Improving Writing Instruction of In-Service Teachers of Students with Disabilities: A Multimedia Approach to Professional Development

---

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

---

by

John Elwood Romig,

B. S., Bob Jones University

M. Ed., Clemson University

May 2018

Abstract

National assessments of writing achievement indicated that a majority of students do not meet grade level expectations (National Center for Education Statistics, 1999; National Center for Education Statistics, 2003; National Center for Education Statistics, 2008; National Center for Education Statistics, 2012). Students with disabilities generally struggle with writing more than students without disabilities (Graham & Harris, 2002; Troia, 2006). The National Commission on Writing (2003) recommended a reform in writing instruction to address the state of writing achievement. Among these recommendations, they recommended professional development in writing instruction and partnerships between universities and school divisions. This study partnered with a local school district to provide professional development in writing instruction. Three 8th grade English teachers participated in this single case multiple baseline research design. The professional development intervention they received, called Content Acquisition Podcast – Professional Development (CAP-PD), consisted of instructional modeling vignettes, customizable curriculum supports, and instructional coaching. CAP-PD was targeted at improving the teachers' use of explicit modeling. The primary dependent variable was the number of modeling implementation markers recorded on the Classroom Teaching Scan that teachers demonstrated during writing instruction. Results demonstrated a functional relationship between the intervention and the number of modeling implementation markers teachers demonstrated. Results maintained above baseline levels for the duration of the study. Teachers indicated they were satisfied by the CAP-PD as a professional development intervention and modeling as an instructional strategy.

Department of Curriculum, Instruction, and Special Education
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, "Improving Writing Instruction of In-Service Teachers of Students with Disabilities: A Multimedia Approach to Professional Development" has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Michael J. Kennedy, Chair

____
Natasha A. Heny

____
Youjia Hua

____
William J. Therrien

__________________________ Date

Dedication

To my grandpa, Elwood Foor, always proud of his "Dr. John."

To my wife, Brianne, always supportive of her "Dr. J."

To my unborn child, always loved by this dad.

Acknowledgements

I am incredibly grateful for the mentorship and friendship of my adviser, Professor Michael Kennedy. Thank you for bearing with my first drafts, adding a second space after all periods, and providing endless pizza. A special thank you to the members of my dissertation committee. In this dissertation and in past projects, Professor Heny was influential in my thinking of writing and writing instruction in a general education context. Professor Hua provided insights into the difficulties of conducting a single case experiment in the area of writing instruction. Professor Therrien provided overall advice about the dependent measures, experimental design, and writing of the dissertation.

Thank you to my colleagues, Hannah Mathews and Emily Wiseman, for their assistance with PLATO scoring. Thank you to Kate Peeples for her assistance in recording observations on short notice, scoring student writing samples, and coding observations with the CT Scan. Thank you to Kiara Samuels for her assistance in scoring student writing samples. Finally, thank you to the Wing Institute for generously funding this study. Without their financial support, this study would not have been possible.

Finally, this dissertation was supported by so many friends and family. The members of my research team past and present (Kat Alves, Wendy Rodgers, Hannah Mathews, Kate Peeples, and Jesse Philips) made going to work every day a pleasure and always jumped to help when needed. A special thank you to my wife, Brianne, who supported me in innumerable ways during this dissertation and my pursuit of this graduate degree; my parents, who supported my education financially and emotionally for over 20 years; and my siblings, who always provided a laugh when needed. Having the support of my family made this dissertation not only possible, but enjoyable.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

According to the National Assessment of Educational Progress (NAEP), the vast majority of secondary students in America struggle with writing. For example, the NAEP writing assessment was administered to samples of 8th and 12th grade students from across the United States approximately every four years dating back to the 1969-1970 school year. This assessment scores students as “below basic”, “basic”, “proficient”, and “advanced.” A score of “proficient” indicates the student has “demonstrated competency” (National Center for Education Statistics, 2011, p. 7). Any score below basic indicates the student is deficient in writing performance according to the NAEP scoring system. On the NAEP, nearly three-fourths of students in 8th and 12th grade failed to reach proficiency on the writing assessment for at least the past four test administrations (since 1998; National Center for Education Statistics, 1999; National Center for Education Statistics, 2003; National Center for Education Statistics, 2008; National Center for Education Statistics, 2012). In addition to results from the NAEP, results from the 2016 SAT indicated that on average students attending public schools score below the threshold for college readiness (College Board, 2016). The plight of writing achievement was summed up by Dana Goldstein in the *New York Times* when she wrote that “poor writing is nothing new, nor is concern about it. More than half of first-year students at Harvard failed an entrance exam in writing – in 1874” (2017, para. 11).

Students With Disabilities

Available evidence suggests that students with disabilities struggle with writing more than their peers without disabilities on writing achievement measures. Recent NAEP administrations included students with disabilities in the administration of the writing assessment. However, results for students with disabilities were not reported separately. That being said, scholars believed students with disabilities struggle with writing more than their peers without disabilities (Graham & Harris, 2002; Troia, 2006).

Very few studies documented the achievement gap in writing between students with and without disabilities. In addition, existing studies were often outdated. However, in one recent example, Gage, Wilson, and MacSuga-Gage (2014) compared the performance of students with emotional and behavioral disabilities (EBD) to their peers without disabilities on the Connecticut State Mastery Test (CMT). The CMT writing assessment included two portions that contributed to the total score. The first portion provided students a writing prompt and 45 minutes to respond to the prompt. This writing sample was assessed using a holistic scoring method, meaning that the score for the assessment indicated the overall strength of the writing sample. Errors in spelling and mechanics were not factored into the holistic score. The second portion of the CMT assessed students' editing and revising abilities. This portion provided students with portions of text with embedded errors and asked students to answer multiple choice questions about the passage to indicate the correct edit or revision. Scores on these two subtests were combined to create the total score for the CMT writing assessment. In the analysis, Gage and colleagues used propensity matching to find peers in the general population otherwise similar to the students with EBD and found that students with EBD performed statistically significantly worse than their matched peers. In other words, students who had an EBD scored lower on the Connecticut writing assessment than their peers who did not have an EBD but were similar in terms of socioeconomic status, race, gender, and other important variables.

Similarly, Englert and Thomas (1987) compared students with learning disabilities (LD) and their peers without disabilities on measures of writing structure. On the writing assessment, they provided students with a starter sentence and a type of paragraph structure (i.e., description, enumeration, sequence, comparison/contrast). Students were asked to finish the paragraph and adhere to the given text structure. Each sentence that students wrote was evaluated in terms of its relationship to the overall topic and text structure. Englert and Thomas found that students with LD scored significantly lower on this measures of writing structure, indicating that students with LD struggled to use a specific text structure when writing. Students in the study were in 3rd-7th grades. There was no treatment effect for grade level, indicating the significant differences in writing performance persist across much of primary school and into secondary school.

Finally, Poplin, Gray, and Larsen (1980) compared the performance of students with LD and those without disabilities in grades 3-8 using the *Test of Written Language (TOWL)*. The *TOWL* produced a written language quotient score that combined subtest scores for vocabulary, thematic maturity, spelling, word usage, style. Students in fifth through eighth grade with LD scored significantly lower on average than their peers on every subtest and the written language quotient of the *TOWL*. These results indicated that students with LD used weaker vocabulary, less mature themes, poor spelling, incorrect word usage, and weak style when writing. Students with LD at each grade level typically performed approximately one standard deviation below the mean for each subtest.

These representative examples of studies comparing students with disabilities to their peers without disabilities indicates a trend in performance: on average, students with disabilities typically struggle with writing more than their peers without disabilities. However, this research base is limited in two ways. First, only a few studies exist that provided supporting evidence of an achievement gap between students with disabilities and their peers. Second, many of these studies are very old and do not reflect changes in curriculum or school contexts that could have an effect on writing achievement today. In summary, writing performance for all students is troubling. The possibility that students with disabilities struggle with writing more than their peers who do not meet grade level standards on average is cause for concern. One avenue for addressing these struggles is the quality and quantity of writing instruction provided to students (Graham & Harris, 2002; Troia, 2006).

Writing Instruction

The present study is designed to improve the writing outcomes for students with disabilities by improving teachers' implementation of an evidence-based instructional practice (i.e., modeling). Teachers will receive a professional development package (i.e., Content Acquisition Podcast-Professional Development; CAP-PD) grounded in a cognitive apprenticeship framework (Collins, Brown, & Newman, 1989).

Current State of Writing Instruction

Graham, Harris, and Larsen (2001) noted that "there is little doubt that children's success as writers is intimately tied to the quality of writing instruction" (p. 75).

However, available evidence from teacher preparation programs indicated that pre-service teachers are not prepared to adequately address students' writing struggles using evidence-based practices. The National Commission on Writing published a report making recommendations for improving student writing outcomes (National Commission on Writing, 2003). Among their recommendations, they called for all pre-service teacher candidates to be taught writing theory and instructional practices. However, over a decade later, teacher preparation programs rarely offer stand-alone writing instruction courses (Myers, et al., 2016). Myers and colleagues also found that teacher educators reported lacking confidence in teaching writing and felt that they did not have adequate time to teach writing. Additionally, Grisham and Wolsey (2011) studied the writing instruction experiences of student teachers and found that the student teachers rarely observed writing instruction in their practicum or internship settings. Results from surveys of in-service teachers supported these findings with most high school (Gillespie, Graham, Kiuhara, & Hebert, 2014) and middle school teachers (Graham, Capizzi, Harris, Hebert, & Morphy, 2014) reporting having little to no preparation to teach writing.

Research is limited in terms of what typical writing instruction looks like in schools. However, available research suggested that writing instruction is often low-quality. Graham et al. (2014) surveyed a random sample of in-service middle school teachers across the country who reported spending little time teaching writing. In the survey teachers also reported using evidence-based practices and adaptations for struggling writers infrequently. Similarly, a national random sample of in-service high school teachers reported commonly using writing activities that require little to no analysis (Gillespie et al., 2014).

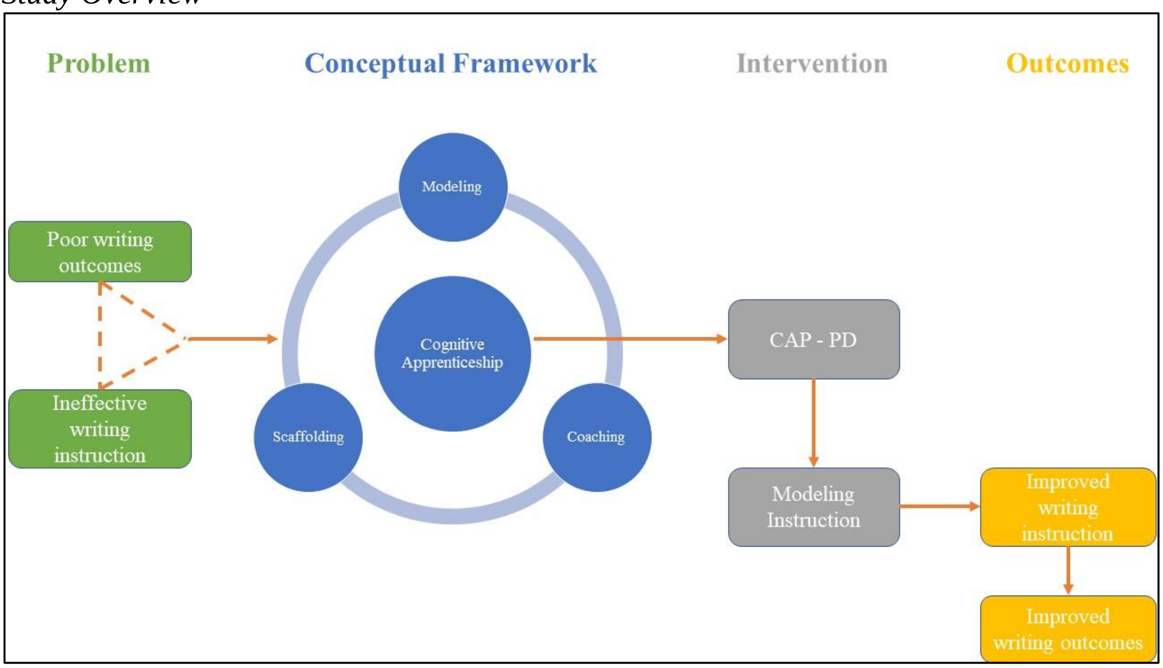
Applebee and Langer (2011) surveyed teachers on their use of writing instruction practices and observed their writing instruction. They found that writing instruction happens infrequently, and when it does happen, it is dominated by activities that involve the teacher doing most of the composing with students filling in blanks, completing worksheets, or using highly formulaic techniques designed to be used for high-stakes assessments. What is even more surprising is Applebee and Langer's observations occurred in schools they perceived as having reputations for excellence in writing instruction. Even in these schools, only half of English classes observed included any writing instruction. Within these observations, only 6.3% of classes included instruction in explicit writing strategies and 5.5% of observations included study of models.

Research on writing instruction is limited because much of it relies on survey research, which may give a limited picture of the instruction that is actually happening in classrooms. With the exception of Applebee and Langer (2011), the field has very little observational evidence of the nature and quality of writing instruction. However, when examining writing outcomes for all students and what is known about the instruction available to them, a reasonable argument can be made that instruction needs to improve to better meet the students' needs.

Conceptual Framework

The current study is designed to improve writing instruction and, consequently, writing outcomes for students with disabilities. The professional development and instruction is grounded in cognitive apprenticeship (Collins et al., 1989). Collins et al. (1989) proposed cognitive apprenticeship as a framework for teaching reading, writing, and mathematics to K-12 students. In recent years, this framework has also been used to guide the development of professional development (PD) packages. This framework is multifaceted and has implications for the content, methods, sequencing, and sociology of instruction. Although the framework is wide-ranging, a cycle of modeling, coaching, and scaffolding form the core of the framework's methods. Although each of these core components has evidence supporting its use in instructional settings, the framework lacks rigorous empirical evidence examining it in its entirety. The current study adds to the research base informing the development of PD grounded in cognitive apprenticeship. See Figure 1 for an illustration of how the framework guides the current study.

Figure 1
Study Overview



Content Acquisition Podcast – Professional Development

CAP-PD is a professional development package grounded in cognitive apprenticeship. CAP-PD has three components: multimedia vignettes, customizable curriculum materials, and instructional coaching based on the Classroom Teaching Scan (CT Scan) (Kennedy, Rodgers, et al., 2017). The multimedia vignettes, or CAP-TVs, are videos incorporating still images and narration that describe instructional practices with embedded modeling videos of teachers demonstrating high-quality implementation of the targeted instructional practice (Kennedy, Rodgers, et al., 2017). Underpinning theory separates CAP-TVs from other readily available videos on the Internet. They are designed using Mayer's (2009) cognitive theory of multimedia learning (CTML). This theory is based on cognitive load theory and is designed to reduce cognitive load and increase learning outcomes. These videos have been experimentally validated for improving in-service and pre-service teacher practice in group experimental and single-case designs (e.g., Kennedy, Hirsch, et al., 2017; Kennedy, Rodgers, et al., 2017). This research base receives a more detailed analysis in chapter two.

The customizable curricular materials, or CAP-Teacher Slides (CAP-TS), are customizable PowerPoint slides designed to prompt teachers' use of evidence-based instructional practices. These slides have images that serve as anchors for teachers and students. These anchor images recur every time a concept is repeated. For example, the same image is used to signal that the teacher is about to ask a review question. The images are organized into an instructional routine that scaffolds teachers in the use of evidence-based practices. The slides include sample narration in the notes section, which teachers can use or ignore. They are also designed using Mayer's (2009) CTML. Kennedy, Rodgers, et al. (2017) provided CAP-TS to teachers. Teachers were not required to use the slides and were encouraged to enhance or alter the slides as they saw fit. Emails from the participating teachers indicated they enjoyed using the materials.

The CT Scan (Kennedy, Rodgers, & Romig, 2017) is an online teacher observation tool developed through an explicit instruction framework. The tool can be viewed at www.classroomteachingscan.com/ctscan/. The CT Scan is primarily a descriptive observation tool that has unique features meant to improve the quality of coaching provided to teachers. The tool records the occurrence, frequency, and duration for a set of customizable teacher behaviors. This data is presented in a series of colored pie graphs and a timeline to aid instructional coaching. CAP-PD uses the CT Scan to describe observed lessons and provide targeted feedback to teachers. Kennedy, Rodgers, et al. (2017) provided the CT Scan coaching to teachers via email.

Each of the components of CAP-PD correspond to the core components of cognitive apprenticeship: modeling (CAP-TV), scaffolding (CAP-TS), and coaching (CT Scan). This study examines the effect of CAP-PD, and by extension cognitive apprenticeship, on teachers' writing instruction. CAP-PD was chosen as the professional development package due to its conceptual and empirical support. However, this professional development package serves as a delivery vehicle for an instructional strategy. When selecting an instructional strategy, it is important that it be as conceptually and empirically supported as the professional development package. Modeling is a practice with conceptual (Collins et al., 1989) and empirical support (Troia, 2006) and was identified as am evidence-based practice for teaching writing to secondary students by the What Works Clearinghouse (Graham et al., 2016). For these reasons, modeling was chosen as the instructional strategy to be delivered to 8th grade English teachers.

Present Study

The present study will use a single case multiple baseline design experiment to examine the effects of CAP-PD on 8th grade English teachers' modeling instruction in writing. The experimental design was developed in accordance with guidelines from the What Works Clearinghouse Single Case Design Standards (Kratochwill et al., 2010). Teachers will be observed daily using the CT Scan. The CT Scan will be the primary dependent variable. Additionally, observations will be scored using the Protocol for Language Arts Teaching Observations (PLATO) to provide descriptive information about the instructional context. Student writing samples will also be collected to provide descriptive context about the participating students. Teachers will be observed a minimum of five times before receiving the intervention (CAP-PD: CAP-TV, CAP-TS, CT Scan coaching). Teachers will receive CT Scan coaching delivered via email daily until the conclusion of the intervention phase. The amount of coaching received will vary due to the nature of the research design, but all teachers will receive a minimum of five coaching emails. Each teacher will remain in the intervention phase until he or she has been observed a minimum of five times. After the intervention phase, teachers will enter the maintenance phase of the intervention. In this phase, they will no longer receive instructional coaching. Teachers will continue to be observed daily until they have been observed for a minimum of five days.

This study is designed to meet the WWC design standards for single case experiments (Kratochwill et al., 2010). In doing so, it will build on the limited empirical evidence examining cognitive apprenticeship for improving teacher practice. Chapter 2 summarizes the literature bases of cognitive apprenticeship, CAP-PD, and writing instruction more specifically while emphasizing the ways this study extends each of these research fields.

CHAPTER II: REVIEW OF THE LITERATURE

The following chapter provides rationale for studying writing instruction for students with disabilities and reviews the empirical literature in effective writing instruction, CAP-PD, and instructional modeling. These reviews were not conducted as literature syntheses or meta-analyses, but they do rely on several recent syntheses in the areas. This chapter begins with a statement of the problem being addressed in this study followed by an analysis of modeling as an evidence-based instructional practice. Finally, this chapter concludes with an overview of CAP-PD and the theory and research supporting it.

Statement of Problem

Data from the National Assessment of Educational Progress (NAEP) administered in 2011 indicated that approximately 75% of eighth- and twelfth-graders did not reach proficiency on this national writing assessment (National Center for Education Statistics, 2012). Results from the 1998, 2002, and 2007 administrations of the NAEP (National Center for Education Statistics, 1999; National Center for Education Statistics, 2003; National Center for Education Statistics, 2008) indicated this poor performance was part of a long-term trend. Although results for students with disabilities were not reported separately in the NAEP results, students with disabilities generally struggle with writing more than their peers without disabilities (Gage, Wilson, MacSuga-Gage, 2014; Graham & Harris, 2002).

A possible hypothesis explaining the poor performance of writing outcomes for students could be the quality of writing instruction provided in schools. A robust evidence base supports the use of explicit instruction for students with disabilities (Archer & Hughes, 2011; Hughes et al., 2017). Specifically within writing, explicit instruction is a key component of effective instruction for struggling writers (Graham & Hebert, 2010; Graham & Perin, 2007; Troia, 2014). However, many students with disabilities are likely not receiving instruction best suited for their educational needs. As of 2014 the majority of students with disabilities (61.8%) spent the 80% or more of their school day in the general education setting, and another 18.9% of students with disabilities spent 40-79% or more of the school day in the general education setting (National Center for Education Statistics, 2016). Observational evidence indicated that instruction available in this setting often did not align to an explicit instruction framework (McKenna, Shin, Ciullo, 2015; Swanson, 2008; Vaughn et al., 2002). Specifically in the area of writing instruction, available observational research indicated that teachers often rely on low-level writing activities and instruction rather than high-quality, evidence-based instruction (Applebee & Langer, 2011).

The lack of high-quality writing instruction can be traced to the lack of preparation to teach writing. A mixed-methods survey by Myers et al. (2016) indicated a stark picture of teacher preparation for writing instruction. Specifically, responses to the survey indicated that the vast majority (72%) of teacher preparation programs did not provide stand-alone methods courses for writing instruction. In response to open-ended items on the survey, some teacher educators felt they lacked adequate time for writing instruction. Grisham and Wolsey (2011) found that teacher candidates reported rarely observed writing instruction occurring in their practicum placements. Similarly, in surveys of secondary teachers, most high school (Gillespie, Graham, Kiuhara, & Hebert, 2014) and middle school teachers (Graham, Capizzi, Harris, Hebert, & Morphy, 2014) reported having little to no preparation to support learning through writing from their preparation program. Based on these results from teacher preparation programs, it is unsurprising that less than half of high school teachers feel inadequately prepared to teach writing (Kiuhara, Graham, & Hawken, 2009) and that the limited evidence available from observing classrooms suggested that writing instruction in schools lacks the desired quality (Applebee & Langer, 2011).

Evidence from the NAEP writing results, observation evidence of writing instruction, and studies of teacher preparation programs point to one conclusive finding: writing instruction must be improved for all students and especially for students with disabilities. The National Commission on Writing (National Commission on Writing, 2003) recognized this need for improved writing achievement and instruction and made several recommendations for the field. One of the recommendations was for schools to provide regular PD opportunities for all teachers to improve writing instruction. Another recommendation was for university and school partnerships to improve teachers' writing instruction. The present study responded to these recommendations by developing a PD package designed to partner with local schools to improve the writing instruction of secondary English teachers.

Evidence-Based Instruction

Writing Instruction

Several reports have identified effective practices for writing instruction (e.g., Graham & Hebert, 2010; Graham & Perin, 2007; Graham et al., 2016; Troia, 2014). A common theme throughout is that writing instruction must be explicit for students with disabilities and students who struggle with writing. Explicit instruction is important for students without disabilities (Graham et al. 2016), but it is especially important for students with disabilities (Graham & Hebert, 2010; Troia, 2014).

