grade student work and provide feedback during a cognitive apprenticeship to support historical literacy. After PD, teachers focus more on key aspects of historical writing, paid more attention to evaluating student historical thinking, considered the quality of student work (not just work completion), and identified students' needs. This study shows the wide ranging effects of meaningful PD. As with Monte-Sano et al. (2017), we ground our study in the work of Grossman and colleagues work (2009) which emphasizes the need for representations, decompositions, and approximations of specific practices and concepts within a given curriculum as the cornerstone of effective PD. Together these concepts give teachers the opportunity to understand and enact new practices as well as develop new pedagogical skills and beliefs.

COVID-19

The current study was conducted during the COVID-19 pandemic. The pandemic influenced this study greatly and caused considerable chaos in the cooperating school district. Among the myriad difficulties that the pandemic caused, the most pertinent for the current research were (a) the COVID pandemic forced districts around the country to adopt a distance learning approach, relying heavily on virtual instruction and (b) the COVID-19 pandemic coincided with a time of major civil unrest in the United States. Protest surrounding the murder of George Floyd, Breonna Taylor, and many others prompted the most significant civil rights demonstrations since the civil rights movements of the 60s and 70s. (c) Due to the pandemic, many districts, including the cooperating district, suspended student data collection causing widespread issues for researchers.

Methods

This qualitative study is part of a larger study that was originally designed to evaluate a curriculum intervention targeting culturally, academically, and linguistically diverse high school students' historical literacy. Although the 2019 Coronavirus pandemic prevented us from assessing student learning data, the year was used to strengthen a university-district research partnership (RPP) and to pilot a fully virtual PD program in the form of webinars and individual virtual meetings. RPPs are long-term collaborations between universities and school districts and purposefully investigate problems of practice and solutions for improving the district (Coburn and Penuel, 2016). The relationship between UMD and the district allowed for rapid and consistent communication, making responding to teachers' specific needs more efficient and effective. We provided 8 hours of PD, including a webinar for each of the historical investigations. Teachers were asked to preview each historical investigation, provide teaching

materials, complete surveys and meet with researchers, providing us data on whether they attempted to model with students or provide feedback on their writing, how they differentiated instruction, their sense of students' learning, and challenges they faced in implementation. Teachers completed a final survey, providing us with a measure of their understanding of our curricular goals.

The curriculum included six historical investigations that were supported by online PD. Within the curriculum, students were taught important historical thinking skills such as corroboration, contextualization, and sourcing as well as specific historical reading and writing skills. Important aspects of the overarching study such as the development of cognitive apprenticeship, descriptions of the tools used, description of the webinars, and descriptions of the specific investigations can be found in De La Paz et al. (in review).

Originally, the main purpose of the current study was to chronicle teacher fidelity, observe how teachers made adaptations to the curriculum, and analyze the important factors that affect how teachers made implementation decisions. This study considers multiple aspects that may affect implementation including (a) contextual factors at the district and classroom level, (b) teacher's beliefs about history and instruction, (c) teacher's content knowledge and PCK, and factored in (d) how components of the PD affected implementation. I also attempted to analyze teacher self-efficacy and their perceived value for the curriculum. I seek to understand if and how teacher's beliefs and understandings changed throughout the PD series and if this resulted in adjustments in implementation. The research questions were:

1) To what extent do teachers' implement the curriculum with fidelity?

2) How do teachers adapt the yearlong curriculum? Are teachers' adaptations responsive to student needs? Do the adaptations facilitate or pose challenges for meeting the overall goals of the historical reading and writing curriculum?
3) In what ways, if any, does the PD provided to teachers change their beliefs and understandings, and how does this change instruction?

Research Design

I sought to answer questions of what (fidelity), how (adaptations), and why (teacher implementation). Qualitative research is more likely to address the how and the why of actions, and it offers the best design for answering my research questions. A qualitative study approach is appropriate when the main purpose of the study is to understand the inner experience of a participant and determine how their culture and the context affect meaning making rather than simply testing variables (Strauss and Corbin, 1990). In qualitative research, the focus is on meaning making and the process through with attitude and beliefs influence and are influenced by the specific context (Merriam and Tisdell, 2016). A quantitative approach to my research questions would greatly limit my ability to fully explore the individual stories of my participants in this unique setting. Through the use of multiple forms of data (survey, coaching sessions, artifacts), qualitative exploration can provide a timeline, characters, and context to help the unique experiences of my cases come alive.

In qualitative research, the researcher is responsible for data collection and analysis meaning there are certain advantages and disadvantages (Merriam 2014; Stake, 2010). For example, the researcher can collect specific forms of data that are of interest, clarify data, check with participants to ensure accuracy, and remain flexible in the pursuit of important goals. Conversely, there is always the threat of bias and the inability to generalize findings. I believe

that the advantages far outweigh the disadvantages for my specific research questions and that a qualitative design is the most advantageous design for this study.

Research Practice Partnership

The research practice partnership (RPP) was formally established between UMD and a large school district in the Northeast United States after the current data collection ended. At this district, there are roughly 6,000 ninth graders with about 2,000 identified as economically disadvantaged according to FARMS data. The district's 2019 demographic information indicated that individuals who identified as White = 51.3%, Black or African American = 21%, and Hispanic or Latino = 17.1%, two or more races = 6.2%, and Asian = 4%. Records indicate 86.3% of black students graduate within four years compared to 92% of white students.

The UMD team interacted with two individuals: the district social studies coordinator (who is board certified in social studies education) and the district's secondary history teaching specialist. The district required that all 72 US history teachers use the curriculum; however, PD was optional. As a result, all teachers only attended the first session of the year during district mandated PD. After this, participants at the PDs varied; however, nine teachers from eight schools attended regularly.

Participants

The participating school district had 72 ninth grade teachers. Teachers who taught a minimum of one class of ninth grade American History qualified for our study, though many teachers taught multiple grade levels. This district offered ninth grade American History at both standard and honors levels, both of which were included in the study. The district required that all 72 teachers attended the first webinar in the PD series which was delivered before the start of the school year. Though the district encouraged teachers to continue to attend with potential

monetary incentives, the remaining PD sessions were voluntary. Attendance dropped after the first webinar. For most of the year 15 teachers attended the virtual PD, however by the end of data collection, only nine teachers remained who had completed all of the requirements for data collection (participating in virtual coaching sessions and providing instructional materials). From these nine teachers, I chose to study five individuals as representative cases (Table 1 shows basic demographic information for each case and their schools).

My selection was purposeful as I sought to identify examples of exceptional teaching (N = 2) as well as teachers who showed high (N = 3) and low fidelity (N = 2) to the curriculum. I wanted to observe the full range of fidelity of implementation to document adherence to the curriculum. I also hoped to capture the varying belief systems, personal characteristics, student populations, and contexts of teachers' who exhibited both high and low fidelity. Additionally, among those who showed high fidelity, I was interested in the variation in how those teachers made adaptations that adhered to the core tenants of the curriculum while being responsive to their specific students. Finally, I was intrigued by the differences that may have been apparent between those teachers that I considered exceptional and those teachers who had struggled at times throughout the study. By observing teachers of varying ability levels and varying degrees of fidelity, I hoped to gather a more accurate picture of implementation and decision making.

Data Sources and Collection

Throughout the study, I collected data from three primary sources: coaching sessions, surveys, and artifacts. Additionally, I kept personal records and notes, member checking emails and all emails and minutes of meetings between members of the UMD team and two district leaders, and the participating teachers (which were stored on an ELMS organizational space with all curriculum materials and PD lesson plans that UMD created for the district). After gaining

consent from teachers. In November 2020, I began collecting responses to the first post PD survey. This was followed by the first series of coaching sessions (see Table 2) in December 2020 and January 2021. It was at this point that I also began asking teachers to submit materials from each of the investigations. Teachers submitted a number of materials including PowerPoints, adapted student materials, video models, etc. The UMD team had provided teachers with these tools, so I was interested in the adaptations they had made to the original materials. By studying these adaptations, I could answer my second research question. Teachers were permitted to submit materials up until the study ended in June 2021 and still receive the full compensation. I conducted our second round of coaching sessions in March and April 2021 and our final round of coaching sessions in June of 2021. After the fourth investigation, I began giving surveys that asked teachers to reflect on their implementation. I gave a total of three surveys reflecting on investigations four, five, and six. Finally, I administered one last survey on teachers' beliefs and plans for implementation in the coming year.

Coaching sessions. The primary form of data from this study came from a series of coaching sessions and the accompanying communications. All teachers who participated in the PD were required to meet with the UMD team at three times during the investigations for personalized coaching sessions. During these sessions, our team asked teachers a series of questions for three important purposes. First, our team sought to determine the specific problems of practice that teachers faced when they implemented the curriculum in their specific classrooms. We provided individualized support to help teachers address these problems and properly support the specific needs of their students. Second, our team asked questions to better understand how teachers understood the curriculum and were making accommodations to their classrooms. If teachers had misconceptions our team used this time to remediate the teacher's

understandings and make adjustments to how teachers were making instructional decisions. Finally, our team gathered information on teacher's beliefs about the curriculum and their student's ability to access the materials. We addressed any beliefs that would be detrimental to the implementation of the curriculum and provided teachers with suggestions as to how to overcome and perceived barriers to instruction.

To prepare for each coaching session De La Paz and I prepared a series of semi-structured questions in collaboration with the UMD team (Appendix 1). The primary purpose was to elicit responses that could open the door to meaningful teacher instruction. We presented the questions and followed specific threads within answers to support teachers in properly implementing the curriculum. The coaching session questions were changed based on trends in the other data such as artifacts or surveys.

I conducted all of the coaching sessions with the help of other members of the UMD team. I conducted 12 coaching sessions in the first round, 10 coaching sessions in the second round, and 9 coaching sessions in the final round. Coaching sessions ranged from the shortest being 23 minutes long (which was cut short due to technical issues) to the longest at 73 minutes long. I audio recorded each coaching session with the permission of the participant and attempted to take as detailed notes as possible without it being distracting to the coaching session. I transcribed the first set of coaching sessions with the help of a fellow graduate student. This work was far too time intensive, so I had the remaining audio transcribed using a transcription service. After each set of coaching sessions, I reviewed the information and created a summary of the initial results, identifying early themes in the data, and making detailed personal notes. I then presented these trends to our district points of contact as a means of member checking.

Surveys. I administered a series of surveys to collect additional information from teachers. Surveys offered an important contribution to overall data collection. Because surveys were far easier to complete than it was to participate in coaching sessions, teachers were more willing to complete them frequently. So, I used surveys as a way to collect more close ended information (years teaching, formal education, etc.) and as a way of collecting more frequent data points. Starting at the beginning of the study, in September I used a survey to collect background information on the teachers including how long they had been teaching, their highest degree, their experience with teaching literacy, and their experience with teaching historical literacy (Appendix 2).

I also collected surveys for each investigation after the fourth (Appendix 3). This was a purposeful decision. It often takes teachers multiple investigations to become familiar with the tools and the new curriculum. By the fourth investigation, teachers could make meaningful and accurate reflections on their own implementation. So, I had teachers reflect on their implementation through surveys in February 2021 after the fourth investigation, April 2021 after the fifth investigation, and June 2021 after the final investigation.

I also administered a beliefs survey that asked teachers to discuss their beliefs about the purpose of history education and historical literacy. I had teachers speculate as to their plans for the coming year which will be the third year of the larger project (Appendix 4). Dr. De La Paz and I collaboratively created each survey throughout the investigations.

Artifacts. I examined the artifacts that teachers submitted to me. Teachers were required to submit a PowerPoint and adapted student materials from each investigation. Most teachers submitted more artifacts including class discussion facilitators, mini lessons they created related to historical literacy, student work samples, videos of their instruction, etc. (Appendix 5 shows

example teacher materials). Teachers were asked to submit materials directly after each investigation so that they could be used as part of the member check with the district, but only two teachers consistently submitted materials immediately after the investigation. Most of the teachers submitted most of their materials at the tail end of the study.

Research notes, and memos. The final source of data that I collected was my own notes and memos. I had worked on two previous projects concerning the implementation of a cognitive apprenticeship so I was familiar with the process of implementing the intervention and recognized that my own notes and memos would be a valuable source of data. Throughout the year, as I attended meetings with our district contacts, meetings with our team, meetings with teachers, and the webinars, I took detailed records to help provide context for the study. Due to the COVID-19 pandemic, the context is a particularly important factor to the current study.

It is prudent to take notes soon after collecting important data such as coaching sessions as it is easy to lose the context (Merriam, 2014). Though I attempted to take notes in the moment, these notes often ended up being limited because I wanted to be present in the coaching sessions to follow important threads and ask branching questions if needed. To combat this drawback, I would make sure to review the transcripts within a week of completing them so that the context of the coaching session was still fresh in my mind. When I reviewed transcripts, I would create excel sheets that would include important information and my thoughts about interesting contextual connections to consider later.

Finally, I was careful to write down any thoughts that I had which might reflect my preconceived notions or potential biases. It was important to me that my own biases or notions not poison the data. For example, after the first coaching session, I was frustrated with one of the teachers who had implemented the first investigation with particularly low fidelity choosing to

ignore most of the materials that I and the team had work hard to create. Because I had put so much time into these materials, I was more frustrated about them not being used than I was considerate as to why the teacher did not use the materials. I made sure to acknowledge this frustration so it would not continue to taint my data collection and eventually my analysis.

Data Analysis

For data analysis, I follow the example set by Harry et al., (2009) and use constant comparative analysis. This is a process where the researcher compares data from one source to another in an iterative process to develop a common set of codes that eventually build to themes (Merriam & Tisdell, 2016). I began with my first set of data, took notes, and built minor themes. As I collected new data, I compared my notes and themes to the new data to refine and build on the previous notes and themes. I repeated this pattern until I recognized no new notes or themes.

In reviewing data, a researcher should first gather an overall impression of the data through an in-depth review of the materials (Cresswell, 2014). I achieved this by reviewing each round of coaching sessions, surveys, and artifacts to create detailed summaries for the district member checking (Appendix 6 shows example slides from one of the PPTs produced for the meetings with the district). I also periodically reviewed my personal notes and memos to make sure that I was understanding my data within the correct context.

From there, Creswell (2014) suggests an initial reading and coding of data based on how each source contributes to the understanding of the issue and how it may relate to a given framework. To this end, I began with my first set of coaching sessions, collecting initial notes in the margins. These notes were organized by my framework as a means of addressing exactly the questions for this study. Broadly, I created notes about the context, teacher beliefs, teacher content knowledge and PCK, teacher self-efficacy and interest, and implementation. Fives and

Buehl's (2012) model of teacher practice helped me to identify these broad categories. These initial notes became my first set of codes. From there, I moved to the next set of data which was the second coaching session. I used constant comparative analysis to find new codes and refine my codes from the initial coaching sessions (Table 3 shows an excel sheet of this process). It was at this point that I transitioned from paper and pencil to a computer based qualitative platform (NVivo). On this platform, I could more easily create my codes, create memos associated with my codes, and refine my codes as I worked with increasing amounts of data (Appendix 7 shows a screen grab of the NVivo software). After analyzing the first two sets of coaching sessions, I analyzed the remaining data using the same iterative process. Both surveys and artifacts were included in this analysis as NVivo allows for both text and picture analysis. Importantly, my own notes were only used as a contextual reference and were not used to create the coding scheme.

Once I had identified a series of codes within the data, I organized those codes using my conceptual framework (Table 4 shows my identified codes organized into categories). I identified six codes pertaining to the context, eight about teacher beliefs, 12 codes on teacher content knowledge and PCK, two codes about self-efficacy and value, and 11 codes on implementation for a total of 39 total codes. Once I had identified my initial codes, I followed the lead of Harry et al. (2009) who described the analysis process as a series of six steps. In the first step, the researcher identifies codes in the data. Second, the researcher identifies categories to organize the codes. Third, the researcher develops themes within individual cases. I developed themes for each of the cases by compiling the codes within NVivo and looking for connections between the coded information. I wrote memos to capture identified themes for each case (Table 5 shows an example of the memo writing process). Fourth, the author validates their themes by continuing to collect data and comparing their initial codes, categories, and themes to the new

data. As I collected my final coaching sessions, I used this data to confirm and validate my initial codes, categories, and themes. I identified no new codes or themes in the final coaching sessions showing that my codes and themes properly captured the data.

In the fifth step, the researcher identifies interrelated explanations by conducting a cross case analysis. For this step, I looked at my themes for each case and identified explanations for the connections between the themes. Additionally, it was at this time that I constructed in-depth teacher profiles which helped me to further find commonalities between cases. The final step was to create or confirm an existing theory. Though I had identified certain themes and explanations, I did not feel that my data supported the construction of a new theory, nor did I feel the data sufficient to properly support my existing framework. As a result, my analysis concludes with the fifth level and a robust series of suggestions for future research (Table 6 shows the complete process used during this study).

Trustworthiness

Qualitative research is often more concerned with the overall trustworthiness of the study's findings as opposed to quantitative research which focuses on validity, reliability, and generalization (Merriam & Tisdell, 2016). I promoted trustworthiness by triangulating my findings with multiple forms of data, using member checking, and working with a fellow graduate student to establish the reliability of codes and themes. Triangulation is the process of corroborating evidence by comparing different forms of data (artifacts v field notes), methods of data collection (coaching session v survey), and types of sources (students v teachers) (Cresswell, 2014). Though I only had a single type of source, the teachers, I was able to collect important contextual data from our district points of context. I collect data using multiple methods including surveys, coaching sessions, and artifacts to ensure multiple different accounts

of the same event. Finally, I used different forms of data including my notes as a means of corroborating findings from different points of view.

I used member checking by review data with the district. After set of coaching sessions, I summarized all of the data collected to that point. I presented our findings to district leaders, to keep them knowledgeable about the progress we were seeing and to ask if they could share any other observations that we did not know about.

As a final strategy for validation, I had a graduate student review the data and research processes. At key moments during the data analysis process, this student was consulted to confirm the validity of analytical processes and the outcomes. Over the course of the data analysis, the graduate student contributed more than 40 hours of support. Initially, the graduate student was employed to look over the data and familiarize herself with the tools used during the data analysis (NVivo). Next, as a reliability check, she was asked to review a subset of the data and develop codes independent of my own code book. We then met to compare the codes and discussed any discrepancies. I used the comparison to make minor adjustments to my own code book. Most codes were found to be similar.

After I had completed coding all the data, the same student and I reviewed my codes and the data and discussed if the codes properly captured the data. We then reviewed my initial themes. As someone who had examined all the data, the student was asked if she felt the themes were valid. The student had no suggestions and felt that the identified themes and explanations properly addressed the data.

After developing the interrelated explanations, I consulted with the same student once more. The student was again asked to confirm that the explanations properly captured the cross case themes in the data. The student confirmed that she felt the cross case explanations were

valid. Finally, the graduate student was asked to use my code book and develop themes from an important subset of the data: the beliefs survey. We compared the themes she had recognized in the beliefs survey data to my themes identified about teachers' beliefs in the whole set of data. We identified similar themes from the data contributing to the validity of the code book and the identified themes.

Results

This qualitative case study explores teacher fidelity of implementation, how teachers make adaptations to the curriculum, and why teachers make instructional decisions about implementation. Effective PD can result in improved teacher capacity and improved student outcomes (Yoon et al., 2007). Effective instruction, however, is metered by teacher fidelity and teacher adaptations. It is, therefore, important to understand if teachers implement with fidelity, how they make adaptations, and why they make instructional decisions. The five high school social studies participants taught at five different schools with varying student populations. The following section is divided into seven subsections. The first subsection describes the context at various time points during the intervention as the COVID-19 pandemic necessitated rapid changes to the educational setting. The next five subsections provide detailed analyses of each case. The final subsection section provides an overall summary of how the data relates to the study's conceptual framework.

Context

I present the context as a series of three time points. Although these time points correspond to two investigations each, they also correspond to major shifts in context that greatly influenced teachers during the study. These major shifts include (a) the beginning of the year and the introduction to online learning, (b) the introduction of the COVID-19 vaccine and the return

to in-person teaching for most of the teachers, and (c) the return of most students to in-person instruction. These major changes greatly impacted teacher communication and implementation. Additionally, the first timepoint corresponds to the first two stages in a cognitive apprenticeship model when teachers first model and scaffold new skills, the second timepoint corresponds to the third and fourth stage when teachers begin to release supports, and the last timepoint corresponds to stages when students are attempting to complete skills with independence, and accomplishing this objective.

Timepoint 1: September-December 2020. The shock of moving to online instruction made the initial implementation of the curriculum particularly difficult. By September 2021, students had already been working on online learning for three months not including the summer. Unfortunately, between the 2019-2020 school year and the 2020-2021 school year, the district had not made concrete plans for online learning, and until just weeks before, was considering in-person instruction. Therefore, many teachers were caught off guard and scrambled to create workable materials. Of the transition to online learning one teacher commented, “And really, it’s sort of poorly done and they [the district] sort of just threw some sort of token stuff at it.” Though the UMD curricular materials were built for the online space, they were initially made for a Microsoft environment rather than for google platforms making the curriculum less effective initially. The transition to the online space also necessitated the reduction of instructional time. History classes went from 90 minutes every day to just 45 minutes three times a week.

Many teachers struggled as they dealt with the reality that students had different access to technology and had different skills regarding the use of technology. Much of the early instruction was focused on supporting student access and use of the online platforms. As one teacher put it,

“I would say in general, most of the challenges were technology based and not content.” Another teacher lamented, “because of the issues that we're having in the virtual learning space and being in a title I community, you know, some of them are on their phones.”

It was into this environment teachers began using the new curriculum. Many teachers struggled with the new content. These teachers often fell back on “traditional” and familiar types of instruction such as lecture, close readings, and CER. Teachers also had a tendency in the early investigations to use “hands-off” instruction likely due to a poor understanding of curricular supports and a lack of capacity in teaching historical literacy skills. One teacher described their writing component, “and then for the actual writing I just gave them a google doc and I just had in a box what was required and had their work.” Despite this, certain teachers showed promising material adaptations and quickly acclimated to the new curriculum and the new online space. In fact, five of the 12 initial teachers participating in PD showed an advanced understanding of the curriculum and cognitive apprenticeship as early as the second investigation. These teachers used engaging warm up questions, integrated vocabulary supports, embedded translations, and provided student feedback. Unfortunately, other teachers did not show as adept of an understanding and made adaptations that fundamentally undercut the major focus of the curriculum which was to foster historical thinking and promote historical literacy. These teachers over scaffolded and presented the new skills as piecemeal and unconnected rather than complete skills to be used fluidly when needed. One teacher described her initial adaptations stating:

> “I'm more used to [being] a social studies teacher, you know, using claim evidence reasoning. So, for us, we're just pushing so much for them to be able to create a claim. Pick one document, you know, out of the many that we look at.

Pull that quote and then give the reasoning to tie it all together. So, to me, that's the most important skill. I don't know if that's what coincides with your model."

Timepoint 2: January-March 2021. This time period was marked by a gradual acclimation to the online platforms. After the initial scramble, teachers became more comfortable with the new constraints of the online learning space, including vastly reduced academic time. Teachers settled into effective three day cycles. One teacher describes, "So I did two days with my reading NearPod, and I have one for writing NearPod." Teachers still struggled with attendance and participation which was a continued issue the entire time during distance learning. Students would often sign on and turn off their cameras which made determining participation virtually impossible. Teachers adopted new more effective platforms for instructions such as Google docs, Nearpod, and Jam boards. For example one teacher described using Google docs for monitoring and feedback as such, "Through the google doc, I said the best way to ask me for feedback was just to write a private comment…then I would read it, highlight it, and gave comments on the document itself." The district worked feverishly to pilot new online platforms, however, much of the progress was made individually or within school units.