Explicit instruction is a multi-faceted instructional framework. Hughes, Morris, Therrien, and Benson (2017) defined five essential components of explicit instruction. First, explicit instruction breaks complex skills into smaller chunks for instruction. Second, explicit instruction uses modeling and think-alouds to highlight important features of content. Third, explicit instruction is highly engaging through the use of prompts that are faded systematically. Fourth, explicit instruction provides ample opportunities for students to respond in multiple ways during instruction. Finally, explicit instruction makes use of practice opportunities that are purposefully designed. The review by Hughes et al. (2017) identified these four components as essential to explicit instruction. However, they found that one of the components was almost synonymous with explicit instruction. The second component described, modeling and thinking aloud, was used synonymously with explicit instruction in over 90% of explicit instruction articles they reviewed. In other words, of all the essential components of explicit instruction, modeling is perhaps the most core to its identity.

Research on writing instruction bears out this relationship between modeling and explicit instruction. Not only is explicit instruction in total an effective framework for teaching writing (Graham & Hebert, 2010; Graham & Perin, 2007; Graham et al., 2016; Troia, 2014), but modeling is a practice with significant research supporting its use (Graham et al. 2016; Troia, 2014). Troia (2014) provided an innovation configuration to assist teacher educators in identifying essential aspects of writing instruction. After reviewing 20 meta-analyses or qualitative syntheses, he identified 36 writing instruction and assessment practices that should be in the repertoire of all teachers. Of these practices, teacher modeling was an essential practice.

Similarly, the What Works Clearinghouse (WWC) published a practice report for teaching writing to secondary students. Of the WWC's three recommendations for teachers, the one with the strongest evidence (i.e., "Strong Evidence" according to WWC standards) was to "explicitly teach appropriate writing strategies using a Model-Practice-Reflect instruction cycle" (Graham et al., 2016).

Graham and Perin (2007) conducted a meta-analysis of writing instruction and found the study of models to be a standalone writing practice with a small positive effect on student writing achievement. Other instructional practices in the meta-analysis included modeling as a component. To illustrate, self-regulated strategy development (SRSD), one of the most widely researched and validated interventions in writing (Graham & Harris, 2003), includes modeling as an essential component of the instructional cycle (Graham & Harris, 1992).

Cross Content Instruction

Modeling is an effective instructional practice not only in the area of writing, but it is also effective across content areas. To illustrate, a wide range of philosophical backgrounds consider modeling an evidence-based instructional practice (i.e., Bandura, 1986; Collins et al., 1989). Further, general education and special education teacher educators support the use of modeling. The high-leverage practice (HLP) movement attempted to identify instructional practices that all beginning teachers should be able to implement after finishing a teacher preparation program that also have a strong likelihood of improving student outcomes (Ball & Forzani, 2009). TeachingWorks High Leverage Practices (n.d.) identified nineteen such practices that can be used across grade levels, subject areas, and contexts. Among them was "explaining and modeling content, practices, and strategies" (p. 1). Similarly, the Council for Exceptional Children and the CEEDAR Center partnered to identify HLPs for special education teachers (McLeskey et al., 2017). Two of the special education HLPs specifically reference modeling as an essential component of the practice. According to McLeskey and colleagues, special education teachers should include modeling when teaching cognitive and metacognitive strategies and when using explicit instruction.

Modeling is an effective practice across contents for students with and without disabilities (McLeskey et al., 2017; TeachingWorks High Leverage Practices, n.d.). It is also an effective practice specific to writing (Graham et al., 2016; Troia, 2014). Therefore, when choosing an instructional practice to be the focus of a PD, modeling is a compelling choice. If the PD is successful and teachers are able to incorporate modeling into their instruction, they have gained a practice that is useful to teach a wide range of students and skills.

Modeling Research Base

As referenced above, the WWC identified modeling as part of a cycle of effective instruction for teaching writing to secondary students. This recommendation had six studies that met WWC's standards without reservations. The following section will review this research base. However, one of the studies excluded students with disabilities and English language learners. Because students with disabilities are the primary focus of this study, Midgette et al. (2008) was not considered in this literature review. The present study builds on the strengths of these studies and improves on their limitations when possible.

Festas et al. (2015) studied the effect of self-regulated strategy development on opinion writing of eighth grade students. Matched pairs of eighth grade teachers in Portuguese schools were randomly assigned to teach using the SRSD instructional process or to continue their standard instruction. Participating students (N = 380) wrote in response to a pretest, posttest, and maintenance (two months post intervention) measure. Writing samples were scored for the number of genre elements and writing output (i.e., number of words written). Results indicated students in the SRSD group included more genre elements on average at posttest and maintenance than the comparison group. Also, students in the SRSD group wrote fewer words on average than students in the comparison group on the posttest. However, significant differences in number of words did not persist on the maintenance measure.

Fitzgerald and Markham (1987) randomly assigned 30 sixth grade students to receive an intervention in writing revision or to read good literature. Instruction in revision was conducted by two trained doctoral students. The intervention cycle included discussion of revision, modeling of revision with think-alouds, and guided practice. The intervention instruction was observed for amount of instructional time, content area of instruction, format of instruction, amount of practice writing, opportunity to revise one's own writing, and type and amount of teacher feedback. The quality of modeling instruction or writing instruction was not assessed during observations. Results indicated students in the experimental group suggested more revisions than peers in the comparison group.

Hubner, Nuckles, and Renkl (2010) examined the effect of writing to learn through learning journals and informed prompting using a 2x2 factorial design. Researchers randomly assigned 70 participants from German secondary schools to one of the four conditions. The instruction for this study lasted for two sessions or approximately 3.5 hours. In two of the experimental conditions, students received an example of a learning journal as a model demonstrating critical features of learning journals. However, no modeling instruction was provided to students. Therefore, this study is not as helpful for informing the present study of modeling instruction.

Kim et al. (2011) conducted a large-scale study examining the effect of cognitive strategy instruction on the analytical writing of Latino dual language learners in grades 6-12. In this study researchers randomly assigned 103 teachers to the experimental or comparison conditions. Participants in the experimental condition participated in the Pathway Project, a comprehensive PD project. PD included instruction in how to model the cognitive strategies, specifically modeling for the students how, when, and why to use the strategies. Project Pathway also included curriculum materials, guided practice, and independent practice. Teachers in the comparison group received a PD focusing on interpreting test data, using test data to improve school scores, and understanding the English Language Arts textbook. Students in the Project Pathway group performed significantly higher on average than students in the comparison group on measures of writing quality.

Page-Voth and Graham (1999) investigated the effect of goal-setting by randomly assigning 30 students with learning disabilities to one of three experimental conditions. The conditions included a goal-setting condition, a goal-setting plus strategy condition, and a control condition. In the goal-setting plus strategy condition, a project-based instructor reviewed the basic parts of a good essay and modeled how to use the goal-setting strategy when writing. Participants received six sessions of instruction. Dependent measures included essay writing samples from each student and a modified self-efficacy scale. Results indicated participants in the goal-setting group and goal-setting plus strategy group included more supporting reasons in their essays when compared to students in the control group. Students in the two experimental groups also wrote longer and higher quality essays than students in the control group. However, differences between the goal-setting group and the goal-setting plus strategy group on the number of essay elements, length or essay, or quality of essay were not statistically significant. Also, there were no statistically significant differences on the measure of students' self-efficacy measure.

One limitation of the studies supporting the WWC's recommendation is that modeling and other components of the intervention were not implemented in isolation. For example, studies of self-regulated strategy development (SRSD) include modeling, but in these studies, modeling is part of a much larger and complex instructional process (e.g., Festas et al., 2015). Other non-SRSD studies (Fitzgerald & Markham, 1987; Kim et al., 2011) that included modeling also included other instructional elements such as curriculum materials, guided practice, discussion, and activating background knowledge. From these studies, it is difficult to make conclusive statements about the effectiveness of modeling on students' writing.

Relatedly, a second limitation is that studies did not provide a detailed description of modeling instruction used in the studies. Presumably, these studies did not provide a detailed description of modeling because modeling was part of a larger intervention package. However, the limitation remains that practitioners and researchers wanting to capitalize on the positive effects of these studies do not have enough information available to adequately repeat the instructional methods. For example, Page-Voth and Graham (1999) only briefly referenced modeling instruction by saying, "one of the five questions… was used to model the strategy used by the goal setting plus strategy group" (p. 233). Obviously, this statement is not enough to replicate in another research study or to inform the type of modeling instruction practitioners should expect to lead to similar results.

A third limitation is that most studies lacked a broad measure of writing instruction or modeling quality. Some studies relied on dependent measures closely aligned to the intervention or treatment validity measures (i.e., Festas et al., 2015; Page-

Voth & Graham, 1999). Fitzgerald and Markham (1987) observed instruction more broadly, describing instructional time, content area of instruction, and other general aspects of instruction without determining quality of instruction. Some did not report observing instruction in the comparison condition at all (i.e., Festas et al., 2015; Fitzgerald & Markham, 1987; Page-Voth & Graham, 1999). Therefore, it is difficult to say what aspects of instruction led to the improved student outcomes.

Defining Modeling

The studies in the WWC practice report were used as a starting place to develop a definition of modeling and to identify markers of high-quality implementation of modeling. However, as discussed in the study limitations, these studies rarely described how modeling was done and focused more attention on what was modeled. Although the content of modeling is certainly important, for the purposes of developing a PD package aimed at improving teachers' modeling, how something was modeled is more valuable.

Because of the lack of modeling description in the studies, the definition and description of modeling's essential components used in this study relied on theory books and practitioner-oriented textbooks. Specifically, Bandura (1986) and Collins et al. (1989) described modeling from a theoretical perspective. Practitioner-oriented textbooks using an explicit instruction framework (e.g., Archer & Hughes, 2011; Mercer & Mercer, 2005) provided descriptions of modeling with varying degrees of empirical support. The descriptions of modeling in these textbooks and theory books were translated into a modeling definition and descriptors (Implementation Markers of modeling). See Table 1 for a list of modeling components and citations supporting each element.

Table 1
Modeling Checklist Items

Checklist item	Examples
Cue students' attention to determining features of skill	Teacher explicitly tells students what features of the model to pay attention to.
Explain when to use modeled skill	Teachers tells students when (i.e., in what settings, scenarios, writing tasks) to use the modeled skill.
Step-by-step demonstration of overt procedure for learning or applying strategy	Teacher provides a step-by-step demonstration of a skill or strategy.
Provide examples and non-examples	Model includes examples of the skill being used and non-examples of the skill.
Monitor understanding	Teacher monitors student understanding of the modeled skill by asking students to demonstrate understanding in some way.
Provide corrective feedback	Teacher provides feedback to students that helps shape their understanding of the skill or process (i.e., more than simply saying "yes" or "good job").
Note organization, relationships, and clues in the new material that elicit learning strategies	Teacher explicitly identifies relationships between new material
Provide multiple models	Teacher provides more than one model of the skill or process.
Gradual release of responsibility	Teacher begins heavily teacher-directed modeling and gradually allows students to contribute ideas to the model as appropriate based on student understanding.
Provide anticipated benefits of modeled skill/process	Teacher tells students why to use the modeled skill or process by explaining the anticipated benefits the skill will have for students' writing.
Connect modeled skill to expected student behavior	Teacher explicitly states that students should repeat the modeled skill in their writing.
Variety of models	Teacher provides a variety of models including models ranging in skill and diversity.
Maintain lively pace	Teacher provides an appropriately brisk pace.

Enthusiasm	Teacher demonstrates enthusiasm when modeling skill (e.g., excited voice, smiling, affect).
Think aloud to demonstrate covert processes	Teacher makes mental processes visible by thinking aloud while modeling.

Measuring Modeling Quality

As mentioned in the WWC studies' limitations, finding a measure of modeling quality was nearly as difficult as finding a description of modeling. None of the studies in the WWC report published a broadly applicable observational measure of teachers' writing instruction. Also, most widely-used teacher observation tools do not measure modeling quality. Of all the teacher observation tools included in the Gates Foundation Measures of Effective Teaching (MET) study (Kane & Staiger, 2012), only one tool, the Protocol for Language Arts Teaching Observation (PLATO; PLATO 5.0, 2017), measured modeling explicitly.

PLATO. The PLATO was developed by Pam Grossman and colleagues at Stanford University to conduct research in language arts classrooms (Grossman, Cohen, & Brown, 2014). PLATO uses an explicit instruction framework to measure instructional quality. When measuring instruction with the PLATO, observers divide lessons into 15-minute segments. The teacher then scores the segment on 13 elements of instruction. Each element is rated on a 4-point scale (1 = "provides no evidence"; 2 = "provides limited evidence"; 3 = "provides evidence with some weaknesses"; 4 = "provides consistent strong evidence"). These segment scores are then averaged to create the PLATO composite score. Therefore, the total composite score is on a 4-point scale.

Because the PLATO uses an explicit instruction framework for evaluating instruction and explicitly includes modeling as one of the 13 components, it is a helpful measure for this study and will provide valuable data as a secondary measure of teacher quality. However, given the nature of the experimental design in this study, PLATO cannot be the primary dependent measure of teacher quality. PLATO's composite score presents two challenges when used in a single-case design experiment. First, because the segment scores are averaged across the lesson, the composite score could potentially mask small yet meaningful improvements in instructional quality. Second, because the composite score is on a 4-point scale, there is little room to detect functional relationships between the independent and dependent variable. Because single-case designs use visual analysis to detect functional relationships, dependent measures must be able to demonstrate meaningful and socially significant (Horner et al., 2005) changes in performance. On a 4-point scale, teachers could make considerable improvement in teaching practice on a number of domains before moving their composite score enough to be detected visually in a single-case design. However, even with these limitations, the PLATO is a high-quality, validated observational tool that was used to provide more context related to the instruction that was not present in the CT Scan.

CT Scan. The Classroom Teaching (CT) Scan (Kennedy, Rodgers, & Romig, 2017) will be used as the primary dependent measure in this study. The CT Scan was developed with an explicit instruction framework by Michael Kennedy and his colleagues at the University of Virginia. The CT Scan is an Internet-based observation tool accessible on a computer or mobile device. It was developed to primarily observe instruction for students with disabilities. However, it is a cross-categorical tool able to observe instruction in all content areas. Although it can be used to observe general or special education classes, research using the CT Scan has focused on observing inclusive general education courses.

Rodgers (2017) used a metaphor of a building to describe how the CT Scan's framework captures instruction. In this metaphor, individual student needs and matching curriculum form the foundation of the building. Instructional strategies (e.g., modeling) form the structure of the building, and teacher-student interactions form the core of the building. Finally, the building exists within the classroom environment (both the school environment and the home environment where students reside).

The CT Scan differs from many traditional observation tools in that it is primarily descriptive, not evaluative. When observing with the CT Scan, observers record the instructional practices used and the CT scan reports the duration and frequency of these teaching behaviors. At the broadest level, observers record the general category of instruction (e.g., general content instruction, writing instruction, mathematics instruction, and classroom management). Under each category, the observer records the specific instructional practice being used (e.g., modeling, activating prior knowledge, facilitating student presentation). Finally, under each practice the observer selects the implementation markers (IMs) that the teacher demonstrated while implementing the practice. The IMs for modeling are listed in Table 1. Providing an explicit cue and explaining expected benefits of the skill are two examples of IMs for modeling instruction. IMs are recorded as present or absent during an instructional practice. Their duration is not recorded.

Although the CT Scan is not primarily an evaluative tool, it does capture quality of instructional practice through the IMs. IMs are developed from the research base supporting a given instructional practice. Quality of each practice is measured by the percent of IMs teachers include as part of each practice. For example, modeling has 16

IMs including providing a think-aloud when modeling (see Festas et al., 2015; Fitzgerald & Markham, 1987; Goeke, 2009; Mercer & Mercer, 2005). Therefore, when observing modeling instruction, an observer would indicate whether or not a teacher made thinking processes visible to students (i.e., think-aloud) along with the 15 other IMs. The percent of IMs indicates the percentage of fidelity that the teacher used a given instructional practice. These IMs can be customized by users. For this study, the IMs will reflect the descriptions of modeling instruction available in explicit instruction textbooks and modeling theory books. See Table 1 for the IMs and the citations supporting their inclusion.

In summary, modeling is the focus of the PD in this study because it has a strong evidence-base for use with students with and without disabilities to teach a wide range of academic skills including writing. However, little research provided insight into how exactly modeling should be implemented. This study attempted to clearly define how modeling should be implemented while simultaneously increasing teachers' use of modeling via a multimedia-based PD package. This PD package, CAP-PD, is based on components of effective PD, grounded in a theoretically-sound framework (cognitive apprenticeship), and has preliminary evidence supporting its use for improving teachers' practice (Kennedy, Rodgers, et al., 2017).

Content Acquisition Podcasts for Professional Development (CAP-PD)

CAP-PD is a multimedia-based PD package combining several elements of effective PD (see Darling-Hammond et al., 2017; Guskey & Yoon, 2009). This package was effective for improving in-service middle school teachers' vocabulary instruction (Kennedy, Rodgers, Romig, Lloyd & Brownell, 2017). The package includes CAP-TV

(Content Acquisition Content Acquisition Podcasts – Teacher Videos), CAP-TS (CAP-Teacher Slides), and instructional coaching based on teacher observations using the Classroom Teaching (CT) Scan. The following sections further describe these components and the research base supporting them.

CAP-T

Content Acquisition Podcasts for Teachers (CAP-T) form the foundation of CAP-PD. They are enhanced podcasts combining still images with narration (Kennedy et al., 2012). Technically, they are not podcasts in the strictest sense of the term because they do not rely on really simple syndication (RSS) for distribution; however, they combine the strengths of traditional podcasts with sound theoretical principles and evidence-based instructional practices (Kennedy, 2013). Mayer's (2009) cognitive theory of multimedia learning (CTML) guides the looks and sounds of CAP-Ts, and principles of explicit instruction (e.g., Archer & Hughes, 2011) guide the design of embedded instructional practices. A link to an example of a CAP-T defining functional behavior assessments can be seen here: https://vimeo.com/111015222.

Cognitive Load Theory. The CTML (Mayer, 2009) and accompanying instructional design principles (Mayer, 2008) are grounded in cognitive load theory (Chandler & Sweller, 1991). Cognitive load theory considers the amount of cognitive resources expended when learning (Sweller, Ayres, & Kalyuga, 2011). Cognitive load theorist argue learning should be designed in such a way to limit the amount of cognitive resources required to process information. The triarchic model of cognitive load includes extraneous processing, intrinsic processing, and germane processing (DeLeeuw & Mayer, 2008). Theoretically, these three factors contribute to the overall cognitive load required for learning. According to DeLeeuw and Mayer (2008), extraneous processing involves all cognitive processes that do not support the learning objective. Distracting or poorly laid out multimedia can increase extraneous processing. CAP-Ts reduce extraneous processing by limiting the amount of text on the screen and including only essential information in the podcasts.

Intrinsic processing is the processing inherent to learning the material. Intrinsic processing can be effected by the complexity of the topic, or “the number of interacting elements that must be kept in mind at one time” (DeLeeuw & Mayer, 2008, p. 223). CAP-Ts limit intrinsic processing by keeping podcasts short with recordings focused on a central idea. Content is also presented in an explicit instruction framework that displays information in a logical and sequential manner (Archer & Hughes, 2011).

Germane processing involves the processing required to organize learned material and connect it to background knowledge (DeLeeuw & Mayer, 2008). Germane processing can be affected by the learner’s motivation to learn the material and background knowledge available. If intrinsic and extrinsic processing are low, more cognitive resources are available for germane processing. When creating CAP-Ts, providing necessary background information and a compelling rationale for learning the material should increase germane load.

CTML. Mayer (2008) identified 12 instructional design principles that have each been experimentally validated to improve learning outcomes. Each of the principles is targeted to reduce extraneous processing, manage intrinsic processing, or foster germane processing (Mayer, 2009). The CTML (Mayer, 2009) organizes these principles into a multimedia instructional design theory. This theory does not have the level of empirical evidence supporting it as the principles do individually. Therefore, more research is needed using this theory to create multimedia materials to determine if these principles collectively reduce cognitive load and lead to improved learning outcomes. The following section describes each of the 12 principles and applies them to CAP-Ts.

Five principles are aimed at reducing extraneous processing (Mayer, 2008). The coherence principle "reduce[s] extraneous material" (p. 763). CAP-Ts achieve this principle by including only material relevant to the specific learning target for each podcast. The signaling principle provides a cue or other mechanism to "highlight essential material" (p. 763). CAP-Ts achieves this principle by proving explicit visual and audio cues when each section is beginning and when important information is being covered. The redundancy principal removes competing narration and on-screen text. On-screen text should be minimal and should not be different from what the learner hears in the narration. CAP-Ts achieve this principal by carefully selecting on-screen text that aids understanding. The spatial contiguity principle places on-screen text and images in close proximity. CAP-Ts achieve this principal by placing on-screen text or labels near accompanying pictures. The temporal contiguity principle "presents corresponding narration and animation at the same time" (p. 763). CAP-Ts achieve this goal by using text and pictures that correspond to the narration.

Three principles are aimed at managing intrinsic processing (Mayer, 2008). The segmenting principle chunks information into "learner-paced segments (p. 765). CAP-Ts achieve this principle by segmenting videos into chunks that are signaled by clear beginning and ending points. Learners are also encouraged to pause CAP-Ts as necessary. Finally, when information would create a CAP-T longer than 15-20 minutes, this material is broken into separate podcasts. The pretraining principle "provide[s] pretraining on the name, location, and characteristics of key components" (p. 765). CAP-Ts achieve this principle by beginning each podcast with an explicit statement of the purpose of the podcast and an advance organizer to guide learning. The modality principle minimizes the amount of text on screen and emphasizes spoken text instead. CAP-Ts accomplish this goal by relying on still images and narration while only rarely including written text when necessary.

Two principles are aimed at fostering germane processing. The multimedia principle presents narration with images rather than narration alone. CAP-Ts achieve this principle by including illustrative images with all narration. The personalization principle "present[s] words [narration] in conversational style rather than formal style" (p. 766). CAP-Ts achieve this principle by using a conversational tone throughout the podcast recording.

Empirical Evidence Supporting CAP-Ts. Although each of Mayer's (2008) principles has research supporting it in isolation, little research has examined the CTML as a coherent theory for reducing cognitive load and improving learning outcomes. Some CAP-T research has begun to fill this research gap. Kennedy et al. (2016) conducted an experiment to teach pre-service teachers essential information regarding functional behavior assessments (FBA). Participants were randomly assigned to watch a CAP-T or participate in a live-lecture. The PowerPoint for the lecture condition was created to simulate a traditional lecture presentation; it did not adhere to the CTML. To assess the effect of instruction using the CTML on participants' perceived cognitive load, participants responded to the NASA-TLX, a six-item survey asking participants questions to respond a scale from low to high (see Hart, 1988). In this study participants in the lecture group reported significantly higher self-reported perceived cognitive load scores on average when compared to the CAP-T group.

Romig et al. (2018) also used the NASA-TLX to measure participants' perceived cognitive load scores. This three-group randomized control trial compared learning via lectures, CAP-TVs, and articles. The lecture condition used a PowerPoint that was identical to the CAP-TV slides, meaning it was also designed using Mayer's (2009) CTML. Results from the NASA-TLX indicated that participants in the CAP-TV group and lecture group did not have statistically significant differences in perceived cognitive load. On average, cognitive load for each of these groups was lower than the article group. These findings indicated that when instruction is guided by the CTML, whether in CAP-T format or not, participants' perceived cognitive load may be lower than when learning via text. This study and Kennedy, Hirsch et al. (2016) are the only CAP-T/TV studies to evaluate the effect of CAP-Ts on reducing perceived cognitive load. However, a more substantial research base supports CAP-Ts for improving knowledge outcomes.

CAP-Ts have strong evidence supporting their use as tools to improve pre-service teacher candidates' content knowledge. Currently, twelve published, experimental or quasi-experimental studies support the use of CAP-Ts for improving pre-service teachers' knowledge or application of content. Beginning CAP-T research compared learning via CAP-Ts to learning via text-based reading assignments. These studies (Driver, Pullen, Kennedy, Williams, & Ely, 2014; Ely, Kennedy, Pullen, Williams, & Hirsch, 2014; Hart & More, 2013; Kennedy, Driver, Pullen, Ely, & Cole, 2013; Kennedy et al., 2012; Kennedy et al., 2016; Kennedy, Newton, Haines, Walther-Thomas, & Kellems, 2012;

Kennedy, Thomas, Aronin, Newton, & Lloyd, 2014) unanimously found that learning via CAP-Ts led to higher levels of comprehension on researcher-created posttests with large effect sizes (d = .72-1.14). Two of these studies examined whether significant differences in comprehension persisted on maintenance measures and found significant differences favoring the CAP-T group (Kennedy et al., 2013; 2016).

In addition to comparisons with text-based learning, research comparing CAP-Ts to lecture-based learning suggested comprehension outcomes were at least equal and, in some cases, were superior for students in the CAP-T group. For example, Kennedy et al., (2016) compared learning via CAP-Ts to a traditional lecture presentation. The lecture presentation used a PowerPoint that violated principles of Mayer's (2009) CTML. Results indicated no significant difference between the two instructional conditions (lecture and CAP-T) on the researcher-created measure of FBA knowledge. Participants in both groups significantly improved from pretest to posttest. Alternatively, in a conceptual replication, Hirsch, Kennedy, Haines, Thomas, and Alves (2015) found that participants in the CAP-T group significantly outperformed peers in a lecture condition on a comprehension and application measure adapted from the previous FBA study.

CAP-TV

These studies showed that CAP-Ts consistently outperformed text-based learning on measures of comprehension. In comparison to lecture-based learning, two studies demonstrated that pre-service teachers in the CAP-T group performed at least as well as peers in a lecture comparison condition, and they outperformed peers on average in another study. Although improving knowledge of teachers is an important outcome, a main goal of teacher training (i.e., pre-service preparation and in-service PD) is to change teaching practice. To that end, researchers developed a variation of CAP-Ts by embedding modeling videos into the CAP-T instructional format. These enhanced CAP-Ts were dubbed CAP-Teacher Videos (CAP-TV) and were aimed at improving teacher practice.

Research into CAP-TVs indicated findings similar to CAP-T research. Ely, Kennedy, Pullen, Williams, and Hirsch (2014) conducted a randomized control trial of pre-service teachers comparing CAP-TVs to text-based learning. This experiment taught pre-service teachers the Intensifying Vocabulary Instruction (IVI) approach to instruction. They found that on average participants in the CAP-TV group significantly outperformed peers reading a practitioner-oriented article on a measure of vocabulary instruction knowledge and, more significantly, on a measure of vocabulary teaching performance (i.e., an IVI fidelity checklist). As indicated by the fidelity checklist, participants in the CAP-TV group included more elements of IVI during storybook reading activities and after reading activities. Effect sizes for CAP-TV group on the teaching performance checklist were moderate to large (d = .65-1.14).

In a follow-up study, Ely, Pullen, Kennedy, and Williams (2015) conducted a single-case experiment teaching in-services teachers to use IVI. Three teachers in this experiment watched a CAP-TV once that explained and modeled IVI. Using a multiple-baseline design, researchers saw increases in performance on the IVI fidelity checklist after teachers received the intervention (CAP-TV). However, according to the What Works Clearinghouse (WWC) standards for single-case designs (Kratochwill et al., 2010), this study did not meet standards with reservations. One teacher had only two baseline data points collected. The WWC standards require that a multiple-baseline design include at least three baseline collection points for each participant in order to "meet standards with reservations" (p. 29).

Romig et al. (2018) conducted a randomized experiment teaching pre-service teachers the essential components of the "model it" stage within self-regulated strategy development (SRSD). This three-group experiment compared CAP-TVs to a lecture-based comparison group and an article-based comparison group. Similarly to previous research on CAP-T and -TVs, results indicated the CAP-TV group significantly outperformed the article comparison group on a researcher-created measure of SRSD knowledge. In comparison to the lecture group, results were similar to findings from Kennedy et al. (2016) in that the CAP-TV group did not significantly outperform the lecture group on the knowledge measure. However, on a researcher-created measure of SRSD implementation of the "model it" stage, the CAP-TV group significantly outperformed the lecture group and the article group. Additionally, with only one exception, the CAP-TV outperformed both the article group and the lecture group on each individual checklist item. These results suggest viewing a CAP-TV can significantly improve pre-service teachers' modeling of writing.

These three studies were the only studies of CAP-Ts or -TVs that examined the effect of solely viewing a CAP-T or -TV on pre- or in-service teacher practice. However, when this limited number of studies is viewed in context of broader CAP-T research, the collective results indicated that CAP-Ts and -TVs lead to superior comprehension when compared to article-based learning and at least equal, and sometimes superior, comprehension outcomes when compared to lecture-based learning. The most promising results from the CAP-TV studies was that viewing a CAP-TV once can have a significant impact on teachers' instruction. However, significant room for growth remained. In Romig et al. (2018) participants in the CAP-TV group earned only half of the possible points on the implementation checklist. In Ely et al. (2014), participants earned approximately 75% of possible points on the fidelity checklist. Participants in Ely et al. (2015) earned a higher percentage of points on the fidelity checklist but still failed to reach 100% fidelity. These results indicated that other supports in combination with viewing a CAP-TV were necessary to achieve high-quality implementation of complex teaching strategies.

CAP-TVs with Supports

Emerging research on CAP-TV supports focused on instructional coaching. Kennedy, Hirsch, et al. (2017) found that their intervention of CAP-TVs plus coaching increased the number of teacher-provided opportunities to respond, praise statements, and precorrections. In comparison to teachers receiving a traditional one-day PD session and no coaching, results also indicated higher rates of on-task behavior for students in the intervention condition. Participants received one coaching session after viewing the CAP-TV.

Kennedy, Rodgers, et al. (2017) combined CAP-TVs with coaching and supplemental instructional materials. This study investigated the ability of this intervention package (i.e., CAP-TV, coaching, and supplemental materials) to increase the number of evidence-based vocabulary practices used by three middle school science teachers. Results of this single-case multiple baseline experiment indicated a functional relationship between the intervention package and the number and duration of evidence-based vocabulary practices.

The coaching provided in these two studies differed in three important ways. First, Kennedy, Rodgers, et al. (2017) delivered coaching via email, and Kennedy, Hirsch, et al. (2017) delivered coaching in-person. Second, Kennedy, Rodgers, et al. (2017) provided daily coaching emails to teachers (range of 8-14 coaching emails per teacher); however, Kennedy, Hirsch et al. (2017) provided one coaching session to participating teachers. Finally, the two studies differed in the teaching practices that were the focus of the PD. Kennedy, Rodgers et al. (2017) focused on improving the number of evidence-based vocabulary practices science teachers used. The targeted evidence-based vocabulary strategies were student-friendly definitions of terms, examples and nonexamples of terms, morphological approaches, semantic relationships among terms, and high-quality discussions of terms. Kennedy, Hirsch, et al. (2017) focused specifically on three evidence-based behavior management practices: teacher-directed opportunities for students to respond, behavior specific praise statements, and pre-corrections.

Although these studies differed significantly in terms of the dosage of coaching provided to teachers, it is also important to note that the difficulty of targeted practices differed significantly between the studies. Although not necessarily intuitive, increasing the number of opportunities for student respond, behavior-specific praise statements, and precorrection statements (i.e., Kennedy, Hirsch, et al., 2017) is decidedly easier than adding five complex instructional strategies using an explicit instruction paradigm (i.e., Kennedy, Rodgers, et al., 2017). Kennedy, Rodgers et al. (2017) did not collect data on student outcomes. However, Kennedy, Hirsch et al. (2017) found that students whose teachers were in the experimental condition had higher rates of on-task behavior on average when compared to peers in the comparison condition. Future CAP-PD research should carefully consider the difficulty of targeted practices within the PD and dosage of coaching necessary to achieve significant gains in teacher outcomes and student outcomes.

As discussed previously, grounding in theory and research has been a strength of CAP-PD from its inception. Mayer's (2009) CTML guides the creation of CAP-TVs and CAP-TSs. Additionally, development of CAP-PD as a whole is guided by the cognitive apprenticeship framework (Collins, Brown, & Newman, 1989). This framework, its components, and limited research base are described below.

Cognitive Apprenticeship

Recent iterations of CAP-PD relied on cognitive apprenticeship (Collins, Brown, & Newman, 1989) as a helpful conceptual framework for designing a PD intervention. This framework posits applications to the content, methods, sequencing, and sociology of learning. Collins et al. (1989) proposed cognitive apprenticeship as a framework for teaching K-12 students the skills of reading, writing, and mathematics. However, as demonstrated below, this framework can also be applied to learning on PD contexts. The following sections summarize Collins et al. (1989) descriptions of content, methods, sequencing, and sociology from a cognitive apprenticeship perspective and apply these aspects to PD learning.

Content. Collins et al. (1989) defined content as "knowledge required for expertise" (p. 477). This definition is helpful for those developing PD and making decisions about what content to include. Within the cognitive apprenticeship framework, content delivered in a PD can be limited to only that necessary for expertise in a given area. Within the realm of content, they described domain knowledge, heuristic strategies, control strategies, and learning strategies. Each of these aspects of content is described below. It is important to remember that this framework was proposed as a framework for teaching K-12 students. Many of these areas are unexplored in PD contexts. Currently, empirical evidence cannot inform to what extent, if any, these aspects of content apply to teacher learning. Although a limitation for the current study, this framework provides several careers worth of research questions to pursue for those interested in teacher PD.

Domain knowledge. Collins et al. (1989) defined domain knowledge as "the conceptual and factual knowledge and procedures explicitly identified with a particular subject matter" (p. 477). In a PD context, this type of knowledge would include information or descriptions about teaching strategies. Collins et al. (1989) argued that this type of knowledge is important yet insufficient to solve problems and carry out domain tasks. In the context of this study, simply describing modeling instruction for teachers would be insufficient for teachers to enact this instructional strategy in their classrooms.

Heuristic strategies. Collins et al. (1989) defined heuristic strategies are "generally effective techniques and approaches for accomplishing tasks" (p. 478). This knowledge might be referred to colloquially as "tricks of the trade" (p. 478). Heuristic strategies do not necessarily work in all contexts, but they can be helpful with they do work. Collins et al. (1989) provide an example of a heuristic strategy for writing from a K-12 student's perspective: writers can ignore syntactical errors and other presentation details to facilitate the flow of writing. This strategy is intended to facilitate idea generation and translation of ideas onto paper without getting bogged down in syntax. In the context of this study, these heuristic strategies would include effective approaches for teachers to implement modeling in the classroom more efficiently.

Control strategies. Control strategies are metacognitive strategies that help students select which strategies to use for a given problem and determine when these strategies are effective or ineffective (Collins et al., 1989). Collins et al. (1989) described control strategies as those that govern monitoring, diagnostic, or remedial components of a strategy. Monitoring strategies are those that "help students to evaluate their progress in a general way" (p. 479). For teachers, monitoring strategies are strategies teachers use to determine progress toward the lesson objective. Monitoring understanding by asking for students to respond to rote and deep questions individually and as a group is one monitoring strategy teachers can employ. Teachers can also solicit more authentic responses from students as a method of monitoring progress towards the lesson's objective.

Diagnostic strategies are those that the problem solver uses to arrive "at a useful analysis of the nature or cause of his difficulties" (p. 479). These strategies can stem from monitoring strategies. Collins et al. (1989) refer to students isolating specific words or phrases hindering reading comprehension as a diagnostic strategy. For teachers, isolating the source of student misunderstanding or struggle is a diagnostic strategy. For example, if a student does not want to write and produces very little text, a teacher could diagnose the problem as a fine motor problem that makes writing laborious. Alternatively, the teacher could diagnose the problem as a lack of useful strategies for brainstorming and idea generation.

According to Collins et al. (1989), diagnostic strategies are only helpful if they lead to remedial strategies. Remedial strategies are “activities that will lead out of the difficulty by introducing new knowledge or providing an alternate tack on the problem” (p. 479). If the teacher identified the source of writing difficulty as lack of strategies for idea generation, the teacher might teach a strategy such as free writing, which allows students to write all thoughts freely as they come.

Learning strategies. The final area of content proposed by Collins et al. (1989) is learning strategies. Learning strategies are “strategies for learning any of the other kinds of content” (p. 479). Teachers are very familiar with learning strategies as they apply to their students. Specifically in the area of writing, a common learning strategy is to receive critiques on a writing sample and the rationale for these critiques (Collins et al., 1989). In a PD context, learning strategies include any strategies the teacher uses to make learning more efficient. These could include note-taking and comprehension checks.

PD research does not have the level of research exploring the various content (domain knowledge, heuristic strategies, control strategies, and learning strategies) outlined in cognitive apprenticeship. However, CAP-PD is limited to only the content necessary for expertise, meaning that only information to help facilitate high-quality implementation of instructional practices is provided in the PD.

Methods. Collins et al. (1989) described six teaching methods divided into three separate groups. The first group constituted the core of cognitive apprenticeship. The teaching methods in the core group are modeling, coaching, and scaffolding. The second group included two teaching methods. These methods were articulation and reflection.

They were "designed to help students both focus their observations of expert problem solving and gain access to (and control of) their own problem-solving strategies" (p. 481). The final teaching method is exploration. This method "is aimed at encouraging learner autonomy" (p. 481), or self-determination. The following section describes each of these teaching methods and applies them to teacher PD.

Modeling. When modeling, an expert performs a task that allows students to observe and develop conceptual models necessary to accomplish the task (Collins et al, 1989). Often the important components of the skill are hidden; that is, they are mental processes that are not visible to a viewer. Therefore, modeling should include making mental processes visible to observers via thinking-aloud whenever appropriate.

In a PD context, modeling is an important method of instruction because the strategies being conveyed to teachers are often complex and not understood via a simple description. The model must demonstrate these skills for teachers to fully understand and enact the strategies. Modeling is also supported by empirical study of teacher PD. Darling-Hammond et al. (2017) found that modeling was a common method in PD studies that had a positive effect on teacher practice and student outcomes. CAP-PD provides modeling via CAP-TVs. An experienced teacher models the targeted practice and component skills. The CAP-TV narration notes essential features of the model.

Coaching. When coaching, an experienced performer observes students performing a task and offers "hints, scaffolding, feedback, modeling, reminders, and new tasks" (Collins et al., 1989, p. 481) as necessary. Collins et al. (1989) described two roles of coaching. First, it may highlight an aspect of a task the student did not notice previously during modeling. Second, it may remind the student of an aspect that was known previously but was overlooked or forgotten. Coaching should be highly-interactive and highly situated, meaning that the coaching stems from immediate problems of practice.

Similar to modeling, coaching is an instructional method with considerable evidence supporting its use as a PD tool. Kraft, Blazar, and Hogan (2017) conducted a meta-analysis on the effect of teacher coaching on instruction and student achievement. They found 44 studies that used a causal research design to study teacher coaching. Across these studies, they found a pooled effect size of .58 on teachers' instruction and .15 on student achievement. Again, cognitive apprenticeship was not proposed as a framework for PD contexts, and it does not have an experimental literature base supporting the various aspects of the framework; however, this component, instructional coaching, has considerable evidence supporting its use with teachers.

CAP-PD provides coaching based on CT Scan outputs. Kennedy, Rodgers et al. (2017) provided daily coaching for teachers via email. Kennedy, Hirsch et al. (2017) provided a single in-person coaching session to teachers. This coaching was not based on the CT Scan. Current research has not evaluated the dosage of coaching necessary to achieve sufficient change in teachers' practice within a CAP-PD framework. Coaching studies included in the meta-analysis of instructional coaching by Kraft et al. (2017) varied widely in terms of the dosage of coaching provided to teachers. Professional development studies that demonstrated a positive effect on student outcomes generally provided sustained coaching or follow-up with a significant number of contact hours between and providers and teachers (Darling-Hammond et al., 2017; Guskey & Yoon, 2009). Therefore, this study will provide daily coaching to teachers. Future research should consider the dosage of coaching necessary to achieve lasting change in teachers' practice.

Scaffolding. Collins et al. (1989) refer to scaffolding as the "supports the teacher provides to help the student carry out a task" (p. 482). Often scaffolding involves the teacher carrying out part of the task for the student (Collins et al., 1989). In a K-12 context, a common example of scaffolding is cue cards (Collins et al., 1989). In a PD context, scaffolding could include sample lesson plans, checklists, or other supplemental materials designed to facilitate teachers' implementation of the desired practice.

CAP-PD has provided varying supports as scaffolds to aid teachers' implementation of instructional practices. Kennedy, Rodgers et al. (2017) provided teachers with PowerPoint slides personalized to the individual terms taught in each teacher's lesson. However, due to the varying nature of writing instruction, creating individual PowerPoints for each teacher and each topic taught would not be feasible. Therefore, teachers will be provided with a PowerPoint template that will guide them through high-quality implementation of the targeted instructional practice (modeling).

Reflection. Collins et al. (1989) described reflection as the process whereby students "compare their own problem-solving processes with those of an expert, another student, and ultimately, an internal cognitive model of expertise" (p. 482-483). Collins et al. (1989) proposed multiple considerations for promoting reflection. First, they described a postmortem approach to reflection whereby an expert analyzes the completed task. Second, they advocated using various recording technologies to allow students to replay their performance or the performance of an expert and compare this performance to the ultimate goal performance for a given task. Since 1989, recording technologies have made incredible advances making recording devices easier to use and more readily accessible. Also, electronic storage allows for larger files to be stored at a much cheaper price than was available almost 30 years ago. Therefore, recording performance and reflecting on the recording should be exponentially easier to implement today than it was in 1989. Collins et al. (1989) noted that the level of detail may vary in reflection; however, they proposed abstracted replay, whereby the defining features of expert performance are highlighted, should be included in most reflections. In a K-12 context, students can watch how they solved a mathematics problem, how they analyzed and discussed a text, or how they thought aloud while brainstorming a writing sample. In a PD context, teachers can watch how they or an expert implement a specific practice, such as modeling.

Via the CT Scan, CAP-PD provides teachers visual outputs depicting lessons and the practices included in the lesson. Teachers can reflect on these practices and their implementation and compare this information to expert implementation of the practices. The coaching emails in CAP-PD will help facilitate this reflection and comparison.

Exploration. Collins et al. (1989) described exploration as the fading of supports that forces students to solve problems on their own. In the area of writing, they proposed that teachers could encourage students “to write an essay defending the most outrageous thesis they can devise” (p. 483) as an example of exploration. The goal of exploration is to provide students with general guidelines or goals and allow students freedom to explore within the given parameters. In a PD context, exploration can be accomplished when supports (e.g., coaching, materials) are faded allowing teachers to implement the given practice to solve a wide variety of problems of practice.

Current iterations of CAP-PD do not fully implement the exploration of cognitive apprenticeship. The design of this study does not allow for fading the supports and coaching provided teachers. The goal for this study is to provide as much coaching as possible to improve teachers' practice. Future research should consider the dosage of coaching necessary to achieve change in instruction and how to effectively fade the supports in a manner that leads to maintained change in teachers' practices.

Sequencing. In the cognitive apprenticeship framework, sequencing referred to the order that skills are taught to students (Collins et al., 1989). Collins et al. (1989) provided three principles to consider when sequencing skills. The first principle is to increase complexity. Increasing complexity requires teachers to design activities so that students must use progressively more complex skills to complete the activities. In a K-12 context, teachers can increase complexity by learning to spell words with short vowel sounds before learning to spell words with long vowel sounds which often have complex word patterns. In a PD context, teachers should learn simple skills before learning more complex skills. However, little research exists to identify the complexity of teaching practices. Any ranking of complexity would be conjectural.

CAP-PD will accomplish this goal by demonstrating a simple concept (e.g., combining two sentences) and progressing towards more complex skills such as incorporating metaphors and imagery in text. Scant empirical evidence is available to inform the complexity of skills in writing. Therefore, these decisions will be made based on feedback from content experts.

The second principle is to increase diversity. In a K-12 context, teachers can increase diversity by providing numerous contexts for implementing a skill. For example, students can write for a variety of purposes: for pleasure, to entertain, to inform. Increasing diversity allows students to know when to use a learned skill and when the learned skill is not appropriate to use. In a PD context, teachers can implement a learned practice (e.g., modeling) with an increasing number of skill areas. For example, a teacher might implement modeling with one skill (e.g., combining sentences with a conjunction) and add the new practice (modeling) to an increasing number of skill areas (e.g, writing a paragraph, providing feedback on a writing sample).

The CAP-TV for this experiment will increase diversity by demonstrating the target practice (modeling) with a single skill (e.g., writing topic sentences). As the CAP-TV progresses, the videos will demonstrate the practice with a more diverse set of skills (e.g., combining sentences, using imagery, and providing support for a position). The CAP-TV will also demonstrate the target practice with diverse instructional settings (one-on-one, small group, whole group).

The third principle is to teach global skills before local skills. When teaching global before local skills, teachers design activities so that students have the opportunity to solve problems or create products, whereby they generate a mental model for the end goal of the smaller skills. To accomplish this goal, teachers must scaffold students toward solving the problem because they have not yet learned the skills necessary to accomplish the goal. Once students see what the completed process looks like, the teacher gradually removes scaffolds and teaches necessary skills. In a PD context, teaching global skills before local skills can be accomplished by modeling a practice for a teacher in its entirety before breaking down the practice into smaller component skills. Also, teaching global before local skills could be accomplished by providing a high level of supportive materials (e.g., lesson plans, student handouts, technology tools) and fading these supports as teachers learn to implement the practice into their teaching repertoire.