The most significant contextual change began in January and ran through March as the majority of teachers received the COVID-19 vaccine. This was a time of particularly strained communication between the district, UMD, and the teachers as the district grappled with the difficult question of returning to school. Ultimately, the district chose to return, and most teachers returned to their classroom by the beginning of March. In person instruction was voluntary, so most students remained home. This created a difficult new organizational issue for teachers as they had to split instruction between in-person and online students all the while wearing PPE for the entire day. One teacher recounts this hectic period saying, "We are partially

back in school. So, March 8th, ninth graders who opted to come in, came in. They're cohorted, so you only would ever possibly have half in class and half on the computer".

Despite these difficulties, teachers showed consistent rapid pedagogical growth. Teacher knowledge of historical thinking and their pedagogical knowledge on how to support historical thinking and writing greatly improved. Pacing improved as most teachers adopted our three-day cycle of day one content, day two reading, and day three writing. Teachers showed more expert use of online platforms and adopted more materials directly from the curriculum rather than relying on their past experiences. Teacher practice improved as they implemented important aspects of the curriculum such as flexible use of historical thinking skills, the use of scaffolds that more accurately supported the curriculum, and student monitoring with feedback. One teacher said of the use of scaffolds in the third investigation, "I gave them the activity, I gave them the Nearpod to work through, but at the same time, my co-teacher opened up a breakout Google meet and worked with several students."

It is also worth noting that during this time period, insurrectionists attempted to overthrow the government during the January 6th insurrection. Though the main contextual influence on the year was COVID-19, this event and the considerable political and social turmoil spurred by the murders of George Floyd and many other unarmed black and brown people had considerable influence. These events likely influenced teachers and students. Moreover, they were also of particular importance as the curriculum was based on historical controversies many of which either implicitly or explicitly concern racial disparities in this country.

Timepoint 3: April-June 2021. During this final time period, as teachers reached the end of the year, they began to establish a sense of normalcy. Many of the students had returned to the classroom. One teacher reported having 18 of her 25 students in person. This period was

also typified by slowed COVID-19 infections rates as a spike in the winter was easing. This considerably reduced the anxiety of many teachers. This, coupled with the oncoming summer break, made many teachers more optimistic than they had been during other periods in the year. Teachers quickly established stable in person routines and by now, had a robust list of online and in person tools to support students. Teachers did begin to express frustration that they still had students out and began to be bothered by the continued focus of online work. One teacher recalled this time period saying, "Yeah and I actually found myself muting my [google] meet and talking to them in the classroom."

Unfortunately, at this point, many of the teachers who had originally engaged in PD had dropped out. I was left with only nine teachers or roughly 12% of the total teachers in the district and about 60% of the teachers originally signed up for PD. The teachers who were left showed an advanced understanding of the curricular goals and had clearly learned several new pedagogical skills. While it might have been due to our ongoing PD through the year, it is equally possible that the teachers who remained were particularly strong or highly motivated to teach historical literacy. The final investigations were exemplified by the teacher's ability to adhere to the core components of the curriculum while being responsive to student's needs. Many teachers drew skills directly from the webinars and began to place greater emphasis on cooperative classroom work, effective and targeted content building, more independence, real time monitoring, and meaningful feedback. As an example, one teacher described a mini lesson on making historical judgments recounting:

> "I shared exemplars of judgments so she (and others) could hear/see how judgments could sound in different contexts - including judgments that were less

explicit but still made clear the source was not just an expert to quote, but a part of the historical picture."

By this point in time, many teachers had successfully removed many of the scaffolds that students required in early investigations. Though I did not have access to student work, teachers reported that many students achieved independent mastery of critical reading, writing, and thinking skills from the curriculum.

Though there were commonalities in the contextual changes throughout the year, each of the cases had unique experiences and implemented curriculum based on their own specific beliefs, content knowledge, PCK, self-efficacy, and value. I outline these factors by providing detailed teacher profiles.

Emily: The Reliant Teacher

Emily showed considerable growth throughout the investigations and was a far different educator from beginning to end. She was often reliant on the curriculum or other teachers to help plan her lessons as she openly admitted to having little experience teaching historical literacy. She, as with other teachers, was influenced greatly by the pandemic and the PD.

Context. Emily was a career educator and had been working in the field for 12 years at the start of the study. She was a white female who had only recently transitioned to the school district participating in the study. This posed an initial problem for Emily as on top of the changes due to the COVID-19 pandemic, she was unfamiliar with the general procedures of the district and had little experience with her fellow co-workers. She described the teachers switching students between semester as such, " And you switch teachers at semesters here, which I'm not use to either. So we have to be on the same page." She felt little agency during the study and often deferred to her cooperating teachers to take the lead in planning despite her extensive

experience in the field. As she said, “I’m not used to the county. So I’m definitely still getting used to even having a head of my US department, for example…so pacing being determined by that person is a new concept to me.”

Emily openly admitted to being unfamiliar with much of the pedagogy introduced during the PD saying, “ I am not going to pretend. I say this a lot, not to the kids, but I’m not well trained in how to teach writing.” She said that she had little training in teaching historical literacy and had used it sporadically during her 12 years as an educator. This may have also contributed to her willingness to take a secondary role in team planning.

During the school year, Emily faced similar challenges to other teachers. The transition to online learning created considerable stress as class periods became shorter and Emily had to rely on several underdeveloped pedagogical skills concerning using online platforms. Emily describe her early issues with technology as such,

> “[the problem] is the kids trying to learn how to function by having the document open and their google form open and possibly the documents themselves open. I’m not totally sure how to deal with that as we are going through it.”

Emily’s school was comparatively more affluent, so she dealt with fewer issues relating to student access to technology. The majority student population identified as white (75%) with the remaining population split between African American students (10%), Hispanic students (9%), and students of two or more ethnicities (6%). Her student’ performance was academically average compared to the other schools in the study with 64% proficiency on the statewide ELA test and 46% proficiency in the statewide mathematics test.

Beliefs. Emily believed that her students' learning was hampered by several rather daunting barriers. She lamented about the lack of instructional time available to students and

believed that many of her students lacked the resources at home to be successful during this academic year. She lamented the alternate versions of histories students were learning at home saying, “many students are told different versions of history and they end up frustrated, confused, or challenged by the events told in school.” She taught both honors and standard and when comparing the two, she was overly pessimistic about the ability of the standard students to complete challenging tasks. She described her students as “too literal” and “lacking self-regulation skills.”

Despite this rather pessimistic outlook for her students, Emily set lofty goals for them. She specifically noted that she wanted student to be, “self-sufficient”, “reflective” and acquire skills such as, “independence with the IREAD strategy”, “thesis writing”, and “making independent judgements”. She believed that critical thinking is the most important skill that students can learn in a history classroom. She felt strongly that it was important to use skills from history to “sift through” the bias in modern media. In this way, her goals aligned closely with the goals of the curriculum. It is difficult to know if she truly held these beliefs before the study as it was clear from her significant pedagogical growth that Emily was learning from attending the PD.

Though she had good intentions, Emily struggled to engage her students and noted that engagement was her chief concern when educating children. She believed that grades and interest are the two primary motivators for students. She expressed how she, “thought that [making the work an assessment grade] would help then do it because they need assessment grades which are 60% of their grade in social studies.”

Content Knowledge and PCK. It was clear by her early implementation as well as her own admission that Emily had little experience teaching historical reading, thinking, or writing.

She did mention having used SHEG and something she described as the "taco method," however, she often relied on domain general teaching strategies such as close reading or CER. Examples from her adapted materials show what she called close readings, which were essentially independent readings with basic comprehension questions like, "What is Washington encouraging African Americans to do when he tell them to cast down their buckets." Emily's pedagogical knowledge grew steadily throughout the investigations as she began to use multiple evidence-based practices for instructing historical literacy. In later investigations, she used techniques mainly drawn from the provided curriculum. By the end of the investigations she describe the analysis of a document as such, "I broke them into four different groups. Each group had a different document and then they sourced, essentially, the judgments answers together on a google slide."

A good example of her progress came in how she graded student work. In early investigations, she focused on grading for domain general structure (does the student have a topic sentence?), participation, and grammar. By the end, she was grading largely based on how students exhibited historical thinking skills such as determining reliability. By the end of the semester, Emily was impressively knowledgeable and exhibited high level planning for challenging tasks such as facilitating discussion, using scaffolding to properly support historical reading and writing, and providing meaningful feedback. By the end of the investigations, she state that the most important parts of her lessons were the "Judgements!!!! I broke the E, A, and D into separate boxes and had students work on a group activity."

Self-Efficacy and Value. Emily commented on more than one occasion about the value of the work. She believed that the curriculum was valuable in that it supported student critical thinking as well as provided support for teachers who may not have the time to create such

intensive materials. She stated, “I am absolutely thankful for all of the materials that you guys provided because they are making it [planning] a lot easier.” Despite this, her self-efficacy was low at least initially. She admitted to being ill-equipped to teach historical literacy.

Implementation. As mentioned, Emily’s implementation changed considerably from the beginning to the end of the investigations. Early on, she relied on familiar forms of instruction such as close reading and CER. She tended to over scaffold her students and often broke skills into their component parts and had students complete what amounted to checklists. She describe her writing instruction saying, “And then I made a templet for them of essentially a lot of the pieces that you have in a full essay. And then they fill in the template…” Her early implementation focused little on historical thinking skills and instead, stressed engagement and domain general skills.

As Emily attended more PD and interacted with the group, her implementation changed considerably. She began to treat skills as more fluid and interconnected. Her instruction shifted away from checklists to focus more on student historical thinking. She describe one mini lesson in class as such, “I used one of the documents as a case study on judgements. We went through it via NearPod, spending extra time on the EAD questions.” She began to integrate writing and adopted several high leverage strategies such as classroom discussions, real time student monitoring, and feedback. Her use of technology improved greatly, and she became an expert on multiple effective online platforms. Emily showed high fidelity to the curriculum and used many tools exactly as they had been provided to her. Her final investigation included an engaging warm up question, a classroom discussion, proper scaffolding for historical reading, thinking, and writing, real time monitoring, and meaningful feedback.

Erma: The Trend Setting Teacher

Erma was an exceptional teacher who acclimated quickly to the demands of the pandemic and the new curriculum. She was a thoughtful educator and leveraged her extensive pedagogical knowledge to design tools that other teachers often adopted.

Context. Like the other teachers in our case study, Erma had extensive experience teaching with 14 years total. She was a white female and during the study, Erma was teaching at a new school that opened the same year as the study. The school was a magnet program that focused specifically on information technology. This likely helped to ease the transition to online learning precipitated by COVID-19. Because of its new status, the district had not yet compiled information for the school at the time of the completion of the study; however, Erma, reported that the school was more affluent than many in the district and that the majority of her students identified as white.

Erma described extensive training in teaching historical reading and writing, and she specifically mentioned post graduate course work that supported her teaching practice. She even describe taking training specifically concerning writing in history saying, “I have participated in at least two professional development series on writing in social studies.” She was a leader in her school and had previously served as team lead at a previous school in the district. Although her school was new, Erma seemed to work closely with her team. Erma often describes team member constructing communally used materials, “I think it was Bailey [another teacher at the school] and some of the team members who put together…NearPod interactive slides for the lesson.” Another study participant, not chosen as a case, also worked at her school and similarly reported the closeness of their working relationship.

Despite her comfort teaching historical literacy, Erma still struggled with the contextual changes that came throughout the year. Although she was rapidly able to accommodate the

transition from 90-minute periods of in person instruction to 45 minute online periods, she still had to grapple with the greatly reduced academic time with students. She stated, "these time constraints are the number one problem that are creating a barrier for students to do this kind of thinking [historical thinking]" Though students had access to technology at home, the transition to full time online instruction came with a steep learning curve. Erma found that much of her early instruction was dedicated to familiarizing students with online platforms. She mentioned that the success of the investigation was often contingent on the online tool, "But the Google docs…was much more functional for the students. So that was a big reason that the second investigation went much better." These issued complicated the initial roll out of the curriculum.

Beliefs. Erma's beliefs stand in contrast to those of Emily. Erma had consistently more faith in her students' ability to complete challenging tasks. She stated of her students regarding their ability to complete the writing in the second investigation, "I don't think that will be a challenge for them to manage…if we revisit it and revisit and revisit they will get it." When students failed, Erma often went back to her own materials and teaching strategies to troubleshoot for future instruction. Erma was an exceedingly thoughtful educator and spent considerable time planning to meet the specific needs of her students. Erma thought of herself as a guide to learning rather than a lecturer and believed that learning was the result of interesting topics, opportunities to respond, and effective scaffolding. She said of herself, "I see myself as a 'guide on the side' who is supporting/scaffolding student to engage with content."

Erma's beliefs about the purpose of historical education were similar to those of Emily. She believed that historical thinking skills are exceedingly important to being a productive member of society. She explicitly mentioned the importance of historical literacy skills to becoming a more informed consumer of media. She stated,

> "I think historical literacy skills are the most important thing we can teach students. In our modern, connected world, student have ready access to…so many sources of information…our job as history teachers is to teach students the skills of sourcing, critical thinking, corroboration and research."

She was likely influenced by the political and social turmoil of this time period as she often discussed students being social activists. She believed that integrating students' interests and making explicit connections to the students' own lives and their political and social realities was the most effective way of engaging students and building skills. She described her attempt to gain students attention as such, "I 'hook' students with a short discussion on the legalization of marijuana providing students the opportunity to respond out loud."

Content Knowledge and PCK. Erma showed impressive content knowledge and often showed understanding of the investigation topics that went beyond the curriculum. This was not in itself surprising as other the teachers in my sample similarly were content experts. More impressive was Erma's considerable pedagogical knowledge. Erma brought to the study skills in teaching both domain general and historical reading and writing. She also demonstrated knowledge of historical thinking and prior experience teaching students in these difficult skills. She discussed her previous experience as such, "In the past, I used the 'OPCVL" method (origin, purpose, content, value, limitations) in which I asked students to identify the values and limitations of the origin, purpose and content of their sources."

While other teachers struggled to acclimate to the reduced academic time, Erma quickly recognize the need for a three-day cycle of investigation implementation, demonstrating a strong understanding of pacing and her students' abilities. Erma even showed advanced knowledge of online tools. Even in early investigations, Erma was already identifying and piloting tools that

would be used extensively by all the teachers in the final investigations. For example Erma introduced other teachers to the idea of splitting students to work on documents in small groups, "I had them familiarize themselves with all four documents, but the only selected two to annotate completely in small groups." Erma's previous knowledge of teaching historical literacy allowed her to quickly integrate new knowledge.

Self-efficacy and Value. Erma mentioned on more than one occasion how valuable the curriculum was as a "time-saver" for teachers. Though she enjoyed teaching with primary sources, she admitted how difficult it is to plan for such lessons. Though Erma never discussed her own self-efficacy, her willingness to pilot new ideas demonstrated a high level of confidence.

Implementation. Erma is an expert teacher who arrived at many of the best practices earlier than other teachers. For example, she used timelines effectively for content instruction and integrated discussion to engage students by using interesting questions in just the second investigation. She described a discussion saying, "And we talked about social media and just spent like five minutes at the start of class. Okay, so now do you know what to trust on social media. How do you know what not to trust…" She was an early adopter of cooperative work even when students were on the computers and continued to refine this teaching practice throughout the investigations to help students grapple with historical thinking skills. Erma often used the curriculum tools with near 100% fidelity but made meaningful adaptations to support her specific students. She expertly faded scaffolds such that by the end of the investigations, many of her students were working at an independent level. She often would create multiple breakout rooms and continue to support struggling students while others worked independently, "I always provided a separate breakout room depending on what I'm doing for kids to just work and get things completed."

Erma was particularly strong at supporting diverse students and was one of few teachers who made truly differentiated materials. She would allow students to choose adjusted versions of materials to meet their level. She described her materials saying, "[the materials] are partly in Spanish and partly in English , so that they [ELL students] could read it and follow along," and even provided different spaces for students who needed more or less direct support. Erma worked closely with many of the other teachers in the study and help them to build capacity in meaningful tools such as the Google suite, Jam board, and NearPod. Erma was truly a trend setter and introduced effective practices such as using meaningful feedback and utilizing cooperative work to help students make and evaluate judgments.

Stacey: The Most Improved Teacher

Stacey struggled with the initial implementation. In her eyes, her students had far too great barriers to learning and could not complete the curriculum. Though this mindset was pervasive, Stacey made considerable strides in her final implementation of the curriculum.

Context. Stacey had 12 years of teaching experience and was also a white female. Throughout her career she had worked as a department chair and during the study, she was the team lead for her grade. She had been teaching at her current school for several years and was comfortable with her team, but interestingly, left the school at the conclusion of the study to teach at a different grade level. Stacey's school was one of the most diverse schools with 41% of students identifying as white, 35% of students as African American, 14% of students as Hispanic, 5% of students as Asian, and the rest being more than one ethnicity. Her school struggled academically with only 37% of students reading proficiency on the statewide reading exam and 24% on the statewide math exam.

Though she never explained her exact training, she claimed that she had post-secondary training in teaching historical reading and writing. This seemingly did not prepare her for the challenges of teaching historical literacy remotely because Stacey struggled the most with the transition to online education. The reduction of academic time caused Stacey considerable issues and it was not until teachers began to return to school that Stacey began to correct issues with pacing. As she put it, "This terrible 45 minute class period. I mean twice a week is just, I never before have we felt this way about just getting through the content…" She often expressed frustration with the time constraints and other district wide mandates during the pandemic. Though Stacey experienced much the same contextual shifts and barriers as other teachers, the changes seemed to disrupt her instruction more. Stacey taught both honors and standard classes and held very different expectations for these two groups of students. She described her classes as such, "Honors were definitely able to do it…but for the co-taught it was certainly a struggle."

Beliefs. Stacey's beliefs were dominated by her views on her students. Stacey seemed to believe that her students' capabilities were largely determined by their innate abilities or by their home lives. She stated, "my school is a Title 1 so we just have a lot of parents who are not active and do not participate." She demonstrated on multiple occasions throughout the investigations that she believed her honors students were capable of completing the curriculum while her standard students were not. As she said, "I know that my honors students will do it…but with co-taught…they need so much support and virtually we cannot provide that for them." She stated that scaffolding was important to education but often demonstrated over scaffolding likely tied to the fact that she did not believe that many of her students were capable of the tasks required in the curriculum. She commented that barriers at home such as a lack of resources were a major contributor to student success or failure. She stated explicitly that her approach to education was

to guide students which seemed to contradict the heavy-handed scaffolding and lecturing she often used in her classroom.

By the end of the study, her stated belief on the purpose of historical education was to help students build critical thinking skills but chose not to elaborate on why those skills were relevant. As she said, "The most important aspects of historical thinking are formulating judgements because it helps students think about the credibility of a historical source." In her mind, motivation was a product of interest and external factors such as grades. This was similar to the beliefs of Emily. Also similar to Emily, Stacey often had trouble with engagement.

Content Knowledge and PCK. Stacey was a particularly interesting case because she had a high level of content expertise. She often spent more time than other teachers establishing the context and produced truly impressive background information for students. Unfortunately, this content expertise hid a seemingly shallow pedagogical base. Again, like Emily, Stacey often relied on domain general forms of instruction and, in early investigations, showed little knowledge of how to support students' historical thinking. She described her early attempts at supporting students as such, "We gave then sentence starters. So actually embedded an organizer into your [UMD] packet and we did background information and then we said create your claim."

Despite this, her expertise in supporting student historical literacy grew immensely during the study. By the end of the study, Stacey showed a knowledge base similar to those teachers who had begun the study with much more training. She began to expertly support student historical thinking. She described one interaction with a student that shows her growth, "I said 'can you trust him.' She said, 'well yes and no.'…I said, 'you can certainly say there are reason to trust and reasons not to…you can certainly show us both perspectives." Stacey

developed skills related to online education platforms such as the Google suite and NearPod as well as discussion forums such as Jam board. She also learned how to support historical reading and writing. She began to use real time monitoring and feedback as well as classroom discussion and cooperative work. Ultimately, it is difficult to tell if any of these skills were completely new to Stacey or if she abandoned them during the hectic days of the early pandemic and later adopted them back. For the purpose of the current analysis, Stacey showed impressive growth throughout the semester.

Self-Efficacy and Value. Initially, Stacey seemed to see little value in the curriculum. She focused her instruction on the topics of each investigation but implemented instruction that fundamentally changed important aspects of the curriculum. As the study moved forward, this seemed to change, and Stacey became more diligent about implementing the curriculum. Stacey likely believes herself to be a capable teacher considering her leadership roles, but she never made mention of her true self-efficacy beliefs.

Implementation. As with her growth in PCK, Stacey showed considerable growth in her implementation. Stacey's initial implementation was the most fraught. She struggled with pacing and the first investigation took nearly two and a half weeks while most teachers completed it in less than half the time. As she said, "so for the first investigation we went all in and we spent, I would say about two and a half to three weeks on it." To compensate, she completed the second investigation in only two days. It is also generous to say that she was actually implementing these investigations as the adaptations she made, made to our curriculum nearly unrecognizable and fundamentally undercut core tenants. She relied on her previous experience and most of her instruction supported content knowledge and domain general reading skills. She provided far too many scaffolds and removed them too slowly, never allowing her students space to build

independent skills. She described her scaffolding in the final investigation as such, "so we heavily prompted them in terms of intro, claim, for the supporting paragraph. On, just giving them those sentence starters throughout then using that judgements piece."

Stacey's implementation improved tremendously. She began to adopt the curricular materials and adjusted her pacing accordingly. She began to integrate feedback and discussion as well as focused more narrowly on supporting historical thinking and student writing. She described a discussion as such,

> "It almost became a philosophical discussion of like what was successful [referring to Reagan's war on drugs] and because I had a couple of kids say, well, it was successful because Reagan wanted to punish people for doing drugs…then some students were like, well, that's not success because success would be putting people in treatment."

Admittedly, Stacey, even in later investigations, still did not set particularly high expectations of her students and still provided far too much scaffolding ultimately preventing individual mastery. Despite this, by the final investigation, Stacey used many of the strategies introduced by the curriculum and fellow teachers, and she implemented them with high fidelity. She describes her adoption of one such practices saying, "I took I think it was Rachel's idea or someone in our group shared just this amazing idea of really presenting it [the controversy] as a story…here are your characters, Arbenz and Eisenhower."

Rachel: The Master Educator and Social Justice Activist

Rachel is a master educator and held strong beliefs as to the purpose of history education. These two points taken together outline her implementation which was both unique from the other teachers and consistently stronger.

Context. Rachel was the most seasoned of the case teachers with 25 total years of experience. She was a white female who held several leadership positions throughout her teaching career. Rachel found herself in a rather unique position among the cases as she was moving school districts and the pandemic year was her first year with the district. All of her previous experience had been in an urban district. Because of this transition and the unfortunate timing, Rachel felt as though she had little voice in her new school. She was consistently frustrated by the school's robust team planning apparatus as she did not feel her opinions were being considered. As she said, " So I was in Baltimore before. So I'm kind of bottom of the totem pole…just like my worst nightmare." Rachel would frequently plan separately from the rest of her team because she did not feel the lessons being produced were pedagogically strong. It is worth noting that by the end of the semester, her school had recognized her expertise and she became an integral part of her school's planning and was even added to the district's curriculum writing team.

Rachel taught at an economically and culturally diverse school. Thirty two percent of students identified as white, 26% of students as African American, 36% of students as Hispanic, 3% of students as Asian, and the remaining students identified as two or more ethnicities. Forty three percent of students received free or reduced lunch, and 36% and 18% of students reached proficiency on the statewide test in reading and math respectively. Rachel is the only case teacher that worked at an IB school which offered advanced placement programs.

Rachel had considerable post-secondary training in teaching historical literacy. Though she did not elaborate, she did mention taking multiple classes and workshops. Rachel taught both honors and standard classes. She experienced similar contextual challenges to the other teachers in the study. She described the time restrictions as such, "It's like a scramble to cover as much as

we possibly can. And it's ridiculous…the kids get nothing out of it because we are moving so fast." Despite this, she acclimated rapidly to the changing environment. Changes that plagued other teachers such as reduced academic time, transition to online instruction, and issues with engagement, posed comparatively fewer issues for Rachel.