CAP-PD will teach global skills before local skills in two ways. First, the CAP-TV will teach global skills before local skills by showing a modeling video of a teacher demonstrating the entire modeling task. After showing the full modeling practice, the modeling video will break the practice down into its component parts and teach each explicitly. Second, the coaching each teacher receives will frame the feedback in terms of the broader goal for instruction.

Sociology. Collins et al. (1989) argued that the social aspects of learning can play an essential role in determining beliefs about learning and expertise. Further, they argued social contexts can play an integral role in determining learners' motivation and confidence. Within the realm of sociology, they described five characteristics crucial to the sociology of learning. These elements – situated learning, culture of expert practice, intrinsic motivation, exploiting cooperation, and exploiting competition – are described below.

Situated learning. Collins et al. (1989) described situated learning as activities that involve authentic contexts that reflect the varied uses a task can serve. In a K-12 context, situated learning involves providing real-world problems or projects for students rather than isolated problems. In a PD context, situated learning could include providing real-world examples and videos to illustrate teaching practices. Also, as teachers incorporate new practices into their classrooms, they are learning in situated environments.

CAP-PD incorporates situated learning by using authentic contexts for filming the modeling videos with the CAP-TV. Also, teachers are observed using the practices in their authentic contexts. Coaching based on these observations is personally applied to their situations.

Culture of expert practice. Collins et al. (1989) described a culture of expert practice as one "in which the participants actively communicate about and engage in the skills involved in expertise" (p. 488). This culture can be helpful for developing expertise because it provides available models for novice learners to follow. In a K-12 context, activities that incorporate teachers and students discussing how they organized a writing sample could promote a culture of expertise. In a PD context, a culture of expertise could be promoted by experienced teachers modeling practices for inexperienced teachers and providing opportunity for interactions with experienced teachers supporting inexperienced teachers.

CAP-PD does not have a group component to it. Therefore, developing a culture of expert practice is difficult. However, CAP-PD promotes expertise by including experts modeling the practices in the CAP-TV. Also, the CAP-TV materials are reviewed by experts to verify the accuracy of the information and modeling in the CAP-TV.

Intrinsic motivation. Collins et al. (1989) recommended promoting intrinsic motivation by having learning related to a coherent or interesting goal. They argued that using the model-coach-fade cycle could contribute to fostering intrinsic motivation. Having students write for publication is one method of instruction that could promote intrinsic motivation in a K-12 context. In a PD context, intrinsic motivation could be achieved by explaining how the PD is related to the broad goal of improving student achievement and the specific outcomes the practice is likely to have.

CAP-PD builds intrinsic motivation in two ways. First, the CAP-TV that describes and models the practices for teachers clearly states the goal of the PD and the effect on student outcomes that these practices should have. Second, the template for the coaching emails includes a section describing the potential benefits of the practices that were the focus of the coaching email. Finally, CAP-PD incorporates the model-coach-fade cycle which may promote intrinsic motivation.

Exploiting cooperation. Collins et al. (1989) advocated cooperative learning environments where students can use the collective resources of the group to solve problems. Some people may not be able to integrate all the skills necessary for solving certain problems. In a K-12 context, pairing students for a writing activity will allow both to learn from each other's strengths and gain practice with various aspects of the task. In a PD context, grouping teachers by grade or content levels could allow participants to pool resources and get more return on the PD investment. Some teachers in the group could be better at lesson planning, behavior management, or instruction. Together this group could help each other improve classroom practice in all areas.

This element of cognitive apprenticeship is not fully implemented in CAP-PD. CAP-PD is designed so that individual teachers can complete the PD without others from their school. There is not a group or cooperative element to CAP-PD.

Exploiting competition. Collins et al. (1989) contended that using comparison and competition can be effective when comparing processes rather than results. For example, students can compare what strategies they used to revise or edit a writing sample rather than comparing the final quality of the writing sample. In a PD context, exploiting competition could include teachers comparing the practices used in the classroom rather than the actual quality of these practices or student outcomes.

CAP-PD exploits competition in a number of ways. First, when delivering coaching teachers are compared to their own past performance. Second, the CT Scan which provides data for coaching is not an evaluative teacher observation tool. It is descriptive and displays the practices teachers use during in a lesson rather than the quality of the practices.

Alignment of CAP-PD and Cognitive Apprenticeship

CAP-PD was designed to align with the cognitive apprenticeship framework by beginning with the teacher viewing a CAP-TV where the new skill is demonstrated by an experienced teacher. This stage is followed by coaching from an experienced teacher. Finally, teachers enact the new practice in the classroom without support from the CAP-PD team. See Figure 2 for the alignment between cognitive apprenticeship and CAP-PD.

Figure 2
Alignment between cognitive apprenticeship and experiment

Theory	Cognitive Apprenticeship		
Component	Modeling	Coaching	Scaffolding
Description	Learner repeatedly observes a more experienced performer execute a task.	Coach provides tips, reminders, and help to assist performance of a task.	Scaffolds, or supports, assist the learner in enacting the task, sometimes by using similar approximations of the task.
Intervention	CAP-PD		
Component	CAP-T	CT Scan coaching	CAP-TS
Description	Teacher views modeling videos.	Teacher receives daily feedback on performance.	Teacher receives PowerPoint slides that prompt use of steps in high-quality modeling.

Collins et al. (1989) proposed cognitive apprenticeship as a model for teaching reading, writing, and mathematics skills to K-12 students. However, the framework as a whole lacks experimental evidence supporting its use for this purpose. Much of the observational research underpinning the framework was focused on mastering a trade (i.e. leather-making). Perhaps because of these roots in occupational apprenticeship, the framework has gained popularity in teacher education. However, the framework lacks experimental evidence supporting its use for changing teacher practice.

Very few PD studies used cognitive apprenticeship as an explicit framework guiding the design of PD. Much of the empirical evidence supporting cognitive apprenticeship lacked the methodological rigor to make causal claims validating the framework nor did they examine the impact of the PD on teacher instruction. For example, Peters-Burton, Merz, Ramirez, and Saroughi (2015) used cognitive apprenticeship to design a PD package promoting inquiry instruction with secondary teachers. Nineteen teachers participated in the year-long PD. All teachers received the PD. Participants responded to measures of motivation and self-regulation, self-efficacy, and knowledge. This study was limited in two ways. First, all teachers received the PD at the same time. Therefore, the study cannot make strong causal claims about the effect of cognitive apprenticeship-based PD. Second, this study did not observe teachers, and therefore, cannot make claims about the effect of the PD on teaching quality.

Therefore, rigorous research examining and validating cognitive apprenticeship as a framework for PD interventions is needed. CAP-PD heavily incorporates the core of cognitive apprenticeship (modeling, coaching, and scaffolding). Preliminary evidence suggested CAP-PD can change middle school teachers' science vocabulary instruction. Replications of this research in other content areas are needed to confirm these findings and explore the applicability of various components of cognitive apprenticeship.

In addition to these studies that were explicitly designed using cognitive apprenticeship, the major components of cognitive apprenticeship have been recognized as effective components of PD. In their synthesis of PD studies that demonstrated a positive effect on student outcomes, Darling-Hammond et al., (2017) identified modeling as a common activity across effective PD studies. Similarly, Kraft et al. (2017) found coaching to be an effective method of improving teacher practice and student outcomes. Therefore, although little research has explicitly used cognitive apprenticeship as a guiding framework, it is reasonable to assume that these components could be effective when combined in a PD intervention.

Research Questions

This study builds on several research bases. First, the study expands CAP-PD to writing instruction, a content area previously not investigated. Second, this study will use multiple observational measures of instruction to fully describe the nature of writing instruction present before and after the PD intervention. This level of description will provide a clearer picture of the instruction that led to any changes in student outcomes.

Second, this study builds on the CAP-PD research base by answering significant questions that remain regarding its effects. First, all studies of CAP-PD have used a researcher-created measure of teacher practice. Generally, these measures were closely aligned to the intervention. This study includes a researcher-created measure and adds a more distal measure of teacher quality (PLATO). Second, these studies have not examined student academic outcomes. Kennedy, Hirsch, et al. (2017) found positive effects for on-task behaviors of students, but the effect on academic outcomes remains unexplored. This study includes a progress monitoring measure of student writing ability.

Finally, this study adds to the limited research base investigating cognitive apprenticeship. Only one study investigated CAP-PD using the cognitive apprenticeship framework (i.e., Kennedy, Rodgers, et al., 2017). This study will add to the empirical research base for using cognitive apprenticeship to improve teacher quality.

The present study intends to improve the writing instruction of secondary teachers by improving the quality of modeling instruction in writing. The intervention is designed using the cognitive apprenticeship framework and will empirically test this theory for improving teacher practice. The study builds on previous CAP-PD research in several important ways. The following research questions will guide the study.

1. What are the effects of CAP-PD on teachers' modeling of writing skills?
    a. What are the effects of CAP-PD on teachers' modeling of writing skills as measured by the CT-Scan?
    b. What are the effects of CAP-PD on teachers' modeling of writing skills as measured by the PLATO?
2. What are the effects of CAP-PD on student writing outcomes?
3. What are teachers' perceptions of social validity for CAP-PD?

CHAPTER III: METHODOLOGY

This study used a single-case multiple baseline across participants design. Single-case research methodology serves "to document causal, or functional, relationships between independent and dependent variables" (Horner et al., 2005, p. 166). Horner described several defining features of this methodology. First, the individual participant serves as the unit of analysis. Second, the independent variable is systematically manipulated by the researcher. Third, the dependent variable is systematically and repeatedly collected from participants.

Participants

I gained approval from the University of Virginia Institutional Review Board (IRB) to conduct this study. The Board granted this study exempt status, meaning that the study was considered standard educational practice. Because of this status, students did not have to sign a consent or assent form. Rather, students were given a notification letter informing them and their parents of the nature of the study. Parents could opt out of the study if they chose. No one opted out of the study.

Following IRB approval, I emailed the superintendent of Fordham County Public Schools. He granted permission for the study to occur in this school division. Then, I emailed the principal of Fordham County High School. He also granted permission for the study to occur at this school. Finally, I met with all four 8th grade English teachers to discuss the study and request their participation in the study. Three of the four teachers agreed to participate in the study. One of the four teachers had a medical leave of absence that inhibited her ability to participate in the study. The three participating teachers signed an IRB consent form.

Teachers

Three teachers participated in the study. The teachers all taught 8th grade at Fordham County High School. Teachers were included if they met the following criteria: (a) taught 8th grade English courses, (b) taught students with disabilities in these courses, (c) consented to being observed daily for the duration of the study, (d) consented to administering a weekly writing prompt to students for the duration of the study. Teachers were given a $350 honorarium for participating in the study. Funds for this stipend were provided from The Wing Institute Graduate Research Stipend.

The three teachers all had bachelor's degrees. Two had earned additional college credits beyond a bachelor's degree but had not completed a master's degree. All three had professional teacher licensure status. Their years of experience ranged from 4-31. Two teachers had taught 8th grade ELA for two years. The third taught 8th grade ELA for 18 years. See Table 2 for full demographic information for the teacher participants.

Table 2
Teacher Demographic Information

Teacher	Years of experience	ELA YOE	8th Grade YOE	8th grade ELA YOE	Education	Teacher Licensure Status
Dottie	31 (10 PT)	31 (10 PT)	18 (3 PT)	18 (3 PT)	Bachelor's degree plus additional credits	Professional
Bill	13	13	2	2	Bachelor's degree	Professional
Nora	4	2	2	2	Bachelor's degree plus additional credits	Professional

Note. PT = part time; YOE = years of experience

Students

Eighth grade students of participating teachers also participated in the study. Teachers completed a demographic survey for each student. Sixty-one students participated in the study. Forty-one percent (*n* = 25) were female. Fifty-nine percent (*n* = 36) were male. Three students (4.9%) received services through a 504 plan. Nine students (14.8%) were identified by the school as “at-risk” of school failure. Thirteen students (21.3%) were students with disabilities and received special education services through an Individualized Education Plan (IEP). Fifty-nine percent (*n* = 36) did not meet any of these criteria. Of the 13 students with disabilities, four were female and nine were male. Nine were white, two were black, and two were identified as multi-racial. See Table 3 for student demographic information.

Table 3
Student Participant Demographic Information

	n	%
Gender		
Female	25	41
Male	36	59
Disability Status		
IEP	13	21.3
504 Plan	3	4.9
At-risk	9	14.8
Race/Ethnicity		
White	44	72.1
Black	8	13.1
Multiple races	7	11.5
Hispanic/Latino	2	3.3

All students responded to a writing prompt on the fifth day of each new phase. The prompts were randomly selected from a bank (Holt, Rinehart, & Winston, n.d.). Students were assigned an identifying number which was used for all demographic data and writing prompts.

Setting

This experiment took place in a rural high school serving students in 8^{th}-12^{th} grades. The high school was classified as rural, distant according to the Virginia Department of Education (2017). This classification indicates that the school was more than 25 miles from an urbanized area and more than 10 miles from an urbanized core. The student population was comprised of 74.3% White students, 15.8% Black students, 4.8% Hispanic students, 4.4% students of two or more races, and .6% Asian students. Thirteen and a half percent of the students were identified as students with disabilities. Twenty six percent of students were considered economically disadvantaged, and 2% were English Language Learners (Virginia Department of Education, 2017).

The classrooms that participated in the study were all general education classrooms that were inclusive to students with disabilities. All three teachers worked with a special education co-teacher at some point throughout the day. Due to scheduling constraints, two teachers were observed during their co-taught class periods. However, one teacher was observed as the solo instructor for the course.

Dependent Variables

This study used one primary observational measure of teacher practice to determine the effect of the independent variable. A second observational measure was used to supplement the data from the primary measure and provide more instructional context. Finally, student writing samples were collected to further supplement the data provided by the two observational measures. These dependent measures are a strength of the study building on past research in writing instruction and CAP-PD in two ways.

First, using multiple observation measures more fully described instruction. As noted in chapter two, studies in the WWC report on effective writing instruction lacked a broad measure of writing quality (Graham et al. 2016). Similarly, studies of CAP-PD used a dependent measure that was also part of the intervention (Kennedy, Rodgers, et al., 2017). The present study used one dependent measure closely aligned to the intervention, and it also used a second teacher observation measure (i.e., Protocol for Language Arts Teaching Observations) to examine changes in teacher practice more distal to the intervention.

Second, the study used a measure of student writing performance to describe the nature of students' writing. The research design of this study did not permit causal claims about this data. However, the student writing samples provided helpful context about the participating students.

Demographic Survey

One important facet of single case research is a rich description of the participants (Kratochwill et al., 2010; Horner et al., 2005). In order to better describe the participants, each teacher completed a demographic survey indicating his or her number of years of experience teaching, number of years of experience teaching within this grade level, highest degree attained, and teacher licensure status. This survey provided relevant information to determine the generalizability of the results to other samples of teachers. See Table 4 for the items on this survey.

Table 4
Teacher Demographic Survey

Survey Item	Response Format
How many years of teaching experience do you have?	Open-ended
How many years have you taught English?	Open-ended
How many years have you taught in your grade level?	Open-ended
What is the highest level of education you have attained?	Multiple Choice: Bachelor's Bachelor's plus additional earned credits Master's Master's plus additional earned credits Doctoral degree
What is your teacher licensure status?	Multiple Choice: Provisional Initial Professional

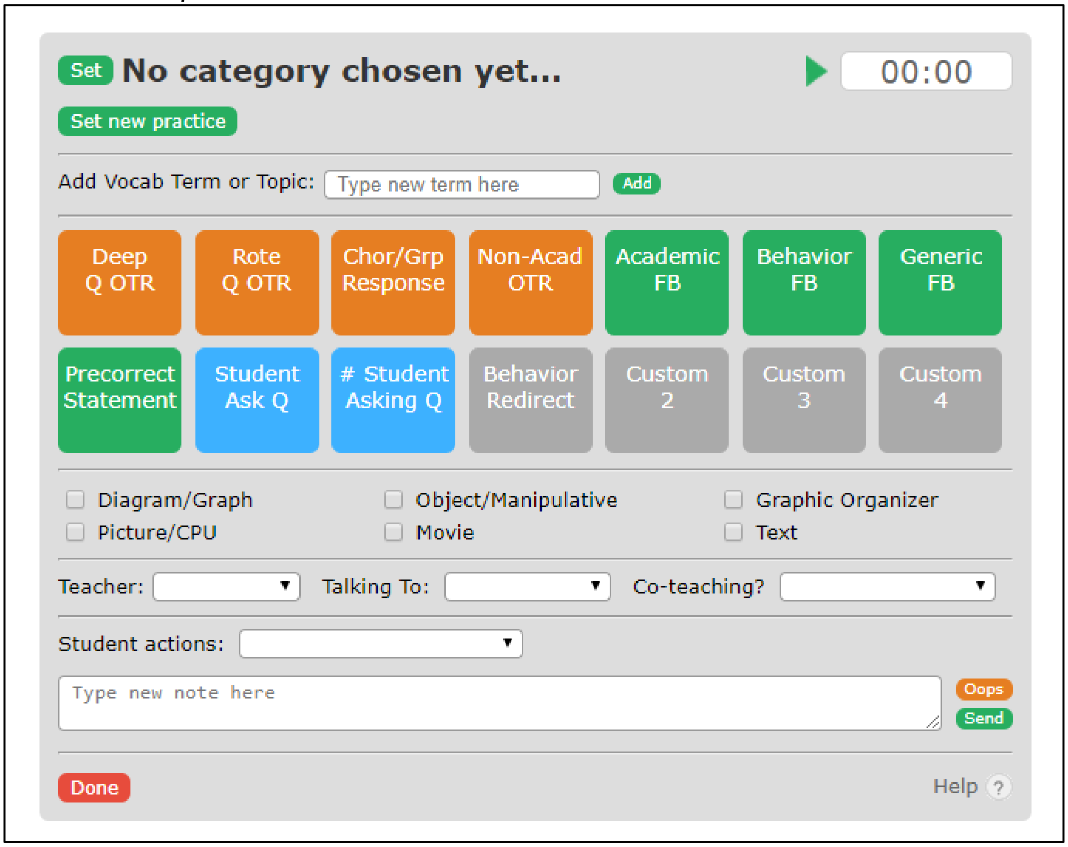
CT Scan

The Classroom Teaching Scan (CT Scan) is an online teacher observation tool that views instruction through an explicit instruction and evidence-based lens. The CT Scan is primarily descriptive, not evaluative. It is a web-based tool that can be viewed at www.classroomteachingscan.com/ctscan. The following section explains the sections of the CT Scan relevant to this study.

When using the CT Scan, observers record duration and frequencies for certain teaching behaviors. Frequency data includes the number of teacher-directed opportunities to respond (OTR), teacher-provided feedback statements (FB), behavior redirects, number of questions asked by students, and the number of students asking questions. Each time a frequency item is observed, the observer clicks the appropriate button on the CT Scan interface, and the tool records the event and the time during the observation that the event occurred. In Figure 3 below, the orange buttons represent four different types of OTRs. From left to right, the options are deep question OTRs, rote question OTRs, choral or group response OTRs, and non-academic OTRs. The CT Scan output records frequency information for the total number of OTRs and for each type of OTR.

The three green buttons to the right of the OTRs are types of FB statements. Academic FB statements are statements that give students specific information on the quality of their work (e.g., "I really like how you used imagery in this paragraph."). Behavior FB statements are statements that give students specific information related to their behavior (e.g., "I like how you all are listening to each other talk and not interrupting during our discussion."). Generic FB statements give students some information about their academic or behavioral performance, but are not specific (e.g., "good job", "yes", "no, that's incorrect."). Buttons on the second row (precorrect statements, student asking question, number of students asking questions, and behavior redirects) will not be used in this study.

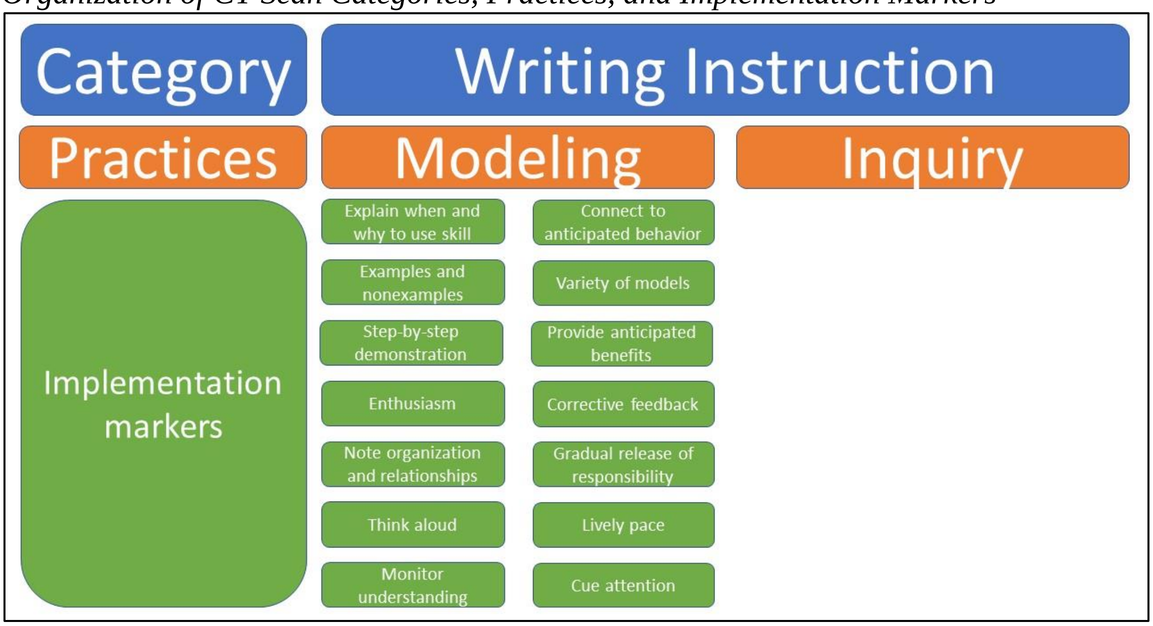
Figure 3
CT Scan Interface



Duration is recorded for several features of instruction. These features include student actions, use of visual aids, the general category of instruction, the specific instructional practice used, and the topic of instruction. When recording student actions, the observer selects options from a dropdown menu. Choices include reading, writing, taking notes, academic conversation with peers, and other school-related student activities.