Beliefs. Rachel held very strong beliefs about the purpose of historical education and was outspoken in her belief that historical education should prepare students to be productive and progressive members of society. As she said, "To me, the most important parts are those that directly affect student's ability to become effective citizens." Rachel believed that connecting the curriculum to students' lives was crucial to promoting motivation. She felt that to properly educate students, instruction must make frequent connections between historical controversies and modern political and social happenings. She said, "I just find it really hard to get kids to do really hard thinking unless they feel like, okay this matters…like in a present sense." She often integrated topics of social activism and systematic injustice into her lessons as central themes.

Like other teachers, she discussed the importance of critical thinking skills. To support students in developing these skills, she saw herself as a facilitator and believed that cooperative work and proper scaffolding are the most effective means of instruction. She insisted that students, "should draw their own conclusions about historical truths," and that, "truth is negotiated," not provided by the teacher. She did acknowledge that students experience various barriers in their home lives, but importantly, did not equate students to the barriers they have faced in the past.

Content Knowledge and PCK. Rachel was uniquely talented and had an extensive tool kit of pedagogical strategies to address the needs of her students. She showed expert content knowledge and strove to provide meaningful and effective content for students. She was

concerned with providing background on both sides of the inquiries. For example she described her thoughts in the second investigation as such,

> "The documents were slated in favor of repeal…I feel like there's a legitimate case to be made that…there was a big drinking problem with a pretty strong link to some ugly stuff happening in families, so I put in that slide."

She began the investigations with an overarching understanding of historical reading, writing, and thinking. She showed her understanding of historical thinking when she described her advice to a student as such, "you got to show your reader that you are a considerate and thought person who's looked at both side. And so that's what the rebuttal shows that you're willing to consider the other side." She quickly adopted meaningful strategies from the curriculum and implemented them using practical and effective online platforms.

Unique about Rachel was her ability for self-evaluation. During coaching sessions, she would often discuss her thoughts about a lesson and make adaptations to improve the lesson for the next implementation. She also showed an unparallel knowledge of differentiation. She differentiated materials by classroom to meet specific student's needs. Her materials would include structured supports for different students. These used scaffolds designed by UMD to support judgments such as prompting questions and statements like, "do you trust this author? Why or why not?" and, "Write why the authors facts and examples are/are not convincing?" Finally, Rachel showed an impressive understanding of historical thinking and created mini lessons to target students' needs and address the most difficult skills. She described one lesson as such, "I shared exemplars of judgments so they could see how judgements could sound sin different content, including ones that were less explicit but still made clear the sources was…part of a historical picture."

Self-Efficacy and Value. Rachel believed the lessons were valuable and she was one of three teachers who completed every investigation, though admittedly some had to be abbreviated. Importantly, Rachel recognized that lessons were only of value if they could be properly adapted for her students and connected to their lives. Though Rachel never explicitly discussed her self-efficacy**,** her readiness to make adaptations to the curriculum suggests that she was confident in her own pedagogical skills.

Implementation. Due in part to Rachel's extensive pedagogical knowledge, her implementation was the most varied and consistently exceptional. More than any other teacher in the study, Rachel was able to adhere to the critical tenants of the curriculum while adapting materials to meet the needs of her students. Unlike other teachers, Rachel showed high fidelity and strong instruction even during the initial hectic days of the early implementation. She adhered closely to the UMD materials saying, "And so I used essentially the guides that you guys have in their, the how 2 write, which I put on the left hand side." This is not to say that her instruction did not become stronger and more nuanced throughout the study.

As the study progressed, Rachel used tools more effectively to support historical literacy. In early investigations, Rachel used discussion to create interest in students while in the later investigations she used discussion to help students make meaningful judgments about primary sources. She described a later discussion remembering a student's comments on the reliability of sources, "Like this one, Arbenz is himself and as he said the speech himself, that could make it a more reputable source if he is telling the truth, but he could be laying making his resignation seem better and more right." While Rachel introduced interesting and topical content in early investigations, in later investigations, she refined her approach and made content instruction highly efficient presenting both sides of an argument equally to engage students in

argumentation. Rachel was an expert at differentiation and often made specialized materials for differing groups of students who needed differential support. Rachel scaffolded students according to need and expertly reduced scaffolds such that by the final investigation many students were working independently. She described her scaffolding during the fourth investigation this way, "I gave them the option of just doing a blank document. But then I also gave them the setup where these prompts…broke it down and they had a little space to write in." Important to her instruction, Rachel made considerable effort to connect instruction to modern political and social movements by engaging students in purposeful discussion.

Steve: The Technical Wizard

Steve showed a unique aptitude for using online systems which benefitted him and his students greatly during the administration of the investigations. Steve showed high fidelity and his use of online platforms made for implementation unlike any other teacher in our study.

Context. Steve had the least experience of any of the case teachers with 11 years of teaching experience. He was the only male teacher. Steve taught at a high achieving high school. His high school had the least racial diversity with 77% of students identifying as white, 5% as African American, 9% as Hispanic, 4% as Asian, and the remaining students as two or more ethnicities. Steve's school was also the most affluent with only 10% of students qualifying for free or reduced lunch. His school also had the highest standardized testing scores.

Steve claimed he was familiar with teaching historical literacy but did not say he had taken any additional training outside of his formal education. He was the team leader for his school and benefitted from a close relationship with the other teachers in the school. He stated that they work closely together to plan all lessons and materials. As he said, "The team that I've

been working with now, we've been working together for a couple of years now…so different writing assignments have kind of been formed collaboratively."

Steve dealt with the same challenges as the other teachers, but these were mitigated in part by his school policy and his students' characteristics. His school held a technology first policy meaning the school had been moving toward integrated technology instruction before the pandemic. Additionally, between his students' academic ability and their access to technology, barriers caused by COVID-19 were slightly less egregious.

Beliefs. Steve believed that the best way to teach students was by allowing students to work with primary documents. He believed in teaching through guidance. As he said, "I like to teach as much as I can through primary sources and document analysis. I want the students to be historians…" This often meant he provided a model then releasing responsibility to his students. He would offer support in the form of feedback only after student had attempted the skill on their own. Steve believed that learning can be achieved through opportunities to respond and interactions with fellow students. Interestingly, he believed that students' home experience could present significant barriers to learning though he made no mention of providing additional scaffolds to students who may experience those barriers. He recognized that his students had certain privileges such as the fact that they, "don't have to have a job to support their family or watch their younger siblings as much as maybe they would have if I taught at Benjamin High [another school in the district]."

Steve believed that the primary purpose of history education is to reinforce the skills introduced by the curriculum such as contextualization, corroboration, and sourcing. He did not provide follow up as to why these skills are important, so it is difficult to tell if these were long held beliefs or ones that he developed during the study. Steve's actions would suggest that he

may have held these beliefs even before the study as even in early interventions he worked to support student historical thinking and writing specifically.

Content Knowledge and PCK. Similar to other teachers, Steve showed a high level of expertise in content knowledge. Though not to the level of Rachel or Erma, Steve showed adequate knowledge of historical literacy that developed considerably through the investigations. He acquired several effective strategies for supporting student historical reading, writing, and thinking. For example Steve used discussion to help student better understand the historical inquiry. He describes on such interaction with students, "Did that have unintended consequences [referring to Reagans war on drugs]? I just kind of changed the idea a little bit. So they're intended or unintended? I made them choose. I explained it slightly and I think a lot of them got it." Steve was seemingly more knowledgeable about supporting writing, or perhaps just more willing to support writing than other teachers in the early investigations. He described his writing support in the fourth investigation as such, "So the writing, we just kind of went over the whole how 2 write guide and tried to discuss what should be in each part of the essay…and we went over the rubric." He did mention that his school had instituted an initiative for teachers to teach writing in history before they began using our curriculum.

What set Steve apart was his extensive knowledge of online platforms. This gave him a significant advantage when instruction transitioned to fully online. Steve was the first teacher to effectively use NearPod and frequently built discussions and cooperative activities into the NearPod. He describes the materials he used in the fourth investigation as such, "I made this more of a self-paced NearPod that they could go through and go back and kind of see it for themselves before they started writing." He also was the first teacher to introduce real time

monitoring through the computer and using feedback. Steve showed flexible implementation, often integrating multiple tools into an easily digestible online platform.

Self-Efficacy and Value. Steve did not discuss his self-efficacy or if he believed the curriculum to be valuable. His actions do, however, give us a hint as to his feelings. Steve often implemented the curriculum with no adaptation other than the use of effective online platforms. This would suggest that he believed the curriculum to be sufficient to support his students' growth in historical reading and writing. Steve was confident in his use of technology but made no mention of his overall efficacy for teaching historical literacy.

Implementation. Steve taught content using the curriculum as well as an impressive array of multimedia tools. Though his initial investigations were largely just modeling and independent work, Steve began to use cooperative work as a means of scaffolding in later investigations. Steve described this cooperative work in the final investigation, "I decided to do a collaborative activity…I gave them time to go through all four documents…And then I had them do this little Socratic seminar." Steve dedicated more time than most teachers to writing instruction and integrated student work examples when modeling. Steve was an early adopter of monitoring and feedback and as early as the second investigation, was using tools on Google suite to monitor student work in real time and provide extensive feedback. As Steve described, "I gave comments on the document itself…and asked them to let me know if they had further questions. I did that as they were working."

In later investigations, Steve began to diversify his instruction. He had students evaluate other student's work as well as their own. He began to integrate self-regulation and even had student set their own writing goals. He described his use of self-evaluation as such, "I kind of asked them to look through it [his feedback] and the how 2 write guide…Does this have all the

components? So it's really trying to get them to evaluate their own writing." In later investigations, Steve used cooperative work as a means of helping students interact and develop historical thinking skills. Steve would often build elaborate NearPods that would integrate curricular supports directly into documents as a way of supporting historical reading, thinking, and writing. Steve released scaffolds earlier than other teachers, however, due in part to his students' familiarity with technology (and printers at home) by his report, his students became increasingly independent in ways that we had hoped.

Discussion

This study's central purpose was to observe teacher fidelity, document how teachers make curricular adaptations, and analyze how important factors such as context, beliefs, and PCK affect implementation. Fidelity of implementation is often one of the expressed goals of intervention studies, however, strict implementation is unwise. Students have different educational needs, and all curricula should be modified in certain ways to best fit specific student populations. Further, not all teachers will receive information in the same way. Important factors within the teacher dictate practice, so better understanding these factors is an important step toward effective curriculum implementation.

Conceptualizing the influences on teacher decision making can open up important avenues for research and practice. Information on contextual factors can enlighten researchers and practitioners on the greatest barriers teachers face in their day-to-day practice. Registering teachers' beliefs can help identify specific ways that teachers engage with information and interpret what inputs are and are not applicable. Observing teacher PCK can help to explain skills that teachers commonly have and ones that may need to be taught in the pursuit of effective

instruction. High fidelity and effective implementation are lofty goals, and the examination of how teachers make decisions is a hugely important undertaking.

In this section, I will discuss four explanations I identified through cross case analysis. I will then take time to explain how PD can better support teachers in the pursuit of high fidelity and effective adapted implementation. Finally, I will discuss the limitations of the study, implications for practice, and future research.

Teacher Fidelity and Implementation: Cross-Case Analysis

A cross case analysis of participants can help to reveal trends about fidelity, how teachers made adaptations across the board, and how critical factors influenced implementation. Four major cross-case explanations emerged from the five participants: (a) teacher implementation fidelity improved over time as did their understanding of curricular goals, (b) teacher PCK exerted considerable control over the tools that teachers used for implementing the curriculum and the adaptations that they made to the curriculum, (c) teachers' beliefs about their students influenced how they used and faded scaffolds, and (d) with little exception, teachers' beliefs about the purposes of history education were well aligned to the underlying intent of UMD's historical reading and writing curriculum.

Teacher Fidelity. In all five cases, teachers showed meaningful improvement in fidelity from the beginning to the end of the study though teachers with more room to improve showed considerably more growth. Emily and Stacey both began the investigations providing instruction that strayed considerably from the overarching goals and strategies used in the curriculum. In fact, for the first and second investigations, Stacey made several implementation choices that fundamentally undercut the core tenants and goals of the curriculum. Emily lamented these early implementation woes saying, "[I'm] not confident - I teach them to write how I do, but not sure

that that's the "right" way." Erma, Rachel, and Steve more quickly understood the goals of the curriculum but did not yet have access to the full range of pedagogical skills that they would by the end of the study. As a result, their fidelity was somewhat limited initially. In the early investigations, teachers used little discussion or collaborative work, and favored teaching content over historical thinking skills. Teachers struggled with pacing, and often offered little support for historical reading and writing. Emily described her early reading lessons as such:

> "We did like a close reading of the one we did together. We did. Actually, maybe I want to say I read through both of them *Souls of Black Folk* and Atlanta compromise. I think maybe I did read through both of them with them. Yeah, we sort of did a close reading together about them."

This changed over the course of the study as teachers attended more PD. PD provided activities for teachers to practice what they were teaching their students. Teacher were given explanations of appropriate use of scaffolds and why it was important to fade them out (to promote student independence). By the final two investigations, all teachers showed an advanced understanding of the curricular goals. As teachers developed more extensive and nuanced PCK, they were able to face each problem with multiple and varied framing tools. Thus, they were able to create instruction that was both effective for the setting and sensitive to the curricular goals.

By the final investigations, the teachers had adopted highly effective teaching techniques and implemented the curriculum with high fidelity. These included a three-day lesson cycle, nuanced support of historical thinking through the use of cooperative activities, and effective writing support using real time monitoring and extensive feedback. Steve described his final writing lesson as such:

> "And then we walked through how 2 write [scaffold]. So, because I told them there's less hand-holding this time, but how you can have your hand held is to have your how 2 write up while you're doing it. It lays it out sentence by sentence, guys. So, we just walked through that again and all the lead-ins for the... that you guys use on the second page for the judgments, I think we're really helpful to the kids. So, I gave them... their challenges this time, which isn't, again, something I got from the webinar was having them cite the document and not the letter."

The increase in teacher fidelity is likely the direct result of two important factors: teacher adjustment to the contextual demands of the COVID-19 pandemic and effective PD. COVID-19 caused fully online instruction with greatly reduced time for social studies (two 60 min lessons per week). Teachers struggled immensely with these shifts which undoubtedly impacted their early implementation. It seems likely that as teachers learned to teach in an online environment (through Google suite tools and Nearpod) they could more readily attend to the district's intent for them to use the curriculum and teach historical reading and writing.

In addition, UMD provided PD that explained (and modeled) how cognitive apprenticeships facilitate student learning. We explained foundational concepts in teaching historical reading and writing, then modeled how teachers could use heuristics (IREAD and H2W) and make their thinking visible to students. In addition, teachers were given the opportunity to engage collaborative discussions with other teachers about the curriculum and instructional decision making (Darling-Hammond et al., 2017: Marx et al., 2020). Moreover, teachers were shown good and weak student work examples and UMD (and participating teachers) collaboratively evaluated students' attempts at corroboration, contextualization, and sourcing. This led to further instructional planning (Monte Sano et al., 2017). Finally, the PD

was responsive to the needs of the teachers (Coburn and Penuel, 2016) and was aligned to address the specific contextual challenges facing instruction during the COVID-19 pandemic. Our current study corroborates findings from Monte-Sano et al., (2014; 2017) that show how effective PD can be used to improve teacher instructional skills for historical literacy.

Teacher PCK, Adaptations, and Implementation. Though teacher PCK became more standardized as the study progressed, initially teachers held a wide range of pedagogical skills including skills for supporting domain general reading and writing, skills for administering online education, and skills for supporting historical literacy including historical thinking. Stacey and Emily showed the least developed PCK within the bounds of the current study, while Erma and Rachel benefitted from extensive pedagogical knowledge for teaching historical literacy, and Steve benefitted from impressive technological skills.

Those teachers, namely Rachel and Erma, who had previous experience teaching historical literacy more quickly understood the goals of the curriculum and more readily adopted the tools. These teachers also were able to make meaningful adaptations to support student understanding. Rachel for example created dedicated mini lessons to support difficult historical thinking skills. She describes the support stating:

> "So, I did a mini lesson on judgments because I was like, 'This is the thing that they are just having such a hard time with.'… And so, we talked through that a good judgment does these things. But then I took them back to the Guatemala essay and said, "Hey, remember what that essay was about? Let's try to evaluate some of the judgments that students made…"

Teachers with less experience, Stacey, and Emily, relied on more familiar forms of teaching such as close reading and using CER, and adopted the curricular materials in later investigations only after attending PD.

Steve had developed a comprehensive knowledge of technological platforms and had integrated these even before the transition to distance learning. He quickly acclimated and began to effectively use NearPod to integrate presentations and scaffolds. Steve described blending physical documents in NearPod:

> "So, the first one I used to the draw it out on their part on Nearpod. What I did was just a screenshot from a word document I used, and I saved this picture. So, I just made that picture the background. And then they were able to write on it and do whatever they want to [it] and the next one. Same deal with the background. And then there is a box to attach a PDF to the directions."

He was also the first teacher to use online platforms for real time monitoring and effective largescale feedback. All the other teachers followed his lead and later adopted many of the effective online tools.

The connection between PCK and implementation is not particularly surprising. Fives and Buehl (2012) acknowledge the importance of PCK as a framing device to help teachers understand specific problems in practice and develop comprehensive plans to address those problems. As teachers developed new and useful pedagogical tools, they could more effectively adhere to the curriculum and support the specific needs of their students. What is particularly promising is the effectiveness of the PD that resulted in the significant changes to PCK.

The PD integrated core components that are widely recognized as effective practices. It was content focused, incorporated active learning, included collaborative work, and used

effective models for practice (Darling-Hammond et al., 2017). The PD was designed to highlight representations, decompositions, and approximations to explain the specific underlying pedagogical practices of the curriculum (Grossman, 2009). Representation of practice provided teachers with examples of expert teaching practice to reveal hidden components of effective implementation. Decomposition of practice allowed teachers to identify the component parts of effective practice. Finally, approximations of practice gave teachers the opportunity to simulate meaningful practice and rehearse, gather feedback, reflect, and improve. Taken together, these three key processes allowed teachers to interact with complex pedagogies and develop needed highly effective skills (Ball & Forzanu, 2009)

Teachers also received continued coaching throughout the year likely contributing to the overall effectiveness of the PD. The coaching sessions served as a platform for individualized coaching and support as well as a convenient means of collecting vital information for the study. During the coaching sessions, the UMD team often helped teachers to identify and address specific gaps in pedagogical knowledge. The UMD team worked with teachers individually to address concerns within their classrooms and build capacity to address their specific student populations. Continued coaching is integral to any effective PD series (Marx et al., 2020). In fact, large school districts such as the District of Columbia Public Schools have established continued coaching as a cornerstone for teacher support and development (DCPS, 2016). The PD series included coaching which helped teachers to apply skills gained during PD to their classrooms.

Teachers' Beliefs and Scaffolding. Interestingly, by the time I had recorded teachers' beliefs at the end of the study, all of the case teachers held similar beliefs as to the overall purpose of historical education. This was in contrast to the teachers' beliefs about their specific

students' abilities. Teachers held varying beliefs as to the impact of students' innate abilities, how students home lives effect of their education, and if students could meaningfully interact with the curriculum. Stacey and Erma both felt more strongly that some of their students' innate abilities would prevent them from accessing the curriculum. Erma and Rachel both mentioned that students often had barriers to education that came from home but did not seem to believe that this prevented any students from accessing materials. Steve seemed to hold a similar mindset to Stacey and Erma, but perhaps because of his high achieving students, he rarely mentioned that his students struggled throughout the study.

The teachers' beliefs about student deficits seemed to dictate teacher implementation to a certain extent. Teachers who indicated that student deficits were insurmountable used more scaffolds and were less likely to remove scaffolds as the students moved through investigations. Stacey was the best example of this type of belief and implementation. She taught both honors and standard, and there was a stark contrast in the difference in implementation. She believed her standard students were often not capable of the tasks. As a result, even as late as the final investigation, she provided highly scaffolded lessons and never required students to write complete essays. As she described it:

> "My co-taught students did not write a five-paragraph essay. They wrote a very highly scaffolded model of four kinds of chunked boxes where they did an introduction, they did two supporting paragraphs and a conclusion. Again, highly scaffolded. So, they are on the complete other end of the spectrum. Nowhere near expecting them to write that five paragraph given the topic on this one."

Conversely, teachers who seemed to believe their students more capable, such as Rachel and Erma, were more likely to reduce meaningful scaffolds over time and allow students to strive

towards mastery. Steve described his effort to reduce scaffolds and release responsibility for analyzing documents stating, “For standard and co-taught, I modeled the first two documents with them, and then I gave them time to do the third and fourth through Nearpod on their own.” Beliefs acted as an important filter for teachers. Those teachers who saw their students as less capable were less likely to deem challenging tasks as relevant and achievable.

Teachers’ beliefs about students and the subsequent implementation can perhaps be explained by the contrast between growth and fixed mindsets. To simplify, a growth mindset is a belief that one can change through effort and perseverance, while a fixed mindset is a belief that innate abilities determine the outcome irrespective of effort or other factors (Dweck, 2016). Teachers such as Stacey and Emily may have had a fixed mindset about their students believing that their innate abilities precluded them from making meaningful growth. This resulted in over scaffolding and ultimately, little independence. Rachel and Erma in contrast likely held a growth mindset of their students and believed that with effort and proper instruction, even students who faced difficult educational barriers could access and achieve within the curriculum. This likely contributed to their meaningful use and reduction of scaffolds pushing students towards mastery.

Consistency in Positive Teachers’ Beliefs. As mentioned, by the end of the study, most teachers reported having similar beliefs about the purpose of historical education. All agreed that the primary goal was to teach students to be critical thinkers. Some teachers such as Emily and Rachel used the more domain general term, critical thinking, while others such as Steve and Stacey directly referenced the historical thinking skills introduced by the curriculum: contextualizing, corroborating, and sourcing. Emily, Erma, and Rachel took this view a step further and asserted the importance of these skills in students' lives specifically when dealing with media bias. As Rachel put it:

> "In our modern, connected world, students have ready access to historic fact and SO MANY sources of information . . . we DO NOT need to be another source of information . . . our job as history teachers is to teach students the skills of sourcing, critical thinking, corroboration & research."

Rachel specifically identified the importance of historical education for teaching social justice. She said:

> "In US History: to know that the US was founded on deeply troubling practices AND remarkably powerful ideas; to know that our political and economic structures have favored some groups over others, not accidentally, and that democracy, if used effectively, is a powerful tool for righting those wrongs."

Interestingly, these beliefs did seem to have direct effects on her practice as much of her content building time was devoted to discussing how historical controversies connect with modern controversies. She, as was the case with Emily and Erma, specifically believed that skills learned in history should be used to promote social activism today. These teachers all promoted discussion as a means of connecting content to modern controversies as early as the second investigation.

It is worth noting that the congruency of beliefs around the purpose of history education may have come as a result of the PD. Teachers' beliefs were recorded at the end of the study after they had experienced PD that taught specifically about the importance of historical thinking. By this point in the investigations, most teachers had adopted similar and consistent practices, many espoused by the PD. This is in contrast to the beginning of the investigations when many teachers used a wide variety of teaching techniques. Teachers may have held different beliefs before PD that changed and became more consistent as did their

implementation. This suggests that effective PD may influence teachers' beliefs as well as their PCK. This finding corroborates work by Lee et al. (2021c) in which the authors implemented a cognitive apprenticeship model in science and observed the changing beliefs of two middle school science teachers. The authors found that after attending PD and administering the model, both teachers' beliefs about the function of writing changed. Both teachers initially believed writing to be a tool for assessment and communication; however, after PD, both teachers acknowledged the key role writing can play in developing scientific reasoning. Lee and colleagues expanded on Guskey's (1986) model of teachers' beliefs and proposed that PD leads to change in practice which leads to change in student outcomes and subsequent change in teachers' beliefs. Monte-Sano et al. (2017) demonstrated that effective PD could help teachers to focus more on key aspects of historical writing, pay more attention to evaluating student historical thinking, consider the quality of student work (not just work completion), and identify students' needs. This study shows the wide ranging effects of meaningful PD.