The most significant distinguishing feature of the CT Scan is its precision in describing the instruction occurring during a lesson. When recording the general category of instruction, the observer selects from preloaded broad categories of instruction (e.g., writing instruction, reading instruction, mathematics instruction, behavior management). Each category of instruction has accompanying instructional practices. For example, some of the practices within writing instruction include cognitive strategy instruction, inquiry instruction, and modeling. When the observer selects an instructional practice, the CT Scan records the practice (e.g., modeling) and corresponding instructional category (e.g., writing instruction). Each instructional practice has accompanying implementation markers. Implementation markers (IMs) are developed specific to an instruction practice and are identified using the best available evidence for the given practice (e.g., meta-analyses, literature syntheses, reports). To better understand the relationship between categories, practices, and implementation markers, see Figure 4 below.

Figure 4
Organization of CT Scan Categories, Practices, and Implementation Markers



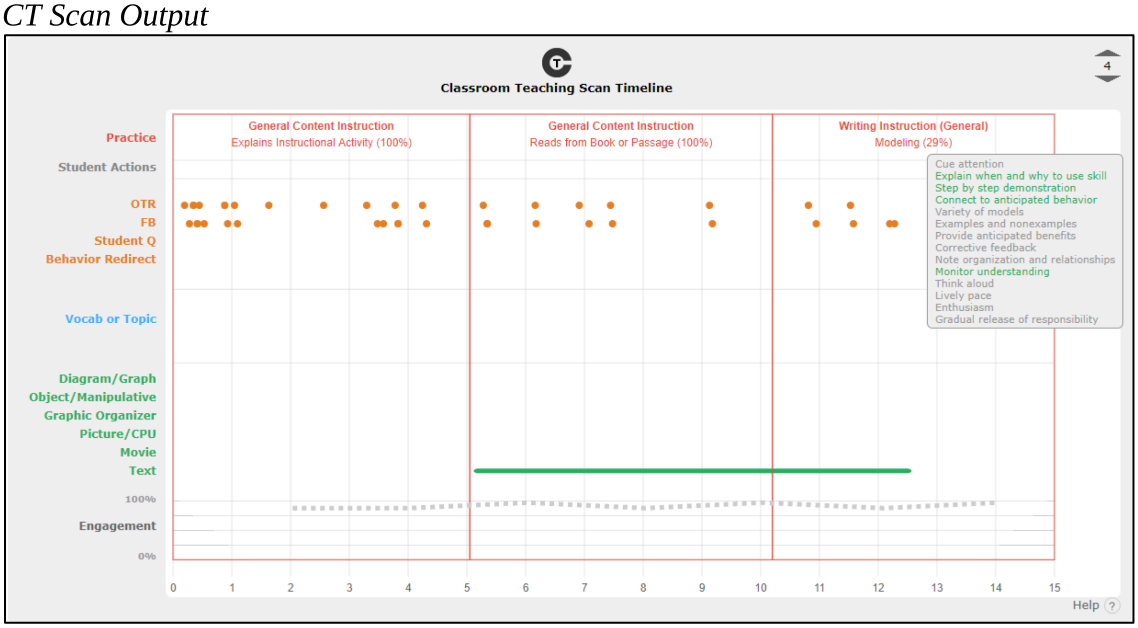
Duration data (OTRs and FBs) are combined with the duration data to create a timeline of instruction as one of the main outputs of the CT Scan. A version of this timeline accompanied the coaching email provided by Kennedy, Rodgers, et al. (2017). See Figure 5 below for a sample output. In this output, the minutes are indicated by vertical lines that are numbered to align with minutes of instruction. The boxes outlined in red represent different instructional practices. The first two practices are "explains instructional activity" and "reads from book or passage." Both of these practices are in the General Content Instruction category.

The final practice is "modeling", and it is in the Writing Instruction (General) category. The percentage next to "modeling" indicates that the teacher demonstrated 29% of the IMs for modeling. The gray box under modeling lists all IMs for this practice. Those highlighted in green indicate the teacher demonstrated them during this iteration of the practice.

The orange dots represent OTRs and FB statements. The green line indicates that the teacher used text as a visual aid during this time. The dotted line near the bottom of the output represents an estimate of the percent of students on-task during the lesson.

Figure 5

CT Scan Output



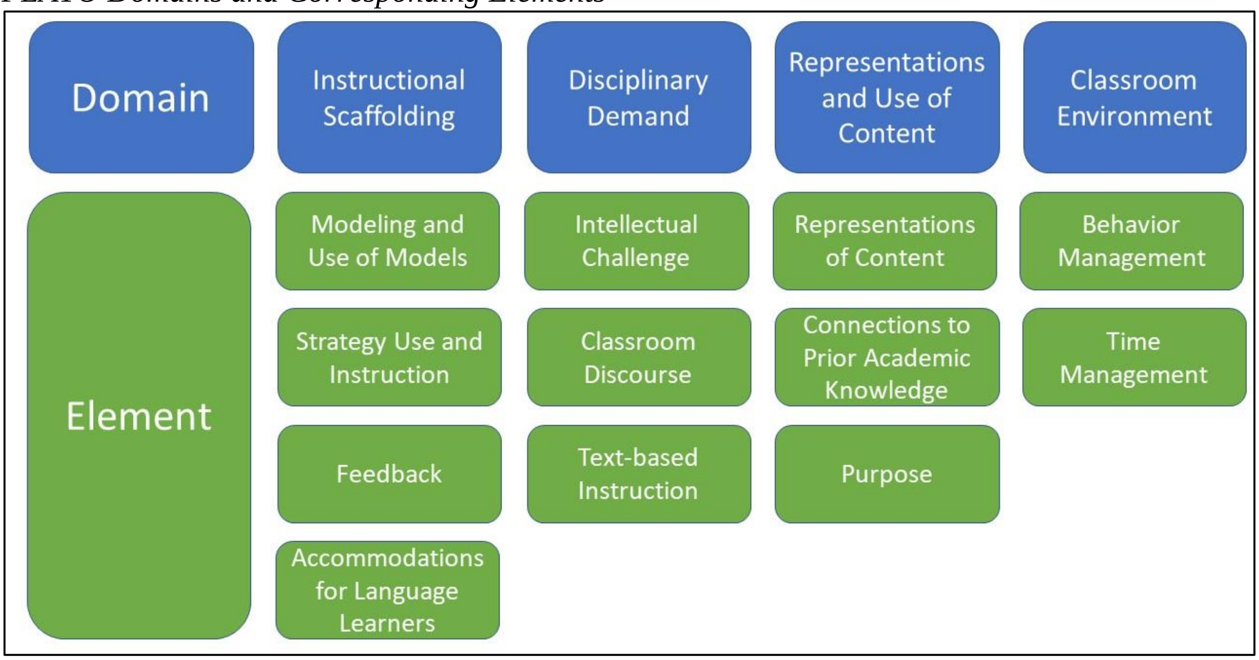
Although the CT Scan does not produce a composite score to evaluate instruction, observers can rate the quality of implementation for each practice using IMs. Because IMs are gleaned from research they represent high-quality implementation of instructional practices as defined by empirical research. The percentage of IMs used during an instructional practice represent the percent of quality the teacher used to implement the practice. The primary dependent variable for this study will be the percent of modeling IMs used in each lesson. Modeling has 14 IMs. See Table 1 for a complete list of IMs for modeling. This dependent variable will guide the phases in the single case design.

Technical properties. The CT Scan is a new observation tool and does not have studies specifically studying its validity and reliability. However, regarding its validity, the tool was developed from literature on effective instruction for students with disabilities and has been reviewed by content area experts. These features of its development support the content and face validity of the tool. Regarding its validity, the CT Scan was used as the primary dependent variable in Kennedy, Rodgers, et al. (2017) with sufficient reliability. In that study, reliability for duration-based data was 87% and agreement on IMs was 85%.

PLATO

Similar to the CT Scan, the Protocol for Language Arts Teaching Observation (PLATO) is a classroom observation tool that views instruction through an explicit instruction lens (PLATO, 2017). Unlike the CT Scan, the PLATO is a domain-specific tool designed to measure instruction in English Language Arts classrooms. The tool includes 12 elements of instruction with a corresponding rubric for each element. The elements are divided into four domains. These domains are Instructional Scaffolding, Disciplinary Demand, Representations and Use of Content, and Classroom Environment. The Instructional Scaffolding domain has four elements: Modeling and Use of Models, Strategy Use and Instruction, Feedback, and Accommodations for Language Learning. The Disciplinary Demand domain has three elements: Intellectual Challenge, Classroom Discourse, and Text-based Instruction. The Representations and Use of Content domain has three elements: Representations of Content, Connections to Prior Academic Knowledge, and Purpose. The Classroom Environment domain has two elements: Behavior Management and Time Management. To more fully understand the relationship between domains and elements, see Figure 6 below.

Figure 6
PLATO Domains and Corresponding Elements



Each element is given a score from 1-4 using a corresponding rubric. A score of "1" indicates that the teacher "Provides almost no evidence." A score of "4" indicates the teacher "Provides consistent strong evidence." Each element's rubric defines what each score means for that given element. When observing instruction using the PLATO, observers divide instruction into 15-minute segments. Observers rate each of the 12 elements for each segment of instruction. After scoring all segments for each element, the observer can calculate average scores for each element and each domain. For example, in an hour-long lesson, the teacher would have four 15-minute segments. If the observer was scoring the Modeling and Use of Modeling element and assigned segment scores of 1, 3, 3, and 4, the average for this element would be 2.75. This element could then be averaged with the other elements in the Instructional Scaffolding domain to create the domain score.

Technical properties. Cor (2011) assessed the technical properties of the PLATO through a generalizability study and found its internal consistency, or reliability, to be high (G- coefficient = .81). Although he did not calculate an overall inter-rater reliability statistic, he also found that most raters scored teachers in similar ways through a FACETS bubble chart. He concluded that the PLATO accurately measures the performance of most, if not all, teachers under certain conditions. First, at least five segments of a lesson should be scored. Second, observers should be trained using the rigorous procedures designed by the PLATO publishers. Third, studies that repeatedly measure the same teachers may have lower reliability because the analysis indicated that the scores on the eighth observation tended to lower reliability. This final point will be a limitation in the current design. Due to the nature of a single case design, teachers will be repeatedly measured using the PLATO more than eight times. However, this limitation is mitigated by the fact that the PLATO is not the primary observation tool in the study.

Kane and Staiger (2012) reported on the validity of the PLATO. The found the tool to be positively correlated with student achievement gains. On state English Language Arts tests, students whose teacher scored in the bottom quartile lost approximately two months of schooling gains compared to students whose teacher scored in the top quartile. On the *Stanford 9* Open-Ended Reading test, students whose teacher scored in the bottom quartile on the PLATO lost approximately 7 months of schooling gains compared to students whose teacher scored in the top quartile.

The PLATO was especially useful for this study for two primary reasons. First, it views instruction through an explicit instruction lens. Therefore, it is well-suited to evaluate instruction aligned with best practices for students with disabilities. Second, one of the 12 elements specifically measures the presence of modeling in a lesson. This element was especially helpful in this study, which was specifically designed to improve the quality of modeling instruction.

CBM-W

Curriculum-based measurement is an assessment tool for monitoring student progress in reading, mathematics, spelling, and written expression (Deno, 1985). Deno and his colleagues created CBM in response to poor technical properties (i.e., validity and reliability) of other progress monitoring measures. Their primary goal was to develop tools with utility for teachers. Therefore, they wanted CBM to be easy and efficient to administer and produce scores that were meaningful and easy to interpret graphically. Standardized administration procedures allow scores on the probes to be compared across administrations. These features of CBM that make it a useful tool for teachers also make them helpful for use in a multiple baseline research design.

When administering a CBM prompt, teachers read from a script that includes directions and a writing prompt. Then, students respond to the prompt for a pre-determined duration. Students' writing is then scored using direct, observable scoring procedures. The most commonly-researched scoring procedures are counting the number of words written, words spelled correctly, correct word sequences, or correct minus incorrect word sequences (CIWS; Romig, Therrien, & Lloyd, 2017). Romig et al. (2017) conducted a meta-analysis examining the criterion validity of these four commonly-researched CBM-W scoring procedures and found that CIWS had the strongest criterion validity for middle school students and was moderately predictive of overall writing ability (r = .59; Romig et al., 2017).

In this scoring procedure, a correct word sequence is defined as two adjacent words that are spelled correctly and are syntactically correct. Words must be capitalized and punctuated correctly to be considered a correct word sequence. All correct word sequences and incorrect word sequences are marked accordingly. The number of incorrect word sequences is subtracted from the number of correct word sequences to create the CIWS score.

For this study, teachers administered a CBM probe on the 5th observation day of each new phase (approximately weekly). Teachers were given a script to read for each administration that included the writing prompt. Students were given a sheet of paper with the prompt at the top of the page and lined space for responding to the prompt on the rest of the page. Students were given seven minutes to respond to the prompt. After students responded to the prompts, teachers removed student names from the writing prompts, assigned each prompt a numeric student number, and gave them to the classroom observer.

Social Validity

Teachers responded to a social validity survey using items adapted from Hirsch (2016). The survey included 5-point Likert-type items and open-ended items. Teachers responded to the survey after completing the study. The survey prompted participants to rate the degree to which the various elements of the CAP-PD contributed to (a) changing their knowledge about modeling instruction, (b) their satisfaction with the PD, (c) the amount of time required to complete the PD, (d) the effectiveness of the intervention on their instruction, and (e) the effectiveness of the intervention on student outcomes. See Table 5 for the social validity survey.

Table 5
Teacher Satisfaction Survey

Previous Knowledge	Unfamiliar				Familiar
On a scale of 1-5 with 1 being unfamiliar and 5 being familiar, how familiar with the components of modeling were you *before* the training?	1	2	3	4	5
On a scale of 1-5 with 1 being unfamiliar and 5 being familiar, how familiar are you *now* with the components of modeling?	1	2	3	4	5
Training	Unsatisfied				Satisfied
On a scale of 1-5 with 1 being unsatisfied and 5 being satisfied, how satisfied are you with the video training you received?	1	2	3	4	5
On a scale of 1-5 with 1 being unsatisfied and 5 being satisfied, how satisfied are you with the email coaching/feedback you received?	1	2	3	4	5
On a scale of 1-5 with 1 being unsatisfied and 5 being satisfied, how satisfied are you with the PowerPoint template slides you received?	1	2	3	4	5
Implementation	Hard				Easy

On a scale of 1-5 with 1 being hard and 5 being easy, to what extent do you think the components of modeling were easy to learn?	1	2	3	4	5
On a scale of 1-5 with 1 being hard and 5 being easy, to what extent do you think the components of modeling were easy to implement?	1	2	3	4	5
How much time do you think it required to implement modeling?	A Lot of Time			Very Little Time	
On a scale of 1-5 with 1 being a great deal and 5 being none, how much preparation time was required to implement modeling?	1	2	3	4	5
On a scale of 1-5 with 1 being a great deal and 5 being none, how much classroom time was required to implement modeling?	1	2	3	4	5
Effectiveness	Not Effective			Very Effective	
On a scale of 1-5 with 1 being not effective and 5 being very effective, how effective do you think modeling was for your students?	1	2	3	4	5
On a scale of 1-5 with 1 being not likely and 5 being very likely, how likely are you to continue using modeling in your instruction?	1	2	3	4	5
On a scale of 1-5 with 1 being not likely and 5 being very likely, how likely are you to recommend modeling to a colleague?	1	2	3	4	5

Observer and Scorer Training

CT Scan. Observers using the CT Scan were involved in the development and creation of the CT Scan. They have completed multiple trainings and have hundreds of hours of observations using the tool. For this study, they participated in an additional two-hour training session to ensure reliability on the modeling practice and IM checklist items. The training consisted of reviewing the definitions of modeling and its IMs.

Then, the observers watched videos using the CT Scan until they reached at least 80% reliability on the modeling practice and IM checklist items.

PLATO. Observers using the PLATO completed standard training provided by the PLATO publishers. This training consisted of a series of in-person or electronic modules lasting approximately 20 hours. After these modules, observers scored a series of five video segments. Reliability was calculated against the PLATO master scores for each component. Observers must reach 80% reliability on each component to receive PLATO certification. After the first round of reliability videos, the observer has a phone call with the PLATO publishers to discuss which components were passed and failed. The PLATO coach gives feedback on failed components. Then, the observer completes a second round of reliability videos and scores the segments only for the PLATO components that were failed in the first reliability test. Again, scores on these segments are compared to the PLATO master scores.

CBM-W. All CBM-W scorers participated in a scorer training session. In this training session, scorers reviewed the definition and scoring procedures for CIWS. Then, scorers double-scored a sample writing probe. Scorers scored CBM-W probes in the study when they achieved at least 80% agreement on the sample probes.

Inter-Observer and Inter-Scorer Reliability

CT Scan. Twenty percent of the observations in each phase were double-coded for reliability. To determine agreement between the two observers, Cohen’s kappa was calculated for each checklist item using procedures from Bakeman and Gottman (1997). Cohen’s kappa considers agreement that is due to chance making it a more precise reliability calculation that the percent of overall agreement. Kappa coefficients were calculated for all implementation markers for each iteration of modeling (i.e., multiple iterations possible in one observation) coded by either observer.

Kappa ranged from 0-1 with an average of .64 across all implementation markers. Of the disagreements, over half (57%) were instances when the primary observer did not code a marker as present, indicating that on average the primary observer slightly underscored observations. Two markers (i.e., lively pace and enthusiasm) had a kappa value of 0. These markers were removed from inclusion in the results. Without these implementation markers included, the average kappa value was .74. Bakeman and Gottman (1997) consider kappa values above .70 to be acceptable.

Reliability for the duration measure was calculated by considering each second of double-scored observations. Second-by-second agreement was calculated dividing the number of agreed seconds (i.e., seconds where both observers agreed that modeling was or was not happening) from the total seconds in the observation and multiplying by 100. Agreement for the duration measure was 87%.

PLATO. Twenty percent of all observations in each phase were double-scored by trained PLATO observers for reliability. To calculate reliability, each scored segment from the two observers was compared. The number of agreed segments was divided by the total number of segments (agreements/total agreements and disagreements) to calculate the percent of agreement (Wyatt, Callahan, & Michael, 1985). Total agreement was 94%. Reliability for each element ranged from 81-100%.

CBM-W. Two scorers scored 20% of all CBM probes in each phase. Using the double-scored prompts, intraclass correlations (ICC) were calculated to estimate reliability. Procedures from Landers (2015) guided ICC calculations. ICC was .958, indicating a high level of reliability.

Experimental Design

This study used a single-case multiple baseline design across participants. Teachers entered the intervention phase individually in random order. The study and experimental design was guided by recommendations the WWC design standards for single-case designs (Kratochwill et al., 2010).

Standard Condition (A)

The baseline condition was modeled after Hirsch (2016). After receiving teacher consent and notifying students and parents of the study, teachers were invited to a meeting to discuss the study schedule and related information. Also, during this meeting the principal investigator provided instructions on how to administer the CBM-W. This training helped to ensure consistent administration of CBM-W probes across teachers.

As part of the initial meeting with teachers, we discussed the observation schedule. Observations occurred when writing was one of the teachers' objectives for the class. Class periods that were dedicated to taking tests, having school pep rallies and other school functions, or are not at all related to writing instruction will not be observed. Baseline observations continued until a minimum of five observations were collected and a predictable pattern of performance was established.

CAP-PD (B)

Teachers entered the intervention phase in random order and entered the phase staggered to meet requirements for a multiple baseline design (Kratochwill et al., 2010;

Horner et al., 2005). The intervention phase continued until a minimum of five observations were collected and a predictable pattern of performance was established.

The independent variable under examination was CAP-PD. CAP-PD included watching a CAP-TV, receiving a CAP-TS template, and receiving instructional coaching based on observations using the CT Scan. The CAP-TV was a three-part video series designed using Mayer’s (2009) cognitive theory of multimedia learning (CTML). The CTML is based on cognitive load theory and is designed to reduce cognitive load and increase learning outcomes in multimedia learning formats. The CAP-TV provided a general description of and rationale for using modeling when teaching writing. The description included why modeling is an important teaching strategy, when to use modeling, and how to implement modeling in the classroom. The section describing how to implement modeling was organized around the CT Scan IMs for modeling. These IMs were identified as markers of high-quality implementation of modeling. Each IM included a description of the IM and videos of teachers demonstrating the IM.

Kennedy, Rodgers, et al. (2017) created CAP-Teacher Slides (CAP-TS) – individual PowerPoint presentations – for every science term taught by teachers in the study to facilitate their use of high-quality vocabulary instruction. These PowerPoint presentations had the targeted vocabulary practices embedded within them. Teachers were encouraged to use and modify the presentations as they saw fit. The present study provided similar supplementary materials to teachers. However, the nature of writing curriculum and instruction made creating a PowerPoint for every possible writing concept prohibitive. Therefore, teachers were given a CAP-TS template, which is a PowerPoint preloaded with slides that guide the teacher through the modeling IMs.

Blanks were provided for teachers to add or modify information. Also, the PowerPoint had instructions at the beginning of the template explaining that use of the template was optional and how to use it.

Instructional coaching was based on the CT Scan and targeted at improving the implementation of teachers' modeling instruction. Coaching was delivered via email. The email used a template used by Kennedy, Rodgers, et al. (2017). The template began with describing something positive about the lesson. Next, the template drew the teacher's attention to the area of improvement. The email explained what happened in the lesson, how to improve that aspect of the lesson, and how the potential improvement will impact students. Graphs of the CT Scan data will accompany the coaching emails to illustrate the information in the coaching email. See Figure 7 for an example of a coaching email based on the sample CT Scan graph in Figure 5.

Figure 7
Sample Coaching Email

Dear Sample Teacher,

Thank you for letting me come and observe your class again today. Attached to this email are graphs representing what the instruction looked like in your class. Below this paragraph are notes highlighting a couple of things you will see in the graphs and to provide feedback and ideas for other ways you can enhance your writing instruction in future lessons. If you have any questions about the graphs or about the feedback in this email, please let me know and I will get back to you as soon as possible.

Observation notes:

Excellent lesson today! One thing I particularly liked to see was your use of modeling to demonstrate how to write a compare and contrast paragraph. According to the timeline, you spent approximately 5 minutes modeling this skill. While modeling, you explained when to use the skill, provided a step-by-step demonstration, monitored student understanding, and explained that you wanted students to use this skill in their writing. It's so important to model the skills that you expect students to use because it is one of the most efficient methods of instruction available.

Another thing you will see on the timeline is that you used 29% of the markers of high-quality modeling. Next time, try including more of the elements of high-quality modeling. The PowerPoint template provides a simple framework for including all 14 elements of high-quality modeling. Including all elements of high-quality modeling will multiply the impact of writing instruction in your students' writing.

Thank you again, and have a great evening!