Professional Development

Importantly and practically, this study informs how PD helps to change factors that affect teacher practice as well as how PD can change fidelity and teacher adaptations over time. I identified four cross-case explanations that speak to our research question: (a) teacher implementation fidelity improved dramatically over time as did their understanding of curricular goals, (b) teacher PCK exerted considerable control over the tools that teachers used for implementing the curriculum and the adaptations that they made to the curriculum, (c) teachers' beliefs about their students dictated how they used and faded scaffolds, and (d) teachers' beliefs about historical education were largely uniform and reflected the overall views of the study.

Each of these findings exemplifies the need for effective PD. I found that through effective PD, even in a fully online space, teachers can be taught to administer interventions with fidelity (even when initially they show reluctance), teachers can improve their PCK, and teachers can change their beliefs. I also found that factors such as beliefs and PCK likely dictate implementation, so if PD can adjust these factors, it can effectively impact teacher practice. This study provides new findings because PD was fully online suggesting that online PD can be an important tool in future studies.

The effect of the PD series was not surprising. As discussed other studies that used cognitive apprenticeships have documented the meaningful effects of PD on beliefs and PCK (Lee et al, 2021; Monte-Sano et al., 2017). More interesting perhaps is the structure of this specific PD series. This study distinguishes itself from others in that it occurred during the COVID-19 pandemic and all PD was administered online. This included the webinar series and the coaching sessions. Though our team followed the same models as Monte-Sano et al. (2017) in designing the PD series a number of critical changes were necessary owing to the online instruction. Important factors of effective PD such as opportunities to collaborate and discuss were fundamentally shifted and our team had to make several creative decisions regarding how to adhere to the core principles of effective PD. These included the use of communal folders for sharing adaptations, the use of online discussion platforms, and the incorporation of online whole group curated discussions.

Additionally. It seems likely that in the online space coaching sessions may have been more important. The expressed purpose of coaching is often to help teachers apply more broad learning to their specific contexts and students (Marx et al., 2020). In this study, teacher's classrooms looked very different depending on their students (specifically their access to

technology), their schools and the instructional decisions made by principals and department leads, and their own expertise. Coaching sessions became critical as teachers attempted to navigate this new environment. Our team worked with teachers to understand the supports, make adaptations, and meet the needs of their students.

Implications for Practice

I employed a case study design, so it is worth considering that all outcomes for the current study were analyzed within a highly contextualized environment. Thus, extrapolation cannot be done without overextending the data. Instead, I provide suggestions for researchers to better understand teacher implementation and how PD can be used to build teacher capacity and potentially change important factors affecting teacher practice.

Perhaps the most important outcome of this study for future practice is the applicability of a fully online PD. As discussed, due in part to the COVID-19 pandemic and the large number of teachers initially enrolled, teacher PD was fully online. Teachers attended seven online PD sessions coupled with continued coaching to help teachers implement the curriculum. This is in contrast to similar studies such as De La Paz et al. (2014) and De La Paz et al. (2017) which used extensive in person training. The less intensive online PD was still successful in improving teacher PCK for implementing the cognitive apprenticeship and potentially improving student historical literacy outcomes (albeit with a small number of teachers). Our study demonstrates that by using effective PD strategies and coaching, online training may be used to promote similar teacher and student learning outcomes. This must however be understood within the context of the study which included highly experienced teachers. More research, with a greater number of participants, is required to confirm the validity of a fully online PD model for improving teacher capacity in historical literacy.

The practicality of the online PD seems to be predicated on a few factors. First, that the PD retains evidence based practices despite being online. These include that the PD be content focused, incorporate active learning, include collaborative work, and use effective models for practice (Darling-Hammond et al., 2017). Many of these practices, however, are employed during in person PD as well. What perhaps makes the online PD more functional is the ease by which teachers can receive continued coaching. Coaching is an effective practice for ensuring teacher instructional change (Marx et a., 2018) and was an important tenant of the PD in this study. During the study, coaching was completed during the coaching sessions showing the flexibility of online PD. Instead of having to gather all teachers in one place, or finding time to meet in person, online PD allows for specific and flexible continued support. It is perhaps worth noting that the most effective PD may be a combination of both in person and online. For larger or more hands on sessions of PD, in person would be the priority. However, as teachers move into their classrooms and their schedules become more hectic, online coaching may become an effective means of continued support.

Limitations

Though the findings of this study represent a significant contribution to the literature, there are certain limitations to consider. First, our cases, though diverse, may not have been representative of the district population of teachers. This is a limitation inherent to case study design but was also of particular concern for our study. The original sample of teachers was over 70, however, because participation in PD was voluntary, this number quickly fell to just nine teachers. From these teachers, I chose the most representative five. All of the teachers had over 10 years of experience and all, but one was currently or had at some point been in a leadership

position such as a team lead or department head. This suggests that our population was likely all strong teachers compared to the general pool.

Second, I was only able to measure teachers' beliefs at the end of the investigations. Initially, this did not seem like a concern; however, as I began to realize that teachers were showing significant changes throughout the investigations, I realized how valuable at least one other measure of teachers' beliefs would have been. Although I was able to document teachers' beliefs through coaching sessions and surveys, I may have been able to better understand changing beliefs if I had a baseline.

Finally, restrictions due to COVID-19 prevented the entirety of our data collection. Although I was able to collect multiple forms of data to ensure triangulation, I was not able to collect the full breadth of what I had planned. Most notably, I was not able to observe implementation in the classroom and instead, had to rely on teacher reporting and materials.

Future Directions

This study provides a meaningful launch point for future research. First, future research could fill in some of the limitations of my current study. I recommend that researchers continue to observe teacher implementation and beliefs but use more diverse teacher samples and employ mixed methods techniques that also consider student outcomes. Ultimately, I would like for this research to leave the realm of purely qualitative and begin to integrate mixed methods to observe how implementation affects student outcomes. Measuring teachers' beliefs at different times in an investigation would provide meaningful data on how teachers' beliefs change and interact with PCK and other factors to influence implementation.

More expansively this study opens a series of important questions regarding PD. Namely, can PD be effective with reduced dosage, and is online PD a viable and effective method? This

study would suggest that with relatively little PD (only 7 total hour-long sessions and up to 3 more hours of coaching during the coaching sessions) that were fully online, teachers were able to learn the necessary skills to effectively implement the curriculum with fidelity. Future research is certainly needed to explore these outcomes. New and struggling teachers would be a meaningful subpopulation to examine using the parameters of this current study. Additionally, a critical component of PD was continued coaching. I do not know how effective the online PD would have been with a larger population of teachers in which continued coaching may not be possible. Finally, it is worth noting that the UMD team is currently exploring a "teach the teacher" model in which researchers instruct a series of teacher leaders to bring the investigations to other teachers in the district. If successful, this model could confirm the long term viability of the intervention in a district.

To conclude, I believe that this study makes a significant contribution to the field. It helps to map the decision-making process used by teachers to implement instruction. It confirms that within our sample, a fully online PD can effectively change teacher PCK, beliefs, and curriculum implementation. Though there are important limitations to this study, it opens the door to exciting new research.

Tables

Table 1: Demographic Information

Teacher	Years Teaching	Historical Literacy Training	School Demographic Information					Gender		Free and Reduced Lunch	Statewide Test Scores	
			White	African American	Hispanic	Asian	Other	M	F		Reading	Math
Emily	12	No	75	10	9	1	6	50	50	22	46	65
Erma	14	Yes	NA									
Stacey	14	Yes	49	26	14	5	6	52	48	44	24	37
Rachel	25	Yes	32	26	36	3	3	48	52	43	36	18
Steve	11	Yes	77	5	8	4	5	51	49	10	54	66

Note: Stacey was teaching at a newly opened school in the district so demographic information had not yet been published.

Table 2: Data Collection by Investigation

Investigation	Investigation 1	Investigation 2	Investigation 3	Investigation 4	Investigation 5	Investigation 6
PD Date	October 8	November 23	January 6	February 22	March 14	May 17
Data Collection Window	September-December 2020		January-March 2021		January-March 2021	
Coaching session		Coaching session 1		Coaching session 2		Coaching session 3
Survey	Teacher Demographics Survey			Implementation Survey 1	Implementation Survey 2	Implementation Survey 3
						Teachers' Beliefs Survey
Artifacts	Artifacts INV 1	Artifacts INV 2	Artifacts INV 3	Artifacts INV 4	Artifacts INV 5	Artifacts INV 6

Table 3: Open Coding Context and Factors Coaching sessions 1

First Coding (INT 1)		Reduced Codes (INT 1)	Final Codes (INT 1)
"Team planning system"	new students	School/district factors	School/district factors Teacher/classroom characteristics Student characteristics Student ability Student participation/engagement Tech issues/barriers Time restrictions/Coverage Other Obstacles classroom application
district requirements	school requirements		
new job	administration		
new school			
co-taught	Spanish ability	Teacher/classroom characteristics	
previous experience	team lead		
previous writing experience	training		
black and Latino	compatible students	Student characteristics	
demographics	"Title one"		
"Housing project"			
"High performance"	student ability	Student ability	
"Hot shot"	struggling students		
honors	student ability		
honors	student ability		
IB	student level		
standard			
attendance	participation	Student participation/engagement	
completion	stamina		
engagement	student engagement		
participation			
computer barriers	tech issues	Tech issues/barriers	
tech barrier	tech		
tech concerns	virtual environment		
time constraints	time restrictions	Time restrictions/Coverage	
time content coverage	content coverage		

at homework	limited home resources	Other Obstacles
community issue	no physical docs	
benefits of virtual learning	pacing	classroom application
extended time	writing requirements	

Table 4: Full Code Book

Context and Factors	Beliefs	Planning, Content Knowlesge, and PCK	Self-Efficacy and Value	Implementation
School/District	Barriers to Learning	Content Knowledge	Self-Efficacy	Content Knowledge
Student Ability	Goals/ Expectations	Discusion and coperative work	Value	Discusion and coperative work
Engagment/ Participation	Historical Thinking	Examples/ Feedback		Examples/ Feedback
Teacher/ Classroom	Scaffolding	Grading		Grading
Time Restriction/ Content Coverage	Srudent Ability	Independent Work		Independent Work
	Student Motivation	Instructional Materials		Lesson Planning
	Teaching Approach	Lesson Planning		Scaffolding
		Scaffolding		Teaching Historical Thinking
		Teaching Historical Thinking		Teaching Reading
		Teaching Reading		Teaching Writing
		Teaching Writing		Technocal Knowledge
		Technocal Knowledge		

Table 5: Level 3:Themes Memo Writing

Teacher Themes	
Code	Memo
Context	Pittman faces similar contextual barriers to other teachers such as the short periods, distanced learning, and the change in students midway. She teaches at an IB school meaning that here students are stronger though she still has a high proportion of ESL students and mention certain racial disparities. Pittman has high degrees of participation likely due to her strong teaching abilities and level of experience. She does not seem too often agree with her cooperating team and is new to this school doing much of her previous work in Baltimore.
Beliefs	Pittman sees her role as a facilitator of historical investigations. She often works in social activism and social justice as a central theme. This corroborates her belief that in order to engage student you need to connect content to their lives. She acknowledges certain home barriers but focuses more on the difficulty of certain skills and how students can be properly taught. Like other teachers she discusses the importance of critical thinking skills.
CK and PCK	Pittman is a master teacher. She has an expert understanding of history and numerous pedagogical skills. Early in the investigations Pittman understood the goals of the cognitive apprenticeship and made adjustments accordingly. Pittman in many ways was the ideal teacher because she grasped the core components but was not afraid to make adjustments when it was merited. Her understanding of scaffolding and reducing scaffolds as well as providing writing support was unparalleled and she was able to support historical thinking, read, and writing by providing meaning explicit instruction and scaffolding.
SE and Value	Pittman has a high degree of self-efficacy which seems to be well earned as she has a number of years of experience and has taken multiple courses to improve her skills. Though she talks little about her thoughts on the curriculum, as evidenced by her high fidelity it is likely that she sees value in the tools and supports
Implementation	It is difficult to encapsulate Pittman's implementation because she is an extremely thoughtful educator often making multiple adaptation to support specific student needs. Of particular note however Pittman often worked hard to make both sides of the argument equal before argumentation adding targeted content. She built useful scaffolds and was one of few teachers who used true differentiation often building multiple forms of the same tool at differing levels. She employed effective mini lessons that targeted specific problem areas. She supported writing through physical supports and had student review previous work presenting grows and grows. Present

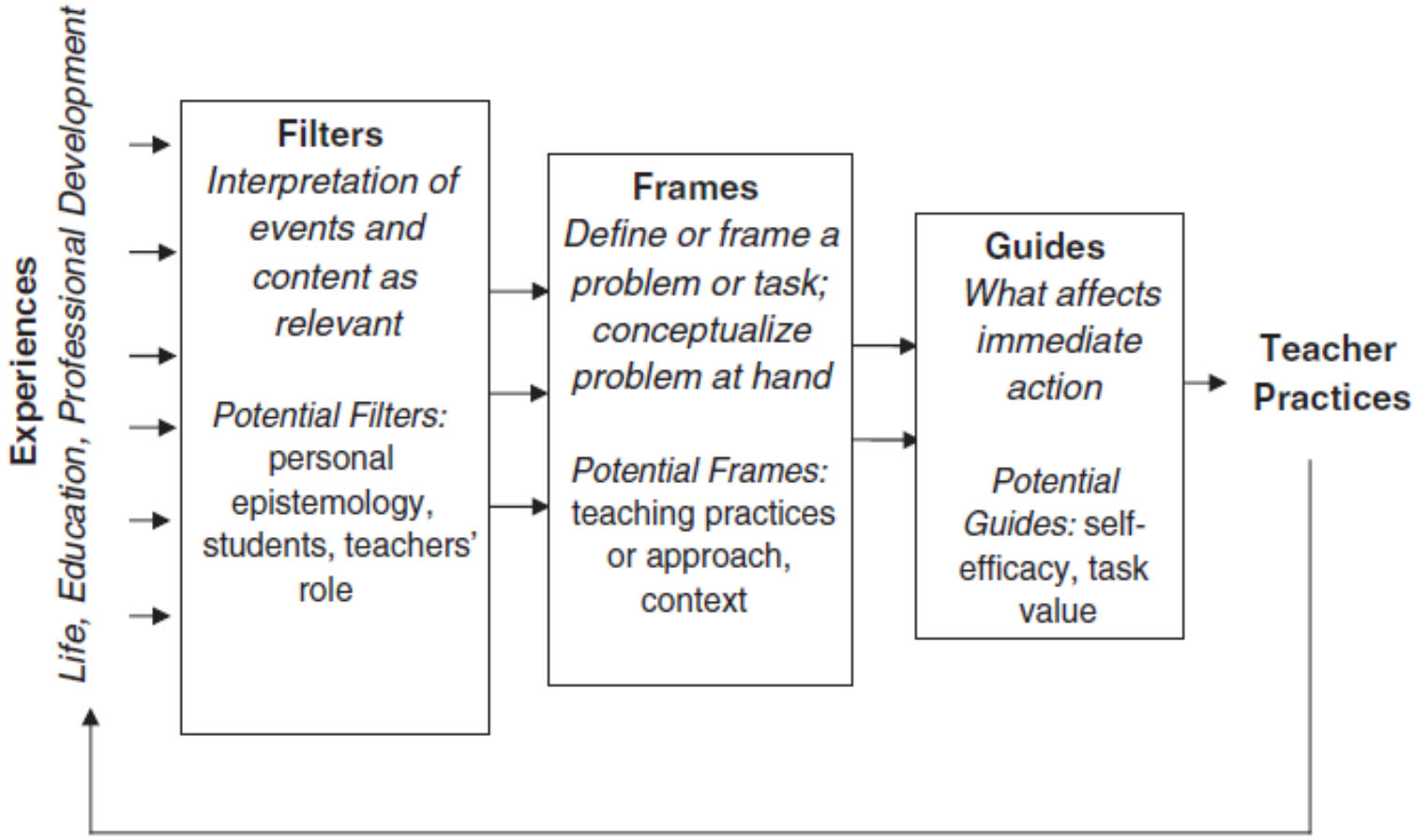
	in her mind at all times was access and social activism so she often integrated supports into materials and used discussion and cooperative work to connect investigations to modern controversies.

Table 6: Data Analysis Map

6. Theory Confirmation	I use the themes and interalatted explinations to support but the assertaions proposed within the conceptual framework about teacher implementation and the effects of PD.				
5. Interrelating Explinations	Teacher implementation fidelity improved over time as did their understanding of the curriculum.	Teacher PCK influenced what tools they used.	Beliefs about student abilities influenced how teachers used scaffolding.	Beliefs about the purpose of history educations were largely uniform.	
4. Testing Themes	The final coaching session and the final survey were used as additional data as well as tests of validity. The completed code book was check against the new data sources and no new codes were identified suggesting that the codes were valid. A similar teachnique confirmed the validity of the categories.				
3. Themes	Themes were created through memo writing by observing the common themes within all of the coded data.				
2. Categories	Context and Factors	Beliefs	Planning, Content Knowledge and PCK	Self-Efficacy and Value	Implementation
1. Open Codes	Compiled based on initial coaching sessions, surveys, and materials.				

Figures

Figure 1: Five and Buehl Implementation Model



Appendices

Appendix 1: Example Coaching session Protocol

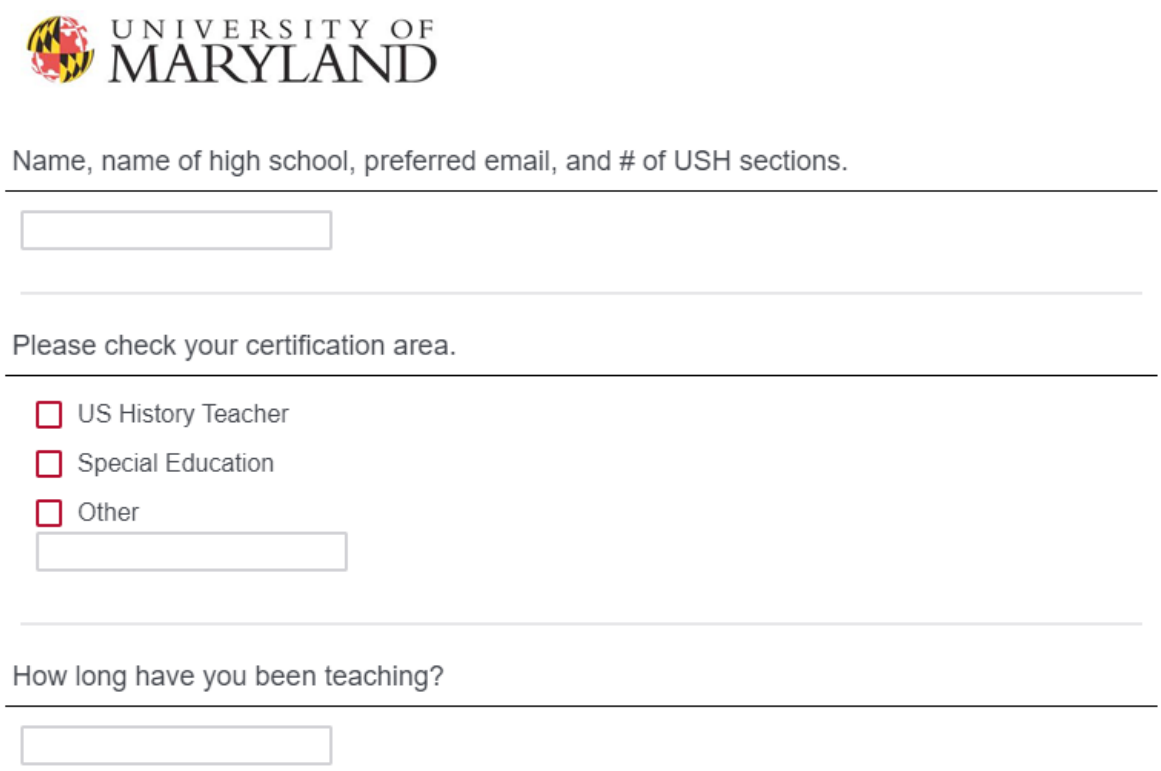
INV 3.4 Coaching session Protocol

1) Which investigation did your team use (INV 3, INV 4, Both)?
2) Could you briefly describe how the investigation went?
3) How many instructional days did you use for the investigation?
    a. Where you able to complete the tasks, you wanted to in that amount of time?
    b. Can you describe what you did for each day of instruction?
4) What choices did you make in using the materials we provided?
5) What materials did you make, to teach the lesson?
6) Did you attempt to model with students? If so, what did you do and how did it go?
7) How many documents did you use with the students?
8) How many (or what %) students attended the lesson?
9) How many (or what %) participated during your lesson? Did more participate than before?
10) How did you monitor student engagement?
11) What homework did you assign, if any? (Why did you decide this?)
12) About how many students analyzed documents using all of IREAD?
13) Were students able to understand reliability?
14) How did you help students to understand reliability?
15) Were any students or, about how many students able to complete the entire essay?
16) What was the quality of the students' annotations?
17) What was the quality of the students' writing?
18) Where students able to make meaningful judgements?
19) How did you support students to make judgements?
20) Did students choose evidence that supported their claims?
21) Did students compose paragraphs that follow H2w?
22) How did you approach grading?
23) What went well? Or, were there any successes? What are your goals for next time?

Other Writing Activities (time permitting)

1) Did you have your students complete any other writing activities?
2) Could you briefly describe the activity?
3) Did you use any of the materials provided by UMD?

Appendix 2: Teacher Demographic Survey



How long have you been teaching US history?

What other subjects have you taught and how long?

Are you a co-teacher USH this year?

- Yes
- No

How confident are you teaching students to read primary sources? Explain.

Are you a co-teacher USH this year?

- Yes
- No

How confident are you teaching students to read primary sources? Explain.

What type of training (if any) have you had for teaching writing in social studies?

How confident are you in teaching students to write historical arguments?

Are you a team lead at your school? Would you be willing to discuss how this program is working at your school, and then communicate with UMD?

Would you like more specific grading rubrics for each investigation?

What other supports can UMD provide, to help you with this program?

Please share additional ideas or concerns. Let us know if you would like other videos or supports.

Appendix 3: Implementation Survey

Name

High School Name

Did you teach any historical reading or writing lessons after finishing Investigation 2? (you may choose multiple activities)

- [ ] Investigation 3
- [ ] UMD writing lesson (from Webinar 3)
- [ ] Other historical reading or historical writing activity (please describe)
- [ ] Did not teach any historical reading or writing lessons

What did you do and how did that go? If you did not teach any of the above, explain why.

We discussed the following student sample (Inv. 3 student example) in the webinar. What do you notice about the student's judgments? Summarize what you observe, focusing on the highlighted text.

Explain what you might say or do if this student was in your class, in Investigation 4.

Write one new idea that you learned after participating in the webinar.

Write one question you have about helping students improve their judgements (especially related to the context or the quality of authors' facts and examples, or about Investigation 4).

Consider your students. Choose one from Investigation 2 or 3 who wrote at least 3 paragraphs.
Describe the class (e.g., honors, standard) and any general characteristics (e.g., motivated, good reader, understands historical controversies, etc.) that led you to choose this student.

Upload de-identified student sample.

Drop files or click here to upload

OR Paste his or her work from google document to this box.

What goal do you have for this student? How will you guide this student in meeting this goal?

Write one goal that you have for yourself during Investigation 4. What supports do you need to accomplish this goal?

Appendix 4: Beliefs Survey

How would you describe your approach to teaching history? Please describe your teaching practices.

In your opinion what are the most significant challenges that you face when teaching history?

What do you think is most important for students to learn in history?

What do you find to be the most significant challenges that students face when learning history?

Please define what historical thinking means to you.

Now that you gave this definition, what aspects of historical thinking are most important for students to learn (and why?)

How do you work to support students' growth in historical thinking skills?

After attending the PD, speaking with us during interviews and gaining experience with the UMD materials this year, we'd like to ask you about what you now understand as the goals, challenges, and important instructional steps for each investigation. Therefore, we ask the following questions. For next year, please share your plans for implementing each investigation, if circumstances for teaching are ideal (students back in the classroom, each student has a Chromebook to work on, things are back to normal, etc.).

Investigation 1

Appendix 5: Example Teacher PowerPoint

Unit 2 Lesson 5 Outcome

Outcome:

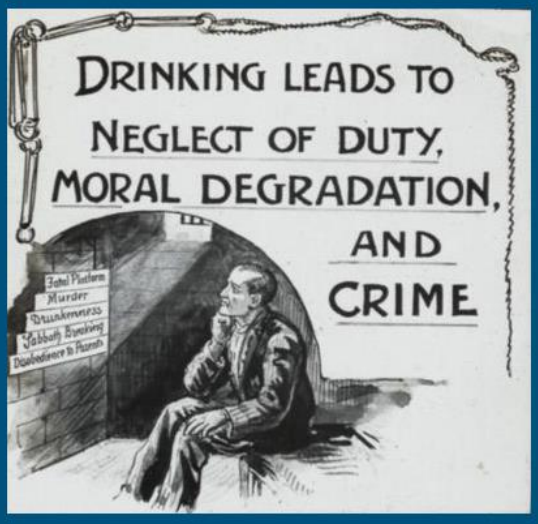
I can write high-school level historical arguments (part 2).

TODAY: I can use IREAD to analyze documents from 1926 arguing for and against Prohibition.

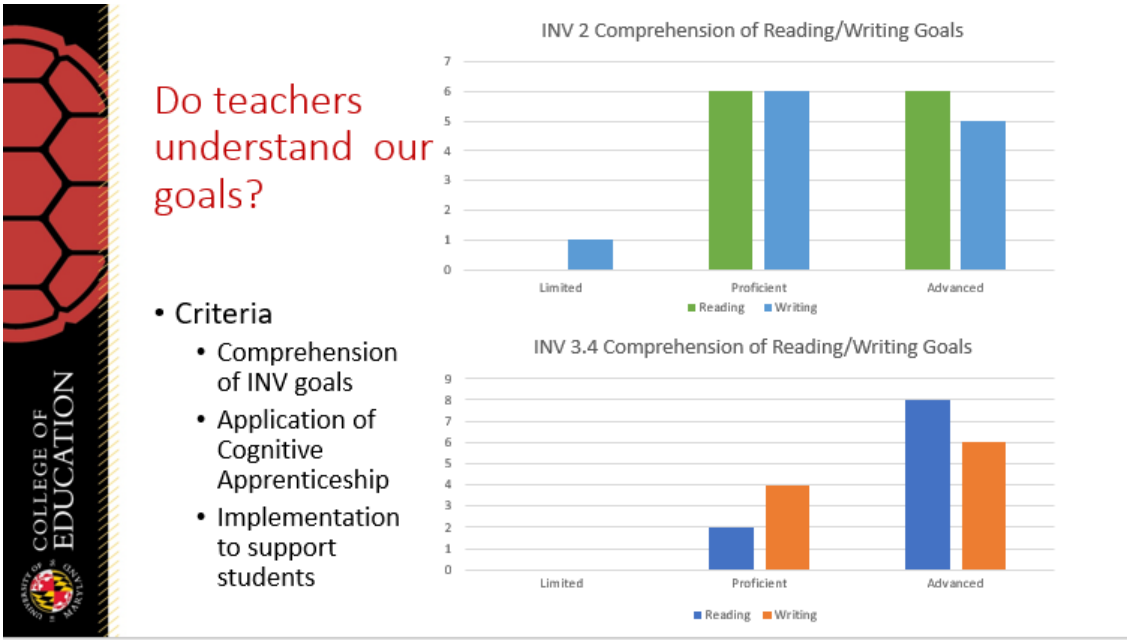
Salir:

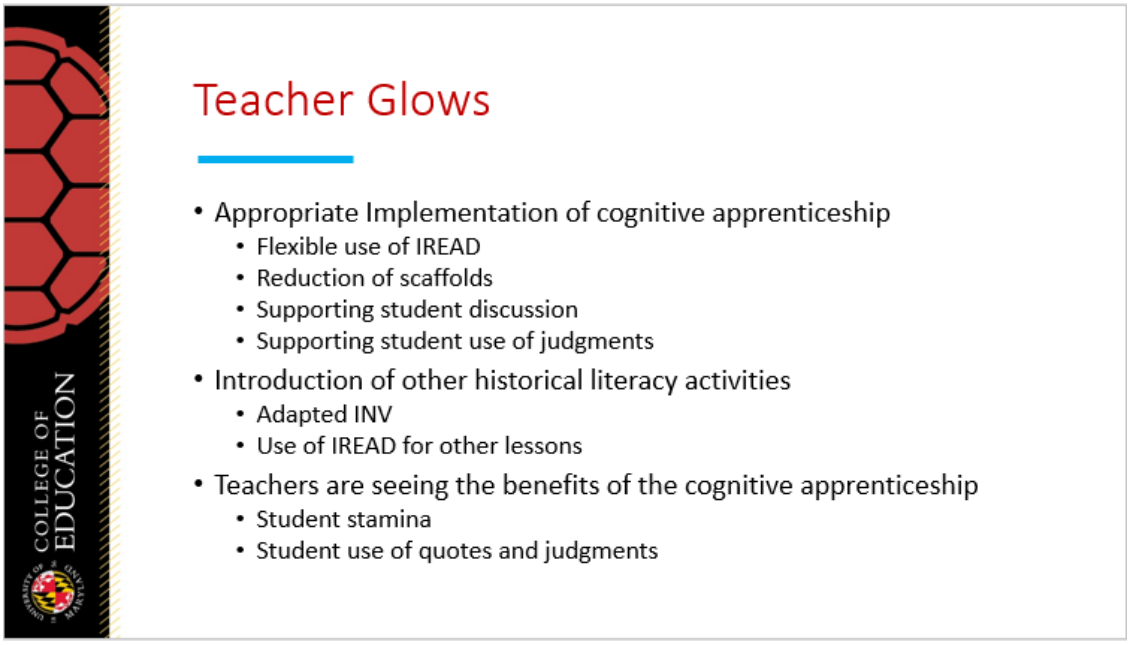
Puedo escribir argumentos históricos de nivel secundario (parte 2).

HOY: Puedo usar IREAD para analizar documentos de 1926 que argumentan a favor y en contra de la Prohibición.

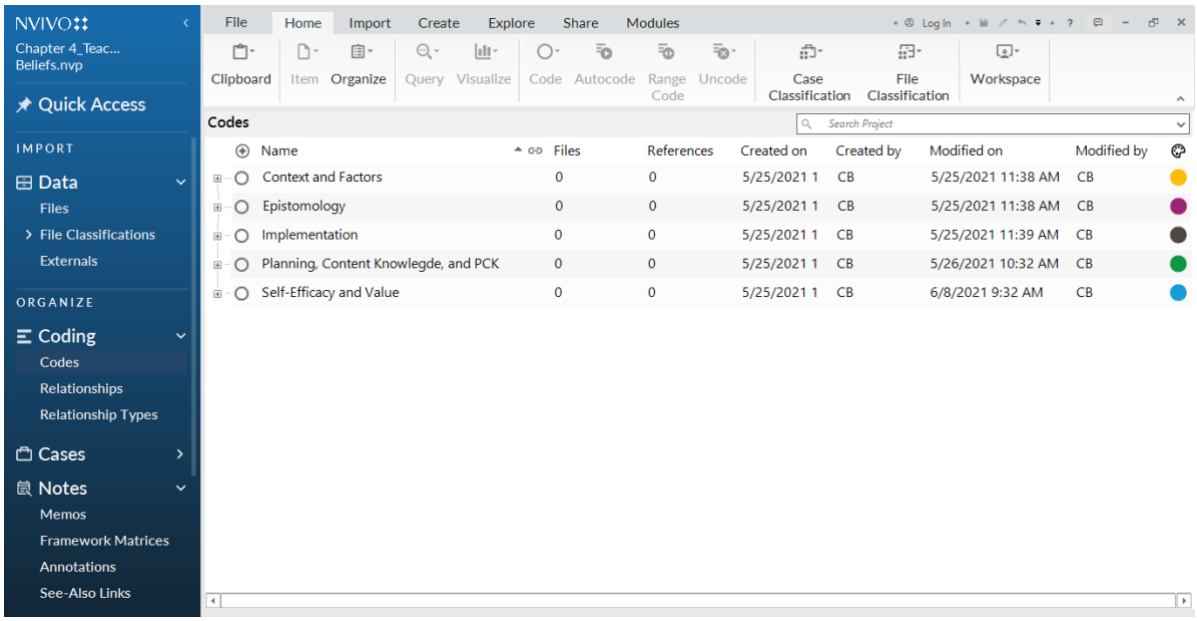


Appendix 6: Member Checking PowerPoint





Appendix 7: NVivo



Chapter 5: Discussion and Implications

This dissertation focuses broadly on supporting teachers and students as they teach and learn disciplinary literacy skills. Each of the three studies contained in this dissertation contributes to this overarching focus. In my first study, in Chapter Two, I observe self-regulated learning and instruction (SRI), which is often viewed as an important element for successful cognitive apprenticeship models of instruction (Brown et al., 1989; Harris & Graham, 2009). In my first study, I completed a formal synthesis on the effects of SRI on motivation outcomes for students with or at risk for LD. In so doing, I attempted to clarify why SRI helps to improve student motivation in addition to its academic benefits (see Antononiou & Souvignier, 2007; Harris et al., 2006; Schunk & Cox, 1986).

In my second study, in Chapter Three, I explored the writing outcomes from a cognitive apprenticeship intervention in middle school science developed by De La Paz, Levin and Felton, who also helped to develop and evaluate De La Paz's earlier work in history. I analyzed the data to determine if students' scientific literacy skills and argumentative writing skills improved on both near and far transfer outcomes. In my final study, in Chapter Four, I documented the implementation of a historical literacy curriculum that used a cognitive apprenticeship model. The curriculum was designed by De La Paz with several important contributors (My primary role was as project director and in so doing, I led three other doctoral students and a masters student in developing curriculum, maintaining our online learning platform, and collecting data that we did as a team. I also served as primary liaison with the participating school district leaders in our research practice partnership).

In my final study, I observed teacher fidelity, recorded teacher adaptations, and analyzed how important factors such as teachers' beliefs and PCK affected the implementation of our

curriculum. I begin this chapter with a brief summary of each study including important findings. Next, I discuss the broader implications for students and teachers. I conclude with limitations of the dissertations and future directions for both research and practice.

Summary and Findings

Each of the three studies contributes to a better understanding of supporting teachers and students as they teach and acquire complex disciplinary skills. The first study affirms the effectiveness of SRI instruction, the second study analyzes the impact of a cognitive apprenticeship model in science, and the third study observes and details the implementation of a similar model in history.

Elements of Effective Self-Regulated Instruction for Promoting Student Motivation

Struggling students from K-12 show significantly lower academic motivation when compared to age range peers (Graham et al., 2017). Importantly, motivation predicts academic achievement (Pajares, 1996). Despite this, motivation is rarely included in intervention studies that include SRI as an outcome of interest (De La Paz & Butler, 2018). A recent literature review by De La Paz and Butler (2018) identified that SRI may have a positive impact on student motivation. In the first study I determined that the trend observed in our previous research remained valid after a more systematic review of the literature on SRI and motivation.

Butler and De La Paz (in press) completed a synthesis based on 29 studies that used an SRI intervention, and observed motivation outcomes for students with or at risk for LD. I used a multistep identification process (online search, hand search, reference search). I coded studies for instructional components, outcomes measures, and a series of indicators of study quality. I had two major purposes for the study. First, to descriptively identify the types of SRI that researchers have used to help address motivation. Second, to quantitatively analyze those studies of

sufficient rigor to observe if interventions using SRI components did, in fact, improve student motivation.

The analysis revealed that there were five types of SRI components in studies that focused on improving motivation including: goal-setting, self-monitoring for task completion, self-talk for task completion, self-talk for progress monitoring, self-talk for emotional regulation, and self-monitoring for emotional regulation. Further, I found that across academic domains and outcome measures, motivation increased after intervention with over 80% of studies reporting positive outcomes. I also recognized that there were significant differences between academic domains. First, there were more studies in reading and writing than in mathematics that met the criteria for inclusion. Second, writing studies reported consistently higher motivation outcomes than did either of the other two domains. Finally, writing studies were more likely to include interventions that integrated multiple forms of SRI than either reading or mathematics.

This study, by establishing the positive effects that SRI has on motivation, lends further credibility to the cognitive apprenticeship models used in the second and third studies. It is well documented that writing motivation predicts writing performance (Graham et al., 2017; Troia et al., 2012). Though disciplinary writing and domain general writing require different skills, it is likely that motivation is similarly important in disciplinary settings. Thus a cognitive apprenticeship model with embedded SRI has the potential to improve academic outcomes and motivation outcomes.

Cognitive Apprenticeship in Middle School Science

Helping students to develop scientific literacy skills is difficult. Students do not naturally think like scientists, nor can they explain the reasoning behind important scientific processes; however, they do have the ability to learn these important skills (De La Paz & Levin, 2018;

Hogan & Maglienti, 2001; Lee & De La Paz, 2021). In the second study, I analyzed data from an intervention that was designed to support students as they learned scientific thinking and argumentative writing skills. The intervention was designed by De La Paz, Levin, and Felton and used a cognitive apprenticeship model to integrate cognitive and linguistic supports into inquiry lessons. The intervention also integrated self-monitoring and goal setting skills during a writing lesson. Students learned to self-evaluate and make revisions to their scientific arguments.

The intervention was composed of four, three-day, investigations on science topics provided by the cooperating teacher using adapted materials from ADI (Sampson et al., 2013a, b). Each investigation began with a central inquiry and required that students conduct an experiment to collect data. Students would then write an argumentative essay to address the central inquiry. A critical component of the intervention was small and large group discussions. Students engaged in small group discussion to construct an initial argument and then in a whole group teacher-led discussion to refine the argument by discussing critical questions that promoted scientific thinking.

The study used a quasi-experimental pre-post control group design, and students were assessed on both near and far transfer outcomes of scientific argumentative writing. I used a multilevel analysis to account for classroom level differences and baseline inequivalences. I found that students in the treatment condition made considerable growth on the near transfer and far transfer outcomes. When compared to the control group, the treatment group scored made statistically more growth on the far transfer measure. As part of the analysis, I speculated as to why our cognitive apprenticeship model supported transfer.

In examining transfer this study contributes to the research on the effectiveness of cognitive apprenticeship models. I analyzed the components of the model designed by De La

Paz, Levin, and Felton and detailed how these components may have contributed to student transfer. Key components such as the use of authentic learning investigations and evidence based science instruction, the emphasis on small and whole group discussion, and the incorporation of evidence based writing instruction helped to mediate learning and support transfer. These key variables helped students to create initial learning that was deep and lasting and helped students to recognize specific cues in both content and context that can promote transfer.

Cognitive Apprenticeship Implementation in High School History

In my final study I shifted focus and attempted to understand how teachers implemented a cognitive apprenticeship in a historical setting. I analyzed qualitative data collected during a yearlong project in which our team at UMD designed a curriculum to support students' historical thinking and argumentative writing. The curriculum included six investigations that began with a central inquiry and had students examine historical sources to create an argument. The cooperating district provided investigation topics and our team developed critical historical questions. Our team designed questions to connect to current events and specific problems in society. For example, the first and last investigations both connected to the issue of systemic racism. The first investigation asked students, "Who was the stronger advocate for African Americans during the 1900s, Booker T. Washington or W.E.B Dubois?" The final investigation asked students, "Did Reagan's intensified war on drugs have its intended consequences?" Teachers were supported through seven professional development (PD) sessions and continued coaching.

The goals of this study were to observe teacher fidelity to the curriculum, document the ways that teachers made adaptations to support their students, and analyze how teachers' beliefs, content knowledge, PCK, and other factors affected their implementation. Additionally, I sought

to observe how PD may affect these critical factors over the course of the study. I collected several data sources including coaching sessions, surveys, and teacher materials to address the goals. I chose five teachers as cases and developed profiles for each teacher.

I found that teachers were able to implement the curriculum with fidelity after completing the yearlong PD. I documented a series of adaptations that teachers made to the curriculum, regarding teachers using scaffolding to support students. I identified a series of cross cutting explanations that help to explain how important factors affect implementation of the curriculum. First, teachers showed improved fidelity throughout of the PD series. Second, through PD, teachers expanded their PCK which affected their adaptations to the curriculum. Third, teachers' beliefs about student ability affected decisions surrounding scaffolding and adaptations. Finally, all teachers held similar beliefs about the purpose of history education by the end of the study. Importantly these themes show that teachers can learn a cognitive apprenticeship model with relatively little training (only 10 total hours between PD and coaching). Furthermore, the findings showed that the PD could alter the way teachers think and teach in the classroom by changing teachers' beliefs and improve teacher PCK

Implications for Practice

The studies in this dissertation outline several important implications for practice including the importance of integrating SRI, the use of cognitive apprenticeship models for teaching disciplinary skills, and the need for effective PD to support teachers when implementing a cognitive apprenticeship model.

Importance of Self-Regulation

This dissertation reaffirms the validity of teaching self-regulation. I identify SRI as an effective form of instruction for promoting student motivation. Eighty percent of studies that

incorporated SRI showed positive change in student motivation. Furthermore there was some indications that certain types of SRI were particularly effective at promoting motivation. In the second study, SRI was integrated into the intervention. The participating teacher modeled self-talk to students as they interacted with the scientific inquiry. As part of the investigations, students also received direct instruction on how to use self-evaluation and make revisions to their arguments. Students wrote stronger arguments that more effectively used evidence and provided justification after discussions and opportunities to revise. In the final study, pedagogically strong teachers who exhibited high fidelity used SRI to support students while they learned how to write historical arguments. For example, teachers had students set goals that addressed specific aspects of historical writing such as creating a rebuttal. Across the three studies, SRI was shown to be an effective and meaningful component of instruction.

Efficacy of the Cognitive Apprenticeship

The latter two studies reiterate the effectiveness of cognitive apprenticeship models for supporting disciplinary literacy. Butler et al. (2021) adds to the research on the usefulness of the model in science and demonstrates how students can acquire skills needed to transfer knowledge. This study also shows the importance of discussion in helping students to think more like scientists. Argumentative discussion is an effective way to promote student learning and can help students to think and reason more like experts (Felton et al., 2019). Discussion was the cornerstone of this study and students engaged in both small and large group discussions. Felton et al. (in review) discusses the important ways that the cooperating teacher in the study shifted "dialogic stances" while supporting student thinking. This teacher was highly responsive to the needs of his students and would oscillate between guiding student thinking, facilitating peer dialogue, and coaching student skills.

The final study demonstrated that with meaningful PD, teachers can learn how a cognitive apprenticeship supports student learning and implement it with a high degree of fidelity while still making meaningful adaptations in materials. This study also shows that key factors such as teachers' beliefs and PCK affect implementation and can be changed through the use of PD. This means that PD can not only teach the curriculum, but also change critical factors that affect implementation and ensure high levels of fidelity. Taken together, these studies show that the cognitive apprenticeship is accessible to teachers with proper training and once administered, it can support disciplinary literacy in the classroom.

Meaningful PD to Support Implementation

Effective PD entails several components including that it is content focused, it provides effective models for implementation, it occurs over a sustained period of time, and it provides ongoing coaching to support teachers (Darling-Hammond et al., 2017). Though not a direct focus of the second study, De La Paz, Levin, and Felton designed highly effective means of administering PD. This PD resulted in high levels of fidelity and highly effective instruction. The teacher had over 90% fidelity on a researcher created measure and delivered lessons that scored an average of 3.8 out of 4 on a scale of instructional quality. In the third study, teachers showed improved fidelity as well as greatly expanded PCK and even changing beliefs after attending the entire PD. Meaningful PD that follows these guidelines and the examples laid out in this dissertation can make teacher practice stronger and implementation more efficacious.

Future Directions

The studies in this dissertation open multiple avenues for future research. To extend my first study, an initial step would be expanding the student population to typically achieving students. Additionally, research into some of the interesting trends identified in the study may

yield important outcomes. For example, it would be interesting to explore why SRI was more commonly used in writing than reading and mathematics, and writing intervention studies showed more consistently positive results. Explorations into these differences could provide important information about the use and usefulness of SRI across academic domains.

With regards to studies incorporating SRI and motivation for struggling students, researchers need to be more explicit about their theory of change and choice in outcome measures in the future. If researchers are to prove that their interventions improve motivation they will need to provide theoretical explanations as to how their specific intervention components may affect motivation. This is similar to what we did in my second study, outlining the potential instructional components that led to successful transfer. Researchers should also be more explicit about their choice of motivation measure, providing a rationale for why they chose the specific motivation constructs.

To extend my second study, researchers could focus initially on addressing some of the gaps in this study. Though the study established that with meaningful instruction students could transfer knowledge, there were certain limitations pertaining to the sample size and study design. Future research needs to confirm the efficacy of cognitive apprenticeship models by using RCTs and expanding the sample size. Focusing on specific populations of learners such as students with disabilities and English language learners could provide insight into what populations a cognitive apprenticeship model can effectively support.

It would be particularly important in the future to detail exactly why our cognitive apprenticeship model was effective in promoting transfer. Though we make speculations, we do not isolate components of the model to examine those speculations. By using more specific outcomes measures, future researchers could identify the specific components that are most

effective. For example, researchers could use outcomes related to student's content knowledge to analyze if certain components like evidence based science instruction are the most effective components of the model.

To extend my final study, researchers could focus on expanding the sample size to examine if trends in teachers' beliefs remain consistent in larger and different teacher populations. Studies could focus on mixed methods designs to examine if the changes in implementation would have discernible effects on students' outcomes. Additionally, researchers could focus on measuring beliefs at different times throughout the investigations to observe more incremental and nuanced growth.

More broadly this study reveals questions as to the nature of PD. Using a fully online PD with only 10 hours of training, teachers were able to successfully implement a cognitive apprenticeship. This is an important finding that demands further investigation. First, future research should examine the dosage. If teachers could achieve effective implementation with 10 hours of training, could other teachers who did not have the same level of background in historical reasoning or years teaching also be effective with this level of support? Additionally, in our study, teachers benefitted from both formal PD and coaching but it was not clear which contributed more to teacher success. Finally, researchers should explore PD models for supporting large numbers of teachers such as a "teach the teacher" model. Though our model was successful it may not be a model that could effectively support an entire school district.