- John

Implementation fidelity. Videos were originally posted to EDpuzzle (www.edpuzzle.com). However, EDpuzzle stopped supporting videos on the Vimeo platform at the beginning of the intervention phase. Therefore, teachers were given a link to view the videos on Vimeo (www.vimeo.com). The videos were password protected to prohibit teachers not in the intervention phase from viewing them. However, distributing the videos on Vimeo rather than EDpuzzle caused the loss of data regarding the number of individual views for each teacher.

The CAP-Teacher Slides (CAP-TS) aspect of the intervention was delivered to the teachers via email when they entered the intervention phase of instruction. Teachers were asked not to share these with other teachers until the conclusion of the study. Each teacher received the same template prompting the use of modeling with IMs.

Finally, coaching emails were delivered to teachers via email daily. These emails were scored for implementation fidelity ensuring they adhere to the coaching template from Kennedy, Rodgers, et al. (2017). The template began with a positive statement about the lesson and a statement about why this positive element is beneficial to students. Next, the template drew the teacher's attention to a single item for improvement. In this study, the areas for improvement were modeling IMs. If helpful, the email noted where on the CT Scan output the corresponding information could be found. After stating what the teacher could improve about the IM, the email indicated the expected benefit to students. See Figure 8 for the checklist which were used to rate implementation fidelity of the coaching emails. Two graduate students scored emails for implementation fidelity. Implementation fidelity was 90% with 100% agreement between the two raters.

Figure 8
Coaching Email Fidelity Checklist

Email Checklist

- Includes a positive statement about an aspect of the observation
    - Indicates where positive element can be found on the CT Scan output
    - Explains why positive element is beneficial to students
- Includes a single element of instruction for improvement
    - Indicates where the element can be found on the CT Scan output
    - Provides tip for implementing element during instruction
    - Explains expected benefit to students

Maintenance Condition (C)

After the CAP-PD phase, observations and data collection continued in the same manner as the baseline and CAP-PD phase. Teachers did not have access to the CAP-TV or receive instructional coaching in the maintenance phase. However, they did have access to the CAP-TS. Teachers continued to administer CBM-W to their students as they had previously. The maintenance phase continued until the final teacher had a minimum of five observations and a predictable pattern of performance was established.

Data Analysis

To make causal claims about the effect of the CAP-PD intervention, recommendations from the *What Works Clearinghouse: Single-Case Design Guidelines* (Kratochwill et al., 2010) guided data analysis. Data was presented in time series graphs with the teacher observation data on the Y-axis and time on the X-axis. Data was analyzed using traditional visual analysis considering changes in the level, trend, and variability in performance (Horner et al., 2005). According to Horner et al. (2005), level is the mean performance for a participant in a phase. The trend is the line of best fit through the data. Variability refers to the consistency of performance around a mean or trend within phases.

A change in level indicates the average performance of the participant was increased or decreased. A change in level could be a positive change, such as increasing the average number of IMs used in a lesson. However, a change in level could also be negative such as decreasing the average number of IMs used in a lesson.

A change in trend indicates the rate of performance was altered. Changes in trend could include increasing the rate of change demonstrated in a previous condition or changing the direction of the trend. For example, a teacher making slow, insufficient progress of modeling implementation could have a slightly positive trend in the baseline condition. When the intervention is introduced, a positive change in trend would increase the previous trajectory and lead to faster high-quality implementation of modeling IMs. Alternatively, in the case of a teacher demonstrating fewer IMs for modeling as the baseline phase progresses, a successful intervention could change the trajectory of the teacher's performance and increase the rate of IMs used in a lesson thus changing the trend of performance from decreasing rate to an increasing rate.

A change in variability indicates the consistency of performance around that mean or trend is altered. For example, if a teacher demonstrates wildly inconsistent percentages of IMs from lesson to lesson, the CAP-PD intervention could demonstrate a positive effect by creating a more stable, predictable pattern of performance. In this case, the teacher might not have a change in level or trend due to the high variability in the baseline condition, but the study was able to demonstrate a positive effect on teaching performance.

When analyzing changes in level, trend, and variability, Horner et al. (2005) note the importance of considering the immediacy of effect, the proportion of overlap in adjacent conditions, the magnitude of changes in dependent variables, and the consistency of data pattern across multiple implementations of independent variable. Data that do not conform to these four considerations can compromise the demonstration of a functional relationship.

Chapter IV: Results

Using the Classroom Teaching (CT) Scan, I conducted direct observations of teachers' writing instruction with special focus on modeling and the use of modeling implementation markers. These implementation markers are listed in Table 1. Data are presented individually for each teacher in Figure 9. Each observation was video-recorded. These video recordings were used have lessons double-scored with the CT Scan for reliability purposes. Two observers trained in the use of the CT Scan watched 20% of the observations for each teacher.

These video recordings were also used to score instruction with the Protocol for Language Arts Teaching Observations (PLATO). Two observers trained in the use of the PLATO scored two videos per phase for each teacher. These PLATO observations occurred on the first and last days of the baseline, intervention, and maintenance phases for each teacher. These data were not collected frequently enough to make experimental claims in a single-case design. Therefore, these data are presented to describe the instructional context of the study.

On the fifth day of each new phase, students wrote in response to a curriculum-based measurement (CBM) writing prompt. The number of correct minus incorrect word sequences (CIWS) was scored for each writing prompt. These data were scored by two doctoral students trained in the scoring of CBM for writing. Twenty percent of the writing prompts were double-scored for reliability purposes.

The social validity survey was delivered to teachers electronically via Qualtrics. The survey included Likert-type items and open-ended questions. Each Likert-type item was followed by an open-ended question that allowed participants to elaborate on the Likert-type item. Participants were forced to respond to the Likert-type items, but they had the option of leaving the open-ended questions blank.

The following sections in this chapter present the results from these four dependent measures. First, I present results from the CT Scan modeling implementation checklist. This measure was the only experimental measure in this study. Then, I present results from the PLATO data. Third, I present results from the students' CBM writing samples. Finally, I present results from the social validity survey.

Teachers' Use of Modeling Implementation Markers

Data from the CT Scan modeling implementation checklist was organized into time series graphs for each teacher. In these graphs, presented in Figure 9, the X axis represents days of observation. These days of observation do not correspond exactly to school calendar days because some days were dedicated to testing, visiting the library, or other non-instructional purposes. The primary Y axis (left side) represents performance on the CT Scan modeling implementation checklist. The secondary Y axis (right side) represents average student performance on the CBM writing prompt as scored by CIWS. A solid vertical line on the graph represents a change from the baseline to intervention conditions, and a dotted vertical line represents a change from the intervention to maintenance conditions.

Kratochwill et al. (2013) present four steps of analysis and six features of the data to consider when determining the presence or absence of a functional relationship between the independent and dependent variable. Step 1 is to establish a predictable and stable data patterns in the baseline phase. Step 2 considers the data patterns within each phase of the experiment checking for within phase patterns. Step 3 examines the immediacy of effect by looking at the data points in adjacent phases that are closest to the phase change line. The fourth and final step is to determine whether there are at least three demonstrations of an effect at different points in time. When examining data patterns within and across phases in these steps, visual analysis considers the level (mean), trend (slope), variability, immediacy of effect, overlap, and consistency of data patterns across similar phases. In the following sections, I describe the data related to the first three steps for each teacher in the order they received the intervention. After describing the performance of each teacher, I discuss the data in relation to the fourth step: demonstration of two replications of effect. I calculated the percent of non-overlapping data as an effect size and presented descriptive statistics to support visual analysis in Table 6.

Figure 9

Graphs of Teachers' Use of Modeling Implementation Markers.

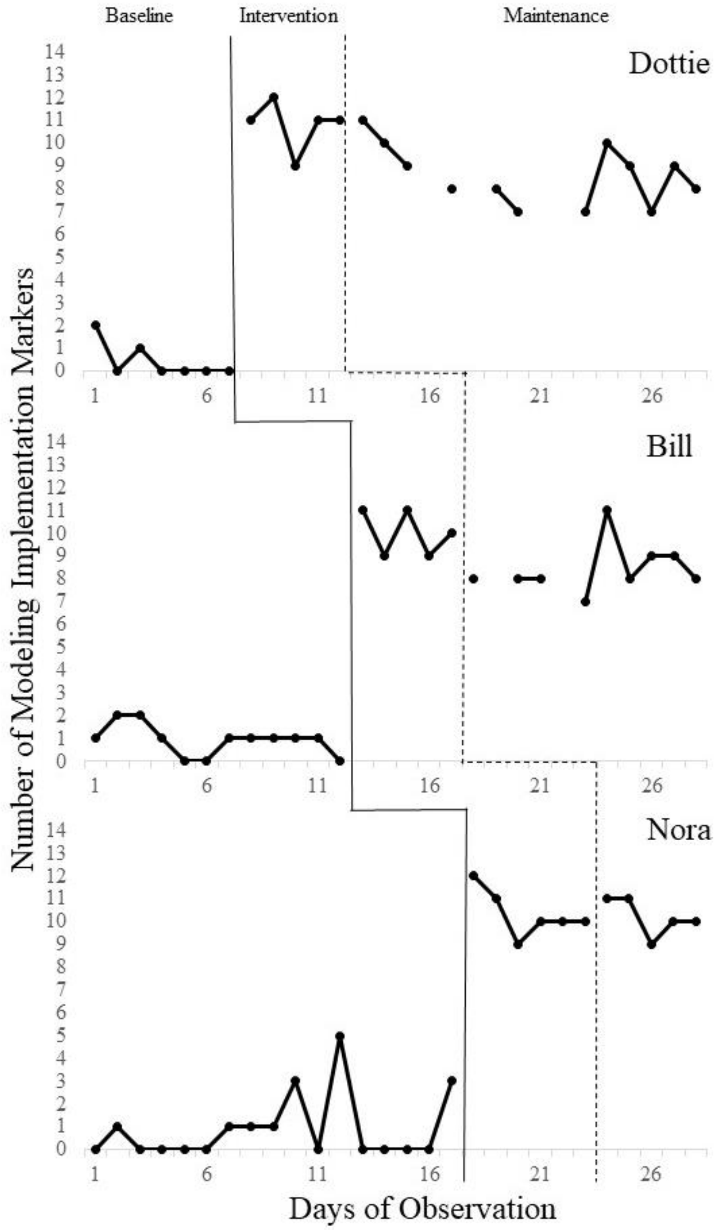


Table 6

Descriptive Statistics for Teachers Across Conditions

	Baseline Condition				Intervention Condition				Effect Size	Maintenance Condition				Effect Size
Teacher	M (SD)	Median (Range)	Mode	Slope	M (SD)	Median (Range)	Mode	Slope	N-ODP	M (SD)	Median (Range)	Mode	Slope	N-ODP
Dottie	0.43 (0.79)	0 (0-2)	0	-0.25	10.80 (1.10)	11 (9-12)	11	-0.10	100%	8.58 (1.31)	8.5 (7-11)	8	-0.11	100%
Bill	0.92 (.7)	1 (0-2)	1	0.05	10.80 (1.10)	10 (9-11)	9	-0.20	100%	8.44 (1.13)	8 (7-11)	8	0.09	100%
Nora	0.88 (1.6)	0 (0-5)	0	-0.08	10.33 (1.03)	10 (9-12)	10	-0.10	100%	10.20 (0.84)	10 (9-11)	10	-0.30	100%

Note. When two or modes are present, the lowest number is displayed here.

Dottie

Regarding the stability of the baseline performance, visual analysis for Dottie indicated very low and stable performance. She used an average of 0.43 modeling implementation markers during writing instruction. Her median score was 0 with a range of 0-2. Her mode score was 0.

Regarding the data patterns within each phase, visual analysis of the baseline phase indicated stable performance with a slight downward trend (slope coefficient = -0.25). Her intervention performance appeared flat with some slight variation for two of the five observations. The slope coefficient for her intervention data was -0.20.

Regarding the immediacy of effect, Dottie had an immediate increase in her level of performance from baseline to intervention. Her four data points closest to the phase change line demonstrated 0 implementation markers for modeling. After receiving the intervention, the next observations demonstrated 11, 13, 9, 11, and 11 implementation markers for modeling respectively. This change in level demonstrated an immediate effect at the time of the intervention. Dottie had no overlapping data when comparing the intervention and baseline phases. The percent of non-overlapping data was 100%.

Bill

Regarding the stability of the baseline performance, visual analysis indicated that Bill had a stable baseline with few implementation markers of modeling. Although somewhat higher than Dottie, baseline performance for Bill was quite low. Descriptive analysis supported this visual analysis. His average score was 0.92, and his mode score was 1. His range of scores was 0-2.

Regarding the data patterns within each phase, his baseline performance appeared to have a recurrent wave trend based on visual analysis. This trend appeared to rise to a peak, plateau, and return to 0 before beginning the rise in performance again. The second rise in performance did not reach the same peak as the first rise. The slope for the baseline phase was -0.08. Visual analysis of the intervention phase indicated a predictable rise and fall in performance on successive days and a slightly decreasing trend overall. His slope was -0.20.

Regarding the immediacy of effect, Bill had an immediate increase in his level of performance based on visual analysis. His most recent baseline observations demonstrated 1, 1, 1, 1, and 0 implementation markers of modeling respectively. His first intervention observation demonstrated 11 implementation markers of modeling, and his intervention data was never lower than 9 (range 9-11). His intervention phase mean was 10, mode was 9, median was 10, and range was 9-11. There was no overlapping data between his baseline and intervention scores. The percent of non-overlapping data was 100%.

Nora

Regarding the stability of baseline performance, Nora had the most variability in baseline performance of the three teachers. However, her baseline data was still relatively stable and predictable. Although she was the only teacher to have an increasing trend in her baseline performance and the highest upper limit in baseline performance (range 0-5), her performance on average was similar to the other two teachers. Her baseline mean was 0.88, mode was 0, and median was 0.

Regarding the data patterns within each phase, visual analysis of her baseline performance indicated a slight increasing trend. Her slope coefficient was 0.08, confirming the visual analysis of a slightly increasing trend. Her intervention performance had a decreasing trend and was somewhat steeper than the other two teachers. Her intervention phase slope coefficient was -0.34. However, her performance appeared to plateau after some initial variability upon receiving the intervention.

Regarding the immediacy of effect, Nora had an immediate increase in the level of her performance upon receiving the intervention based on visual analysis. Prior to receiving the intervention, her most recent baseline observations received scores of 0, 0, 0, 0, and 3. Immediately after receiving the intervention, her next observations received scores of 12, 11, 9, 10, 10, and 10. There was no overlapping data between her baseline and intervention performance. The percent of non-overlapping data was 100%.

Replications of Effect

To have a functional relationship between an independent and dependent variable in a single case design, the study must demonstrate two replications of an effect (Kratochwill et al., 2013). As demonstrated in the previous paragraphs, Dottie, the first teacher to receive the intervention, demonstrated a functional relationship between receiving the CAP-PD intervention and an immediate increase in the level of performance. This effect was replicated by Bill and Nora. All three teachers demonstrated a similar effect after receiving the intervention. Performance ranged between 9 and 12 implementation markers for all teachers in the intervention phase. Each teacher had a similar decreasing trend in intervention performance, and no overlapping data between baseline and intervention.

Maintenance of Effect

Because the observation data indicated a functional relationship between the intervention and the number of implementation markers for modeling, further visual analysis was conducted to determine whether the effect was maintained over time. Dottie had a downward trend in the maintenance phase. The slope of this trend was steeper than her downward trend in the intervention condition. She had no overlapping data with the baseline condition. Also, she had 50% non-overlapping data in the maintenance condition when compared to the intervention condition, indicating the half the maintenance data was below the lowest intervention data point. Dottie had some missing observations in the maintenance condition due to absences from school.

Bill had a slight drop in the level of his performance. However, his performance was fairly stable (slope = 0.09) for the duration of the maintenance phase. His maintenance data was higher than his baseline data and had no overlap with the baseline phase. His maintenance data had 66.67% of non-overlapping data in comparison to the intervention phase, indicating that one-third of the data was lower than the lowest intervention data point.

Nora had the shortest maintenance phase. The pattern in her maintenance data looked very similar to her intervention data. She had no overlap between her baseline and maintenance phase. She had 100% overlap between her intervention and maintenance phases, indicating that her maintenance performance was completely within the range established during the intervention.

Percent of Class Period Spent Modeling

In addition to the number of modeling implementation markers, I also graphed the percent of class time spent modeling. The percent of class time spent modeling was not a research question of interest. However, I graphed this data to determine whether the demonstrated functional relationship between the intervention and number of implementation markers had any residual effect on the percent of time spent on modeling.

As indicated by Figure 10, the percent of time spent modeling was more variable during the baseline condition than the number of implementation markers. If this dependent measure was the experimental measure for the study, the baseline condition for Dottie would have been too variable to begin the intervention after the 7th observation. Specifically, the 7th observation demonstrated a substantial spike in the percent of time spent modeling. However, even with the variable baseline condition, Dottie had apparent change in level and spent more time modeling during the intervention than she did during the baseline condition on average. The intervention condition did have some overlap with the baseline condition.

Bill had a baseline pattern similar to the pattern present for the number of implementation markers in a class period. During the intervention, he had an immediate increase in the level of time spent modeling. However, this change in level does not appear to be different from the pattern established during the baseline phase. His intervention data had 100% overlap with his baseline data.

Nora had a variable baseline similar to the number of implementation markers during the baseline phase. During the intervention phase, she had an immediate increase in the amount of time spent modeling. However, the data for this phase had a steep downward trend and substantial overlap between the baseline and intervention phases. Fifty percent of her intervention data overlapped with her baseline data.

Based on this data, it does not appear that the intervention had a functional relationship with the percent of time during a class period that teachers spent modeling. In order to establish a functional relationship, at least two replications of an effect must be demonstrated. Each of the teachers had data features that did not support a functional relationship. See Table 7 for the descriptive statistics for time spent modeling during the observation. Because there was no functional relationship between the intervention and the intervention, there was no effect to maintain. Therefore, no further visual analysis of the maintenance data was conducted.

Figure 10
Percent of Time Spent Modeling

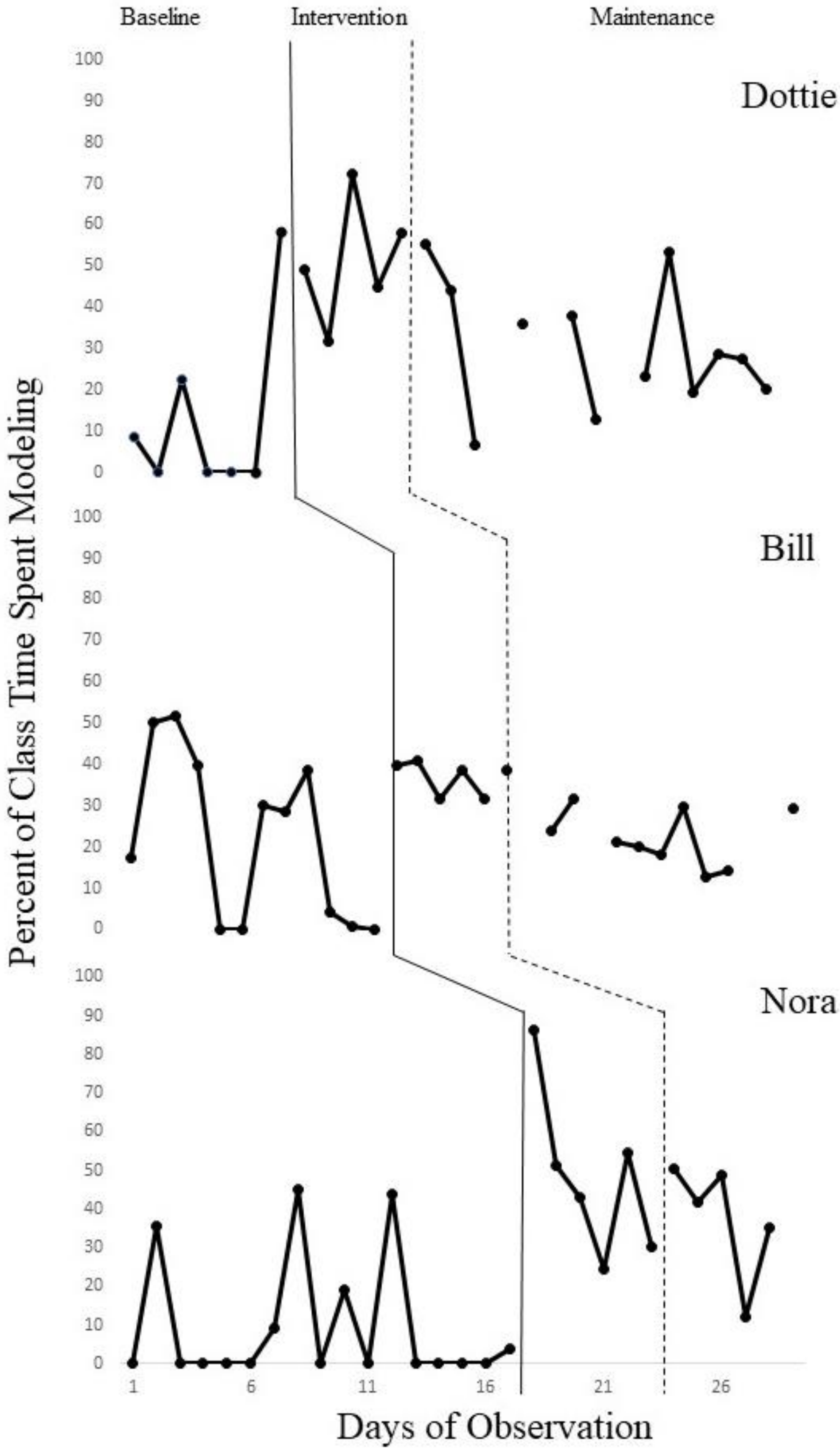


Table 7
Amount of Time Spent Modeling per Observation Across Phases

	Average Percent			Average Minutes		
	Baseline	Intervention	Maintenance	Baseline	Intervention	Maintenance
Dottie	12.79	51.20	28.44	10.54	42.61	21.74
Bill	21.76	36.48	23.75	19.14	27.06	18.01
Nora	8.71	48.35	37.68	6.65	36.76	26.92

PLATO

Two researchers coded video-recorded observations using the PLATO. The researcher who coded all observations was blind to the condition of the participants and the design and purpose of the study. The second observer coded a subset of the observations for reliability purposes. She was blind to the condition of the participants while scoring, but she was aware of the design and purpose of the study. These observations were not collected enough times to lead to experimental claims. However, these data serve two purposes. First, they describe the nature of instruction more fully than the CT Scan modeling checklist. Second, they point to areas for future research.