Tables

Chapter 2 Table 1: Study Design and Domain

Study	Design	Domain	Study Quality	Sample Size	Percentage Struggling Learners	Grade	Dosage (hrs.)	Dependent Measure
Antononiou and Souvignier (2007)	RCT	Reading	Medium	73	100%	5th-8th	29	SE
Bandura and Schunk (1981)	RCT	Mathematics	Low	41	100%	1st-3rd	~3.5	SE
Berkeley, Mastropieri, and Scruggs (2011)	RCT	Reading	High	59	100%	7th-9th	6	CA
Buzza and Dol (2015)	QED	Mathematics	Low	16	100%	10th	NR	SE, I
Cerar (2012)	SCD	Writing	High	7	100%	6th-8th	22	SE
Cuenca-Carlino, Freeman-Green, Stephenson, & Hauth (2016)	SCD	Mathematics	High	6	100%	8th	10	SE
Cuenca-Carlino and Mustian (2013)	SCD	Writing	High	9	100%	6th-8th	~15	SE
Cuenca-Sanchez, Mastropieri, Scruggs, and Kidd (2012)	RCT	Writing	High	21	100%	6th	16.5	SE
Daniel (2003)	SCD	Mathematics	Low	18	100%	6th-8th	8	SE, CA
Ennis and Jolivette (2014)	SCD	Writing	High	6	100%	9th	~7	SE, IM
Garcia and De Caso (2006)	RCT	Writing	Medium	60	100%	5th-6th	~8	SE
García-Sánchez and Caso-Fuertes (2005)	RCT	Writing	Medium	191	100%	5th-6th	~23	SE
García-Sánchez and Fidalgo-Redondo (2006)	RCT	Writing	Medium	121	100%	5th-6th	~20	SE
Graham and Harris (1989)a	RCT	Writing	Medium	33	66%	5th-6th	6	SE
Graham and Harris (1989)b	SCD	Writing	Medium	3	100%	6th	4.5	SE
Graham, Harris, and Mason (2005)	RCT	Writing	High	73	100%	3rd	~5.8	SE

Harris, Graham, and Mason (2006)	RCT	Writing	High	66	100%	2nd	6.3	IM
Macarthur and Philippakos (2010)	SCD	Writing	Medium	6	100%	6th-8th	~7.5	SE
Mills (2012)	SCD	Writing	High	10	100%	8th	~7	SE
Miranda, Villaescusa, and Vidal-Abarca (1997)	RCT	Reading	Medium	80	75%	5th-6th	~17	CA
Nelson and Manset-Williamson (2006)	RCT	Reading	High	20	100%	4th-8th	20	SE, CA, A
Page-Voth and Graham (1999)	RCT	Writing	High	30	100%	7th-8th	NR	SE
Rooney (1997)	RCT	Reading	Medium	91	100%	3rd-5th	~2	SE
Sawyer, Graham, and Harris (1992)	RCT	Writing	High	33	66%	5th-6th	~5	SE
Schunk (1983)	RCT	Mathematics	Low	30	100%	4th-5th	5	SE
Schunk (1985)	RCT	Mathematics	Low	40	100%	6th	3.5	SE
Schunk and Cox (1986)	RCT	Mathematics	Medium	90	100%	5th-10th	3.5	CA
Schunk and Rice (1993)	RCT	Reading	Medium	44	100%	5th	5	SE
Sexton, Harris, and Graham (1998)	SCD	Writing	Medium	6	100%	5th-6th	~7	CA, E

*RCT: Randomized Control Trial, SCD: Single Case Design, QED: Quasi Experimental Design, SE: Self-Efficacy, CA: Causal Attributions, E: Effort, A: Affect, IM: Intrinsic Motivation, I: Interest NR: not reported

Chapter 2 Table 2: Frequency of Occurrence of Self-Regulated Instructional Components

	Sample Size (Total Treatment Conditions)	Goal Setting	Self-Monitoring Task Completion	Self- Talk for Task Completion	Self-Talk for Progress Monitoring	Self-Monitoring for Emotional Regulation	Self-Talk for Emotional Regulation
All Studies	46	51%	54%	56%	33%	5%	35%
Reading	13	23%	54%	69%	15%	8%	8%
Writing	23	78%	78%	87%	70%	9%	83%
Math	10	71%	71%	14%	14%	0%	0%
Single Component	11	27%	18%	45%	0%	0%	0%
Multi Component	35	71%	77%	77%	54%	9%	83%

Note: Percentages were calculated by dividing number of treatment conditions in which a specific self-regulated instructional component is present by total number of treatment conditions.

Chapter 2 Table 3: Effect Size by Treatment Condition

Study	Domain	Dependent Measure	IV	ES	Maintenance ES
Antononiou and Souvignier (2007)	Reading	SE	Treatment	-0.37	0.38
Berkeley, et al. (2011)	Reading	CA (success)	RCS	-0.03	0.09
			AT	0.83	0.39
		CA (failure)	RCS	-0.04	0.04
			AT	0.27	0.39
Cuenca-Sanchez, et al. (2012)	Writing	SE	Treatment	1.03	
*García-Sánchez and De Caso (2005)	Writing	SE	Treatment	0.51	0.31
García-Sánchez and Fidalgo-Redondo (2006)	Writing	SE	SRSD	0.76	
			SCM	1	
Graham, Harris, and Mason (2005)	Writing	SE (planning)	SRSD	0.2	
			SRSD+	0.38	
		SE (generating)	SRSD	0.25	
			SRSD+	0.1	
Harris, Graham, and Mason (2006)	Writing	IM	SRSD	1.14	
			SRSD+	0.35	
		E	SRSD	0.72	
			SRSD+	0.25	
Miranda, Villaescusa, and Vidal-Abarca (1997)	Reading	CA (for effort)	SI	0.63	1.17
			AT	0.86	1.2
Nelson and Manset-Williamson (2006)	Reading	SE	Treatment (EC)	-0.46	
		CA (success)	Treatment	0.58	
		CA (failure)	Treatment	1.3	
		A (positive)	Treatment	-0.12	
		A (negative)	Treatment	-0.45	
Page-Voth and Graham (1999)	Writing	SE	Goals	-0.3	
			Goals+Strategy	-0.07	
Rooney (1997)	Reading	SE	Self-Regulation	0.7	
			MF	0.26	
			SR+MF	0.36	

		CA	Self-Regulation	0.44
			MF	0.25
			SR+MF	0.27
Schunk and Cox (1986)	Math	SE	Continuous	0.02
			Discontinuous	0.91

Note: *Garcia-Sanchez and De Caso (2005) reported multiple other outcomes but did not provide information to calculate ES. Outcomes: SE=self-efficacy, CA=casual attributions, A=affect, E=Effort, IV: RCS= reading comprehension strategy, AT=attribution retraining, SRSD=self-regulated strategy development, SCM=social cognitive model, SRSD+=self-regulated strategy development plus peer support, SI=strategy instruction, SR= self-regulation, MF=model feedback, treatment=authors offered no specific name for the intervention

Chapter 3 Table 1: Student Demographic Information

	Full Sample (*N=110*)		Control Condition (*N=38*)		Treatment Condition (*N=72*)	
Variable	Mean	SD	Mean	SD	Mean	SD
Hispanic Status	79%		61%		89%	
Special Ed	29%		21%		33%	
Gender	43%		47%		42%	
TOWL6	15.35	4.8	17.15	5.26	14.4	4.31
TOWL7	8.45	3.23	9.97	2.59	7.642	3.28
Pretest Total	0.82	1	1.34	1.15	0.54	0.8
Pretest Claim	0.57	0.63	0.84	0.64	0.43	0.62
Pretest Evidence	0.17	0.4	0.29	0.52	0.11	0.32
Pretest Reasoning	0.07	0.26	0.21	0.41	0	0

Chapter 3 Table 2: Near Transfer (Kinetic Energy) Outcomes

	Initial CER		Revised CER					
Outcome	Mean	SD	Mean	SD	DF	t-score	p-value	ES
CER Total	0.15	0.36	1.54	1.79	80	3.17	<0.003**	1.08
Claim	0.51	0.69	0.66	0.73	92	1.48	0.142	0.21
Evidence	0.22	0.51	0.43	0.69	90	2.88	0.006*	0.34
Reasoning	0.18	0.44	0.61	0.79	96	4.46	<0.001**	0.67

Chapter 3 Table 3: Far Transfer (Purple Loosestrife) Outcomes

	Pretest		Posttest					
Outcome	Mean	SD	Mean	SD	DF	t-score	p-value	ES
CER								0.67
Total	0.54	0.8	1.39	1.61	142	4.74	<0.001***	
Claim	0.43	0.62	0.67	0.67	142	3.03	0.034*	0.37
Evidence	0.11	0.32	0.49	0.69	142	4.19	<0.001***	0.71
Reasoning	0	0	0.24	0.54	142	3.69	<0.001***	0.63

Chapter 3 Table 4: Baseline Equivalence

	Control		Treatment				
Variable	Mean	SD	Mean	SD	DF	t-score	p-value
Hispanic Status	0.61	0.5	0.89	0.32	108	3.7	<0.001***
Special Ed	0.21	0.41	0.33	0.47	108	1.35	0.18
Gender	0.47	0.51	0.42	0.5	108	0.56	0.57
TOWL6	17.15	5.26	14.4	4.31	108	2.95	0.004***
TOWL7	9.97	2.59	7.642	3.28	108	3.8	<0.001***
Pretest Total	1.34	1.15	0.54	0.8	108	4.27	<0.001***
Pretest Claim	0.84	0.64	0.43	0.62	108	3.27	0.002**
Pretest Evidence	0.29	0.52	0.11	0.32	108	2.25	0.03*
Pretest Reasoning	0.21	0.41	0	0	108	4.34	<0.001***

Chapter 3 Table 5: Unconditional Model

Parameter	Estimate	Std. Error	df	t	Significance	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.42	0.165	8.28	8.59	<0.001	1.04	1.79

Chapter 3 Table 6: Treatment Control Comparison

	Total		Claim		Evidence		Reasoning	
Variable	RC	SE	RC	SE	RC	SE	RC	SE
Intercept	-0.602	0.574	0.245	0.26	-0.511	0.257	-0.242	0.205
Cohort	0.676*	0.303	-0.145	0.133	0.402**	0.129	0.261*	0.11
Hispanic	0.195	0.322	0.111	0.145	0.073	0.144	0.054	0.116
Special Ed	-0.031	0.286	-0.067	0.129	0.051	0.13	-0.083	0.102
Gender	-0.426	0.25	-0.1	0.113	-0.143	0.112	-0.128	0.089
TOWL6	-0.021	0.044	0.017	0.02	-0.003	-0.02	-0.015	0.016
TOWL7	0.171**	0.063	0.011	0.028	0.078**	0.028	0.061**	0.11
Pretest Total CER	0.621***	0.142						
Pretest Claim			0.477***	0.092				
Pretest Evidence					0.229	0.144		
Pretest Reasoning							0.443*	0.19
Goodness of Fit								
-2 Log Likelihood		360.76		186.48		186.07		135.76
Treatment ES		1.95		-0.31		0.93		0.57

Chapter 4 Table 1: Demographic Information

Note: Stacey was teaching at a newly opened school in the district so demographic information had not yet been published.

Teacher	Years Teaching	Historical Literacy Training	School Demographic Information					Gender		Free and Reduced Lunch	Statewide Test Scores	
			White	African American	Hispanic	Asian	Other	M	F		Reading	Math
Emily	12	No	75	10	9	1	6	50	50	22	46	65
Erma	14	Yes	NA									
Stacey	14	Yes	49	26	14	5	6	52	48	44	24	37
Rachel	25	Yes	32	26	36	3	3	48	52	43	36	18
Steve	11	Yes	77	5	8	4	5	51	49	10	54	66

Chapter 4 Table 2: Data Collection by Investigation

Investigation	Investigation 1	Investigation 2	Investigation 3	Investigation 4	Investigation 5	Investigation 6
PD Date	October 8	November 23	January 6	February 22	March 14	May 17
Data Collection Window	September-December 2020		January-March 2021		January-March 2021	
Coaching session		Coaching session 1		Coaching session 2		Coaching session 3
Survey	Teacher Demographics Survey			Implementation Survey 1	Implementation Survey 2	Implementation Survey 3
						Teachers' Beliefs Survey
Artifacts	Artifacts INV 1	Artifacts INV 2	Artifacts INV 3	Artifacts INV 4	Artifacts INV 5	Artifacts INV 6

Chapter 4 Table 3: Open Coding Context and Factors Coaching sessions 1

First Coding (INT 1)		Reduced Codes (INT 1)	Final Codes (INT 1)
"Team planning system"	new students	School/district factors	School/district factors
district requirements	school requirements		Teacher/classroom characteristics
new job	administration		Student characteristics
new school			Student ability
co-taught	Spanish ability	Teacher/classroom characteristics	Student participation/engagement
previous experience	team lead		Tech issues/barriers
previous writing experience	training		Time restrictions/Coverage
black and Latino	compatible students	Student characteristics	Other Obstacles
demographics	"Title one"		classroom application
"Housing project"			
"High performance"	student ability	Student ability	
"Hot shot"	struggling students		
honors	student ability		
honors	student ability		
IB	student level		
standard			
attendance	participation	Student participation/engagement	
completion	stamina		
engagement	student engagement		
participation			
computer barriers	tech issues	Tech issues/barriers	
tech barrier	tech		
tech concerns	virtual environment		
time constraints	time restrictions	Time restrictions/Coverage	
time content coverage	content coverage		

at homework	limited home resources	Other Obstacles	
community issue	no physical docs		
benefits of virtual learning	pacing	classroom application	
extended time	writing requirements		

Chapter 4 Table 4: Full Code Book

Context and Factors	Beliefs	Planning, Content Knowlesge, and PCK	Self-Efficacy and Value	Implementation
School/District	Barriers to Learning	Content Knowledge	Self-Efficacy	Content Knowledge
Student Ability	Goals/ Expectations	Discusion and coperative work	Value	Discusion and coperative work
Engagment/ Participation	Historical Thinking	Examples/ Feedback		Examples/ Feedback
Teacher/ Classroom	Scaffolding	Grading		Grading
Time Restriction/ Content Coverage	Srudent Ability	Independent Work		Independent Work
	Student Motivation	Instructional Materials		Lesson Planning
	Teaching Approach	Lesson Planning		Scaffolding
		Scaffolding		Teaching Historical Thinking
		Teaching Historical Thinking		Teaching Reading
		Teaching Reading		Teaching Writing
		Teaching Writing		Technocal Knowledge
		Technocal Knowledge		

Chapter 4 Table 5: Level 3:Themes Memo Writing

Teacher Themes	
Code	Memo
Context	Pittman faces similar contextual barriers to other teachers such as the short periods, distanced learning, and the change in students midway. She teaches at an IB school meaning that here students are stronger though she still has a high proportion of ESL students and mention certain racial disparities. Pittman has high degrees of participation likely due to her strong teaching abilities and level of experience. She does not seem too often agree with her cooperating team and is new to this school doing much of her previous work in Baltimore.
Beliefs	Pittman sees her role as a facilitator of historical investigations. She often works in social activism and social justice as a central theme. This corroborates her belief that in order to engage student you need to connect content to their lives. She acknowledges certain home barriers but focuses more on the difficulty of certain skills and how students can be properly taught. Like other teachers she discusses the importance of critical thinking skills.
CK and PCK	Pittman is a master teacher. She has an expert understanding of history and numerous pedagogical skills. Early in the investigations Pittman understood the goals of the cognitive apprenticeship and made adjustments accordingly. Pittman in many ways was the ideal teacher because she grasped the core components but was not afraid to make adjustments when it was merited. Her understanding of scaffolding and reducing scaffolds as well as providing writing support was unparalleled and she was able to support historical thinking, read, and writing by providing meaning explicit instruction and scaffolding.
SE and Value	Pittman has a high degree of self-efficacy which seems to be well earned as she has a number of years of experience and has taken multiple courses to improve her skills. Though she talks little about her thoughts on the curriculum, as evidenced by her high fidelity it is likely that she sees value in the tools and supports
Implementation	It is difficult to encapsulate Pittman's implementation because she is an extremely thoughtful educator often making multiple adaptation to support specific student needs. Of particular note however Pittman often worked hard to make both sides of the argument equal before argumentation adding targeted content. She built useful scaffolds and was one of few teachers who used true differentiation often building multiple forms of the same tool at differing levels. She employed effective mini lessons that targeted specific problem areas. She supported writing through physical supports and had student review previous work presenting grows and grows. Present

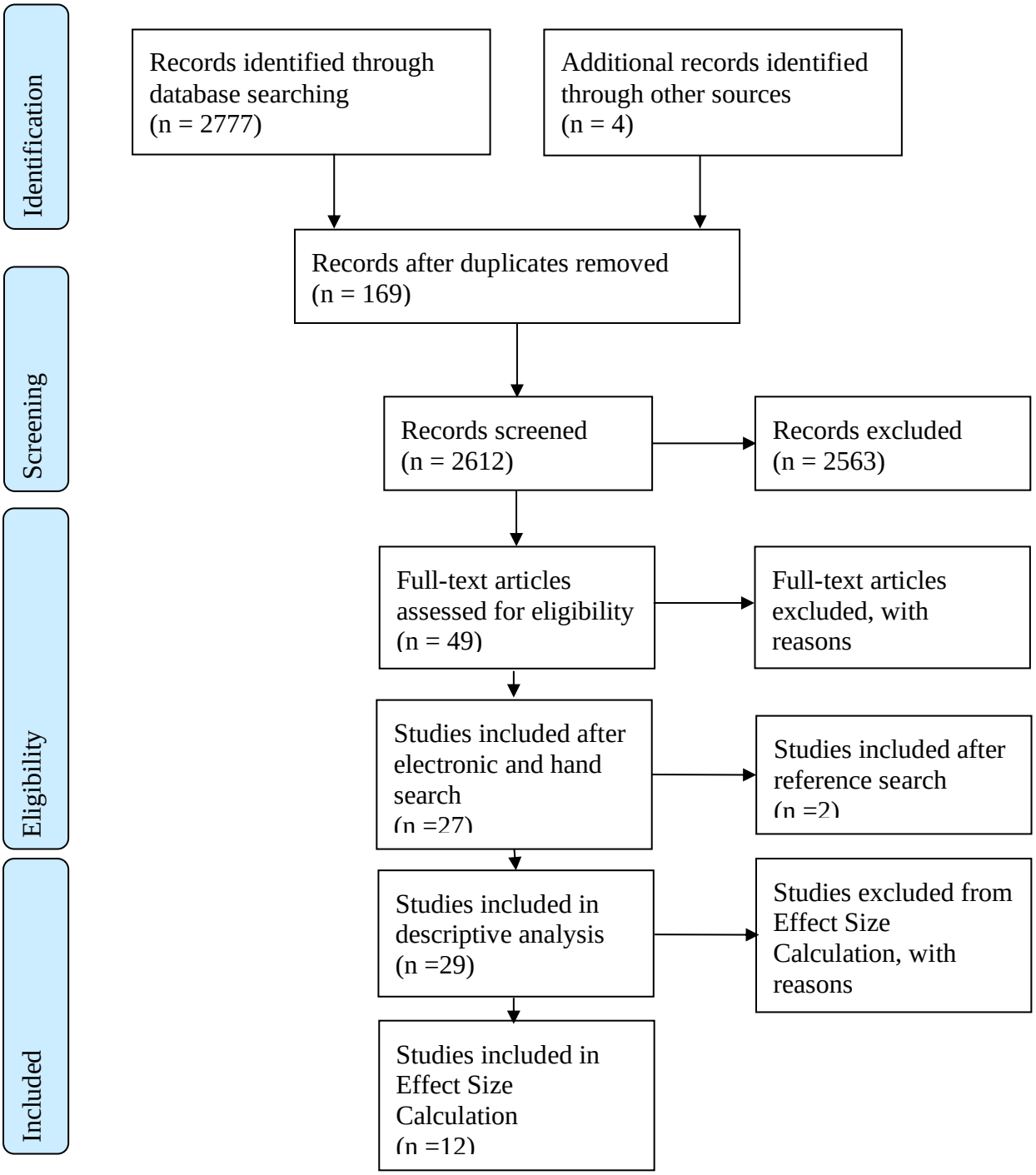
	in her mind at all times was access and social activism so she often integrated supports into materials and used discussion and cooperative work to connect investigations to modern controversies.

Chapter 4 Table 6: Data Analysis Map

6. Theory Confirmation	I use the themes and interalatted explinations to support but the assertaions proposed within the conceptual framework about teacher implementation and the effects of PD.				
5. Interrelating Explinations	Teacher implementation fidelity improved over time as did their understanding of the curriculum.	Teacher PCK influenced what tools they used.	Beliefs about student abilities influenced how teachers used scaffolding.	Beliefs about the purpose of history educations were largely uniform.	
4. Testing Themes	The final coaching session and the final survey were used as additional data as well as tests of validity. The completed code book was check against the new data sources and no new codes were identified suggesting that the codes were valid. A similar teachnique confirmed the validity of the categories.				
3. Themes	Themes were created through memo writing by observing the common themes within all of the coded data.				
2. Categories	Context and Factors	Beliefs	Planning, Content Knowledge and PCK	Self-Efficacy and Value	Implementation
1. Open Codes	Compiled based on initial coaching sessions, surveys, and materials.				

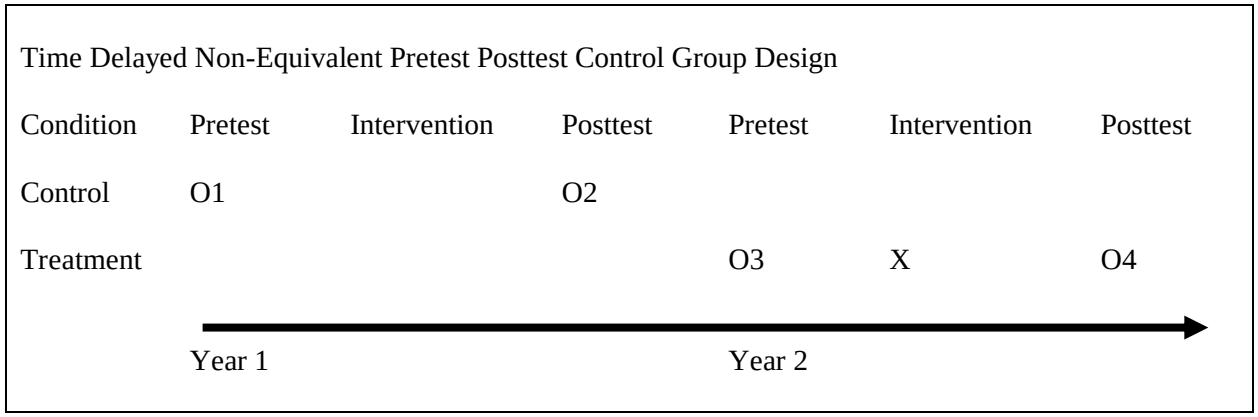
Figures

Chapter 2 Figure 1: Search Protocol



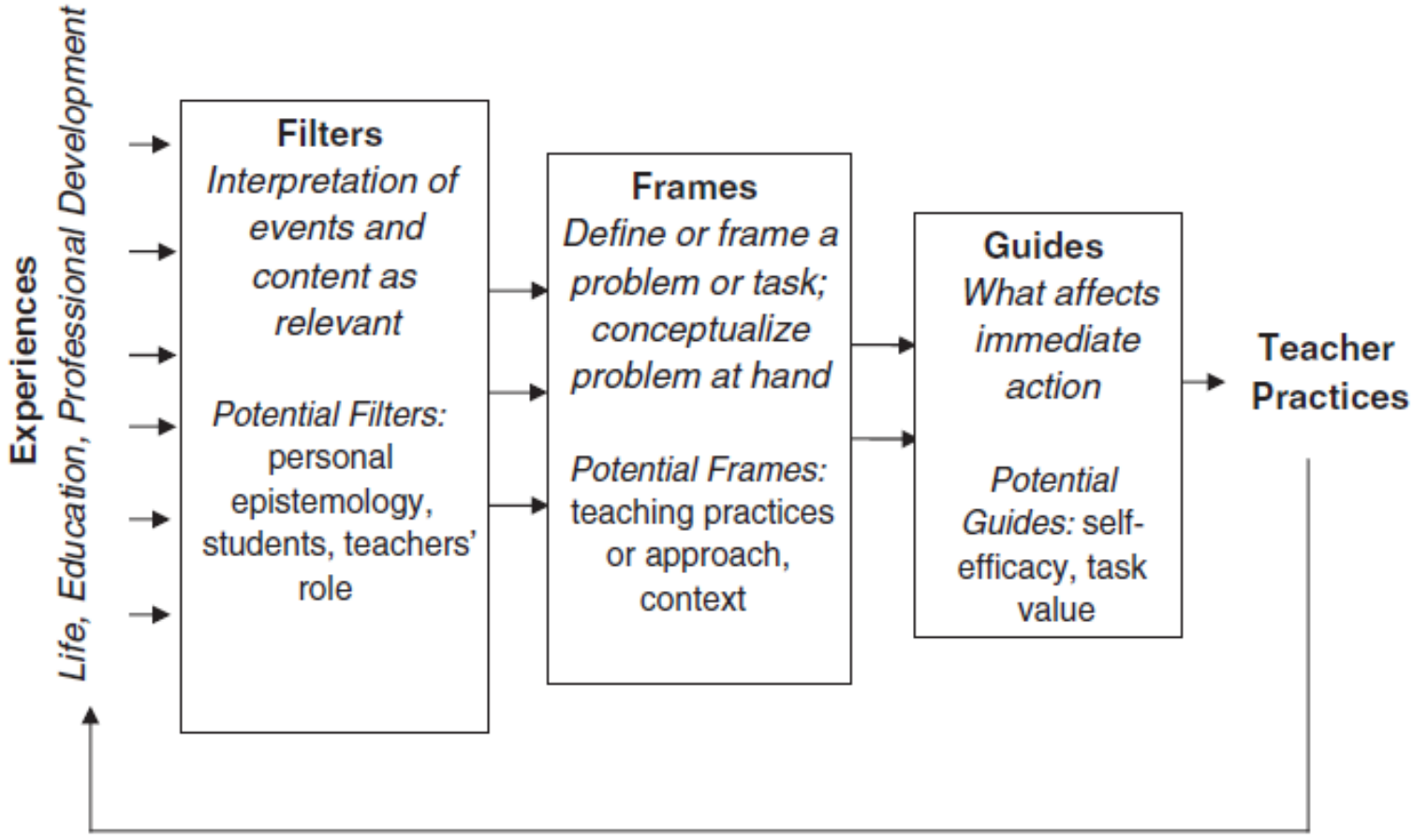
Hand search was conducted using the following journals: Journal of Special Education, Learning Disabilities: A Contemporary Journal of Special Education, Exceptional Children, Behavior Disorders, Learning Disabilities Quarterly, Journal of Learning Disabilities, Contemporary Education Psychology.