In general, PLATO scores were fairly low. Across all observations and all elements, the mode score was 1, the mean was 1.81, and the median was 2. All teachers scored relatively high in both elements under the Classroom Environment domain. Bill averaged 2.25 for Time Management and 3.33 for Behavior Management across all observations. Nora averaged 2.75 for Time Management and 3.50 for Behavior Management across all observations. Dottie averaged 3.00 for Time Management and 4.00 for Behavior Management across all observations. Alternatively, Text-Based Instruction was a relative weakness for all teachers with all teachers receiving a score of 1 for every segment of instruction scored with the PLATO. A score of 1 for Text-Based Instruction indicates that there were “no opportunities for students to engage in the writing process, or the students' writing [was] formulaic (e.g., fill-in-the-blank, recopying), or less than a paragraph in length of connected text" (PLATO, 2017, p. 13).

In terms of the content of the observations, Grammar/Spelling was the primary Content Domain for 34 out of 38 scored segments. Other Content Domains included Writing Instruction: Drafting (twice), Writing Instruction: Organizing, and Word Study/Vocabulary/Decoding.

Again, these data were not collected frequently enough to make claims about a functional relationship in a single case multiple baseline design, and the sample size was too small to make group experimental analysis appropriate. However, the data do indicate that future research using the PLATO or other observational tool may be warranted. See Table 8 for the results for each element per observation and teacher. The paragraphs below explain some of the notable trends.

Table 8

PLATO Element Scores for Each Teacher

	Dottie												Nora												Bill											
	O1		O2		O3		O4		O5		O6		O1		O2		O3		O4		O5		O6		O1		O2		O3		O4		O5		O6	
Modeling and Use of Models	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	2	1	2	3	1	1	2	1	1
Strategy Use and Instruction	1	1	1	2	3	1	1	1	1	1	1	1	1	3	1	1	1	1	1	2	1	2	1	1	2	1	2	2	1	1	4	2	1	2	1	1
Feedback	2	2	2	1	1	1	2	2	1	2	1	2	2	2	2	2	1	2	2	2	1	2	1	2	2	2	2	2	1	2	2	2	1	2	1	2
Accommodations and Language Learning	1	1	2	1	2	1	2	2	1	2	1	2	1	2	2	2	1	1	2	2	2	2	2	2	1	2	3	1	2	2	3	3	1	2	2	3
Intellectual Challenge	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	3	1	1	1
Classroom Discourse	2	1	2	2	2	2	1	2	1	2	1	1	2	1	2	2	1	1	2	2	1	3	1	2	2	1	2	2	1	2	2	2	1	2	2	2
Text-based Instruction	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Representations of Content	2	2	1	2	3	2	2	1	1	2	1	1	3	3	1	3	2	1	2	3	3	3	1	2	3	2	3	2	2	2	3	3	1	2	2	2
Connections to Prior Academic Knowledge	1	1	1	2	2	1	2	2	1	1	1	1	1	2	1	2	1	2	2	1	1	1	2	2	2	1	2	1	2	1	2	2	1	2	2	2
Purpose	1	1	1	2	3	1	2	2	2	2	1	1	1	2	1	2	1	3	3	2	2	4	3	3	1	1	2	2	1	1	2	2	2	2	2	2
Time Management	2	2	2	2	2	3	1	3	3	3	1	3	3	2	4	2	2	3	3	4	3	4	1	2	3	2	3	4	2	3	4	3	3	3	3	3
Behavior Management	3	4	3	3	4	4	4	3	4	3	2	3	4	3	4	3	4	4	4	4	4	4	2	2	4	4	4	4	4	4	4	4	4	4	4	4

Note. Shaded observations indicate baseline observations

Bill had only one baseline observation scored with the PLATO. However, in comparison to the baseline observation, Bill scored higher on the Modeling and Use of Models element in the first intervention observations. However, remaining observations returned to the baseline level of performance for this element. Similarly, the Strategy Use and Instruction element was higher than baseline performance shortly after the intervention. The performance level for this element also returned to baseline levels. Alternatively, performance on the Connections to Prior Knowledge and Purpose elements had relatively sustained improvement for most of the intervention observations.

Dottie had two baseline observations scored with the PLATO. Post-intervention scores indicated higher performance on the Modeling and Use of Models element. However, this increased performance was somewhat inconsistent with some post-intervention segments receiving low scores. No other element had a clear trend partially due to the variability of the data both pre- and post-intervention.

Nora had the three baseline observations and the most of any participant in the study. Post-intervention this teacher scored higher on average for the Purpose element than she had during the baseline observations. This teacher also had a slight increase in scores for the Time Management domain post-intervention.

The 12 PLATO elements are organized into four domains. Domain scores are created by averaging the individual element scores. Examining these domains made some of the differences from baseline to intervention clearer. See Table 9 for PLATO domain scores.

Nearly every score for Bill in Instructional Scaffolding post-intervention was higher than his baseline scores. The post-intervention scores for Representations and Use of Content were generally higher than the baseline observations. His post-intervention scores for Disciplinary Demand and Classroom Environment stayed relatively the same as the baseline scores.

Nora had relatively stable scores for Instructional Scaffolding, Disciplinary Demand, and Classroom Environment from baseline to post-intervention. On average, scores for Representations and Use of Content post-intervention were slightly higher than baselines scores.

Dottie had some scores for Instructional Scaffolding that were higher than baseline scores, but on average there did not appear to be a clear difference between baseline and post-intervention segments. Scores for Disciplinary Demand, Representations and Use of Content, and Classroom Environment were all relatively stable from baseline to post-intervention.

Table 9

PLATO Dimension Scores for Teacher

Observation	Instructional Scaffolding	Disciplinary Demand	Representations and Use of Content	Classroom Environment
		Bill		
1	1.25	1.33	1.33	2.5
1	1.25	1.00	1.33	3.00
2	1.75	1.33	1.00	2.50
2	1.25	1.33	2.00	2.50
3	2.00	1.33	2.67	3.00
3	1.00	1.33	1.33	3.50
4	1.50	1.00	2.00	2.50
4	1.50	1.33	1.67	3.00
5	1.00	1.67	1.33	3.50
5	1.50	1.33	1.67	3.00
6	1.00	1.00	1.00	1.50
6	1.50	1.00	1.00	3.00
		Nora		
1	1.25	1.33	1.67	3.50
1	2.25	1.00	2.33	2.50
2	1.50	1.33	1.00	4.00
2	1.75	1.33	2.33	2.50

3	1.25	1.00	1.33	3.00
3	1.25	1.00	2.00	3.50
4	1.50	1.33	2.33	3.50
4	1.75	1.33	2.00	4.00
5	1.50	1.00	2.00	3.50
5	2.00	2.00	2.67	4.00
6	1.25	1.00	2.00	1.50
6	1.50	1.33	2.33	2.00
		Dottie		
1	1.50	1.33	2.00	3.50
1	1.50	1.00	1.33	3.00
2	2.00	1.33	2.33	3.50
2	1.75	1.67	1.67	4.00
3	1.25	1.00	1.67	3.00
3	1.75	1.33	1.33	3.50
4	3.00	1.33	2.33	4.00
4	2.00	1.33	2.33	3.50
5	1.00	1.67	1.33	3.50
5	2.00	1.33	2.00	3.50
6	1.25	1.33	2.00	3.5
6	1.75	1.33	2.00	3.5

Note: Baseline observations are indicated by darker shading.

Curriculum-Based Measurement

Students completed a CBM writing prompt on the fifth day of each phase. Students wrote in response to the persuasive writing prompt for seven minutes. These prompts were then scored for the number of correct minus incorrect word sequences (CIWS). Prompts were scored by three graduate students trained in the scoring of CBM for writing. Twenty percent of the prompts were double-scored for reliability purposes. These data were not collected frequently enough to make experimental claims in a multiple baseline design. Therefore, these data for all students are presented descriptively in Table 10. The data for only students with disabilities are presented in Table 11.

Although these data are presented for descriptive purposes only, it should be noted that most of the CBM data collected in the intervention phase had significant overlap with the baseline data. Variation in performance appeared to be due to the nature of the writing prompt rather than the phase of the intervention. Also of note, these data have very large standard deviations, indicating a large amount of variability within each teacher's set of students for each prompt. Because CIWS is a relatively new CBM-W scoring procedure, normative scores do not exist 8th grade students. Therefore, it is difficult to make comparisons between this sample and a broader population of students.

Table 10
CBM-W Results for All Students

Teacher	Prompt 1 *M* (sd)	Prompt 2 *M* (sd)	Prompt 3 *M* (sd)	Prompt 4 *M* (sd)	Prompt 5 *M* (sd)	Prompt 6 *M* (sd)
Bill	33.15 (31.31)	44.50 (32.16)	48.65 (34.58)	36.32 (33.14)	32.83 (26.09)	35.44 (29.26)
Nora	45.85 (26.80)	59.21 (31.77)	69.32 (31.81)	49.32 (37.24)	46.35 (28.26)	42.20 (16.28)
Dottie	41.68 (41.69)	46.25 (32.74)	57.69 (45.71)	-1.21 (35.72)	36.53 (29.90)	42.53 (26.49)

Note. Baseline data are indicated by darker shading.

There was also a substantial amount of overlap in the baseline and intervention data for students with disabilities. In comparison to their peers, students with disabilities had substantially lower CBM scores on average. Again, these data had very large standard deviations relative to the size of the means.

Table 11
CBM-W Results for Students with Disabilities

Teacher	Prompt 1 *M* (sd)	Prompt 2 *M* (sd)	Prompt 3 *M* (sd)	Prompt 4 *M* (sd)	Prompt 5 *M* (sd)	Prompt 6 *M* (sd)
Bill	2.60 (14.35)	23.00 (29.05)	14.80 (18.43)	5.20 (4.32)	9.80 (12.62)	4.50 (18.95)
Nora[a]	5 (*)	1 (*)	24 (*)	10 (*)	-	6 (*)
Dottie	28.67 (47.59)	34.00 (30.16)	29.33 (48.01)	36.00 (38.56)	12.60 (28.27)	14.40 (16.89)

Note. Baseline data are indicated by darker shading. [a]Nora had only one student with an IEP in her class. Therefore, no standard deviation is available for these prompts. This student was absent for prompt 5.

Social Validity

Teachers completed a 24-item social validity survey. Twelve items asked participants to respond to a Likert-type prompt on a scale of 1-5. After each of these items, participants had the opportunity to respond to an open-ended prompt further explaining the rating they gave. For example, the first item asked participants, "on a scale of 1-5 with 1 being not familiar at all and 5 being very familiar, how familiar with the components of modeling were you before the training?" After that item, teachers had an open response box with a prompt that said, "in the space below, please feel free to elaborate on your familiarity with the modeling components before the training. You may leave the box blank if you wish." See Table 12 for each participant's numeric responses to the social validity survey.

Table 12
Social Validity Survey Results by Teacher

Item	Teacher		
	Bill	Nora	Dottie
On a scale of 1-5 with 1 being not familiar at all and 5 being very familiar, how familiar with the components of modeling were you before the training?	2	1	2
On a scale of 1-5 with 1 being not familiar at all and 5 being very familiar, how familiar are you now with the components of modeling?	4	4	5
On a scale of 1-5 with 1 being very dissatisfied and 5 being very satisfied, how satisfied are you with the video training you received?	4	3	5
On a scale of 1-5 with 1 being very dissatisfied and 5 being very satisfied, how satisfied are you with the email coaching/feedback you received?	5	4	5
On a scale of 1-5 with 1 being very dissatisfied and 5 being very satisfied, how satisfied are you with the PowerPoint template you received?	5	3	5
On a scale of 1-5 with 1 being very difficult and 5 being very easy, to what extent do you think the components of modeling were easy to learn?	5	4	4

On a scale of 1-5 with 1 being very difficult and 5 being very easy, to what extent do you think the components of modeling were easy to implement in your classroom?	4	4	4
On a scale of 1-5 with 1 being none at all and 5 being a great deal, how much preparation time was required for you to implement modeling?	4	3	4
On a scale of 1-5 with 1 being none at all and 5 being a great deal, how much classroom time was required for you to implement modeling?	3	4	4
On a scale of 1-5 with 1 being not at all effective and 5 being very effective, how effective do you think modeling was for your students?	4	4	5
On a scale of 1-5 with 1 being very unlikely and 5 being very likely, how likely are you to continue using modeling in your instruction?	5	5	5
On a scale of 1-5 with 1 being very unlikely and 5 being very likely, how likely are you to recommend modeling to a colleague?	5	4	5

In general teachers reported high levels of satisfaction with modeling as an instructional strategy and with CAP-PD as an intervention. All teachers reported low familiarity with the components of modeling before the intervention (range 1-2) and high familiarity with the components of modeling after the intervention (range 4-5). Dottie said that she knows the components of modeling, but needs to be more deliberate in using them in her instruction. However, Nora said that using the components of modeling became second nature when teaching a new concept.

Teachers reported that the components of modeling were easy to learn and easy to implement in the classroom. Bill indicated that having access to watch the CAP-TVs repeatedly made the information easier to learn. Dottie said that the "direct guidance" (coaching emails) made learning easier.

Teachers also reported that they thought modeling was effective for their students. Dottie said that, "When modeling was done correctly, [her] students had no problem" with the new skill being taught. Bill said that he was “a big fan of modeling” and thought it helped his students. All teachers reported they were very likely to continue using modeling in their instruction and would likely recommend modeling to a colleague.

Teachers generally reported very high levels of satisfaction with each component of CAP-PD (i.e., CAP-TV, CAP-TS template, and CT Scan coaching). Specifically regarding the CAP-TV, teachers Bill, Nora, and Dottie rated their satisfaction as a 4, 3, and 5, respectively. Bill said he was “a visual learner, so [he] liked things to be modeled.” As mentioned previously, Bill also said he thought having access to repeatedly viewing the CAP-TV made learning the information easier. Nora rated her satisfaction with the CAP-TV and CAP-TS as a 3. In the open response sections, she indicated that she wanted more support for using the CAP-TS in her lesson preparation. She also indicated that she would have liked the CAP-TS to be used in the CAP-TV so she could see a model of the intervention materials being prepared and used.

All three teachers reported high levels of satisfaction with the email coaching (range 4-5). Teacher Bill said, “it was nice to know where [he] need[ed] to improve, even if being criticized (constructively), isn’t easy to take” (parenthetical is original). Dottie said she “needed the feedback to help adjust [her] time and direction.” Nora said she would have liked to receive coaching on how to prepare lessons using the CAP-TS. The focus on coaching was on instructional practice in the classroom. At times the coaching emails directed teachers to the CAP-TS as an aid in their teaching, but the coaching did not provide specific support for how to prepare lessons using the CAP-TS.

Two of three teachers reported very high satisfaction with the CAP-TS. Nora rated her satisfaction as a 3. She said she wanted more “explanation or coaching” for how to use the CAP-TS. Bill said the CAP-TS "made it easier and more effective" for him to teach. Dottie said they are still using the CAP-TS in their planning as a team.

Two of three teachers indicated they spent a large amount of time preparing lessons to incorporate modeling. Interestingly, the teacher who reported the most struggle with using the CAP-TS reported spending the least amount of time preparing the lessons. In the open response section, Bill indicated the preparation requirements were high at the beginning of the intervention but declined as he became more comfortable with the new instructional strategy and materials. Dottie said that any extra preparation time that was required was "worth the effort."

Chapter V: Discussion

This study examined the effects of a professional development package designed with a cognitive apprenticeship framework. Cognitive apprenticeship is a multifaceted conceptual framework for teaching skills to K-12 students (Collins et al., 1989). This framework has several features to it (methods, content, sociology, and sequencing). However, the core of the framework is a cycle of modeling, coaching, and scaffolding. As outlined in chapter two, the individual components of this framework had strong empirical support, yet with few exceptions, the framework as a whole lacked rigorous empirical support.

In one such exception, Kennedy, Rodgers et al. (2017) used cognitive apprenticeship to create a professional development package, Content Acquisition Podcast – Professional Development (CAP-PD), to improve middle school teachers' science instruction. CAP-PD aligned to the cognitive apprenticeship framework by providing modeling (instructional vignettes), coaching (performance feedback via email), and scaffolding (customizable curriculum materials). This single case multiple baseline research design demonstrated a functional relationship between the intervention and three teachers' use of evidence-based vocabulary practices.

The present study built on the findings of Kennedy, Rodgers, et al. (2017) in two important ways. First, this study added a distal teacher observation measure to provide more context to the instruction. Second, this study included a measure of student academic achievement. In addition to these improvements, this study extended the CAP-

PD literature base to study English teachers' writing instruction. This extension made some changes to CAP-PD necessary. Specifically, rather than providing separate customizable curriculum materials (Content Acquisition Podcasts-Teacher Slides; CAP-TS) for each lesson, this study provided teachers with one customizable template that they could use to plan and implement during all lessons. The results of this study are discussed below and are followed by a discussion of how this study added to the cognitive apprenticeship framework.

CT Scan

In the previous chapter, two types of data were presented from the CT Scan. First, the number of implementation markers for modeling, the experimental dependent measure for this study, indicated a functional relationship between the intervention and the quality of teachers' modeling as defined by the implementation marker checklist. Second, the duration of modeling in each lesson was examined to determine whether there was a functional relationship between the intervention and instructional time.

Number of Implementation Markers

Each teacher had an immediate increase in the level of performance at the phase change line. This marked a clear improvement in the quality of modeling provided by the teachers. As demonstrated by the teachers' baseline data for implementation markers and instructional time spent modeling, teachers were doing some modeling before the intervention. However, the quality of the modeling greatly improved from an explicit instruction perspective after receiving the intervention. Although teachers demonstrated many more implementation markers of modeling post-intervention, some implementation markers remained difficult for teachers to implement. The "enthusiasm" and "lively pace" markers were rarely demonstrated by teachers. In addition, although not as rare, "step-by-step demonstrations" were also not used frequently by the teachers in this study.

Although the improved instruction maintained above baseline levels for the duration of the study, the maintenance phase for two teachers was on a downward trend. The long-term effects of this intervention were not a research question answered by this study, but they are worth considering in future research. It is also important to consider what supports could be provided to teachers to help maintain or improve performance. Specifically, allowing teachers to have access to the modeling videos and providing intermittent coaching to teachers support teachers in the maintenance of performance over a longer period of time.

Time Spent Modeling

Although it was not the experimental dependent variable for this study, data for the amount of time spent modeling was presented in the previous chapter to look for a possible residual relationship between the intervention and instructional time. In accordance with single case experimental designs (Kratochwill et al., 2013), the intervention did not appear to have a functional relationship with the amount of instructional time spent modeling. Interestingly, this finding is counter to what teachers reported in the social validity survey with multiple teachers indicating the felt like they were spending a lot of time modeling for students. The fact that this new instruction did not require any additional instructional time should be viewed as a positive result from the study. Teachers are pressed for instructional time as it is and could be reluctant to implement a new instructional strategy if it is going to create an even more crowded schedule. This study demonstrated that teachers were able to improve their instruction without a secondary effect on their instructional time.

However, although a functional relationship between the intervention and time spent modeling was not demonstrated in this study, the possibility of such a relationship cannot be ruled out. The means for time spent modeling were quite different for teachers from baseline to intervention conditions. If these means were present in a larger group research design, they could be different enough to be considered statistically significant.

PLATO

Two significant technical issues made interpreting the PLATO scores difficult. First, floor effects could have limited the amount of growth demonstrated by the teachers. A floor effect occurs when actual performance falls below the bottom threshold of an assessment and, therefore, hinders the accuracy of the results (Ary, Jacobs, & Sorenson, 2010). Much of the instruction scored by the PLATO in this study received very low scores. The mode score for all segments was a 1 (the lowest possible score). Therefore, it is possible that the teachers' growth was masked by a floor effect.

A second, related issue was that the PLATO might not be sensitive enough to detect the changes in teacher performance present in this study. The PLATO has little evidence examining its sensitivity to growth over a short period of time. It is possible that the improvements the teachers made were too small to be reflected in the PLATO element scores and a more powerful intervention is necessary to be reflected in PLATO scores.

One practical issue that could have limited the growth demonstrated in PLATO scores was the segmenting of instruction and the standardized approach to scoring the observations. The first and last observation from each phase was scored for each teacher. For these observations, the first two segments of each lesson were scored. Therefore, the first 30 minutes of instruction was scored by the PLATO. In many cases, the teachers' modeling instruction occurred outside of the first 30 minutes. Scoring all segments for these observations could have revealed more growth for teachers by including the instruction that was designed based in the professional development intervention.

These technical and practical problems could have limited the ability of the PLATO to reveal growth made by the teachers. However, even if the PLATO had varied between baseline and intervention phases, there were not enough observations scored with the PLATO to make causal claims in a multiple baseline design. Although these data cannot be used experimentally, they did provide helpful descriptive information about the nature of instruction occurring in these classrooms.

Teachers provided very little text-based instruction. All teachers received a score of 1 for the Text-Based Instruction element for every segment during baseline and post-intervention. These data indicate that students were producing very little, if any, text during writing instruction. Anecdotally, I observed a limited number of days dedicated to students writing for a specific project or prompt. However, when these activities did occur, they fell on days that were not scored with the PLATO.

These data have several implications for use of the PLATO in future CAP-PD studies. First, an immediate next step for this study is to score all remaining segments of the PLATO. This scoring would ensure that any instruction that was based on CAP-PD was analyzed by the PLATO. Second, future CAP-PD studies should continue to use the PLATO with diverse teachers. Finding teachers with higher baseline PLATO scores may lead to more recognizable improvements in PLATO scores. Finally, using the PLATO in a group experimental design may be more appropriate than a single case design. A group experimental design would allow for analysis to detect small by significant changes in teacher performance.

Social Validity

Teachers generally reported very high satisfaction with CAP-PD as a professional development intervention and with modeling as an instructional strategy. They also noted helpful feedback that could improve CAP-PD in the future. Specifically, future modeling videos should be as similar to the participants' context as possible in terms of curricular materials. It is also important to consider ways of modeling the non-instructional strategies (e.g., planning) that must be done in order for instructional strategies to be done effectively.

CBM-W

Similarly to the PLATO data, CBM-W data was not collected frequently enough to support experimental claims. However, the intervention writing prompts did not vary substantially from the baseline prompts. Descriptively, the CBM-W data reinforced the need for improving writing instruction for students with disabilities by illustrating the gap in writing fluency between students with disabilities and their peers. Specifically, the writing samples from students with disabilities scored substantially lower on the CIWS scoring procedure than their peers without disabilities.

Implications

Cognitive Apprenticeship

In light of these findings, this study built support for cognitive apprenticeship as a conceptual framework for designing professional development. As discussed in chapter two, cognitive apprenticeship has little rigorous support for its use as a whole in professional development contexts. Peters-Burton et al. (2015) designed a professional development package to promote secondary teachers' inquiry instruction. In this study, all teachers received the professional development package at the same time (no control group), and there was no measure of teacher practice. This study was representative of the cognitive apprenticeship literature base in professional development contexts.