Chapter 3 Figure 1: Experimental Design



Note: O1: Control pretest, O2, Control posttest, O3: treatment pretest, X: intervention, O4: treatment posttest

Chapter 4 Figure 1: Five and Buehl Implementation Model



Appendices

Chapter 4 Appendix 1: Example Coaching session Protocol

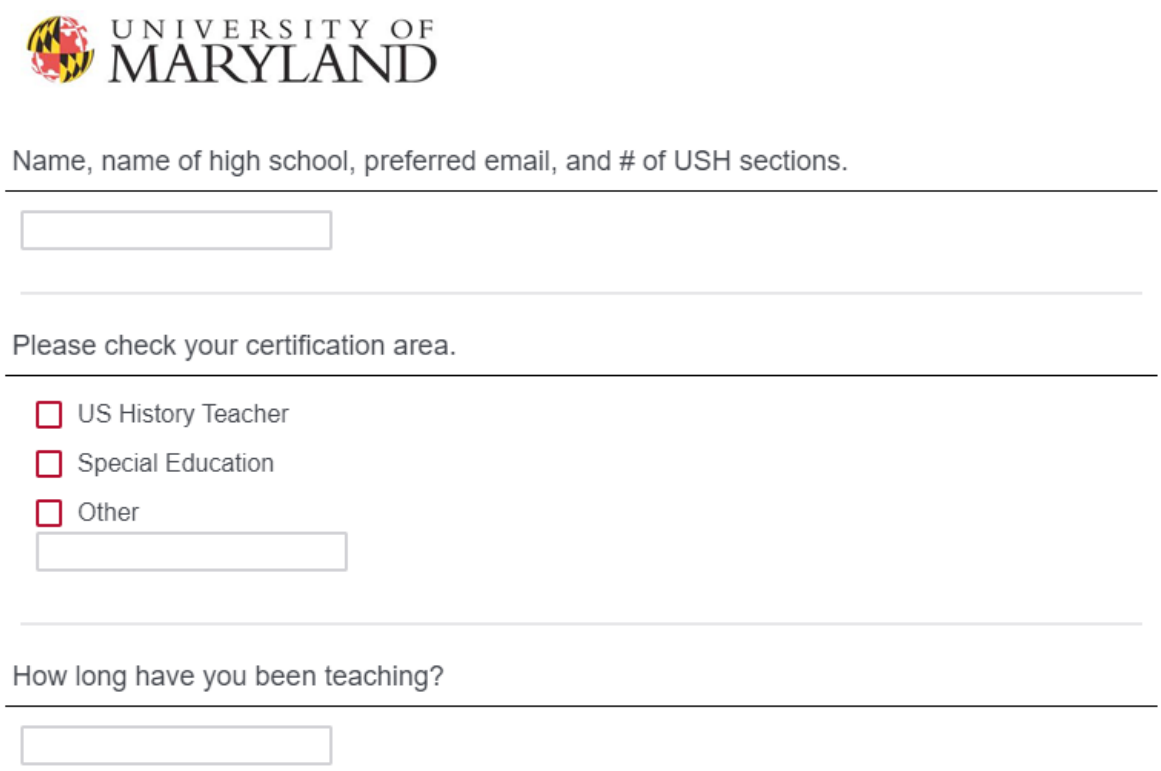
INV 3.4 Coaching session Protocol

24) Which investigation did your team use (INV 3, INV 4, Both)?
25) Could you briefly describe how the investigation went?
26) How many instructional days did you use for the investigation?
    a. Where you able to complete the tasks, you wanted to in that amount of time?
    b. Can you describe what you did for each day of instruction?
27) What choices did you make in using the materials we provided?
28) What materials did you make, to teach the lesson?
29) Did you attempt to model with students? If so, what did you do and how did it go?
30) How many documents did you use with the students?
31) How many (or what %) students attended the lesson?
32) How many (or what %) participated during your lesson? Did more participate than before?
33) How did you monitor student engagement?
34) What homework did you assign, if any? (Why did you decide this?)
35) About how many students analyzed documents using all of IREAD?
36) Were students able to understand reliability?
37) How did you help students to understand reliability?
38) Were any students or, about how many students able to complete the entire essay?
39) What was the quality of the students' annotations?
40) What was the quality of the students' writing?
41) Where students able to make meaningful judgements?
42) How did you support students to make judgements?
43) Did students choose evidence that supported their claims?
44) Did students compose paragraphs that follow H2w?
45) How did you approach grading?
46) What went well? Or, were there any successes? What are your goals for next time?

Other Writing Activities (time permitting)

4) Did you have your students complete any other writing activities?
5) Could you briefly describe the activity?
6) Did you use any of the materials provided by UMD?

Appendix 2: Teacher Demographic Survey



How long have you been teaching US history?

What other subjects have you taught and how long?

Are you a co-teacher USH this year?

- Yes
- No

How confident are you teaching students to read primary sources? Explain.

Are you a co-teacher USH this year?

- Yes
- No

How confident are you teaching students to read primary sources? Explain.

What type of training (if any) have you had for teaching writing in social studies?

How confident are you in teaching students to write historical arguments?

Are you a team lead at your school? Would you be willing to discuss how this program is working at your school, and then communicate with UMD?

Would you like more specific grading rubrics for each investigation?

What other supports can UMD provide, to help you with this program?

Please share additional ideas or concerns. Let us know if you would like other videos or supports.

Appendix 3: Implementation Survey

Name

High School Name

Did you teach any historical reading or writing lessons after finishing Investigation 2? (you may choose multiple activities)

- [ ] Investigation 3
- [ ] UMD writing lesson (from Webinar 3)
- [ ] Other historical reading or historical writing activity (please describe)
- [ ] Did not teach any historical reading or writing lessons

What did you do and how did that go? If you did not teach any of the above, explain why.

We discussed the following student sample (Inv. 3 student example) in the webinar. What do you notice about the student's judgments? Summarize what you observe, focusing on the highlighted text.

Explain what you might say or do if this student was in your class, in Investigation 4.

Write one new idea that you learned after participating in the webinar.

Write one question you have about helping students improve their judgements (especially related to the context or the quality of authors' facts and examples, or about Investigation 4).

Consider your students. Choose one from Investigation 2 or 3 who wrote at least 3 paragraphs.
Describe the class (e.g., honors, standard) and any general characteristics (e.g., motivated, good reader, understands historical controversies, etc.) that led you to choose this student.

Upload de-identified student sample.

Drop files or click here to upload

OR Paste his or her work from google document to this box.

What goal do you have for this student? How will you guide this student in meeting this goal?

Write one goal that you have for yourself during Investigation 4. What supports do you need to accomplish this goal?

Appendix 4: Beliefs Survey

How would you describe your approach to teaching history? Please describe your teaching practices.

In your opinion what are the most significant challenges that you face when teaching history?

What do you think is most important for students to learn in history?

What do you find to be the most significant challenges that students face when learning history?

Please define what historical thinking means to you.

Now that you gave this definition, what aspects of historical thinking are most important for students to learn (and why?)

How do you work to support students' growth in historical thinking skills?

After attending the PD, speaking with us during interviews and gaining experience with the UMD materials this year, we'd like to ask you about what you now understand as the goals, challenges, and important instructional steps for each investigation. Therefore, we ask the following questions. For next year, please share your plans for implementing each investigation, if circumstances for teaching are ideal (students back in the classroom, each student has a Chromebook to work on, things are back to normal, etc.).

Investigation 1

Appendix 5: Example Teacher PowerPoint

Unit 2 Lesson 5 Outcome

Outcome:

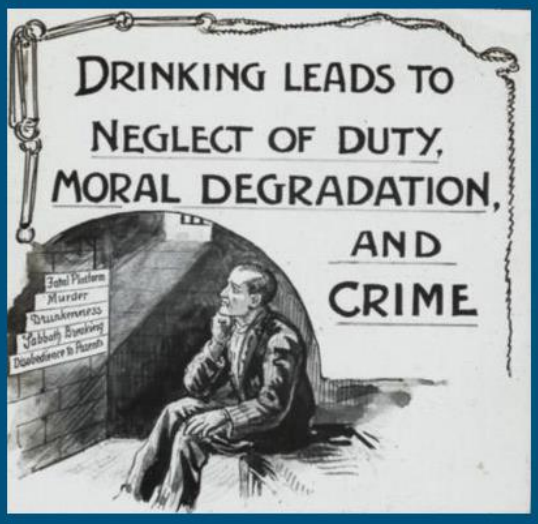
I can write high-school level historical arguments (part 2).

TODAY: I can use IREAD to analyze documents from 1926 arguing for and against Prohibition.

Salir:

Puedo escribir argumentos históricos de nivel secundario (parte 2).

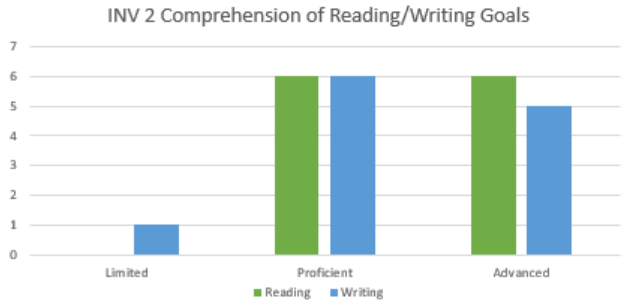
HOY: Puedo usar IREAD para analizar documentos de 1926 que argumentan a favor y en contra de la Prohibición.



Appendix 6: Member Checking PowerPoint

Do teachers understand our goals?

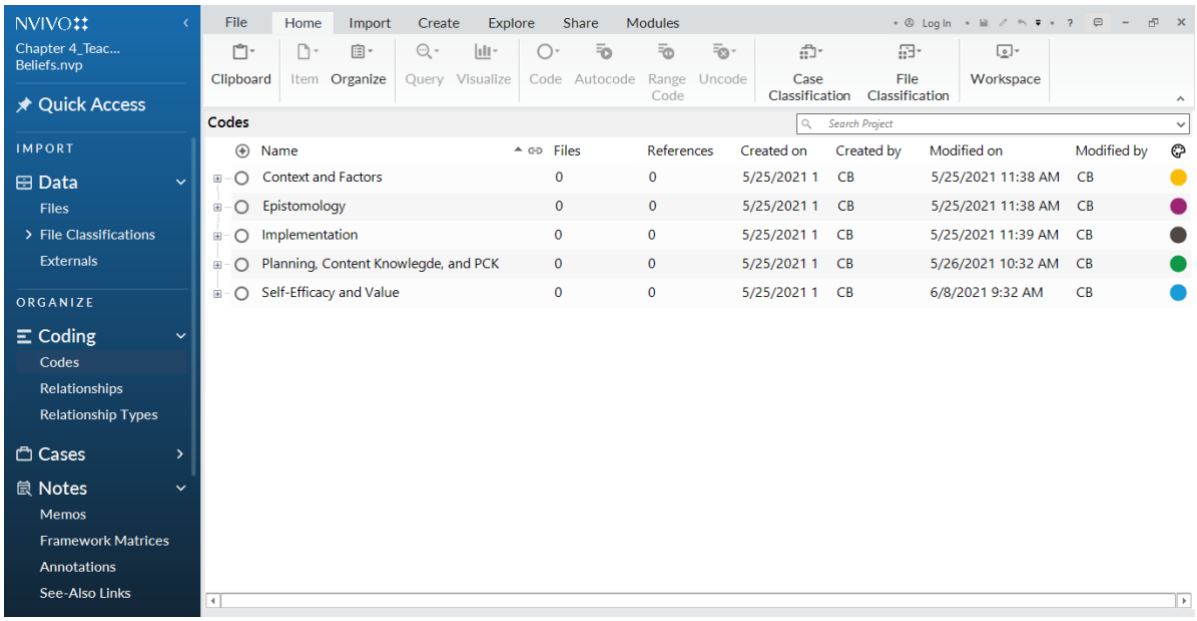
- Criteria
  - Comprehension of INV goals
  - Application of Cognitive Apprenticeship
  - Implementation to support students



Teacher Glows

- Appropriate Implementation of cognitive apprenticeship
  - Flexible use of IREAD
  - Reduction of scaffolds
  - Supporting student discussion
  - Supporting student use of judgments
- Introduction of other historical literacy activities
  - Adapted INV
  - Use of IREAD for other lessons
- Teachers are seeing the benefits of the cognitive apprenticeship
  - Student stamina
  - Student use of quotes and judgments

Appendix 7: NVivo



References

Adkins M. H., & Gavins M. V. (2012) Self-regulated strategy development and generalization instruction: Effects on story writing and personal narratives among students with severe emotional and behavioral disorders. *Exceptionality*, 20(4), 235-249. doi: 10.1080/09362835.2012.724625

Afflerbach, P. A., & VanSledright, B. (2001). Hath! Doth! What! The challenges middle school students face when they read innovative history text. *Journal of Adolescent and Adult Literacy, 44(8),* 696–707. doi: 40018742

Alexander, P. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. Maeher & P. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 4, pp. 213–250). JAI Press.

Anolli, A., Antonietti, A., Crisafulli, L. & Cantoia, M. (2001). Accessing source information in analogical problem-solving. *Quarterly Journal of Experimental Psychology*, *54A*, 237–261. doi: https://doi.org/10.1080/02724980042000093

*Antononiou, F., & Souvignier, E. (2007). Strategy instruction in reading comprehension: An intervention study for students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 5(1), 41-57. doi: 10.17161/fec.v43i8.6690

Anolli, A., Antonietti, A., Crisafulli, L. & Cantoia, M. (2001). Accessing source information in analogical problem-solving. *Quarterly Journal of Experimental Psychology*, *54A*, 237–261. doi: 10.1080/02724980042000093

Bain, R. (2006). Rounding up unusual suspects: Facing the authority hidden in the history classroom. *Teachers College Record*, *108*(10), 2080–2114. doi:/10.1111/j.1467-9620.2006.00775.x

Ball, D. L., & Forzani, F. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497–511. doi: 10.1177/0022487109348479

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis. *Review of Educational Research*, *74*(1), 29–58. doi: https://doi.org/10.3102/00346543074001029

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior* (Vol 4. Pp. 71-81). Academic Press.

*Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, 41(3), 586–598. doi: 10.1037/0022-3514.41.3.586

Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637. doi: 10.3102/00346543074001029

Barton, K. (2005). Primary sources in history: Breaking through the myths. *Phi Delta Kappan, 86*(10), 745-753. doi: 10.1177/003172170508601006

*Berkeley, S., Mastropieri, M. A., & Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *Journal of Learning Disabilities*, 44(1), 18-32. doi:10.1177/0022219410371677

Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995-2006: A meta-analysis. *Remedial and Special Education*, 31(6), 423-436. doi: 10.1177/0741932509355988

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching, 49*(1), 68-94. doi: https://doi.org/10.1002/tea.20446

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional context. *Science Education*, 94(5), 765-793. doi: 10.1002/sce.20402

Booth, M. (1983). Skills, concepts, and attitudes: the development of adolescent children's historical thinking. *History and Theory, 22*(4), 101-117. doi: 10.2307/2505218

Bransford, J., Brown, A. L., & Cocking, R. (Eds.). (1999). *Learning and transfer. In How people learn: Brain, mind, experience, and school* (pp. 39–66). Washington, DC: National Academy Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32–42. doi: 10.3102/0013189X018001032

Butler, C. (2021) *Implementation of a historical literacy curriculum: Teacher fidelity, adaptations and practice* [Unpublished manuscript]. Department of Counseling, Higher Education, and Special Education, University of Maryland.

Butler, C., De La Paz, S., Levin, D., & Felton, M. (2021). *The effects of a cognitive apprenticeship on transfer of argumentative writing in middle school science* [Manuscript submitted for publication]. Department of Counseling, Higher Education, and Special Education, University of Maryland.

Butler, C., & De La Paz, S. (in press). Synthesis on the impact of self-regulated instruction on motivation outcomes for student with or at risk for learning disabilities. *Learning Disabilities Research and Practice.*

*Buzza, D. C., & Dol, M. (2015). Goal setting support in alternative math classes: effects on motivation and engagement. *Exceptional Education International*, 25(1), 35-66. doi: 10.5206/eei.v25i1.7716

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*. 40(2), doi: 10.1186/1748-5908-2-40

Carter, M., Ferzli, M., & Wiebe, E. N. (2007). Writing to learn by learning to write in the disciplines. *Journal of Business and Technical Communication, 21*(3), 278–302. doi: https://doi.org/10.1177/1050651907300466

Chang, N. M. (2006). *Learning to discriminate and generalize through problem comparisons* (Unpublished doctoral dissertation). Pittsburgh, PA: Carnegie Mellon University.

Center on Education Policy (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era.* Washington DC: Center on Education Policy.

*Cerar, N. (2012). Teaching students with emotional and behavioral disorder how to write persuasive essays fluently. (Unpublished doctoral dissertation). George Mason University, Fairfax, Va.

Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist, 47*(3), 177–188. doi: 10.1080/00461520.2012.695709

Coburn, C. E., & Penuel, P. R. (2016). Research-practice partnerships in education: Outcomes, Dynamics, and open questions. *Education Researcher*, 45(1), 48-54. doi: 10.3102/0013189X16631750

Collins, A., Brown, J.S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6–11, 38–46.

Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects. (2010). National Governors Association Center for Best Practices and Council of Chief State School Officers. Accessed July 7, 2014. Retrieved from http://www.corestandards.org/.

Conradi, K., Jang, B. G., & McKenna, M. C. (2014). Motivation terminology in reading research: A conceptual review. *Educational Psychology Review*, 26(1), 127–164. doi: 10.1007/s10648 013-9245-z

Council for Exceptional Children. (2014). *Council for Exceptional Children Evidence Based Practices in Special Education*. Retrieved from https://www.cec.sped.org/

Creswell, J. (2014). *Educational research: planning, conducting, and evaluating quantitative and qualitative research. N*ew York: Pearson College Division.

Chu, L., Sampson V., Hutner T. L., Rivale S., Crawford, R. H., Baze, C. L., & Brooks, H., S. (2019) Argument-driven engineering in middle school science: A exploratory study of changes in engineering identity over an academic year. *Journal of Pre-College Engineering Education Research*, 9(2), 72-84. doi: 10.7771/2157-9288.1249

*Cuenca-Carlino, Y., Freeman-Green, S., Stephenson, G. W., & Hauth, C. (2016). Self-Regulated strategy development instruction for teaching multi-step equations to middle school students struggling in math. *The Journal of Special Education*, 50(2), 75-85. doi:10.1177/0022466915622021

*Cuenca-Carlino, Y., & Mustian, A. L. (2013). Self-regulated strategy development: Connecting persuasive writing to self-advocacy for students with emotional and behavioral disorders. *Behavioral Disorders*, 39(1), 3-15. doi:10.1177/019874291303900102

*Cuenca-Sanchez, Y., Mastropieri, M. A., Scruggs, T. E., & Kidd, J. K. (2012). Teaching students with emotional and behavioral disorders to self-advocate through persuasive writing. *Exceptionality,* 20(2), 71-93. doi:10.1080/09362835.2012.669291

*Daniel, G., E. (2003). Effects of cognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. (Unpublished doctoral dissertation). Ohio State University, Columbus, Oh.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

De La Paz, S. (1999). Self-regulated strategy instruction in regular education settings: Improving outcomes for students with and without learning disabilities. *Learning Disabilities Research and Practice, 14*, 92-106. doi: 10.1207/sldrp1402_3

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, *97*(2), 139–156. doi:10.1037/0022-0663.97.2.139

De La Paz, S. & Butler, C. (2018). Promoting motivated writers: Suggestions for teaching and conducting research with students with learning disabilities and struggling learners. Learning Disabilities: A Multidisciplinary Journal, 23 (2) 56-69. doi: 10.18666/LDMJ-2018-V23-I2-9064.

De La Paz, S., Butler, C., Otarola, J., & Harris, T. (in review) *Improving students' academic literacy through a research practice partnership*]Paper Presentation]. AERA Conference 2022, San Diego, United States.

De La Paz, S., & Felton, M. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Journal of*

Contemporary Educational Psychology, 35, 174-192. doi: 10.1016/j.cedpsych.2010.03.001

De La Paz, S., Felton, M., Monte-Sano, C., Croninger, B., Jackson, C., Deogracias, J., & Hoffman, B.P. (2014). Developing historical reading and writing with struggling adolescent readers: implementation and learning outcomes. *Theory and Research in Social Education*, 42, 228-274. doi: /10.1016/j.cedpsych.2010.03.001

De La Paz, S. & Graham, S. (2002). Explicitly teaching strategies, skills and knowledge: Writing instruction in middle school classrooms. Journal of Educational Psychology, 94, 687-698. doi: 10.1037/0022-0663.94.4.687Eccles, J. S., &Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109-132. doi: 10.1146/annurev.psych.53.100901.135153

De La Paz, S. & McCutchen, D. (2017). Learning to write. R. E. Mayer and P. A. Alexander (Eds.) *Handbook of Research on Learning and Instruction*. (2nd ed.). New York: Routledge Press.

De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Worland, K. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52 (1), 31-52. doi: /10.1002/rrq.147

De La Paz, S. & Levin, D. (2018). Beyond "they cited the text:" Middle school students and teachers' written critiques of scientific conclusions. *Research in Science Education*, 48 (6), 1433-1459. doi: /10.1007/s11165-016-9609-8

De La Paz, S. & Sherman, C. (2013). Revising instruction in inclusive settings: Effects for English learners and novice writers. *Learning Disabilities Research and Practice*, 28, 129–141. DOI: 10.1111/ldrp.12011

De La Paz, S. & Wissinger, D. R. (2017). Improving the historical knowledge and writing of students with or at risk of ld. *Written Communication*, 50(6), 658-671. doi: 10.1177/0022219416659444

De La Paz, S., Wissinger, D. R., Gross, M. & Butler, C. (2021). Strategies that promote historical reasoning and contextualization: a pilot intervention with urban high school students. *Reading and Writing*. doi: 10.1007/s11145-021-10183-0

District of Columbia Public School. (2016). *LEAP Handbook.* Washington DC: Office of Instructional Practice.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312. doi: 10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Duschl, R. A., & Osborne, J. (2002) Supporting and promoting argumentation discourse in science education. *Studies in Science Education,* 38(1), 39-72. doi: 10.1080/03057260208560187

Dweck, C. S. (2008). *Mindset*. New York: Ballantine Books.

*Ennis, R. P., & Jolivette, K. (2014). Using self-regulated strategy development for persuasive writing to increase the writing and self-efficacy skills of students with emotional and behavioral disorders in health class. *Behavioral Disorders*, 40(1), 26-36. doi:10.17988/0198-7429-40.1.26

Engle, R. A. (2006). *Findings about the nature of the whale group's transfer* [Supplement]. Available from the author.

Engle, R., A., & Conant. F. R. (2002). Guidelines for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners' classroom. *Cognition and Instruction, 20,* 399–483. doi: 10.1207/S1532690XCI2004_1

Engle, R. A., Nguyen, P. D., & Mendelson, A. (2011). The influence of framing on transfer: Initial evidence from a tutoring experiment. *Instructional Science,* 39(1), 603–628. doi: 10.1007/s11251-010-9145-2

Engle, R. A., Lam, D. P., Meyer, X. S., & Nix, S. E. (2012). How does expansive framing promote transfer? Several proposed explanations and a research agenda for investigating them. *Educational Psychologist, 47*(3), 215-231. doi: 10.1080/00461520.2012.695678

Felton, M., Crowell, A., Garcia-Mila, M., & Villarroel, C. (2019). Capturing deliberative argument: An analytic coding scheme for studying argumentative dialogue and its benefits for learning. *Learning, Culture and Social Interaction,* doi: 10.1016/j.lcsi.2019.100350

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, 92(4), 694–702. doi: https://doi.org/10.1037/0022-0663.92.4.694

Ferretti, R. P.,MacArthur, C.,& Okolo, C.M. (2001). Teaching for historical understanding in inclusive classrooms. *Learning Disability Quarterly*, *24*, 59–71. doi: /10.1037/0022-0663.92.4.694

Fitchett, P., & Vanfossen, P. (2013). Survey on the status of social studies: Development and analysis. *Social Studies Research & Practice, 8(1),* 1-23.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 471–499). American Psychological Association. doi 10.1037/13274-019

Ford, M. J. (2008). 'Grasp of practice' as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, 17(2), 147-177. doi: 10.1007/s11191-006-9045-7

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, 30(1), 1–32. doi: https://www.jstor.org/stable/4129768

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disability by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*, 216–227. doi: 10.2307/1593674

*Garcia, J. N., & De Caso, A. (2006). Changes in writing self-efficacy and writing products and processes through specific training in the self-efficacy beliefs of students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 4(1), 1-27. doi: 10.1163/9781780529295_021

*García-Sánchez, J.-N., & Caso-Fuertes, A.-M. (2005). Comparison of the effects on writing attitudes and writing self-efficacy of three different training programs in students with learning disabilities. *International Journal of Educational Research*, 43(4-5), 272–289. doi: 10.1016/j.ijer.2006.06.006

*García-Sánchez, J., & Fidalgo-Redondo, R. (2006). Effects of two types of self-regulatory instruction programs on students with learning disabilities in writing products, processes, and self-efficacy. *Learning Disability Quarterly*, 29(3), 181-211. doi:10.2307/30035506

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, *9*(3), 672–693. doi: https://doi.org/***10.1111/tops.12278***

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology,* 15(1), 1–38. doi: 0.1016/0010-0285(83)90002-6

Gee, J. P., & Green, J. L (1998). Discourse analysis, learning, and social practice: A methodological study. *Review of Research in Education*, *23,* 119–169.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, *9*(3), 672–693. doi: https://doi.org/*10.1111/tops.12278*

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds), *Handbook of Educational Psychology.* (pp. 457-478). Erilbaum.

Graham, S., Collins, A. A., & Rigby-Wills, H. (2017). Writing characteristics of students with learning disabilities and typically achieving peers. *Exceptional Children*, *83*(2), 199–218. doi:10.1177/0014402916664070

Graham, S., Delay, G. S., Aitken, A. A., Harris, K. R., & Robinson K. H. (2018). Do writing motivational beliefs predict middle school students' writing performance. *Journal of Research in Reading*, 118, 82-104. doi: 10.1111/1467-9817.12245

*Graham, S., & Harris, K. R. (1989a). Components analysis of cognitive strategy instruction: Effects on learning disabled students' compositions and self-efficacy. *Journal of Educational Psychology*, 81(3), 353–361. doi: 10.1037//0022-0663.81.3.353

*Graham, S., Harris, K. R. (1989b). Improving learning disabled students' skills at composing essays: Self-instructional strategy training. *Exceptional Children*, 56(3), 201-214. doi:10.1177/001440298905600305

Graham, S., Harris, K. R., Kuihara, S. A., & Fishman, E. J. (2017). The relationship among writing behavior, writing motivation, and writing performance with young developing writers. *Elementary School Journal.* 118(1), 82-104. doi: 10.1086/693009

*Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology,* 30(2), 207-241. doi:10.1016/j.cedpsych.2004.08.001

Graham, S., Hebert, M., & Harris, K.R., (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523–547. doi: https://doi.org/10.1086/681947

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Education Psychology,* 104(4), 879-896. doi: 10.1037/a0029185

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., Schneider, S. A., Madden, S., & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal, 48*(3), 647–717. doi: 10.3102/0002831210384839

Grooms, J., Enderle, P. J., Hutner, T., Murphy, A., & Sampson, V. (2016). *Argument-Driven Inquiry in Physical Science: Lab Investigations for Grades 6-8 (Vol. 4)*. NSTA Press.

Grossman p., Hammerness, K,, & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice,* 15(2), 273-289. doi:10.1080/ 13540600902875340

Guskey, T. R. (1986) Staff development and the process of teacher change. *Educational Researcher, 15*(5), 5-12. doi: 10.3102/0013189X015005005

Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research*, 57(2), 283-319. doi: 10.1177/0735633117748416

Harris, K. R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly*, 22, 251–262. doi: 10.2307/1511259

Harris, K. R., Graham, S. (2009). Self-regulated strategy development in writing: Premises evolution, and the future. *British Journal of Educational Psychology*, 6, 113-135. doi: 10.1348/978185409X422542

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: effects of self-regulated strategy development with and without peer support. *American Educational Research Journal*, 43(2), 295-340. doi:10.3102/00028312043002295

Harry, B., Sturges, K. M., & Klingner, J. K. (2005). Mapping the process: An exemplar of process and challenge in grounded theory analysis. *Educational Researcher*, 34(2), 3-13. doi: 10.3102/0013189X034002003

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching 38*(6), 663- 687. doi: 10.1002/tea.1025

Holt, T. (1990). Thinking historically: narrative, imaginative, and understanding. New York: College Entrance Exam Board.

Horn, S. K. (2007). Motivating basic writers through self-assessment and goal-setting. *NADE Digest*, 3(2), 1-11. doi: EJ1097707

Huijgen, T., van Boxtel, C., van de Grift, W., & Holthuis, R. (2017) Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory & Research in Social Education*, 45(1), 110-144. doi: 10.1080/00933104.2016.1208597

Hutner, T. L., Sampson, V., LaMee, A., FitzPatrick D., Batson, A., & Aguilar-Landaverde, J. (2019). Argument-driven inquiry in physics. Arlington, Va: National Science Teaching Association.

Kiuhara, S., Graham, S., & Hawken, L. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136–160. doi:10.1037/a0013097

Kober, N. (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*(5), 810-824. doi: 10.1002/sce.20395

LaDuke, A., Lindner, M., & Yanoff, E. (2016). Content, disciplinary, and critical literacies in the c3 and common core. *Social Studies Research & Practice, 11(3),* 96–111.

Lee, Y., & De La Paz, S., (2021a). Writing scientific explanations: Effects of a cognitive apprenticeship for students with LD and English Learners. *Exceptional Children*, 1-18, doi: 10.1177/0014402921999310

Lee, Y., & De La Paz., S. (2021b). Science writing research for students with and at risk for learning disabilities, and English learners: A systematic review. *Learning Disability Quarterly*. 44(4), 261-274. doi: 10.1177/07319487211018213

Lee, Y., Levin, D., & De La Paz, S. (2021c) "Now I've seen what they can do": How implementing a cognitive apprenticeship can impact middle school science teachers' beliefs and practice. *Science Educator.* 28(1), 10-18 doi: EJ1303123

Lee, J., & Swan, K. (2013). Is the common core good for social studies? Yes, but... *Social Education, 77(6),* 327-330.

Leming, J., Ellington, L., Schug, M., & Dieterle, D. (2009). Social studies in our nation's high schools: A national random survey of high school social studies teachers' professional opinions, values, and classroom practices. Washington DC: The Center for Survey Research and Analysis.

Leinhardt, G. (2000). Lessons on teaching and learning in history from Paul's pen. In P.N. Stearns, P. Seixas, & S. S. Wineburg (Eds.), *Knowing, teaching and learning history: National and international perspectives* (pp. 223-245). New York, NY: New York University Press.

Levesque, S. (2009). Thinking historically: Educating students for the twenty-first century. Toronto, Canada: University of Toronto Press.

Levin, D., De La Paz, S., Lee, Y., & Escola, E. N. (2021). Use of cognitive apprenticeship models of instruction to support middle school students' construction and critique of written scientific explanations and arguments. *Learning Disabilities a Multidisciplinary Journal.* 26(1). doi: 10.18666/LDMJ-2021-V26-I1-10380

Marx, T., Klein, E., Colpo, A., Walden-Doppke, M., Reinhardt, E., & Butler, C. (2020). Coaching for ongoing professional learning within tiered support models. Washington, DC: National Center on Intensive Intervention, Office of Special Education Programs, U.S. Department of Education.

McConachie, S. (2010). *Disciplinary literacy: A principle-based framework*. San Francisco: Jossey-Bass.

*Macarthur, C. A., & Philippakos, Z. (2010). Instruction in a strategy for compare–contrast writing. *Exceptional Children*, 76(4), 438-456. doi:10.1177/001440291007600404

McNeil, K.L. & Krajcik, J. S. (2012). Supporting grade 5-8 students in constructing explanations in science: the claim, evidence and reasoning framework for talk and writing. Boston: Pearson.

Merriam, S. B. (2014). *Qualitative research: A guide to design and implementation*. San Francisco: John Wiley & Sons.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation.* San Francisco: John Wiley & Sons.

Mestre, J. (2003). Transfer of learning: Issues and research agenda: Report of a workshop held at the National Science Foundation.

Miller, S. P., Butler, F. M., & Lee, K. (1998). Validated practices for teaching mathematics to students with learning disabilities: A review of literature. *Focus on Exceptional Children*, *31*(1), 1–24. doi: 10.17161/FOEC.V31I1.6763

*Mills, S. (1999). The effects of instruction in peer-revision on the persuasive writing of students with emotional and behavioral disabilities. (Unpublished doctoral dissertation). George Mason University, Fairfax, Va.

*Miranda, A., Villaescusa, M. I., & Vidal-Abarca, E. (1997). Is attribution retraining necessary? Use of self-regulation procedures for enhancing the reading comprehension strategies of children with learning disabilities. *Journal of Learning Disabilities*, 30(5), 503-512. doi:10.1177/002221949703000506

Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal*, *45*(4), 1045–1079. doi:10.3102/0002 831208319733

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *The Journal of the Learning Sciences*, 19(4), 539-568. doi: /10.1080/10508406.2010.481014

Monte-Sano, C. (2011). Beyond reading comprehension and summary: Learning to read and write by focusing on evidence, perspective, and interpretation. *Curriculum Inquiry*, *41*(2), 212–249. doi:10.1111/j.1467-873X. 2011.00547.x

Monte-Sano, C. (2016). Argumentation in history classrooms: A key path to understanding the discipline and preparing citizens. *Theory into Practice*, 55(4), 311-319. doi: /10.1080/00405841.2016.1208068

Monte-Sano, C., De La Paz, S., & Felton, M. (2014). Implementing a disciplinary literacy curriculum for US history: Learning from expert middle school teachers in diverse classrooms. *Journal of Curriculum Studies*, 46, (4), 540-575. doi: /10.1080/00220272.2014.904444

Monte-Sano, C., De La Paz, S., Felton, M., Worland, K., Yee, L.S., & Carey, R. L. (2017) Learning to teach disciplinary literacy across diverse eighth-grade history classrooms within a district-university partnership. *Teacher Education Quarterly*, *44 (4)* 98-124.

Morgan, P. L., Fuchs, D., Compton, D. L., Cordray, D. S., & Fuchs, L. S. (2008). Does early reading failure decrease children's reading motivation? *Journal of Learning Disabilities, 41*(5), 387-404. doi:10.1177/0022219408321112

Moje, E. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A Call for change. *Journal of Adolescent & Adult Literacy, 52(2),* 96-107. doi: jstor.org/stable/20111747

National Council for the Social Studies. (2013). College, career, and civic life (C3) framework for social studies state standards. Silver Spring, MD: NCSS.

National Council for the Social Studies (2013). The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history. Silver Spring, MD: Author.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2013). Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.

*Nelson, J. M., & Manset-Williamson, G. (2006). The impact of explicit, self-regulatory reading comprehension strategy instruction on the reading-specific self-efficacy, attributions, and effect of students with reading disabilities. *Learning Disability Quarterly*, 29(3), 213-230. doi:10.2307/30035507

Nokes, J.D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education, 38*(4), 515–544. doi:10.1080/00933104.2010.10473438

Nokes, J. & De La Paz, S. (2018). Writing and Argumentation in History Education. Metzger, S. A. and McArthur Harris, L. (Eds.) *International Handbook of History Teaching and Learning*. Wiley-Blackwell.

Nokes, J. D., Doles, J. A., & Hacker, D. J. (2007) Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, 99(3), 492-504. doi: /10.1037/0022-0663.99.3.492

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(5977), 463-466. doi: 10.1126/science.1183944

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 944-1021. doi: 10.1002/tea.20035

O'Shea, A., Booth, J. L., Barbieri, C., McGinn, K. M., Young, L. K., & Oyer, M. H. (2017). Algebra performance and motivation differences for students with learning disabilities and students of varying achievement levels. *Contemporary Educational Psychology,* 50(1), 80-96. doi:10.1016/j.cedpsych.2016.03.003

* Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, 91(2), 230-240. doi:10.1037/0022-0663.91.2.230

Parker, W. C., Mosborg, S., Bransford, J., Vye, N., Wilerson, J., & Abbott, R. (2011). Rethinking advanced high school coursework: Tackling the depth/breadth tension in the AP US Government and Politics course. *Curriculum Studies, 43(4),* 533-559. doi: 10.1080/00220272.2011.584561

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 5437-578. doi: 10.3102/00346543066004543

Pajares, F., & Cheong, Y. F. (2003). Achievement goal orientations in writing: A developmental perspective. *International Journal of Educational Research*, 39 (4-5), 437-455. doi: 10.1016/j.ijer.2004.06.008

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24(2), 124–139. doi: 10.1006/ceps.1998.0991

Pajares, F., & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research*, 90(6), 353-360. doi: 10.1080/00220671.1997.10544593

Perkins, D. N. & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist*, 47(3), 248-258. doi: 10.1080/00461520.2012.693354

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31(6), 459–470. doi: 10.1016/S0883-0355(99)00015-4

Rangland, R.G. (2007). Changing secondary teachers' views of teaching American history. *The History Teacher*, *40*(2), 219–246. doi: 10.2307/30036989

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115(3), 381-400. doi: 10.1037/0033-2909.115.3.381

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction, 30(1),* 86-112. doi: 10.1080/07370008.2011.634081

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–574. doi: 10.1037/0022-0663.99.3.561

*Rooney, J. (1997). The effects of story grammar strategy training on the comprehension, self-efficacy and attributions of learning disabled students. (Unpublished doctoral dissertation). The City University of New York, New York, NY.

Russ, R., Scherr, R., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499-525.

Russell, W. B., & Waters, S. (2010). Instructional methods for teaching social studies: A survey of what middle school students like and dislike about social students instruction. *Journal for Liberal Arts and Sciences, 14(2),* 7-14.

Ryan, R. M., & Deci, E. L. (2000) Self-determination theory and facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*. 55(1) 68-78. doi: 10.1037/0003-066X.55.1.68

Sampson, V., Enderle, P., & Grooms, J. (2013a). Argumentation in science education: Helping students understand the nature of scientific argumentation so they can meet the new science standards. *The Science Teacher*, 80(5), 30–33.

Sampson, V.; Enderle, P.; Grooms, J.; Witte, S. (2013b). Writing to learn and learning to write during the school science laboratory: Helping middle and high school students develop

argumentative writing skills as they learn core ideas. *Science Education.* 97(5), 643-670. doi: 10.1002/sce.21069

Sampson, V., & Murphy, A. (2019). Argument-driven inquiry in third-grade science. Arlington, Va: National Science Teaching Association.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanation. *Cognition and Instruction, 23*(1), 23–55. doi: 10.1207/s1532690xci2301_2

Savery, J. (2006). Overview of problem-based learning: definitions and distinctions. *Interdisciplinary Journal of Problem-Based Learning, 1(1),* 9-20. doi: 10.7771/1541-5015.1002

*Sawyer, R. J., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology*, 84(3), 340–352. doi: 10.1037/0022-0663.84.3.340

Saye, J. & the Social Studies Inquiry Research Collaborative (SSIRC). (2013). Authentic pedagogy: Its presence in social studies classrooms and relationship to student performance on state-mandated tests. *Theory and Research in Social Education, 41(1),* 89–1. doi: 0.1080/00933104.2013.756785

Schiefele, U., Schaffner, E., Moller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, 47(4), 427–463. doi: 10.1002/RRQ.030

*Schunk, D. H. (1983). Developing children's self-efficacy and skills: The roles of social comparative information and goal setting. *Contemporary Educational Psychology,* 8(1), 76-86. doi: 10.1016/0361-476X(83)90036-X

*Schunk, D. H. (1985). Participation in goal setting: Effects on self-efficacy and skills of learning disabled children. *Journal of Special Education,* 19(3), 307-317. doi: 10.1177/002246698501900307

Schunk, D. H., & Cox, P. D. (1986). Strategy training and attributional feedback with learning disabled students. *Journal of Educational Psychology*, 78(3), 201-209. doi:10.1037//0022-0663.78.3.201

*Schunk, D. H., & Rice, J. M. (1993). Strategy fading and progress feedback. *The Journal of Special Education*, 27(3), 257–276. doi: 10.1177/002246699302700301

Schunk, D. H., & Zimmerman, B. J. (1994). Self-regulation of learning and performance: Issues and educational applications. Erlbaum.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (2012). *Motivation and self-regulated learning: Theory, research, and applications*. Routledge. https://doi.org/10.4324/9780203831076

Schwartz, D., L. & Nasir, N. (2003). Transfer of learning. In W. Guthrie (Ed.), *Encyclopedia of education* (2nd ed., pp. 1449–1452). NY: Macmillan.

Seixas, P. (1998). Student teachers thinking historically. *Theory & Research in Social Education, 26*(3), 310-341. doi: 10.1080/00933104.1998.10505854

Seixas, P. (2006). *Benchmarks of historical thinking: A framework for assessment in Canada*. Vancouver, CA: Center for the Study of Historical Consciousness. Retrieved from archive.history benchmarks.ca/documents/benchmarks-historical-thinking-frameworkassessment- Canada

*Sexton, M., Harris, K. R., & Graham, S. (1998). Self-regulated strategy development and the writing process: Effects on essay writing and attributions. *Exceptional Children*, 64(3), 295-311. doi:10.1177/001440299806400301

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78(1),* 40-59. doi: 10.17763/haer.78.1.v62444321p602101

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32(1*), 7-18. doi: 10.1097/TLD.0b013e318244557a

Stahl, C. (1996). What happens when students read multiple source documents in history. *Reading Research Quarterly, 31(4),* 430-456. doi: jstor.org/stable/748185

Stake R. E. (2010). *Qualitative Research: Studying how things work.* New York: Guilford Press; 2010.

Stoel, G. L., van Drie, J. P., & van Boxtel, C. (2015). Teaching towards historical expertise. Developing a pedagogy from fostering causal reasoning in history. *Journal of Curriculum Studies*, 47(1), 49-76. doi: /10.1080/00220272.2014.968212

Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research*. Newbury Park, CA: Sage.

Thacker, E. S., Lee, J. K., & Friedman, A. M. (2016). Teaching with the C3 Framework: Surveying teachers′ beliefs and practices. *The Journal of Social Studies Research,* 41(2), 89-100. doi: 10.1016/j.jssr.2016.08.001

The Nation's Report Card (2019). How did US students perform on the most recent assessment? Retrieved from https://www.nationsreportcard.gov/

Thorndike, E. L. (2009). *Educational psychology*. Charleston, SC: Biblio- Life. (Original work published 1903).

Toste, J. R., Didion, L., Peng, P., Filderman, M. J., & McClelland, A. M. (2020) A meta-analytic review of the relations between motivation and reading achievement for k-12 students. *Review of Educational Research*, 90(3), 420-456. doi: 10.3102/0034654320919352

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2012). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing, 26*(1), 17–44. doi:10.1007/ s11145-012-9379-2

U. S. Department of Education. (2010). Twenty-ninth annual report to Congress on the implementation of the Individuals with Disabilities Education Act, Parts B and C. 2007. Retrieved June 12, 2012, from http://www2.ed.gov/about/reports/annual/osep/2007/parts-b-c/

Vacca, R. T., Vacca, J. L., & Mraz, M. (2014). *Content area reading: Literacy and learning across the curriculum* (11th ed.). New York, NY: Pearson.

VanSledright, B. (2010). The challenge of rethinking history education: On practices, theories, and policy. New York: Routledge.

VanSledright, B., & Kelly, C. (1998). Reading American history: The influence of using multiple sources on six fifth graders. *The Elementary School Journal, 98*(3), 239-265. doi: jstor.org/stable/1002259

VanSledright, B., Kelly, T., & Meuwissen, K. (2006). Oh, the trouble we've seen: Researching historical thinking and understanding. In K. Barton (Ed.), *Research methods in social studies education: Contemporary issues and perspectives* (pp. 207-233). Greenwich, CT: Information Age Publishing.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, 41(1), 19–31. doi: 10.1207/s15326985

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology*, 89(3), 420–432. doi: 10.1037/0022-0663.89.3.420

Wineburg, S. (1991). Historical problem solving: A study on the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 81(1), 73-87. doi: 10.1037/0022-0663.83.1.73

Wineburg, S. (2001). Historical thinking and other unnatural acts: charting the future of teaching the past. Philadelphia: Temple University Press.

Wineburg, S. (2006). A sobering big idea. *Phi Delta Kappan, 87(5),* 401-402. doi: 10.1177/003172170608700514

Wineburg, S., & Wilson, S. M. (1991). Models of wisdom in the teaching of history. *The History Teacher, 24(4),* 395-412. doi: 10.2307/494699

Wissinger, D. R., De La Paz, S., & Jackson, C. (2021). The effects of historical reading and writing strategy instruction with fourth- through sixth-grade students. *Journal of Educational Psychology, 113*(1), 49–67. doi: 10.1037/edu0000463

Wissinger, D. R., & De La Paz, S. (2020) Effects of discipline-specific strategy instruction on historical writing growth of students with writing difficulties. *Journal of Learning Disabilities*, 53 (3), 199-212. doi: /10.1177/0022219420904343

Wong, B., & Butler, D. (Ed. 4). (2013) *Learning about learning* disabilities. Elsevier Inc.

Wong, B. Y. L., Butler, D. L., Ficzere, S. A., & Kuperis, S. (1997). Teaching adolescents with learning disabilities and low achievers to plan, write, and revise compare-and-contrast essays. *Learning Disabilities Research & Practice, 12*(1), 2–15. doi: 1997-03105-001

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

Zimmerman, B. J. (1997). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 3–21). Erlbaum.

Zimmerman, B., & Reisemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22(1), 73–101. doi: 10.1006/ceps.1997.0919

Zimmerman, B. J., & Ringle, J. (1981). Effects of model persistence and statements of confidence on children's self-efficacy and problem solving. *Journal of Educational Psychology*, 73(4), 485–493. doi: 10.1037/0022-0663.73.4.485