As mentioned previously, Kennedy, Rodgers et al. (2017) designed CAP-PD using cognitive apprenticeship as a guiding conceptual framework. This study continued the use of cognitive apprenticeship as a guiding framework in CAP-PD research. Based on the results of this study, the core features of cognitive apprenticeship now have two rigorous empirical studies that demonstrated a functional relationship between the use of the framework and improved teacher practice as defined by increased use of evidence-based practices for students with disabilities.

Some important holes in the cognitive apprenticeship literature base remain. First, the impact on student academic outcomes is largely unexplored. Although this study collected student writing samples, these data were not collected in a way that could be analyzed experimentally. Second, both Kennedy, Rodgers, et al. (2017) and this study used only the core features (modeling, coaching, and scaffolding) of cognitive apprenticeship in the development of the professional development activities. These core features are only one subset of a broader set of methods which are situated in an even larger framework. It will be important to study these other aspects of the framework in future studies. Specifically, the sociology of instruction is an important part of the framework that has been unexplored in Kennedy, Rodgers, et al. (2017) and this study.

Research

This study has two primary implications for CAP-PD research. First, this study extended CAP-PD into the area of evidence-based writing instruction. Previous studies of CAP-PD had focused on vocabulary instruction. Although the area of writing instruction poses new challenges not present in the vocabulary instruction domain, this study demonstrated that CAP-PD can be used to increase the quality of middle school teachers' writing instruction as defined by the CT Scan.

Second, this study extended the participants' examined in CAP-PD to include middle school English teachers. Previous studies of CAP-PD studied science teachers exclusively. The extension to include English teachers is important because teachers from different content areas may bring different conceptual backgrounds to professional development activities that cause them to respond differently.

This study is the second study of CAP-PD to demonstrate a functional relationship between CAP-PD and increased use of evidence-based practices for students with disabilities in middle school general education classrooms. However, these studies had some differences between them that have been discussed previously. Moving forward, it will be important to conduct conceptual replications of these studies to confirm their findings.

Practice

These results also have implications for teachers and teacher educators. First, the materials of this study, other than the coaching, are freely available to teachers. Teachers may benefit from accessing the materials and incorporating them into their instruction. Second, teacher educators should consider using a cognitive apprenticeship framework in the design of teacher preparation coursework and professional development.

Specifically, teachers need to have good models of effective instruction. Models could include videos or demonstrations of effective teaching, but it could also include mentor teachers that pre-service teachers see teach on a regular basis. Special attention should be paid to the models of instruction that teachers receive. Also, teachers need to have effective coaching on their instruction. This coaching should include micro-feedback on small teaching moves in addition to the more typical broad-based instructional feedback. Finally, teachers need scaffolds and supports to implement high-quality instruction. These scaffolds could include curricular materials, as in this study, or they could include other supports such as reduced class sizes for novice teachers.

Policy

This study further supports previous notions of effective professional development (Darling-Hammond et al., 2017; Guskey & Yoon, 2009). Specifically, teachers need sustained follow-up after an initial professional development experience. Policy-makers should support teacher professional development by providing sufficient funding to give teachers meaningful models, coaching, and scaffolds to sustain improved instruction.

Limitations

The results of this study should be viewed in light of some limitations. One possible threat to the internal validity of the study was observer drift. I conducted nearly all of the observations with the CT Scan. I was also responsible for delivering the intervention and analyzing the data. Therefore, I knew the condition of all participants at each point of the study. Although reliability for the implementation marker checklist was acceptable, future studies should use observers that are blind to the condition of the participants.

A second possible threat to the internal validity of the study was diffusion of treatment. All three teachers taught in the same school and were accustomed to planning together. At the beginning of the study, they agreed that they would not plan together when a teacher was in the intervention phase. However, it is possible that as the teachers received the intervention, some of the effects were diffused to other teachers via discussion of the days' lesson plans and implementation. Although there was no direct evidence for diffusion of treatment, it cannot be completely ruled out.

Although the conditions that led to possible diffusion of treatment are one possible limitation, it is also possible that these conditions will lead to sustained use of the intervention materials and instructional strategies in the future. In the social validity survey, one teacher indicated the teachers continue to use the CAP-TS. Anecdotally, some of the teachers asked if the CAP-TVs would be available after the conclusion of the study. It is possible that as the teachers work together to plan lessons they will continue their own professional development and improve their instruction as a group. This arrangement is actually more aligned to the cognitive apprenticeship framework than the individualized nature of the intervention provided through the study.

Future Research

This study provided dozens of avenues for future research. I highlighted a few of these avenues here. First, I present areas of research regarding the dependent measures. Second, I present areas of research for modeling as an instructional practice. Finally, I present areas of the CAP-PD intervention to research.

Dependent Measures

Regarding the dependent measures, three dependent measures for this study need further research for use in single-case experimental designs. First, as mentioned previously, the PLATO lacks evidence documenting its sensitivity to growth over very short periods of time. Future research should examine whether it is feasible to document growth in single case experimental designs.

Second, the CT Scan has been used in single case experimental designs previously (Kennedy, Rodgers, et al., 2017). Although, the CT Scan does have limited evidence for reliability and validity broadly, this observation tool needs more evidence supporting its reliability and validity in a variety of settings. The bulk of the research with the CT Scan has been conducted in science classrooms. This study developed a new practice (modeling) and corresponding implementation markers. The new practice and markers have no previous evidence to support their use. Future research should investigate how these codes are associated with other teacher observation measures and student outcomes.

Finally, although CBM in writing has scant evidence supporting its sensitivity to growth over longer periods of time, it has even less evidence supporting its use over very short periods of time. It is possible that essay writing is a skill that is not conducive to demonstrating student growth in a single case multiple baseline design. However, analytic rubrics or holistic ratings might be more sensitive to growth over short periods of time. Also, we have few norms of CBM writing performance with which to orient the performance of these students to a broader population.

Modeling

As discussed in chapter two, the research base on modeling as an instructional strategy for writing has very little guidance for specifically how teachers should model skills. The modeling checklist developed for this study was based on methods textbooks and explicit instruction textbooks. However, these sources were not necessarily based on research. Researchers need an understanding of how teachers should model. Ideally, researchers would identify the primary components of modeling that are required to be done in order to expect significant improvement in student outcomes and a secondary set of components that teachers have more freedom to incorporate as appropriate.

To identify these essential components of modeling, researchers need to recruit large samples of teachers representing diverse instructional ability levels. Participating teachers need to be observed with particular attention being paid to specifically how they implement modeling in the classroom. These observation data need to be linked with student writing assessments to examine associations between components of modeling and student outcomes.

CAP-PD

Finally, this study most directly identified areas for researching the CAP-PD intervention. A core principle of single-case research is a reliance on replication of studies to build the generalizability of findings. This study needs to be replicated with some limitations improved upon. Primarily, someone should conduct the observations who is blind to the condition of each participant. Also, a replication could recruit teachers from multiple schools and allow teachers at the same school to receive the intervention simultaneously. This change would limit the diffusion of treatment threat to internal validity and could potentially capitalized on the proposed benefits of the sociological aspects of the cognitive apprenticeship framework.

Aside from replicating the results of this study, the social validity survey identified some improvements that can be made to the CAP-PD intervention in the future. One of the teachers repeatedly indicated she needed more support with using the CAP-TS. One way of providing this support that she identified was to incorporate the CAP-TS into the modeling videos in the CAP-TV. This suggestion is certainly worth investigating empirically in the future. Additionally, it raises other questions regarding exactly how close the modeled instruction needs to be to the participants' context. For example, does participants' instruction improve more or at a faster rate if the model uses the same curriculum, facilities, or similar student population as will be available in the target instructional setting. Perhaps participants' instruction would improve if the model teacher was personally close to the target teachers. Recruiting model teachers who the participating teachers know and respect (instructional leaders, school- or district-level coaches, master teachers) might increase buy-in from participating teachers.

Related to supporting the use of the CAP-TS, more research needs to be done on how teachers use this resource for planning, designing, and incorporating lessons. This study was the first to use the CAP-TS template. Previous studies provided teachers with already completed PowerPoint presentations and gave them the option to edit the presentations as necessary. Providing a model of a completed presentation to aid teachers' use of the template might be worth investigating in the future.

Finally, more research needs to be done into the coaching methods for CAP-PD. All studies of CAP-PD that used the CT-Scan for observation have delivered coaching to teachers via email. Email has the advantage of standardizing the coaching for teachers, efficiently delivering coaching to teachers, and focusing the coaching to a targeted instructional practice. However, there may be advantages to in-person coaching. For example, Nora indicated she would have liked coaching on how to incorporate the CAP-TS into her planning and instruction. This may have come up in an in-person coaching session and could have been addressed earlier. Additionally, some teachers may feel that email coaching is less confrontational because there is not a person delivering the feedback face-to-face. However, some teachers may feel that email coaching is cold and lacks positive feelings towards the teacher. These are all questions that can be answered empirically in future research.

Summary

Writing achievement of students with disabilities is an area in significant need of improvement. Addressing writing instruction through professional development is one avenue for closing the achievement gap between students with disabilities and their peers. This study designed a professional development package, CAP-PD, in accordance with empirical evidence effective professional development and the cognitive apprenticeship framework. The results from the three teachers participating in the single case multiple baseline research design demonstrated a functional relationship between the CAP-PD intervention and the number of modeling implementation markers as recorded by the CT Scan.

Replication of results is a foundational principle of single case research designs. These results should be replicated in other single case and group research designs. Investigating the long-term effects of the intervention on teachers' instruction may require the use of group research designs. Additionally, significantly improving students' writing takes a long time, making group research designs potentially more feasible than single case designs.

References

Applebee, A., & Langer, J. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal, 100*, 14–27. Retrieved from http://0-search.proquest.com.source.unco.edu/docview/875295901?accountid=12832

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. New York: Guilford.

Ary, D., Jacobs, L. C., & Sorenson, C. K. (2010). *Introduction to research in education.* Belmont, CA: Wadsworth Cengage Learning.

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

Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Engelwood Cliffs, NJ: Prentice Hall.

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*, 29-58.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York, NY: Cambridge University Press.

Birnbrauer, J. S., Peterson, C. R., & Solnick, J. V. (1974). Design and interpretation of studies of single subjects. *American Journal of Mental Deficiency, 79*, 191-203.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing,* *learning, and instruction: Essays in honor of Robert Glasser* (453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

College Board. (2016). 2016 college bound seniors: Total group profile report. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/sat/total-group-2016.pdf

Cor, K. (2011). *The measurement properties of the PLATO rubric*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

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

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100,* 223–234.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Driver, M. K., Pullen, P. C., Kennedy, M. J., Williams, M. C., & Ely, E. (2014). Using instructional technology to improve preservice teachers' knowledge of phonological awareness. *Teacher Education and Special Education, 37,* 309-329. doi: 10.1177/0888406414537902

Englert, C. S., & Thomas, C. C. (1987). Sensitivity to text structure in reading and writing: A comparison between learning disabled and non-learning disabled students. *Learning Disabilities Quarterly, 10,* 93-105.

Ely, E., Kennedy, M. J., Pullen, P., Williams, M. C., & Hirsch, S. E. (2014). Improving instruction of future teachers: A multimedia approach that supports implementation of evidence-based vocabulary practices. *Teaching and Teacher Education*, *44*, 35-43. doi: 10.1016/j.tate.2014.07.012

Ely, E., Pullen, P., Kennedy, M. J., & Williams, M. C., (2015). Using multimedia to improve evidence-based vocabulary instruction in elementary classrooms. *Journal of Special Education Technology*, *30*, 59-72.

Festas, I., Oliveira, A. L., Rebelo, J. A., Damião, M. H., Harris, K., & Graham, S. (2015). Professional development in self-regulated strategy development: Effects on the writing performance of eighth grade Portuguese students. *Contemporary Educational Psychology, 40*, 17-27. doi:10.1016/j.cedpsych.2014.05.004

Fitzgerald, J., & Markham, L. R. (1987). Teaching children about revision in writing. *Cognition and Instruction*, *4*(1), 3–24. http://files.eric.ed.gov/fulltext/ED282220.pdf

Gage, N. A., Wilson, J., MacSuga-Gage, A. S. (2014). Writing performance of students with emotional and/or behavioral disabilities. *Behavioral Disorders*, *40*, 3-14.

Gillespie, A., Graham, S., Kiuhara, S., & Hebert, M. (2014). High school teachers use of writing to support students’ learning: A national survey. *Reading and Writing*, *27*, 1043–1072. doi:10.1007/s11145-013-9494-8

Goldstein, D. (2017, August 2). Why kids can’t write. *The New York Times.* Retrieved from https://www.nytimes.com/2017/08/02/education/edlife/writing-education-grammar-students-children.html?mcubz=1

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J., Olson, C. B., & Smither-Wulsin, C. (2016). Teaching secondary students to write effectively (NCEE 2017-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov

Graham, S., Capizzi, A., Harris, K., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing*, *27*, 1015–1042. doi:10.1007/s11145-013-9495-7

Graham, S., & Harris, K. R. (2002). Prevention and intervention for struggling writers. In M. Shinn, G. Stoner, & H. Walker (Eds.), *Interventions for academic and behavior problems II: Prevention and remedial approaches* (pp. 589-610), Bethesda, MD: National Association of School Psychologists.

Graham, S. & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, S. Graham (Eds.), *Handbook of learning disabilities* (323-344). New York: Guilford Press.

Graham, S., Harris, K. R., & Larsen, L. (2001). Prevention and intervention of writing difficulties for students with learning disabilities. *Learning Disabilities Research & Practice*, *16*, 74-84.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading. A Carnegie Corporation Time to Act Report.* Washington, DC: Alliance for Excellent Education.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476.

Grisham, D. L., & Wolsey, T. D. (2011). Writing instruction for teacher candidates: Strengthening a weak area. *Literacy Research and Instruction, 50,* 348-364. doi:10.1080/19388071.2010.532581

Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English language arts: Variations in PLATO scores by content and context. In T. J. Kane, K. A. Kerr & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 303-331). San Francisco, CA: Jossey-Bass.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development?. *Phi Delta Kappan, 90,* 495-500.

Hart, S. G. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183.

Hart, J. E., & More, C. M. (2013). Investigating the impact of technology on pre-service teacher knowledge of autism spectrum disorder. *Education and Training in Autism and Developmental Disabilities, 48,* 504-513.

Hirsch, S. E. (2016). Professional development in practice: Improving novice teachers' use of evidence-based classroom management practices (Doctoral dissertation). Retrieved from Libra.

Hirsch, S. E., Kennedy, M. J., Haines, S., Thomas, C. N., & Alves, K. D. (2015). Improving teacher candidate knowledge and application of functional behavioral assessments using multimedia. *Behavioral Disorders, 41*, 38-50. doi: 10.17988/0198-7429-41.1.38

Holt, Rinehart, & Winston. (n.d.). Writing prompts for middle school. *Holt Online Essay Scoring.* Retrieved from https://my.hrw.com/support/hos/hostpdf/hostmsprompts.pdf.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*(1), 18–29.

Hughes, C. A., Morris, J. A., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research and Practice*. Advance online publication. doi: 10.1111/ldrp.12142

Kane, T.J. & Staiger, D.O. (2012). *Gathering feedback for teaching*. Seattle, Washington: Bill and Melinda Gates Foundation.

Kaji, A. H., & Lewis, R. J. (2015). Noninferiority trials: Is a new treatment almost as effective as another?. *JAMA*, *313*, 2371-2372.

Kennedy, M. J. (2013). Integrating evidence-based practices and instructional design principles: Introducing content acquisition podcasts. www.teachingld.org/technologies.

Kennedy, M. J., Alves, K. D., Miciak, J., Romig, J., Mathews, H. M., & Thomas, C. N. (2016). Evaluating the relationship between naturalistic content acquisition podcast views and course performance. *Teacher Education and Special Education, 39,* 293-307.

Kennedy, M. J., Driver, M. K., Pullen, P. C., Ely, E., & Cole, M. T. (2013). Improving teacher candidates' knowledge of phonological awareness: A multimedia approach. *Computers & Education. 64,* 42-51. doi: 10.1016/j.compedu.2013.01.010

Kennedy, M. J., Ely, E., Thomas, C. N., Pullen, P. C, Newton, J. R, Ashworth, K, Cole, M, T., & Lovelace, S. P. (2012). Using multimedia tools to support teacher candidates' learning. *Teacher Education and Special Education, 35,* 243-257. doi: 10.1177/0888406412451158

Kennedy, M. J., Hirsch, S. E., Dillon, S. E., Rabideaux, L., Alves, K. D., & Driver, M. K. (2016). Using content acquisition podcasts to increase student knowledge and to reduce cognitive load. *Teaching of Psychology, 43,* 153-158.

Kennedy, M. J., Hirsch, S. E., Rodgers, W. J., Bruce, A., & Lloyd, J. W. (2017). Supporting high school teachers' implementation of evidence-based classroom management practices. *Teaching and Teacher Education. 63,* 47-57. doi: 10.1016/j.tate.2016.12.009

Kennedy, M. J., Newton, J. R., Haines, S., Walther-Thomas, C., & Kellems, R. O. (2012). A triarchic model for teaching "introduction to special education": Case studies, content acquisition podcasts, and effective feedback. *Journal of Technology and Teacher Education, 20,* 251-275.

Kennedy, M. J., Rodgers, W. J., & Romig, J. E. (2017). Classroom teaching scan [Computer software]. Retrieved from www.classroomteachingscan.com/ctscan

Kennedy, M. J., Rodgers, W. J., Romig, J. E., Lloyd, J. W., & Brownell, M. T. (2017). The impact of a multimedia professional development package on inclusive science teachers' vocabulary instruction. *Journal of Teacher Education, 68,* 213-230. doi: 10.1177/0022487116687554

Kennedy, M. J., Thomas, C. N., Aronin, S., Newton, J. R., & Lloyd, J. W. (2014). Improving teacher candidate knowledge using content acquisition podcasts. *Computers & Education. 70,* 116-127. doi:10.1016/j.compedu.2013.08.010

Kim, J. S., Olson, C. B. Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., Collins, P., & Land, R. E. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed latino english language learners in grades 6-12. *Journal of Research on Educational Effectiveness, 4,* 231-263.

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

Kraft, M.A., Blazar, D., Hogan, D. (2017). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. Brown University Working Paper.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Landers, R. N. (2015). Computing intraclass correlations (ICC) as an estimate of interrater reliability in SPSS. *The Winnower*, 2:e143518.81744, 1-4. doi:10.15200/winn.143518.81744

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*, 760-769. doi:10.1037/0003-066X.63.8.760

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

McLeskey, J., Barringer, M-D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., Lewis, T., Maheady, L., Rodriguez, J., Scheeler, M. C., Winn, J., & Ziegler, D. (2017). High-leverage practices in special education. Arlington, VA: Council for Exceptional Children & CEEDAR Center.

McKenna, J. W., Shin, M., & Ciullo, S. (2015). Evaluating reading and mathematics instruction for students with learning disabilities: A synthesis of observation research. *Learning Disability Quarterly, 38,* 195-207. doi:10.1177/0731948714564576

Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth-and eighth-grade students. *Reading and Writing, 21*(1-2), 131-151.

Myers, J., Scales, R. Q., Grisham, D. L., Wolsey, T. D., Dismuke, S., Smetana, L., … Martin, S. (2016). What about writing? A national exploratory study of writing instruction in teacher preparation programs. *Literacy Research and Instruction, 55,* 309-330. doi:10.1080/19388071.2016.1198442

National Center for Education Statistics. (1999). The NAEP 1998 writing report card for the nation and the states. (No. NCES 1999-462). Washington, D.C.: Institute for Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/nationsreportcard//pdf/main1998/1999462.pdf

National Center for Education Statistics. (2003). The nation's report card: Writing 2002 (No. NCES 2003-529). Washington, D.C.: Institute for Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2008). The nation's report card: Writing 2007 (No. NCES 2008-468). Washington, D.C.: Institute for Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2012). The nation's report card: Writing 2011 (No. NCES 2012-470). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012470

National Center for Education Statistics. (2016). Digest of education statistics 2015 (No. NCES 2016-014). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/pubs2016/2016014.pdf

National Commission on Writing. (2003, April). The neglected "R": The need for a writing revolution. Retrieved April 23, 2008, from http://www.collegeboard.com/prod_downloads/writingcom/neglectedr.pdf

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*, 230–240.

Perez-Burton, E. E., Merz, S. A., Ramirez, E. M., & Saroughi, M. (2015). The effect of cognitive apprenticeship-based professional development on teacher self-efficacy of science teaching, motivation, knowledge calibration, and perceptions of inquiry-based teaching. *Journal of Science Teacher Education, 26*, 525-548.

Peters-Burton, E. E., Merz, S. A., Ramirez, E. M., & Saroughi, M. (2015). The effect of cognitive apprenticeship-based professional development on teacher self-efficacy of science teaching, motivation, knowledge calibration, and perceptions of inquiry-based teaching. *Journal of Science Teacher Education, 26*, 525-548.

PLATO 5.0: Protocol for language arts teaching observations. (2017). Stanford: Center to Support Excellence in Teaching.

Poplin, M. S., Gray, R., & Larsen, S. (1980). A comparison of components of written expression abilities in learning disabled and non-learning disabled students at three grade levels. *Learning Disability Quarterly*, *3*, 46-53. doi:10.2307/1510674

Rodgers, W. J. (2017). Specialized instruction in co-taught high school classes (Doctoral dissertation). Retrieved from Libra. (https://doi.org/10.18130/V3M633)

Romig, J. E., Sundeen, T., Thomas, C. N., Kennedy, M. J., Philips, J., Peeples, K., Rodgers, W. J., & Mathews, H. M. (2018). Using multimedia to teach self-regulated strategy development to pre-service teachers. *Journal of Special Education Technology*. https://doi.org/10.1177/0162643417746373

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *Journal of Special Education, 51*, 72-82.

Smyth, J. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. Journal of Consulting and Clinical Psychology, 66, 174–184.

Swanson, E. A. (2008). Observing reading instruction for students with learning disabilities: A synthesis. *Learning Disability Quarterly, 31*, 115-133.

Sweller, J., Ayers, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

TeachingWorks High Leverage Practices.(n.d.) retrieved from http://www.teachingworks.org/work-of-teaching/high-leverage-practices

Troia, G. (2014). *Evidence-based practices for writing instruction* (Document No. IC-5). Retrieved from University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center website: http://ceedar.education.ufl.edu/tools/innovation-configuration/

Vaughn, S., Levy, S., Coleman, M., & Bos, C. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education, 36*, 2–13. doi:10.1177/00224669020360010101

Wyatt, W. J., Callahan, E. J., & Michael, J. (1985). Variations in methods of calculating reliability of duration data. *Psychology in the Schools, 22,* 473-479.